Parametric Methods for Time Series Discrimination

By

Andrew Joseph Grant

A thesis submitted to Macquarie University

for the degree of Doctor of Philosophy

Department of Statistics

July 2018

MACQUARIE
University
SYDNEY · AUSTRALIA

Contents

Abstract

In this thesis we consider the problem of determining whether two or more independent time series have been generated by the same underlying stochastic process, or by the same mechanism. There is an extensive literature on comparing time series from univariate stationary processes on the basis of their second order properties, that is, their dependence structures over time. These existing methods are nonparametric and are based on comparing periodograms or sample autocovariances. They are generally limited by requiring equal sample sizes and Gaussian assumptions. We introduce a parametric approach which involves fitting parametric models to the time series and comparing model parameters. The parametric approach avoids the limitations of the nonparametric and simulations are used to show that it results in a more powerful test. We also show how to extend the parametric approach to compare time series from multivariate stationary processes.

A further extension is to compare time series which are from stochastic processes which contain periodic components. Such time series are typically modelled using mixed models which are made up of a deterministic periodic component and a stationary stochastic component. We develop tests for whether two or more time series have been generated by processes with periodicities at the same fixed frequencies and stationary components with the same second order properties. In order to extend the procedures to the multivariate case we first develop novel methods for frequency estimation in the multivariate mixed model.

Statement of Candidate

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

The following paper resulting from this work has been published in a refereed journal.

Grant, A. J. and Quinn, B. G. (2017). Parametric spectral discrimination. *Journal of Time Series Analysis*, 38(6), 838–864.

______________________________ Date: ____________________

Andrew Joseph Grant

Acknowledgements

I wish to express enormous gratitude to my teacher, Professor Barry Quinn, who supervised this thesis. His patient guidance and willingness to give his time and impart his wisdom has made this an immense learning experience and allowed me to produce work well beyond what I envisaged at the start.

My appreciation also goes to the members of the Department of Statistics, including academic and professional staff as well as my fellow research students. The department provides an excellent environment to develop as a researcher and is very enjoyable to be a part of.

I thank my parents, Margaret and Greg, for always encouraging me to aim high, safe in the knowledge I could never fall far.

Finally, I thank Caitlin and Evelyn for their unwavering support, without which this would not have been possible, and Edward, for providing the final bit of motivation and inspiration needed to bring this work to fruition.

1

Introduction

This thesis is concerned with methods for comparing time series in order to determine if they have been generated by the same source, or by the same underlying mechanism. The need to compare, or in other words, to discriminate between, time series arises in a wide variety of applications. For example, time series discrimination can be used for fault detection in a mechanical system by comparing a vibration signal to a reference signal which is known to be in working order (Bassily et al., 2009; Jin, 2011). Climate data from different locations may be compared to determine if different regions display the same weather patterns (Lund et al., 2009). Comparing the physical properties of wireless signals can be used to enhance the security of wireless networks (Tugnait, 2013). In economics, it may be of interest to know if asset prices behave differently in different time periods (Coates and Diggle, 1986) or across different markets.

A time series is a single, partial, realisation of a stochastic process, or sequence of random variables, observed over time. We wish to compare, therefore, the statistical properties of the underlying stochastic processes of the time series being analysed. In many practical applications, it is sufficient to compare stochastic processes on the basis of their second order properties, that is, their dependence structure over time. This is commonly measured by the covariances between the random variables separated by fixed time lags, known as the autocovariances. An alternative measure of the dependence structure is the discrete Fourier transform of the autocovariances, known as the spectral density. We therefore require a test which compares the autocovariances, or equivalently, the spectral densities, of the stochastic processes generating the time series.

In practice, it will often be more relevant to test whether the stochastic processes have the same autocorrelations rather than the same autocovariances. This would be the case if, for example, a signal was recorded by two devices at different distances. The recordings would differ in scale, but we would wish that our test did not discriminate between the two. If two processes have the same autocorrelations, this is equivalent to the ratio of the spectral densities being constant, and we would say that the processes have the same spectral shape.

The topic of this thesis was originally studied in my Master of Research (MRes) thesis. The MRes thesis reviewed the existing literature on time series discrimination, which is mostly focused on the comparison of two time series from univariate stationary processes. That is, stochastic processes whose statistical properties remain the same over time. Almost all of the existing methods are nonparametric in that they are not based on fitting models to the time series. Instead, tests are based on either sample autocovariances or the periodogram, which is proportional to the squared modulus of the discrete Fourier transform of a time series.

The original work on the topic was by Coates and Diggle (1986) who proposed tests for comparing spectral densities based on computing the periodograms of the two time series at the Fourier frequencies, which can be done efficiently using the fast Fourier transform algorithm. To test for differences in spectral shape, they took the logarithm of the ratios of these values and used their range as a test statistic. It was shown that the test based on this range statistic has very low power. However, there has been a number of papers in recent years which have developed further tests based on comparing the values obtained by evaluating periodograms at the Fourier frequencies. A common approach, for example, is to consider distance measures between these values from each time series. These new tests have generally been shown to be improvements on those of Coates and Diggle (1986).

It is easy to see the appeal in using the periodogram as a basis for spectral discrimination, since smoothing the periodogram leads to a consistent estimator of the spectral density. Furthermore, the random variables obtained by computing the periodogram at a fixed set of points are asymptotically independent and identically distributed as multiples of chi-squared random variables. This property can be used to derive asymptotic theory of a periodogram based test statistic. However, if the periodogram is computed at a set of frequencies that is not fixed as the sample size increases, such as the Fourier frequencies, these asymptotic properties do not hold except in the case where the underlying processes are Gaussian and white. The asymptotic theory for the periodogram based tests therefore requires these fairly strong assumptions to be made. A further drawback to the nonparametric tests is that they generally require the sample sizes of the time series to be equal, although some tests have been adjusted to account for unequal sample sizes.

The MRes thesis examined a parametric approach which was proposed by Quinn (2006). Under this approach, autoregressions are fitted to the two time series. If two autoregressive processes have the same spectral shape then they have the same autoregressive parameters. Therefore, the null hypothesis of interest is that the two time series are from autoregressions with the same autogressive parameters. The alternative hypothesis is that the autoregressive parameters are not all the same. A test statistic was derived using a likelihood ratio procedure based on Gaussian likelihood functions. Since it is not assumed that the time series are from processes which are Gaussian, it is in fact a pseudo-likelihood ratio procedure. In order to compute the test statistic, autoregressive models must be fitted to two time series with the same autoregressive parameters but potentially different residual variances and sample sizes. The use of a parametric approach was motivated by the fact that parametric methods are generally expected to be more powerful than nonparametric. It also allows for time series with unequal sample sizes.

The MRes thesis presented algorithms for computing the pseudo-likelihood ratio test statistic and examined its distribution using simulations. The simulations suggested that when the time series are generated by autoregressions with known orders, the test statistic follows a chi-squared distribution even when the processes are not Gaussian. When the time series are generated by autoregressions of unknown orders, it was shown how to estimate the orders using information criteria. Simulations suggested that in this case the test statistic still follows a chi-squared distribution, which is expected since the orders are estimated using consistent procedures. However, when the time series are not generated by autoregressions, the test does not perform well in simulation studies. An important extension of the parametric approach therefore is to generalise it to the case of more general processes, since we do not wish to assume that the time series truly are autoregressive. A further extension is to the comparison of multivariate time series, which has not been studied widely in the literature.

The conclusion to the MRes thesis also discussed an extension to the case of time series which are generated by mixed models with both periodic and stationary components. These models are used for phenomena with periodicities which commonly arise in, for example, signal processing and astronomy. The periodic signal is typically modelled as the sum of sinusoids and the stationary component may be modelled, for example, using an autoregression. Of central importance in the analysis of these models is the estimation of the frequencies of the periodic components. In the discrimination context, we may wish to determine whether independent time series have periodic components at the same fixed frequencies. Depending on the application, we may also wish to incorporate the comparison of the stationary components, or these may be considered to be entirely background noise.

In this thesis we focus on the parametric approach for time series discrimination, concentrating in particular on the aforementioned extensions. The outline of the thesis is as follows.

In Chapter 2 we give background information relating to the models used throughout the thesis. We also introduce some of the notation and statistical theory that is commonly used.

In Chapter 3 we consider methods for discriminating between time series from univariate stationary processes. We begin with an overview of the existing nonparametric methods. We describe the parametric approach based on fitting autoregressions, with orders estimated using information criteria, and demonstrate the problems which arise when the time series are not autoregressive, as identified in the MRes thesis. We then propose a modification to the parametric test which fits autoregressions where the order is fixed at a function of the sample sizes. It is shown that this fixed order autoregressive approach results in a test which performs well even when the underlying processes are not autoregressive. We also establish the asymptotic properties of the estimators of the model parameters and the asymptotic distribution of the test statistic using weaker assumptions than Gaussianity. A simulation study compares the parametric approach with nonparametric alternatives and demonstrates that the parametric test is more powerful. Finally, we show how to extend the procedure for comparing more than two time series.

In Chapter 4 we develop a parametric test based on the same pseudo-likelihood ratio procedure as before but where we fit autoregressive-moving average models to the time series. This requires a procedure for fitting autoregressive-moving average models to two time series with the same model parameters but with potentially different residual variances and sample sizes. The procedure we develop is based on an extension of the Hannan–Rissanen algorithm (Hannan and Rissanen, 1982; Hannan and Kavalieris, 1984a). It is shown using simulations that this new test is more powerful than that which fits autoregressions if the true autoregressive and moving average orders are known. However, when the orders need to be estimated, which will be the case in practice, the simulations suggest that there is no advantage in fitting autoregressive-moving average models over fitting fixed order autoregressions of sufficient length.

In Chapter 5 we consider the comparison of two or more multivariate stationary processes. We develop tests using a similar pseudo-likelihood ratio procedure used in previous chapters for three different null hypotheses. The first is that the time series have been generated by vector autoregressions with the same autoregressive parameters. The second is that the spectral densities of the underlying stochastic processes differ only by the multiplication of a constant. The third, which is the most general, is that the corresponding components of each of the vector processes have the same spectral shape. The tests are based on fitting vector autoregressions to the time series. We establish asymptotic properties of the parameter estimators under the null hypotheses and demonstrate the performance of the tests using simulations.

In Chapter 6 we consider the estimation of frequency in time series with periodic components. Although frequency estimation for univariate time series from periodic processes is a widely studied problem, little attention has been given to the case where the time series are multivariate. In the multivariate, or multichannel, model that we consider in this chapter, the time series is generated by a vector process where each element of the periodic component has the same fixed frequencies but potentially different amplitudes and phases. We develop novel methods for the estimation of these fixed frequencies and establish the asymptotic properties of the estimators. We also demonstrate the performance of the estimation techniques in a simulation study.

In Chapter 7 we develop methods for discriminating between time series from processes which contain periodic components, both in the univariate and multivariate cases. We consider two different null hypotheses. The first is that the time series are from processes which have periodic components at the same fixed frequencies and stationary noise with the same spectral shape. The second is that the time series are from processes which have periodic components at the same fixed frequencies and with possibly independent noise. We motivate the techniques by considering the comparison of two univariate time series from processes with periodic components. We then generalise the procedures for the case of more than two time series and then for the case of multivariate time series.

We finish in Chapter 8 with a summary of the thesis and some discussion of areas for future research.

2
Background

In this chapter we give many of the definitions and notations that will be used throughout the thesis. We introduce standard time series models and provide some statistical theory which is useful in establishing asymptotic properties of estimators. Much of the background material in this chapter, except where otherwise referenced, can be found in standard time series textbooks, for example, Priestley (1981). Note that we mostly consider univariate random variables in this chapter. The multivariate generalisations are often straightforward, and where they are not they will be given in the thesis as needed. For detail on the multivariate versions of the background material, see, for example, Reinsel (1993).

2.1 Stationary Stochastic Processes

A stochastic process is a sequence of random variables, $\{X_t\}$, where t is some directional variable. Generally, t is considered to represent time, although in practice it can represent something else, such as space. In this thesis, we will assume that $\{X_t\}$ is a discrete time stochastic process, that is, that t is an integer. We will also assume implicitly that the process has second order moments. A time series is a single partial realisation of a stochastic process, observed at $t = 0, \ldots, T-1$ where T is the sample size. We denote by $\{X_t\}$ both a time series and its underlying stochastic process, and will rely on the context to differentiate between the two.

Definition 2.1 *A stochastic process is stationary if the joint distribution of* $X_0, X_1, \ldots X_s$ *and* $X_t, X_{t+1}, \ldots X_{t+s}$ *is the same for all* s *and* t*.*

Definition 2.2 *A stochastic process is weakly stationary if, for some constant* μ*,* $E(X_t) = \mu$ *for all* t *and* $\operatorname{cov}(X_t, X_{t+s}) = \operatorname{cov}(X_0, X_s)$ *for all* s *and* t*.*

We always assume that a stationary process is also ergodic. Loosely speaking, ergodicity ensures that the sample mean of any given realisation of a stochastic process will converge to the true mean of the stochastic process as the size of the sample increases (see, for example, Priestley, 1981, Section 5.3.6). In general, ergodicity is required in order to obtain estimators with good properties in time series analysis. However, it cannot be verified in practice since we only ever see a single realisation.

We are often particularly concerned with the second order properties of a weakly stationary stochastic process, that is, its dependence structure over time. This dependence structure can be measured by the autocovariances or autocorrelations. An alternative measure of the dependence structure is the spectral density.

Definition 2.3 *The autocovariances of* $\{X_t\}$ *are given by*

$$\gamma(j) = \gamma(-j) = \operatorname{cov}(X_t, X_{t-j}).$$

Definition 2.4 *The autocorrelations of* $\{X_t\}$ *are given by*

$$\frac{\gamma(j)}{\gamma(0)}.$$

Definition 2.5 *The spectral density of* $\{X_t\}$ *is*

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-ij\omega}.$$

Because $f(\omega)$ *is periodic with period* 2π*, and even, it need only be defined for values* $\omega \in [0, \pi]$*.*

The spectral density is the discrete Fourier transform of the autocovariance sequence. Given the spectral density, we can obtain the autocovariances by

$$\gamma(j) = \int_{-\pi}^{\pi} e^{ij\omega} f(\omega)\, d\omega.$$

Definition 2.6 *The stochastic process* $\{\varepsilon_t\}$ *is white noise if* $E(\varepsilon_t) = 0$ *and*

$$E(\varepsilon_s \varepsilon_t) = \begin{cases} \sigma^2, & \text{if } s = t \\ 0, & \text{otherwise} \end{cases},$$

for all s *and* t.

That is, a white noise process is a sequence of uncorrelated random variables. The term 'white' comes from the fact that the spectral density of $\{\varepsilon_t\}$ is constant for all ω. A stochastic process which is not white is referred to as coloured.

Definition 2.7 *The stochastic process* $\{\varepsilon_t\}$ *is Gaussian if any subset of the* $\{\varepsilon_t\}$ *is jointly normally distributed.*

If a white noise process is stationary and Gaussian, then the random variables are independent and identically distributed (i.i.d.). If a weakly stationary process is i.i.d. then this implies it is white noise. In time series analysis, we often require more structure than is given by white noise. However, it may not be desirable to make the fairly strong assumptions of independence and Gaussianity. The condition given in the following definition lies between the two.

Definition 2.8 *The stochastic process* $\{\varepsilon_t\}$ *is a sequence of martingale differences if* $E(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$ *for all* t*, where* $\mathcal{F}_t$ *is the* σ*-field generated by* $\varepsilon_t, \varepsilon_{t-1}, \ldots$.

The assumption that $\{\varepsilon_t\}$ is a sequence of martingale differences is often the weakest assumption that can be made to get good properties of estimators of time series model parameters. The following theorems relating to martingale differences therefore will be useful when establishing asymptotic properties of the estimators that we derive in this thesis. The first is the central limit theorem for sequences of martingale differences and is due to Billingsley (1961). The second is the law of the iterated logarithm for martingale differences and is due to Stout (1970).

Theorem 2.1 (Martingale central limit theorem) *Let* $\{\varepsilon_t\}$ *be a stationary sequence of martingale differences such that* $E\left(\varepsilon_t^2\right)$ *is finite. Then the distribution of* $T^{-1/2} \sum_{t=0}^{T-1} \varepsilon_t$ *converges to the normal distribution with mean zero and variance* $E\left(\varepsilon_t^2\right)$.

Theorem 2.2 (Law of the iterated logarithm for martingale differences) *Let* $\{\varepsilon_t\}$ *be a stationary sequence of martingale differences with* $E\left(\varepsilon_t^2\right) = 1$. *Then*

$$\limsup_{T \to \infty} (2T \log \log T)^{-1/2} \sum_{t=0}^{T-1} \varepsilon_t = 1$$

almost surely.

2.2 Autoregressive-Moving Average Models

According to the Wold decomposition theorem, any weakly stationary process, $\{X_t\}$, can be represented by the sum of a purely deterministic and a purely non-deterministic component. We assume that the deterministic component is constant at zero, that is, that $\{X_t\}$ has zero mean. We can therefore represent $\{X_t\}$ by

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \tag{2.1}$$

where

$$a_0 = 1, \qquad \sum_{j=0}^{\infty} a_j^2 < \infty$$

and $\{\varepsilon_t\}$ is white. Since (2.1) has an infinite number of parameters, it is difficult to work with in practice. However, it can be shown that (2.1) can be approximated arbitrarily well by a finite parameter process satisfying an equation of the form

$$X_t + \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q}, \tag{2.2}$$

subject to certain conditions on $\{\beta_j\}$ and $\{\alpha_j\}$ which we discuss below. A process satisfying an equation of this form is known as the autoregressive-moving average process of orders p and q, denoted ARMA(p, q). When $q = 0$ the process is said to be an autoregression of order p, denoted AR(p), and when $p = 0$ the process is said to be a moving average process of order q, denoted MA(q). Thus, $\beta_1, \ldots, \beta_p$ are referred to as the autoregressive parameters, and $\alpha_1, \ldots, \alpha_q$ are referred to as the moving average parameters. The class of ARMA(p, q) processes therefore provides a way of modelling stationary processes using finite parameter linear models.

It is often convenient to write (2.2) using backshift notation. Let z be the backshift operator such that $z^j X_t = X_{t-j}$. We define the polynomials

$$b_\beta(z) = 1 + \sum_{j=1}^{p} \beta_j z^j \qquad \text{and} \qquad a_\alpha(z) = 1 + \sum_{j=1}^{q} \alpha_j z^j.$$

Then (2.2) can be written as

$$b_\beta(z) X_t = a_\alpha(z) \varepsilon_t. \tag{2.3}$$

In order for (2.2), or equivalently (2.3), to be a good approximation to (2.1), the following conditions must be met.

Condition 2.1 *$b_\beta(z)$ and $a_\alpha(z)$ have no common zeros.*

Condition 2.2 *The zeros of $b_\beta(z)$ all lie outside the unit circle.*

Condition 2.3 *The zeros of* $a_\alpha(z)$ *all lie outside the unit circle.*

If Condition 2.1 is not met then the common factors would cancel in (2.3). Conditions 2.2 and 2.3 ensure that the model has a stationary solution, and that it is causal and invertible.

It follows from Definitions 2.3 and 2.5 that the spectral density of a stationary process that follows an ARMA(p,q) process is given by

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{\left|a_\alpha\left(e^{-ij\omega}\right)\right|^2}{\left|b_\beta\left(e^{-ij\omega}\right)\right|^2}.$$

We can therefore estimate the spectral densities of stationary processes using ARMA(p,q) model parameter estimators.

2.3 The Periodogram

A widely used function in time series analysis, particularly in spectral analysis, is the periodogram. The periodogram is defined by

$$I_{T,X}(\omega) = \frac{2}{T}\left|\sum_{t=0}^{T-1} X_t e^{-i\omega t}\right|^2.$$

It is easy to see the appeal in using the periodogram for spectral analysis, since the spectral density is the discrete Fourier transform of the autocovariance sequence. However, the periodogram is not a consistent estimator of the spectral density. It can be shown though that by smoothing the periodogram, using for example, a mean or median smoother, a consistent estimator of the spectral density can be obtained.

Another appeal of the periodogram is that it can be efficiently computed at the set of frequencies

$$\omega_j = \frac{2\pi j}{T}, \qquad j = 1, \ldots, \left\lfloor \frac{T-1}{2} \right\rfloor, \tag{2.4}$$

where $\lfloor k \rfloor$ is the largest integer smaller than or equal to k, using the fast Fourier transform algorithm. The frequencies given by (2.4) are referred to as the Fourier frequencies.

For a given frequency λ, the random variable

$$\frac{I_{T,X}(\lambda)}{2\pi f(\lambda)}$$

has asymptotically the chi-squared (χ^2) distribution with two degrees of freedom. Furthermore, for any fixed set of frequencies $\lambda_1, \ldots, \lambda_m$, the random variables

$$\frac{I_{T,X}(\lambda_j)}{2\pi f(\lambda_j)}, \qquad j = 1, \ldots, m,$$

are asymptotically independent and have the χ^2 distribution with two degrees of freedom. This follows from Theorem 2.6 below. Note that this is not true for the whole set of Fourier frequencies, except for the case where $\{X_t\}$ is Gaussian white noise.

Suppose that $\{X_t\}$ is generated by (2.1) with $\{\varepsilon_t\}$ a sequence of martingale differences and $E\left(\varepsilon_t^2\right) = \sigma^2$. Let

$$C_T(\omega) = \sum_{t=0}^{T-1} \varepsilon_t \cos(\omega t) \qquad \text{and} \qquad S_T(\omega) = \sum_{t=0}^{T-1} \varepsilon_t \sin(\omega t).$$

The following theorems are due to An et al. (1983).

Theorem 2.3

$$\limsup_{T\to\infty} \max_{\omega} \frac{I_{T,X}(\omega)}{2\pi f(\omega)\log T} \leqslant 1$$

almost surely.

Theorem 2.4

$$\limsup_{T\to\infty} \max_{\omega} \frac{C_T^2(\omega)}{\sigma^2 T \log T} \leqslant 1 \qquad \textit{and} \qquad \limsup_{T\to\infty} \max_{\omega} \frac{S_T^2(\omega)}{\sigma^2 T \log T} \leqslant 1$$

almost surely.

The periodogram is discussed in further detail in Chapter 6, where it is shown how it arises in the estimation of fixed frequencies in periodic processes.

2.4 Theorems Used for Establishing Asymptotic Properties of Estimators

In this section we provide two theorems that will be widely used throughout the thesis in order to prove the strong consistency and central limit theorems of estimators. The first is Lemma 1 of Wu (1981), who gave a sufficient condition for the strong consistency of non-linear least squares estimators. The sufficient condition in fact applies to any estimator which is the minimiser or maximiser of a function.

Let $S_T(\theta)$ be a function of a parameter θ depending on sample size T, such that $S_T(\theta)$ diverges to ∞ at rate $O(T)$ for all θ. Denote the minimiser of $S_T(\theta)$ with respect to θ by $\widehat{\theta}_n$ and the true value of θ by θ_0.

Theorem 2.5 *Suppose, for any $\delta > 0$,*

$$\liminf_{T\to\infty} \inf_{|\theta-\theta_0|\geqslant\delta} \{S_T(\theta) - S_T(\theta_0)\} > 0$$

almost surely. Then $\widehat{\theta}_T \to \theta_0$ almost surely as $T \to \infty$.

The next theorem is originally due to Hannan (1973a) and later generalised by Hannan (1979). We give here a version of the theorem which is Theorem 4 of Quinn and Hannan (2001). It relates to the asymptotic distribution of the real and imaginary parts of random variables of the form

$$T^{-(2k+1)/2} \sum_{t=0}^{T-1} t^k \varepsilon_t e^{-it\omega_j}, \qquad \omega_j = \frac{2\pi j}{T}, \tag{2.5}$$

where $\{\varepsilon_t\}$ is stationary with mean zero and spectral density $f(\omega)$ which is continuous at a fixed frequency ω_0. Random variables of this form arise often in time series analysis. For example, they have been seen in the previous section in the periodogram.

Theorem 2.6 *Let $k = 0, 1, \ldots, K-1$ and consider the $2Km$ quantities that are the real and imaginary parts of (2.5), for m values of ω_j nearest to ω_0, $0 < \omega_0 < \pi$. Then the distribution of these $2Km$ random variables converges to the distribution of a vector of normal random variables with zero means. For different values of j the limiting random variables are independent and the imaginary terms are independent of the real. For fixed j, and for each of the real and imaginary components, the K limiting random variables have covariance matrix with (k,l)th entry $\pi f(\omega_0) / (k+l+1)$, $k, l = 0, \ldots, K-1$.*

2.5 The Kronecker Product and Vec Operator

When dealing with vectors and matrices we will often make use of the Kronecker product and the vec operator. These are defined below, as well a theorem which connects the two. The definitions and theorem can be found in, for example, Neudecker (1969).

Definition 2.9 *Let A and B be $m \times n$ and $p \times q$ matrices, respectively. Then the Kronecker product $A \otimes B$ is the $mp \times nq$ matrix with (i,j)th block $a_{ij}B$, where a_{ij} is the (i,j)th element of A.*

Note that, for matrices A, B, C and D of appropriate dimensions,

$$(A \otimes B)(C \otimes D) = AC \otimes BD.$$

Definition 2.10 *Let A be an $m \times n$ matrix. Then $\operatorname{vec} A$ is the mn-dimensional column vector produced by stacking each column of A on top of each other, in order from left to right.*

Theorem 2.7 *Let A, B and C be matrices of appropriate dimensions. Then*

$$\operatorname{vec}(ABC) = \left(C' \otimes A\right) \operatorname{vec} B.$$

3

Autoregressive Spectral Discrimination

3.1 Introduction

The problem of determining whether two independent time series are realisations of the same weakly stationary stochastic process was first considered by Coates and Diggle (1986), who proposed nonparametric tests based on the periodogram to compare theoretical spectral densities. Two null hypotheses were considered. The first was that the spectral densities of the underlying processes are the same. The second was that the ratio of the spectral densities is constant, that is, that the two processes have the same spectral shape.

Suppose we have two time series from the same source. If they were collected at different distances from the source, they would have the same autocorrelation structure but would differ in amplitude. Similarly, the spectral densities might not be exactly equal due to calibration problems in the measuring devices. It is therefore of more interest to consider whether the two samples come from underlying processes with the same spectral shape.

Consider two univariate, stationary stochastic processes, $\{X_t\}$ and $\{Y_t\}$, assumed to have zero means. Their spectral densities are

$$f_X(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_X(j)\, e^{-i\omega j} \qquad \text{and} \qquad f_Y(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_Y(j)\, e^{-i\omega j},$$

where $\gamma_X(j)$ and $\gamma_Y(j)$ are the autocovariances of $\{X_t\}$ and $\{Y_t\}$, respectively. The null hypothesis is that $f_X(\omega)/f_Y(\omega)$ is the same for all $\omega \in (0, \pi)$, and the alternative hypothesis is its complement.

The tests of Coates and Diggle (1986) were based on the ratios of the periodogram ordinates at the Fourier frequencies. To compare spectral shape they considered the range of the logarithm of the periodogram ratios, with a large range indicating a departure from the null hypothesis. The test requires Gaussian assumptions and was shown to have very low power. Further work by Diggle and Fisher (1991) produced a test using the normalised cumulative periodogram based on Bartlett's test for white noise (Bartlett, 1954).

There has been considerable interest in the problem in recent years. Lund et al. (2009) and Jin (2015) proposed testing for equality of autocovariances and autocorrelations up to a chosen lag. The obvious drawback to this method is that an arbitrary, finite number of autocovariances must be chosen to test. Jin and Wang (2016) derived a new test which compares the first r autocorrelations for $r = O(\log T/\log\log T)$, where T is the sample size. Dette et al. (2011) and Preuß and Hildebrandt (2013) considered measures of distance between periodogram ordinates. Fokianos and Savvides (2008) and Jin (2011) proposed semiparametric methods, which fit parametric models to $\log\{f_X(\omega)/f_Y(\omega)\}$. Other tests based on the periodogram have been given by Tugnait (2013), Lu and Li (2013) and Decowski and Li (2015). These nonparametric and semiparametric tests generally require Gaussian assumptions and most require equal sample sizes.

A less obvious, but natural, approach to discriminating between spectral densities is parametric, where the processes are modelled as, for example, long-order autoregressions. There is a long history of fitting long-order autoregressions to stationary processes in the hope of estimating spectral shape (see, for example, Durbin, 1959 and Durbin, 1960). Let $\{X_t\}$ and $\{Y_t\}$ satisfy

$$X_t + \beta_{X,1}X_{t-1} + \cdots + \beta_{X,p_X}X_{t-p_X} = \varepsilon_t$$

and

$$Y_t + \beta_{Y,1}Y_{t-1} + \cdots + \beta_{Y,p_Y}Y_{t-p_Y} = u_t,$$

respectively, for some parameters $\beta_{X,1}, \ldots, \beta_{X,p_X}$ and $\beta_{Y,1}, \ldots, \beta_{Y,p_Y}$, and orders p_X and p_Y. It is assumed that the innovation processes, $\{\varepsilon_t\}$ and $\{u_t\}$, are independent sequences of martingale differences with

$$E\left(\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right) = \sigma_\varepsilon^2 \qquad \text{and} \qquad E\left(u_t^2 \mid \mathcal{G}_{t-1}\right) = \sigma_u^2,$$

where $\mathcal{F}_t$ and $\mathcal{G}_t$ are the σ-fields generated by $\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$ and $\{u_t, u_{t-1}, \dots\}$, respectively. The spectral densities of $\{X_t\}$ and $\{Y_t\}$ are

$$f_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi \left|1 + \sum_{j=1}^{p_X} \beta_{X,j} e^{-ij\omega}\right|^2} \quad \text{and} \quad f_Y(\omega) = \frac{\sigma_u^2}{2\pi \left|1 + \sum_{j=1}^{p_Y} \beta_{Y,j} e^{-ij\omega}\right|^2}.$$

Thus, $f_X(\omega)$ and $f_Y(\omega)$ will have the same shape if and only if the autoregressive parameters are equal. The null and alternative hypotheses are then

$$H_0 : \beta_{X,j} = \beta_{Y,j} \ \forall j$$
$$H_A : \exists\, j \text{ such that } \beta_{X,j} \neq \beta_{Y,j}.$$

Quinn (2006) briefly outlined a likelihood ratio approach for testing this hypothesis, and a similar hypothesis for mixed sinusoidal and autoregressive processes. Also proposed was an iterative procedure for estimating parameters under H_0, as well as an indication of how to estimate the sinusoidal and autoregressive orders. Several simulations were carried out, for the case of two sinusoids added to first order autoregressive noise. The results were inconclusive, and few details were given, as the paper and content were geared towards an Engineering audience. An alternative method for comparing autoregressive parameters has been given by De Souza and Thomson (1982).

The parametric approach has a number of advantages over the nonparametric. It does not require Gaussian assumptions and allows for unequal sample sizes. Also, since it is based on the likelihood ratio principle, it is expected to be asymptotically more powerful than the nonparametric tests, especially when the spectral densities are those of autoregressions. So, for example, the test will be more powerful than any based on an increasing number of autocovariances, such as that of Jin and Wang (2016), when the processes are autoregressive, or near-autoregressive.

In this chapter we derive the likelihood ratio statistic and propose novel methods for its computation. We prove the strong consistency and central limit theorem of the estimators, and the asymptotic distribution of the test statistic, under H_0. The properties of the test statistic are then examined using simulations. It is shown that the test performs well when the time series are from autoregressions, but not when the processes are not autoregressive. A modification to the test is proposed which is shown to improve its performance when the time series do not come from autoregressions. The modified test is then compared with nonparametric alternatives using simulations and is shown to have higher power. Much of the material in this chapter has been published in Grant and Quinn (2017).

3.2 The Likelihood Ratio Test

The test statistic will be derived using the likelihood ratio principle. Denoting the maximised log-likelihood functions under H_0 and H_A as $\widehat{l}_0$ and $\widehat{l}_A$, respectively, the test statistic is

$$\Lambda = 2\left(\widehat{l}_A - \widehat{l}_0\right).$$

When the autoregressive orders are known, Λ will be shown to follow, asymptotically, a χ^2 distribution with degrees of freedom the difference between the number of parameters estimated under H_A and under H_0, even without Gaussian assumptions. The autoregressive orders are not known, however, and must be estimated. The effect on the asymptotic distribution is examined in Section 3.5.

Although no assumptions are made about the distributions of $\{\varepsilon_t\}$ and $\{u_t\}$, the likelihood functions are derived as though these processes are Gaussian. Hence we are using a pseudo-likelihood, or Gaussian likelihood, procedure. We avoid the preperiod value problem by conditioning the joint distributions of $\{\varepsilon_t\}$ and $\{u_t\}$ on the first p_X and p_Y values, respectively, remaining fixed at their observed values. The conditional Gaussian log-likelihood functions are therefore

$$l_X\left(\beta_X, \sigma_\varepsilon^2\right) = -\frac{T_1}{2}\log\left(2\pi\sigma_\varepsilon^2\right) - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=p_X}^{T_1-1}\left\{b_{\beta_X}(z)\,X_t\right\}^2$$

and

$$l_Y\left(\beta_Y, \sigma_u^2\right) = -\frac{T_2}{2}\log\left(2\pi\sigma_u^2\right) - \frac{1}{2\sigma_u^2}\sum_{t=p_Y}^{T_2-1}\left\{b_{\beta_Y}(z)\,Y_t\right\}^2,$$

where T_1 and T_2 are the sample sizes of $\{X_t\}$ and $\{Y_t\}$, respectively.

3.3 Parameter Estimation

3.3.1 Parameter Estimation Under the Alternative Hypothesis

Under H_A, $\{X_t\}$ and $\{Y_t\}$ are independent autoregressions, and their parameters can be estimated using the usual techniques to maximise the conditional Gaussian log-likelihoods. Quite often the Levinson–Durbin approach (Levinson, 1947; Durbin, 1960) is used to estimate model parameters rather than maximum likelihood. Asymptotically this will have no impact. Summing the two individually maximised log-likelihoods, we have

$$\widehat{l}_A = -\frac{T_1+T_2}{2} - \frac{T_1}{2}\log\left(2\pi\widehat{\sigma}_{\varepsilon;A}^2\right) - \frac{T_2}{2}\log\left(2\pi\widehat{\sigma}_{u;A}^2\right),$$

where $\widehat{\sigma}_{\varepsilon;A}^2$ and $\widehat{\sigma}_{u;A}^2$ are whichever estimators are used of σ_ε^2 and σ_u^2, respectively, under H_A.

3.3.2 Parameter Estimation Under the Null Hypothesis

Under H_0, we have a particular case of two autoregressions with the same autoregressive parameters but with potentially different innovation variances and sample sizes. We thus have

$$X_t + \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} = \varepsilon_t$$

and

$$Y_t + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} = u_t,$$

where $\beta = \begin{bmatrix} \beta_1 & \cdots & \beta_p \end{bmatrix}'$ are the common parameters and p is the common order. The conditional Gaussian log-likelihood is

$$l_0\left(\beta, \sigma_\varepsilon^2, \sigma_u^2\right) = -\frac{T_1}{2}\log\left(2\pi\sigma_\varepsilon^2\right) - \frac{T_2}{2}\log\left(2\pi\sigma_u^2\right) - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=p}^{T_1-1}\left\{b_\beta(z)X_t\right\}^2 - \frac{1}{2\sigma_u^2}\sum_{t=p}^{T_2-1}\left\{b_\beta(z)Y_t\right\}^2.$$

Instead of maximising this, we reparametrise and use a profile likelihood approach. Letting $\sigma_\varepsilon^2 = \lambda\sigma_u^2$, we rewrite the conditional Gaussian log-likelihood as

$$l_0\left(\beta, \sigma_\varepsilon^2, \lambda\right) = -\frac{T_1+T_2}{2}\log\left(2\pi\sigma_\varepsilon^2\right) + \frac{T_2}{2}\log\lambda - \frac{1}{2\sigma_\varepsilon^2}\left[\sum_{t=p}^{T_1-1}\left\{b_\beta(z)X_t\right\}^2 + \lambda\sum_{t=p}^{T_2-1}\left\{b_\beta(z)Y_t\right\}^2\right].$$

This is maximised with respect to β, for fixed λ, by

$$\widehat{\beta}_\lambda = \begin{bmatrix} \widehat{\beta}_{\lambda,1} & \cdots & \widehat{\beta}_{\lambda,p} \end{bmatrix}' = -C_\lambda^{-1}c_\lambda,$$

where C_λ is the $p \times p$ matrix with (i,j)th element

$$(T_1+T_2)^{-1}\left(\sum_{t=p}^{T_1-1}X_{t-i}X_{t-j} + \lambda\sum_{t=p}^{T_2-1}Y_{t-i}Y_{t-j}\right)$$

and c_λ is the $p \times 1$ vector with ith element

$$(T_1+T_2)^{-1}\left(\sum_{t=p}^{T_1-1}X_tX_{t-i} + \lambda\sum_{t=p}^{T_2-1}Y_tY_{t-i}\right).$$

We then have $\widetilde{l}_0\left(\sigma_\varepsilon^2, \lambda\right) = l_0\left(\widehat{\beta}_\lambda, \sigma_\varepsilon^2, \lambda\right)$, which is maximised with respect to σ_ε^2, for fixed λ, by

$$\widetilde{\sigma}_{\varepsilon;\lambda}^2 = (T_1+T_2)^{-1}\left[\sum_{t=p}^{T_1-1}\left\{b_{\widehat{\beta}_\lambda}(z)X_t\right\}^2 + \lambda\sum_{t=p}^{T_2-1}\left\{b_{\widehat{\beta}_\lambda}(z)Y_t\right\}^2\right].$$

We thus obtain the profile log-likelihood

$$\breve{l}_0(\lambda) = \widetilde{l}_0\left(\widetilde{\sigma}_{\varepsilon;\lambda}^2, \lambda\right) = -\frac{T_1+T_2}{2} - \frac{T_1+T_2}{2}\log\left(2\pi\widetilde{\sigma}_{\varepsilon;\lambda}^2\right) + \frac{T_2}{2}\log\lambda.$$

By maximising $\breve{l}_0(\lambda)$ with respect to λ, we may obtain estimators of all the parameters. Denoting the maximiser by $\widehat{\lambda}$, the estimator of β is

$$\widehat{\beta} = \widehat{\beta}_{\widehat{\lambda}} = \left[\begin{array}{ccc} \widehat{\beta}_{\widehat{\lambda},1} & \cdots & \widehat{\beta}_{\widehat{\lambda},p} \end{array} \right]' = -C_{\widehat{\lambda}}^{-1} c_{\widehat{\lambda}}. \tag{3.1}$$

Then $\widehat{\sigma}^2_{\varepsilon;0} = \widetilde{\sigma}^2_{\varepsilon;\widehat{\lambda}}$ and $\widehat{\sigma}^2_{u;0} = \widehat{\sigma}^2_{\varepsilon;0}/\widehat{\lambda}$, where $\widehat{\sigma}^2_{\varepsilon;0}$ and $\widehat{\sigma}^2_{u;0}$ are the estimators of σ^2_ε and σ^2_u, respectively, under H_0.

Any optimisation procedure can be used to find $\widehat{\lambda}$. One such iterative procedure is as follows. Given a current estimate of λ, denoted by $\widetilde{\lambda}$, β is estimated by $\widehat{\beta}_{\widetilde{\lambda}}$. Then λ is re-estimated by

$$\frac{T_2 \widetilde{\sigma}^2_{\varepsilon;\widetilde{\lambda}}}{\sum_{t=p}^{T_2-1} \left\{ b_{\widehat{\beta}_{\widetilde{\lambda}}}(z) Y_t \right\}^2}$$

and the process repeats until convergence. An initial value of $\widetilde{\lambda}$ can be obtained from the estimates under the alternative hypothesis, that is, by letting $\widetilde{\lambda} = \widehat{\sigma}^2_{\varepsilon;A}/\widehat{\sigma}^2_{u;A}$.

3.3.3 Asymptotic Properties of the Estimators

In this section we establish the strong consistency and central limit theorem of the estimators under H_0. Proofs of Theorems 3.1 and 3.2 as well as Lemma 3.1 are given in the Appendix.

Let C_X and C_Y be the $p \times p$ matrices with (i,j)th elements

$$T_1^{-1} \sum_{t=p}^{T_1-1} X_{t-i} X_{t-j} \qquad \text{and} \qquad T_2^{-1} \sum_{t=p}^{T_2-1} Y_{t-i} Y_{t-j},$$

respectively. Letting Γ_X and Γ_Y be the $p \times p$ matrices with (i,j)th elements given by $\gamma_X(|i-j|)$ and $\gamma_Y(|i-j|)$, respectively, note that

$$C_X \to \Gamma_X = \sigma^2_\varepsilon \Omega \qquad \text{and} \qquad C_Y \to \Gamma_Y = \sigma^2_u \Omega$$

almost surely, where Ω depends only on β. Let $\theta = \left[\begin{array}{ccc} \beta' & \sigma^2_\varepsilon & \sigma^2_u \end{array} \right]'$, and let $l_0(\theta)$ be maximised at $\widehat{\theta} = \left[\begin{array}{ccc} \widehat{\beta}' & \widehat{\sigma}^2_\varepsilon & \widehat{\sigma}^2_u \end{array} \right]'$. Denote the true parameter values by $\theta_0 = \left[\begin{array}{ccc} \beta_0' & \sigma^2_{\varepsilon 0} & \sigma^2_{u0} \end{array} \right]'$, where $\beta_0 = \left[\begin{array}{ccc} \beta_{0,1} & \cdots & \beta_{0,p} \end{array} \right]'$.

Theorem 3.1 *$\widehat{\theta} \to \theta_0$ almost surely as $T_1, T_2 \to \infty$.*

Lemma 3.1 is required in the proof of Theorem 3.2. It is almost obvious, and straightforward to prove.

Lemma 3.1 *Let $\{\xi_t\}$ and $\{\zeta_t\}$ be independent sequences of random variables, each converging in distribution to the standard normal. Let*

$$Z_{T_1,T_2} = \sqrt{\frac{T_1}{T_1+T_2}} \xi_{T_1} + \sqrt{\frac{T_2}{T_1+T_2}} \zeta_{T_2}.$$

Then the distribution of Z_{T_1,T_2} *converges to the standard normal as* $T_1, T_2 \to \infty$.

Theorem 3.2 *Assume that* $E\left(\varepsilon_t^3 \mid \mathcal{F}_{t-1}\right)$ *and* $E\left(u_t^3 \mid \mathcal{G}_{t-1}\right)$ *are constants, and that* $E\left(\varepsilon_t^4\right) < \infty$ *and* $E\left(u_t^4\right) < \infty$. *Then, as* $T_1, T_2 \to \infty$, *the distribution of*

$$\begin{bmatrix} (T_1+T_2)^{1/2}\left(\widehat{\beta}-\beta_0\right) \\ T_1^{1/2}\left(\widehat{\sigma}_\varepsilon^2-\sigma_{\varepsilon 0}^2\right) \\ T_2^{1/2}\left(\widehat{\sigma}_u^2-\sigma_{u0}^2\right) \end{bmatrix}$$

converges to the normal distribution with mean zero and covariance matrix

$$\Sigma = \begin{bmatrix} \Omega^{-1} & 0 & 0 \\ 0 & \left(\sigma_{\varepsilon 0}^2\right)^4 \upsilon & 0 \\ 0 & 0 & \left(\sigma_{u0}^2\right)^4 \eta \end{bmatrix},$$

where υ *and* η *are given by*

$$\upsilon = \frac{1}{\left(\sigma_{\varepsilon 0}^2\right)^4} E\left(\varepsilon_t^4\right) - \frac{1}{\left(\sigma_{\varepsilon 0}^2\right)^2} \qquad \textit{and} \qquad \eta = \frac{1}{\left(\sigma_{u0}^2\right)^4} E\left(u_t^4\right) - \frac{1}{\left(\sigma_{u0}^2\right)^2}.$$

If $\{\varepsilon_t\}$ *and* $\{u_t\}$ *are Gaussian then*

$$\Sigma = \begin{bmatrix} \Omega^{-1} & 0 & 0 \\ 0 & 2\left(\sigma_{\varepsilon 0}^2\right)^2 & 0 \\ 0 & 0 & 2\left(\sigma_{u0}^2\right)^2 \end{bmatrix}.$$

3.3.4 Order Estimation

Although we have implicitly assumed above that the orders are known, this will, of course, not be the case in general. We can estimate the orders using information criteria. In what follows, we assume that the processes are autoregressions. Under H_A, the estimated order of $\{X_t\}$, $\widehat{p}_X$, is the minimiser of

$$\begin{aligned} \phi(k) &= -2\widehat{l}_X^{(k)} + kg(T_1) \\ &= T_1 \log \widehat{\sigma}_{\varepsilon;A}^{2(k)} + kg(T_1) \end{aligned}$$

over $k = 0, \ldots, K$, where $\widehat{l}_X^{(k)}$ is the maximised log-likelihood and $\widehat{\sigma}_{\varepsilon;A}^{2(k)}$ is the estimator of σ_ε^2 assuming the order is k, $g(T)$ is a chosen penalty function and K is assumed to be greater than or equal to p_X. This procedure is easily incorporated into the Levinson–Durbin algorithm. The estimated order of $\{Y_t\}$, $\widehat{p}_Y$, is computed in the same way. Common choices of information criterion are AIC where $g(T) = 2$ (Akaike, 1969); BIC, where $g(T) = \log T$ (Akaike, 1978; Schwarz, 1978); and HQIC, where $g(T) = \kappa \log\log T$ for $\kappa > 2$ (Hannan and

Quinn, 1979). It is known that AIC will not produce a consistent estimator and will overestimate the order with a probability of approximately 0.2883 (Shibata, 1976; Quinn, 1988). Both BIC and HQIC provide consistent estimators, with HQIC tending to underestimate the order less often than BIC.

Under H_0, the information criterion we minimise is

$$\phi\left(k\right)=-2\widehat{l}_0^{(k)}+kg\left(T_1,T_2\right),$$

where k is the number of parameters being estimated, equal to the common order. The estimated order, $\widehat{p}$, is thus the minimiser of

$$\phi\left(k\right)=T_1\log\widehat{\sigma}_{\varepsilon;0}^{2(k)}+T_2\log\widehat{\sigma}_{u;0}^{2(k)}+kg\left(T_1,T_2\right)$$

over $k=0,\ldots,K$, where K is assumed to be greater than or equal to p. AIC, BIC or HQIC can then be applied. Here, we use $g\left(T_1,T_2\right)=g\left(T_1+T_2\right)$.

3.3.5 An Alternative Procedure Based on the Levinson–Durbin Algorithm

A more practically efficient approach to parameter estimation under H_0 is to make use of the properties of Toeplitz matrices, which allows us to use the Levinson–Durbin algorithm (see, for example, Hannan and Rissanen, 1982 or Hannan and Kavalieris, 1984b). We shall replace the computations in (3.1) with Toeplitz versions of the statistics in the following manner. For $i=0,\ldots,p$, put

$$\widetilde{c}\left(i\right)=\left(T_1+T_2\right)^{-1}\left(\sum_{t=i}^{T_1-1}X_tX_{t-i}+\lambda\sum_{t=i}^{T_2-1}Y_tY_{t-i}\right),$$

where $\lambda=\widehat{\sigma}_{\varepsilon;0}^2/\widehat{\sigma}_{u;0}^2$ and $\widehat{\sigma}_{\varepsilon;0}^2$ and $\widehat{\sigma}_{u;0}^2$ are current estimates of σ_ε^2 and σ_u^2, respectively. For a given order p, consider the estimator of β

$$\widehat{\beta}^p=-\widehat{\Gamma}_p^{-1}\widehat{\gamma}^p,$$

where $\widehat{\gamma}^p$ is the $p\times1$ vector with ith element given by $\widetilde{c}\left(i\right)$, and $\widehat{\Gamma}_p$ is the $p\times p$ matrix with (i,j)th element given by $\widetilde{c}\left(\left|i-j\right|\right)$. We can then use the Levinson–Durbin algorithm, because of the structures of $\widehat{\Gamma}_p^{-1}$ and $\widehat{\gamma}^p$. The procedure needs to be iterated for fixed p, as $\widehat{\sigma}_{\varepsilon;0}^2$ and $\widehat{\sigma}_{u;0}^2$ need to be updated.

At each step in the algorithm, we define the information criterion as

$$\phi\left(p,\lambda\right)=\left(T_1+T_2\right)\log\widehat{\sigma}_{\varepsilon;0}^{2(p)}-T_2\log\lambda+pg\left(T_1+T_2\right).$$

We put $\widehat{p}_\lambda$ equal to the minimiser over $p=0,\ldots,K$ and

$$\psi\left(\lambda\right)=\phi\left(\widehat{p}_\lambda,\lambda\right).$$

We then define $\widehat{\lambda}$ to be the minimiser of $\psi\left(\lambda\right)$ and $\widehat{p}$ to be equal to $\widehat{p}_{\widehat{\lambda}}$.

3.4 The Test Statistic

The test statistic is

$$\begin{aligned}\Lambda &= 2\left(\widehat{l}_A - \widehat{l}_0\right)\\ &= T_1 \log\left(\frac{\widehat{\sigma}^2_{\varepsilon;0}}{\widehat{\sigma}^2_{\varepsilon;A}}\right) + T_2 \log\left(\frac{\widehat{\sigma}^2_{u;0}}{\widehat{\sigma}^2_{u;A}}\right).\end{aligned}$$

Theorem 3.3 gives the asymptotic distribution of Λ for a given p.

Theorem 3.3 *Under H_0 the distribution of Λ converges to the χ^2 distribution with p degrees of freedom as $T_1, T_2 \to \infty$.*

If we compute the test statistic at the true orders, we reject H_0 at significance level α when Λ is greater than the $100\,(1-\alpha)$th percentile of the χ^2 distribution with p degrees of freedom. If we instead estimate the orders, the degrees of freedom of the χ^2 distribution need to be estimated by $\widehat{p}_X + \widehat{p}_Y - \widehat{p}$. We use BIC or HQIC since they give consistent estimators.

In order for the likelihood ratio test to give realistic results, H_0 must be a subspace of H_A. This will be the case when the order under H_0 is less than or equal to the orders under H_A. That is, when $p \le \min(p_X, p_Y)$. In practice, if the algorithm incorrectly estimates the orders, it is possible that this will not be true even when H_0 is true. This could occur if one or both of the orders under H_A are underestimated, or if the common order under H_0 is overestimated. This will produce a test statistic that is too small and potentially negative. The simplest way to get around this is to ensure that $\widehat{p} \le \min(\widehat{p}_X, \widehat{p}_Y)$.

3.5 Simulations

The first simulation study shown in this section examines the distribution of the test statistic when the autoregressive orders are known. The test was applied to time series which were simulated from either the AR(1) processes

$$X_t + 0.5X_{t-1} = \varepsilon_t \qquad \text{and} \qquad Y_t + 0.5Y_{t-1} = u_t,$$

the AR(2) processes

$$X_t + 0.5X_{t-1} + 0.5X_{t-2} = \varepsilon_t \qquad \text{and} \qquad Y_t + 0.5Y_{t-1} + 0.5Y_{t-2} = u_t,$$

or the AR(3) processes

$$X_t + 0.5X_{t-1} + 0.5X_{t-2} + 0.5X_{t-3} = \varepsilon_t \qquad \text{and} \qquad Y_t + 0.5Y_{t-1} + 0.5Y_{t-2} + 0.5Y_{t-3} = u_t.$$

Table 3.1: Distributions of $\{\varepsilon_t\}$ and $\{u_t\}$, where $N\left(0, \sigma^2\right)$ denotes the normal distribution with mean zero and variance σ^2, $\text{Exp}(\mu)$ denotes the exponential distribution with mean μ and $t\,(v)$ denotes the t-distribution with degrees of freedom v.

Innovations	i	ii	iii	iv	v	vi
$\{\varepsilon_t\}$	$N\,(0,1)$	Exp(1)	$t\,(4)$	$N\,(0,1)$	$N\,(0,1)$	Exp(1)
$\{u_t\}$	$N\,(0,4)$	Exp(2)	$t\,(4)$	Exp(2)	$t\,(4)$	$t\,(4)$

Table 3.2: Summary of test statistics from simulations when p is known for the different distributions of the innovations (inn.) given by Table 3.1.

	$p=1$			$p=2$			$p=3$		
Inn.	Mean	Var	Type I	Mean	Var	Type I	Mean	Var	Type I
i	0.979	1.894	0.047	1.974	3.948	0.051	2.989	5.944	0.048
ii	1.009	2.010	0.050	2.002	4.100	0.051	2.962	5.904	0.049
iii	1.010	2.032	0.052	1.995	4.022	0.051	2.995	6.123	0.051
iv	1.031	2.095	0.054	2.006	3.878	0.048	3.012	5.899	0.050
v	1.003	1.981	0.050	1.983	3.960	0.050	3.000	6.005	0.050
vi	0.981	1.971	0.050	1.978	3.928	0.048	3.000	6.205	0.052

For each process, six sets of 10,000 pairs of time series were generated, where $\{\varepsilon_t\}$ and $\{u_t\}$ were simulated from the distributions given in Table 3.1. The sample sizes were $T_1 = 1,000$ and $T_2 = 2,000$. Table 3.2 gives the means and variances of the resulting test statistics. Also shown are the Type I error rates at the 0.05 significance level, that is, the proportion of test statistics which were greater than the 95th percentile of the χ^2 distribution with p degrees of freedom where p is the true order.

As shown in Table 3.2, in each set of simulations, the means and variances are consistent with the test statistics following the χ^2 distribution with p degrees of freedom. That is, the means are close to p and the variances are close to $2p$. The Type I error rates are close to 0.05 in each case. These results agree with the assertion that the test statistic will follow a χ^2 distribution with p degrees of freedom when the time series have come from autoregressions of order p. The results also suggest that the test is robust to non-Gaussianity and unequal sample sizes.

The test was then run over simulations from the same autoregressive processes as before, but this time p was unknown and was estimated using BIC with $K = 8$. Table 3.3 gives the

Table 3.3: Summary of test statistics from simulations when p is unknown for the different distributions of the innovations (inn.) given by Table 3.1.

	$p = 1$			$p = 2$			$p = 3$		
Inn.	Mean	Var	Type I	Mean	Var	Type I	Mean	Var	Type I
i	1.111	3.375	0.061	2.152	5.530	0.061	3.140	7.260	0.063
ii	1.145	3.599	0.063	2.093	4.826	0.057	3.202	7.800	0.069
iii	1.144	3.889	0.062	2.125	5.514	0.061	3.085	7.481	0.057
iv	1.157	3.557	0.067	2.122	5.304	0.063	3.064	6.800	0.055
v	1.110	3.202	0.061	2.105	5.232	0.061	3.063	7.289	0.058
vi	1.169	3.935	0.064	2.127	5.341	0.060	3.113	7.383	0.062

means, variances and Type I error rates of the resulting test statistics.

As shown in Table 3.3, the means, variances and Type I error rates are higher than their theoretical values. These were inflated due to a small number of incorrect order estimates. In each case approximately 1.5% of simulations had at least one order estimated incorrectly.

The next simulation study examines the power of the test. It will be interesting at this stage to consider a case where the processes are not autoregressions. Thus, along with simulating from autoregressions of varying order, we also simulated from a moving average process of order 1. We are not expecting the power in this case to be very high, since we are not estimating the true order of an autoregression.

The test was applied to four sets of 10,000 pairs of time series. The first set of time series were simulated from the AR(1) processes

$$X_t + 0.5X_{t-1} = \varepsilon_t \qquad \text{and} \qquad Y_t + \beta Y_{t-1} = u_t.$$

The second set of time series were simulated from the AR(2) processes

$$X_t + 0.5X_{t-1} + 0.5X_{t-2} = \varepsilon_t \qquad \text{and} \qquad Y_t + 0.5Y_{t-1} + \beta Y_{t-2} = u_t.$$

The third set of time series were simulated from the AR(3) processes

$$X_t + 0.5X_{t-1} + 0.5X_{t-2} + 0.5X_{t-3} = \varepsilon_t \qquad \text{and} \qquad Y_t + 0.5Y_{t-1} + 0.5Y_{t-2} + \beta Y_{t-3} = u_t.$$

The fourth set of time series were simulated the MA(1) processes

$$X_t = \varepsilon_t + 0.5\varepsilon_{t-1} \qquad \text{and} \qquad Y_t = u_t + \beta u_{t-1}.$$

The four null spectral densities are described in Figure 3.1. In each set, $T_1 = T_2 = 1,000$ and $\{\varepsilon_t\}$ and $\{u_t\}$ were simulated from normal distributions with variances 1 and 4, respectively.

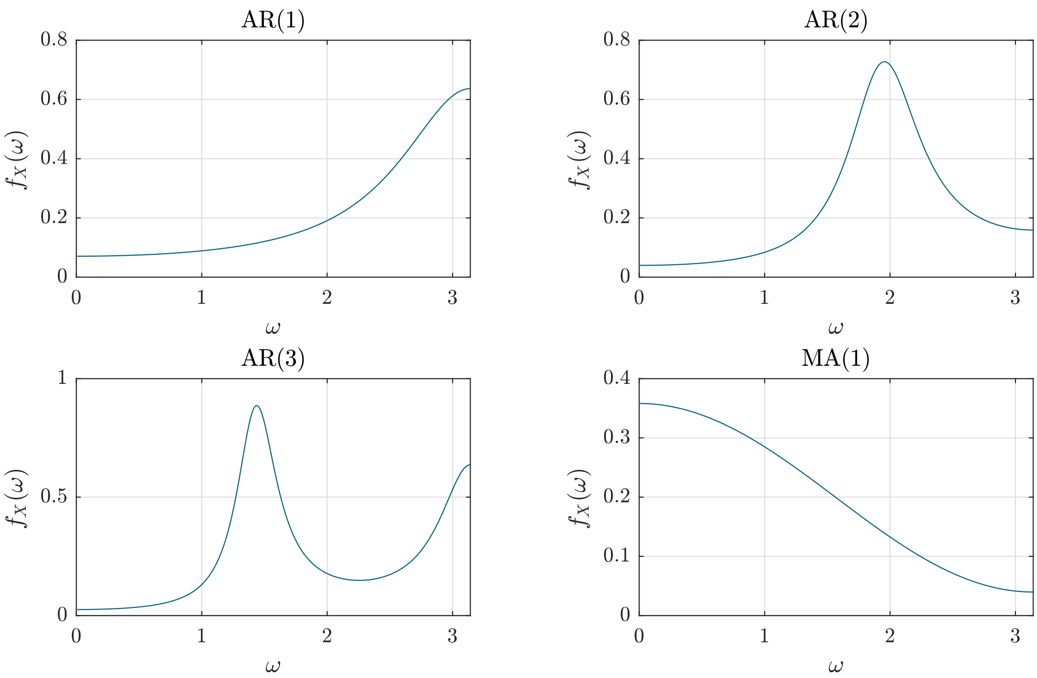

Figure 3.1: Spectral densities for the four $\{X_t\}$ processes described in Section 3.5.

The orders were estimated using BIC with $K = 8$. Figure 3.2 shows the empirical power, that is the proportion of times the test rejected H_0 at the 5% significance level, for $\beta = 0.01, 0.02, \ldots, 0.99$.

The plots in Figure 3.2 show that the test performs well when the time series are generated by autoregressions. The Type I error rates, that is when $\beta = 0.5$, are slightly above 0.05. This is expected and is due to a small number of incorrect order estimates. When the time series are generated by moving average processes, the Type I error rate is close to 0.6. There is clearly a problem when the time series have not come from autoregressions. In the next section we propose a modification to the test to correct this.

3.6 Fixed Order Autoregressive Approximation

Using long-order autoregressions to estimate the parameters in autoregressive-moving average (ARMA) models has a long history (see, for example, Durbin, 1959, Durbin, 1960, and Hannan and Rissanen, 1982). However, it was shown in the previous section that the size of the parametric test is incorrect, most likely because the autoregressive order is being estimated. The proposed solution is to fix the autoregressive order, under both H_0 and H_A, at some sufficiently large number $n = n_{T_1,T_2}$, where n is of lower order than $\min(T_1, T_2)$. The test then rejects H_0 at significance level α when the test statistic is greater than the $100\,(1-\alpha)$th

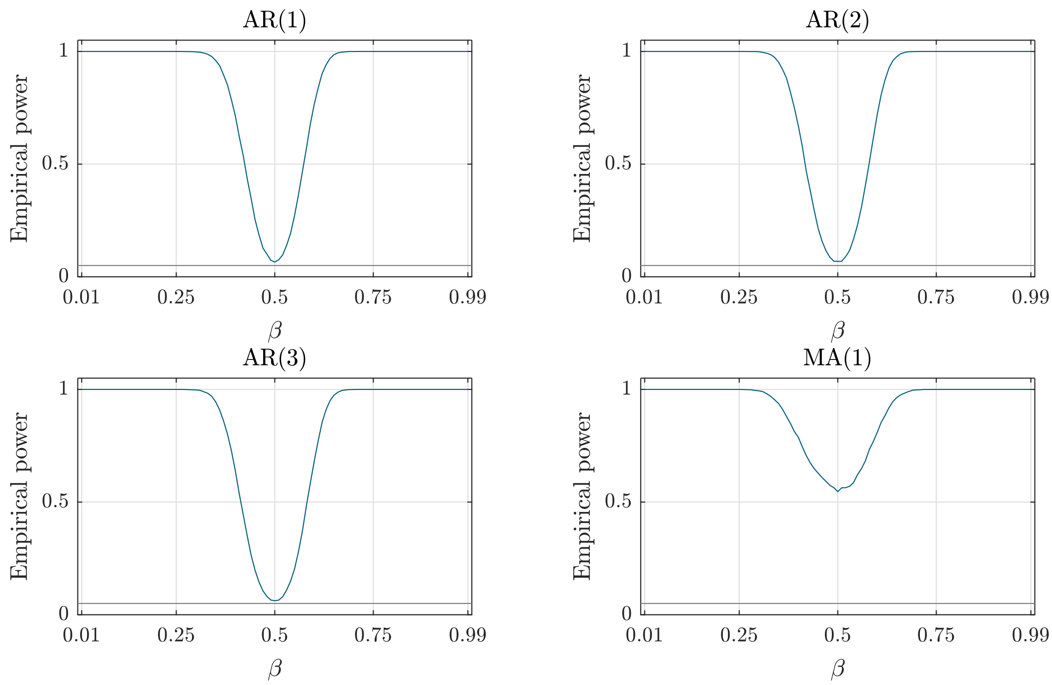

Figure 3.2: Empirical power of the likelihood ratio test when the orders were estimated using BIC. The lowest horizontal line in each plot indicates the significance level of 0.05.

percentile of the χ^2 distribution with n degrees of freedom. By fitting autoregressions of sufficiently large order, the dependence structure of the underlying processes should be captured regardless of their true nature. There will of course be a reduction in the power of the test when the processes really are autoregressions.

An et al. (1982) discuss fitting autoregressions to approximate more general processes by increasing the order of the autoregression with the sample size. Following their Theorem 6, we set $n = \lfloor (\log T_{\min})^v \rfloor$, where $T_{\min} = \min(T_1, T_2)$ and $v > 1$.

To examine the behaviour of the fixed order method for the ARMA case, the test was applied to 10,000 pairs of time series which were simulated from the MA(1) processes

$$X_t = \varepsilon_t + \beta\varepsilon_{t-1} \qquad \text{and} \qquad Y_t = u_t + \beta u_{t-1},$$

as well as 10,000 pairs of time series which were simulated from the MA(2) processes

$$X_t = \varepsilon_t + 0.5\varepsilon_{t-1} + \beta\varepsilon_{t-2} \qquad \text{and} \qquad Y_t = u_t + 0.5u_{t-1} + \beta u_{t-2}$$

for different sample sizes and for $\beta = 0.05, 0.10, \ldots, 0.95$. In each case, $\{\varepsilon_t\}$ and $\{u_t\}$ were simulated from normal distributions with variances 1 and 4, respectively, $T_1 = T_2 = T$ and $v = 1.1$. Figure 3.3 shows the Type I error rates for varying values of β and for different sample sizes. For the MA(2) case, the Type I error rates are plotted only for $\beta = -0.45, \ldots, 0.95$, since this is the range in which the processes are invertible.

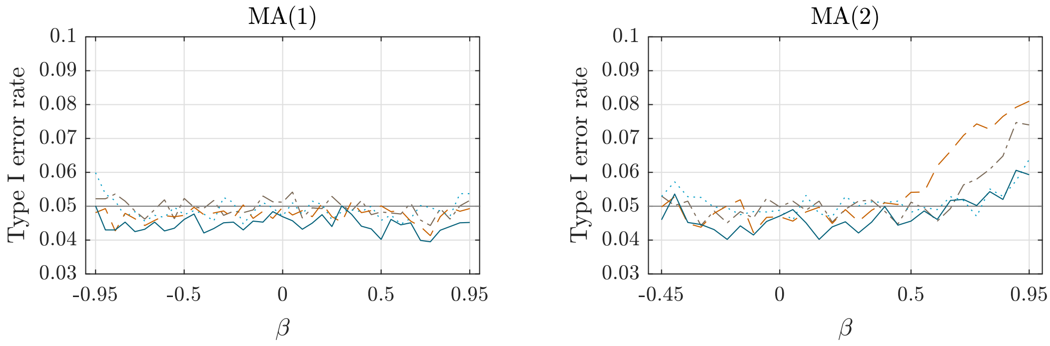

Figure 3.3: Type I error rates using the fixed order method when the time series are from MA(1) and MA(2) processes with sample size $T = 125$ (solid), $T = 500$ (dashes), $T = 1,000$ (small dashes) and $T = 2,000$ (dot-dash).

The Type I error rates are now close to 0.05, which is a significant improvement over the results shown in Figure 3.2. At either end of the plots, where the roots of the auxiliary equations are close to the unit circle, the Type I error rates spike up above the 0.05 line. By increasing v we would reduce the Type I error rate in these cases, however this may also reduce the power of the test. In the simulations which follow we set $v = 1.1$. As long as the roots of the auxiliary equations are not too close to the unit circle, this should retain an acceptable Type I error rate without a loss of power.

To compare the power of the fixed order method with the previous method which estimated the orders using BIC, both tests were applied to 10,000 pairs of time series which were simulated from the same four processes as in Section 3.5. Figure 3.4 shows a comparison of the empirical powers for $\beta = 0.01, 0.02, \ldots, 0.99$.

The fixed order method performs much better for the moving average case. For the autoregression cases, the Type I error rates of the fixed order method are at or below 0.05. There is some loss of power in the autoregression cases, as expected.

3.7 Comparison With Nonparametric Tests

In this section we show the results of another power study which compares the parametric test using the fixed order method with the range test of Coates and Diggle (1986), the normalised cumulative periodogram test of Diggle and Fisher (1991) and the test for equal autocovariances of Lund et al. (2009). Since the nonparametric tests require equal sample sizes, we suppose that $T_1 = T_2 = T$.

The range test has test statistic

$$R_T = \max_{j=1,\ldots,q} \log \left\{ \frac{I_{T,X}(\omega_j)}{I_{T,Y}(\omega_j)} \right\} - \min_{j=1,\ldots,q} \log \left\{ \frac{I_{T,X}(\omega_j)}{I_{T,Y}(\omega_j)} \right\},$$

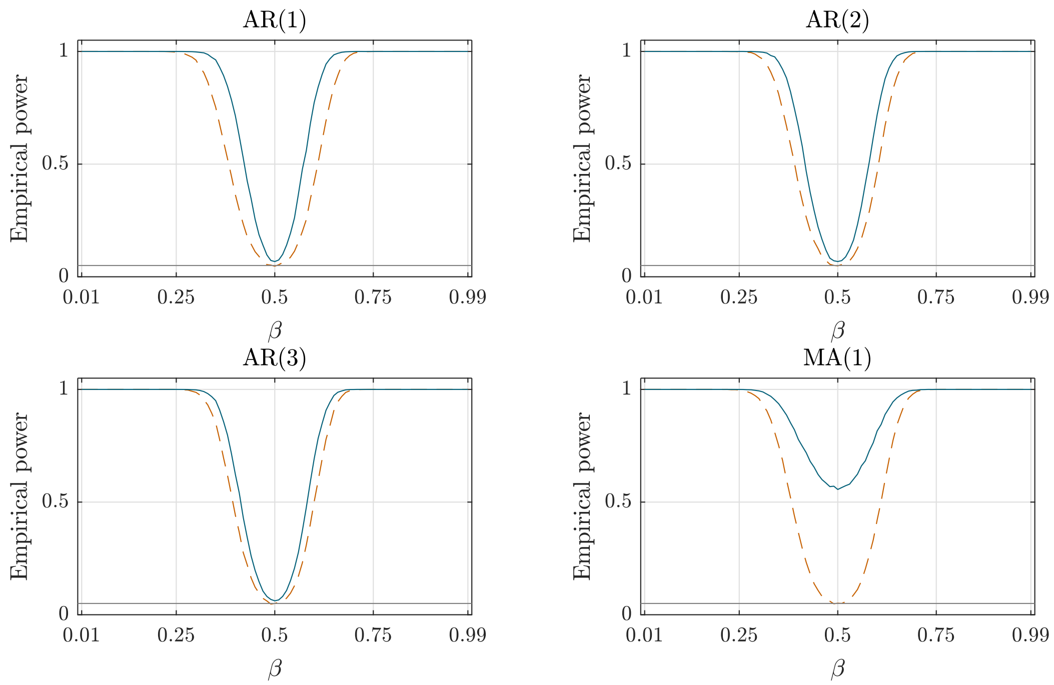

Figure 3.4: Empirical power of the likelihood ratio test when the orders were estimated using BIC (solid) and when using the fixed order method (dashes). The lowest horizontal line in each plot indicates the significance level of 0.05.

where $I_{T,X}(\omega_j)$ and $I_{T,Y}(\omega_j)$ are the periodograms of $\{X_t\}$ and $\{Y_t\}$, respectively, $\omega_j = 2\pi j/T$ and $q = \lfloor (T-1)/2 \rfloor$. The null hypothesis is rejected when R_T is large. Coates and Diggle (1986) expressed the distribution of R_T under H_0 in terms of an infinite sum and computed critical values for q up to 45. We can in fact compute an asymptotic distribution under i.i.d. Gaussian assumptions. Let $W_T = R_T - 2\log q$. Then, under H_0, it is shown in the Appendix that the cumulative distribution function (cdf) of W_T converges to

$$F_W(w) = 2e^{-w/2} K_1\left(2e^{-w/2}\right),$$

as $T \to \infty$, where $K_1(x)$ is the modified Bessel function of the second kind of order one. At the 0.05 significance level, the critical value for W_T is 4.4644, in agreement with the computations of Coates and Diggle (1986).

The normalised cumulative periodogram test has test statistic

$$D_T = \max_{j=1,\ldots,q} \left| F_X(\omega_j) - F_Y(\omega_j) \right|,$$

where

$$F_X(\omega_j) = \sum_{k=1}^{j} I_{T,X}(\omega_k) \Big/ \sum_{k=1}^{q} I_{T,X}(\omega_k), \qquad F_Y(\omega_j) = \sum_{k=1}^{j} I_{T,Y}(\omega_k) \Big/ \sum_{k=1}^{q} I_{T,Y}(\omega_k),$$

$\omega_j = 2\pi j/T$ and $q = \lfloor (T-1)/2 \rfloor$. The null hypothesis is rejected when D_T is large. However, the distribution of D_T is not known. For this power study, we define the critical value to be that which cuts off the largest 5% of simulated values of D_T when H_0 is true. That is, we fix the significance level at 0.05 which allows us to compare the power properties with other tests.

The test for equal autocovariances has null and alternative hypotheses

$$
\begin{aligned}
&H_0 : \gamma_X(i) = \gamma_Y(i), \ i = 0, \ldots, L, \\
&H_A : \exists\, i \in \{0, \ldots, L\} \text{ such that } \gamma_X(i) \neq \gamma_Y(i)
\end{aligned}
$$

where L is the number of autocovariances considered. The test is based on Bartlett's formula (Bartlett, 1946). Since we wish to test for differences in spectral shape, we begin by dividing the time series by their sample standard deviations. The sample autocovariances are

$$
\widehat{\gamma}_X(i) = T^{-1} \sum_{t=i}^{T-1} X_t X_{t-i} \qquad \text{and} \qquad \widehat{\gamma}_Y(i) = T^{-1} \sum_{t=i}^{T-1} Y_t Y_{t-i}.
$$

Let $\widehat{\gamma}(i) = \{\widehat{\gamma}_X(i) + \widehat{\gamma}_Y(i)\}/2$. Assuming that $\{\varepsilon_t\}$ and $\{u_t\}$ are Gaussian, the test statistic is

$$
C_T = \frac{T}{2} \triangle \widehat{\gamma}' \widehat{W}^{-1} \triangle \widehat{\gamma},
$$

where $\triangle\widehat{\gamma}$ is the $(L+1) \times 1$ vector with elements given by $\widehat{\gamma}_X(i) - \widehat{\gamma}_Y(i)$, $i = 0, \ldots, L$, and $\widehat{W}$ is the $(L+1) \times (L+1)$ matrix with (i,j)th element

$$
\sum_{k=-\lfloor T^{1/3} \rfloor}^{\lfloor T^{1/3} \rfloor} \{\widehat{\gamma}(k)\widehat{\gamma}(k-i+j) + \widehat{\gamma}(k+j)\widehat{\gamma}(k-i)\}.
$$

The null hypothesis is rejected at significance level α when C_T is greater than the $100(1-\alpha)$th percentile of the χ^2 distribution with $L+1$ degrees of freedom. We put $L = 10$ as suggested in Lund et al. (2009).

Each test was applied to four sets of 10,000 pairs of time series which were simulated from the same processes as in Section 3.5 with sample sizes of $T = 128$ and $T = 1,024$. Note that the parametric test used orders of 5 and 8, respectively, for these sample sizes. Figure 3.5 and Figure 3.6 compare the empirical powers for $\beta = 0.01, 0.02, \ldots, 0.99$.

As expected, the range test has very low empirical power in all cases. For the AR(1) and AR(2) cases, the empirical powers of the parametric and other nonparametric tests are very close. For the AR(3) case, the parametric test has higher empirical power than the normalised cumulative periodogram test. The test for equal autocovariances has Type I error rates of 0.17 and 0.11 for the AR(3) case. As expected, this test does not perform as well for higher order processes, since only a finite number of autocovariances are considered. For the MA(1) case, the parametric test has the highest empirical power.

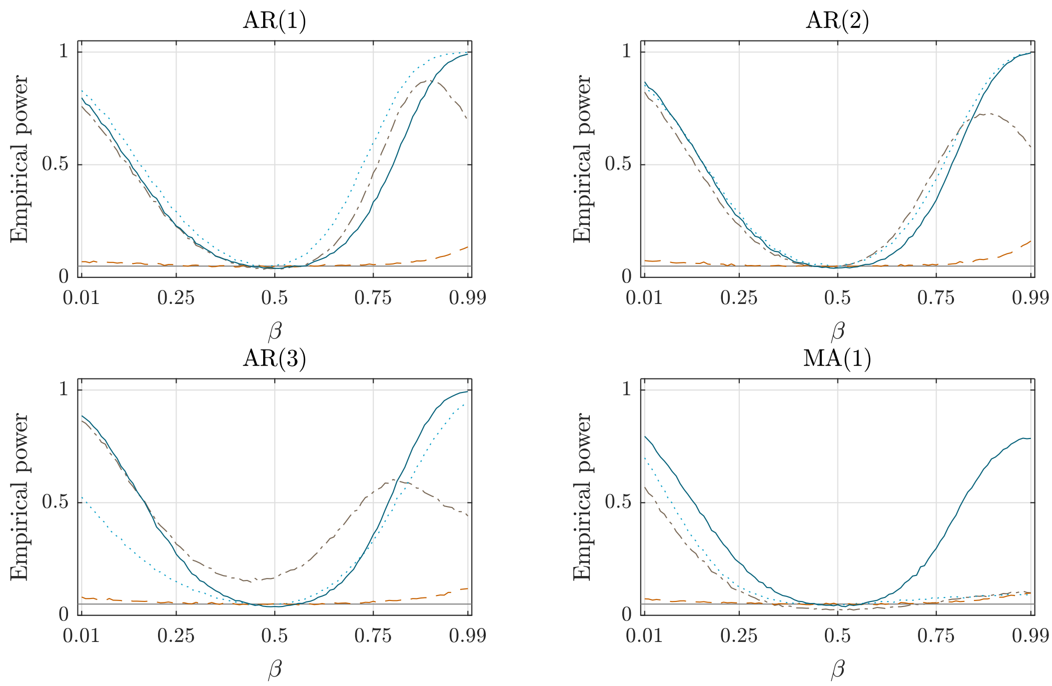

Figure 3.5: Comparison of empirical power of the parametric test (solid), range test (dashes), normalised cumulative periodogram test (small dashes), and test for equal autocovariances (dot-dash), for $T = 128$. The lowest horizontal line in each plot indicates the significance level of 0.05.

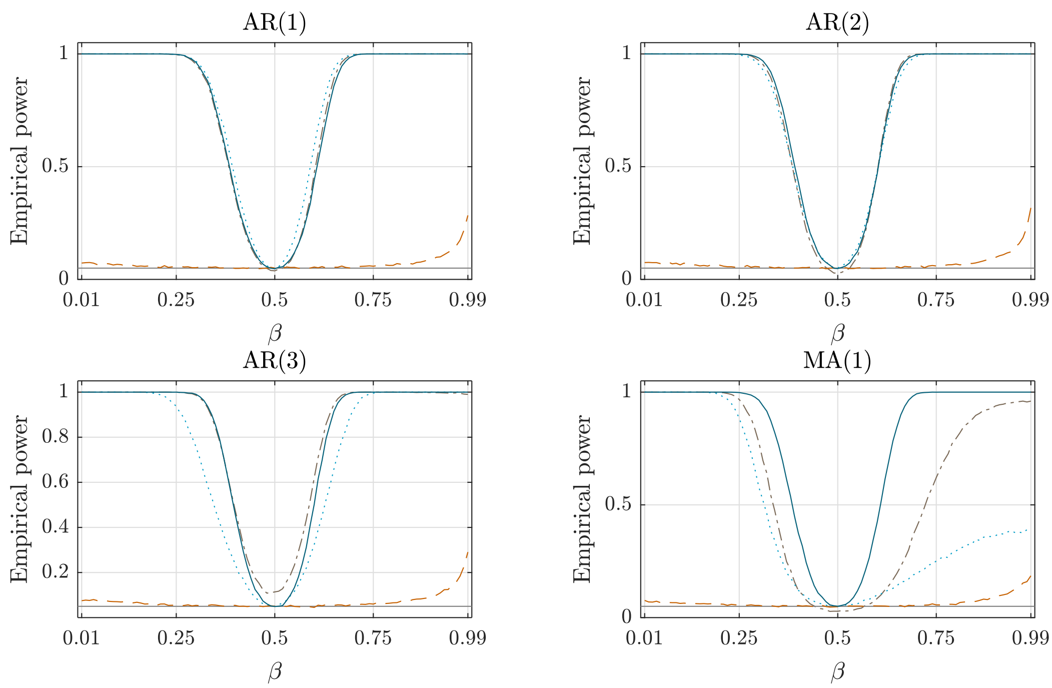

Figure 3.6: Comparison of empirical power of the parametric test (solid), range test (dashes), normalised cumulative periodogram test (small dashes) and test for equal autocovariances (dot-dash), for $T = 1,024$. The lowest horizontal line in each plot indicates the significance level of 0.05.

3.8 Comparing More Than Two Time Series

In this section we extend the parametric test using the fixed order method for comparing more than two time series. Let $\{X_{k,t}\}$, $k = 1, \ldots, n$, be independent, univariate, stationary time series, assumed to have zero mean. We wish to test the hypothesis that the n spectral densities have the same shape. That is, that the ratio of any two of the spectral densities is constant.

We fit autoregressions of order $p = \lfloor (\log T_{\min})^v \rfloor$, where T_k is the sample size of $\{X_{k,t}\}$, $T_{\min} = \min(T_1, \ldots, T_n)$ and $v > 1$. That is, for $k = 1, \ldots, n$, we fit

$$X_{k,t} + \beta_{k,1} X_{k,t-1} + \cdots + \beta_{k,p} X_{k,t-p} = \varepsilon_{k,t}$$

for some parameters $\beta_k = \begin{bmatrix} \beta_{k,1} & \cdots & \beta_{k,p} \end{bmatrix}'$. As before, it is assumed that the $\{\varepsilon_{k,t}\}$ are independent sequences of martingale differences with

$$E\left(\varepsilon_{k,t}^2 \mid \mathcal{F}_{k,t-1}\right) = \sigma_k^2,$$

where $\mathcal{F}_{k,t}$ is the σ-field generated by $\{\varepsilon_{k,t}, \varepsilon_{k,t-1}, \ldots\}$.

The spectral density of $\{X_{k,t}\}$, $k = 1, \ldots, n$, is

$$f_{X_k}(\omega) = \frac{\sigma_k^2}{2\pi \left|1 + \sum_{j=1}^p \beta_{k,j} e^{-ij\omega}\right|^2}.$$

Thus, the spectral densities will have the same shape if and only if the autoregressive parameters are all equal for all k. The null hypothesis is then

$$H_0 : \beta_1 = \cdots = \beta_n$$

and the alternative hypothesis, H_A, is its complement.

The conditional Gaussian log-likelihood for $\{X_{k,t}\}$, $k = 1, \ldots, n$, is

$$l_k\left(\beta_k, \sigma_k^2\right) = -\frac{T_k}{2} \log\left(2\pi\sigma_k^2\right) - \frac{1}{2\sigma_k^2} \sum_{t=p}^{T_k - 1} \left\{b_{\beta_k}(z) X_{k,t}\right\}^2.$$

Under the alternative hypothesis, these can be maximised separately or, alternatively, the Levinson–Durbin algorithm can be used to compute estimators of β_k and σ_k^2. The maximised conditional Gaussian log-likelihood under H_A is then

$$\widehat{l}_A = -\frac{1}{2}\{1 + \log(2\pi)\} \sum_{j=1}^n T_j - \frac{1}{2} \sum_{j=1}^n T_j \log \widehat{\sigma}_{j;A}^2,$$

where $\widehat{\sigma}_{k;A}^2$ is whichever estimator is used of σ_k^2 under H_A, $k = 1, \ldots, n$.

Under H_0 the conditional Gaussian log-likelihood is

$$l_0\left(\beta,\sigma_1^2,\ldots,\sigma_k^2\right)=-\frac{1}{2}\sum_{j=1}^{n}\left[T_j\log\left(2\pi\sigma_j^2\right)-\frac{1}{\sigma_j^2}\sum_{t=p}^{T_k-1}\left\{b_\beta\left(z\right)X_{j,t}\right\}^2\right],$$

where $\beta=\left[\begin{array}{ccc}\beta_1 & \cdots & \beta_p\end{array}\right]'$ is the common parameter vector. Let C_n be the $p\times p$ matrix with (i,j)th element

$$\left(\sum_{m=1}^{n}T_m\right)^{-1}\sum_{k=1}^{n}\frac{1}{\sigma_k^2}\sum_{t=p}^{T_k-1}X_{k,t-i}X_{k,t-j}$$

and let c_n be the $p\times 1$ vector with ith element

$$\left(\sum_{j=1}^{n}T_j\right)^{-1}\sum_{k=1}^{n}\frac{1}{\sigma_k^2}\sum_{t=p}^{T_k-1}X_{k,t}X_{k,t-i}.$$

Then $l_0\left(\beta,\sigma_1^2,\ldots,\sigma_k^2\right)$ is maximised with respect to β, for fixed $\sigma_1^2,\ldots,\sigma_n^2$, by

$$\widehat{\beta}_{\sigma_1^2,\ldots,\sigma_n^2}=-C_n^{-1}c_n,$$

and with respect to σ_k^2, for fixed β,by

$$\widetilde{\sigma}_{k;\beta}^2=T_k^{-1}\sum_{t=p}^{T_k-1}X_{k,t}^2+2c_n'\beta+\beta'C_n\beta,$$

$k=1,\ldots,n$. The maximised conditional Gaussian log-likelihood under H_0 is then

$$\widehat{l}_0=-\frac{1}{2}\left\{1+\log\left(2\pi\right)\right\}\sum_{j=1}^{n}T_j-\frac{1}{2}\sum_{j=1}^{n}T_j\log\widehat{\sigma}_{j;0}^2,$$

where $\widehat{\sigma}_{k;0}^2$ is the estimator of σ_k^2 under H_0, $k=1,\ldots n$. The parameters can therefore be estimated iteratively as follows. Given estimates for σ_k^2, $k=1,\ldots,n$, estimate β by

$$\widehat{\beta}=\widehat{\beta}_{\widehat{\sigma}_{1;0}^2,\ldots,\widehat{\sigma}_{n;0}^2}.$$

Then update the estimates of σ_k^2, $k=1,\ldots,n$, by

$$\widehat{\sigma}_{k;0}^2=\widetilde{\sigma}_{k;\widehat{\beta}}^2$$

and repeat until convergence. In practice, we use $\widehat{\sigma}_{k;A}^2$ as an initial estimate of $\widehat{\sigma}_{k;0}^2$, $k=1,\ldots,n$.

The test statistic is

$$\begin{aligned}\Lambda&=2\left(\widehat{l}_A-\widehat{l}_0\right)\\&=\sum_{j=1}^{n}T_j\log\left\{\frac{\widehat{\sigma}_{j;0}^2}{\widehat{\sigma}_{j;A}^2}\right\}\end{aligned}$$

and H_0 is rejected at significance level α when Λ is greater than the $100\left(1-\alpha\right)$th percentile of the χ^2 distribution with $p\left(n-1\right)$ degrees of freedom.

3.9 Appendix

In what follows, where convergence is indicated, it will mean convergence in the almost sure sense, unless otherwise indicated.

3.9.1 Proof of Theorem 3.1

For a given β, the values of σ_ε^2 and σ_u^2 that maximise $l_0(\theta)$ are given by

$$\widehat{\sigma}_\varepsilon^2(\beta) = T_1^{-1}\sum_{t=p}^{T_1-1}\{b_\beta(z)X_t\}^2 \qquad \text{and} \qquad \widehat{\sigma}_u^2(\beta) = T_2^{-1}\sum_{t=p}^{T_2-1}\{b_\beta(z)Y_t\}^2,$$

respectively. Let

$$\begin{aligned} S(\beta) &= l_0\left\{\beta, \widehat{\sigma}_\varepsilon^2(\beta), \widehat{\sigma}_u^2(\beta)\right\} \\ &= -\frac{T_1}{2}\log\left\{2\pi\widehat{\sigma}_\varepsilon^2(\beta)\right\} - \frac{T_2}{2}\log\left\{2\pi\widehat{\sigma}_u^2(\beta)\right\} - \frac{1}{2\widehat{\sigma}_\varepsilon^2(\beta)}\sum_{t=p}^{T_1-1}\{b_\beta(z)X_t\}^2 \\ &\quad - \frac{1}{2\widehat{\sigma}_u^2(\beta)}\sum_{t=p}^{T_2-1}\{b_\beta(z)Y_t\}^2 \\ &= -\frac{T_1}{2}\log\left\{2\pi\widehat{\sigma}_\varepsilon^2(\beta)\right\} - \frac{T_2}{2}\log\left\{2\pi\widehat{\sigma}_u^2(\beta)\right\} - \frac{T_1+T_2}{2}. \end{aligned}$$

Then

$$S(\beta) - S(\beta_0) = -\frac{T_1}{2}\log\left\{\frac{\widehat{\sigma}_\varepsilon^2(\beta)}{\widehat{\sigma}_\varepsilon^2(\beta_0)}\right\} - \frac{T_2}{2}\log\left\{\frac{\widehat{\sigma}_u^2(\beta)}{\widehat{\sigma}_u^2(\beta_0)}\right\}.$$

But

$$\begin{aligned} b_\beta(z)X_t &= X_t + \sum_{j=1}^{p}\beta_j X_{t-j} \\ &= X_t + \sum_{j=1}^{p}\beta_{0,j}X_{t-j} + \sum_{j=1}^{p}(\beta_j - \beta_{0,j})X_{t-j} \\ &= \varepsilon_t + \sum_{j=1}^{p}(\beta_j - \beta_{0,j})X_{t-j}. \end{aligned}$$

Thus

$$\{b_\beta(z)X_t\}^2 = \varepsilon_t^2 + 2\varepsilon_t\sum_{j=1}^{p}(\beta_j-\beta_{0,j})X_{t-j} + \left\{\sum_{j=1}^{p}(\beta_j-\beta_{0,j})X_{t-j}\right\}^2$$

and so

$$\widehat{\sigma}_\varepsilon^2(\beta) = T_1^{-1}\sum_{t=p}^{T_1-1}\varepsilon_t^2 + 2T_1^{-1}\sum_{t=p}^{T_1-1}\varepsilon_t\sum_{j=1}^{p}(\beta_j-\beta_{0,j})X_{t-j} + T_1^{-1}\sum_{t=p}^{T_1-1}\left\{\sum_{j=1}^{p}(\beta_j-\beta_{0,j})X_{t-j}\right\}^2.$$

Now,

$$\begin{aligned}
&\frac{\widehat{\sigma}_{\varepsilon}^{2}(\beta)}{\widehat{\sigma}_{\varepsilon}^{2}(\beta_0)} \\
&= 1 + \frac{\widehat{\sigma}_{\varepsilon}^{2}(\beta) - \widehat{\sigma}_{\varepsilon}^{2}(\beta_0)}{\widehat{\sigma}_{\varepsilon}^{2}(\beta_0)} \\
&= 1 + \frac{2T_1^{-1}\sum_{t=p}^{T_1-1}\varepsilon_t\sum_{j=1}^{p}(\beta_j - \beta_{0,j})X_{t-j} + T_1^{-1}\sum_{t=p}^{T_1-1}\left\{\sum_{j=1}^{p}(\beta_j - \beta_{0,j})X_{t-j}\right\}^2}{\widehat{\sigma}_{\varepsilon}^{2}(\beta_0)} \\
&= 1 + \frac{2T_1^{-1}\sum_{j=1}^{p}(\beta_j - \beta_{0,j})\sum_{t=p}^{T_1-1}\varepsilon_t X_{t-j} + T_1^{-1}\sum_{t=p}^{T_1-1}\left\{\sum_{j=1}^{p}(\beta_j - \beta_{0,j})X_{t-j}\right\}^2}{\widehat{\sigma}_{\varepsilon}^{2}(\beta_0)}.
\end{aligned}$$

However, as $T_1 \to \infty$, by the strong law of large numbers,

$$\widehat{\sigma}_{\varepsilon}^{2}(\beta_0) = T_1^{-1}\sum_{t=p}^{T_1-1}\varepsilon_t^2 \to \sigma_{\varepsilon 0}^2$$

and

$$T_1^{-1}\sum_{t=p}^{T_1-1}\varepsilon_t X_{t-j} \to E(\varepsilon_t X_{t-j}) = 0,\ j = 1, \ldots, p, \tag{3.2}$$

since

$$E(\varepsilon_t X_{t-j}) = E\{E(\varepsilon_t X_{t-j} \mid \mathcal{F}_{t-1})\} = E\{X_{t-j}E(\varepsilon_t \mid \mathcal{F}_{t-1})\} = 0.$$

Also, as $T_1 \to \infty$,

$$\begin{aligned}
T_1^{-1}\sum_{t=p}^{T_1-1}\left\{\sum_{j=1}^{p}(\beta_j - \beta_{0,j})X_{t-j}\right\}^2 &= (\beta - \beta_0)' C_X (\beta - \beta_0) \\
&\to (\beta - \beta_0)' \Gamma_X (\beta - \beta_0).
\end{aligned}$$

Thus

$$\frac{\widehat{\sigma}_{\varepsilon}^{2}(\beta)}{\widehat{\sigma}_{\varepsilon}^{2}(\beta_0)} \to 1 + (\beta - \beta_0)' \Omega (\beta - \beta_0) \geqslant 1 + |\beta - \beta_0|^2 \omega_{\min},$$

where $\omega_{\min}$ is the smallest eigenvalue of Ω. Let $\delta > 0$. If $|\beta - \beta_0| \geqslant \delta$ then

$$\lim \frac{\widehat{\sigma}_{\varepsilon}^{2}(\beta)}{\widehat{\sigma}_{\varepsilon}^{2}(\beta_0)} \geqslant 1 + \delta^2\omega_{\min} > 1.$$

Therefore

$$\lim_{T_1\to\infty} -\frac{T_1}{2}\log\left\{\frac{\widehat{\sigma}_{\varepsilon}^{2}(\beta)}{\widehat{\sigma}_{\varepsilon}^{2}(\beta_0)}\right\} = -\infty.$$

Similarly,

$$\lim_{T_2\to\infty} -\frac{T_2}{2}\log\left\{\frac{\widehat{\sigma}_{u}^{2}(\beta)}{\widehat{\sigma}_{u}^{2}(\beta_0)}\right\} = -\infty,$$

and so

$$\lim_{T_1, T_2\to\infty} S(\beta) - S(\beta_0) = -\infty.$$

Consequently,

$$\liminf_{T_1,T_2\to\infty} \inf_{|\beta-\beta_0|\geqslant\delta} \{S(\beta_0)-S(\beta)\}>0$$

and the sufficiency condition of Theorem 2.5 is met. Thus $\widehat{\beta}\to\beta_0$.

Now,

$$\widehat{\sigma}_\varepsilon^2\left(\widehat{\beta}\right)-\sigma_{\varepsilon 0}^2=\left\{\widehat{\sigma}_\varepsilon^2\left(\widehat{\beta}\right)-\widehat{\sigma}_\varepsilon^2\left(\beta_0\right)\right\}+\left\{\widehat{\sigma}_\varepsilon^2\left(\beta_0\right)-\sigma_{\varepsilon 0}^2\right\}.$$

Also,

$$\begin{aligned}\widehat{\sigma}_\varepsilon^2\left(\widehat{\beta}\right)-\widehat{\sigma}_\varepsilon^2\left(\beta_0\right)&=2T_1^{-1}\sum_{t=p}^{T_1-1}\varepsilon_t\sum_{j=1}^{p}\left(\widehat{\beta}_j-\beta_{0,j}\right)X_{t-j}+T_1^{-1}\sum_{t=p}^{T_1-1}\left\{\sum_{j=1}^{p}\left(\widehat{\beta}_j-\beta_{0,j}\right)X_{t-j}\right\}^2\\&=2T_1^{-1}\left(\widehat{\beta}-\beta_0\right)'\sum_{t=p}^{T_1-1}\varepsilon_t\begin{bmatrix}X_{t-1}\\ \vdots\\ X_{t-p}\end{bmatrix}+\left(\widehat{\beta}-\beta_0\right)'C_X\left(\widehat{\beta}-\beta_0\right)\\&\to 0,\end{aligned}$$

from (3.2) and since

$$0\leqslant\left(\widehat{\beta}-\beta_0\right)'C_X\left(\widehat{\beta}-\beta_0\right)\leqslant\left(\widehat{\beta}-\beta_0\right)'\left(\widehat{\beta}-\beta_0\right)E_{\max},$$

where $E_{\max}$ is the largest eigenvalue of C_X which converges to the largest eigenvalue of $\Gamma_X<\infty$. Hence

$$\widehat{\sigma}_\varepsilon^2\left(\widehat{\beta}\right)-\sigma_{\varepsilon 0}^2\to 0.$$

Similarly,

$$\widehat{\sigma}_u^2\left(\widehat{\beta}\right)-\sigma_{u0}^2\to 0.$$

Therefore $\widehat{\sigma}_\varepsilon^2\to\sigma_{\varepsilon 0}^2$ and $\widehat{\sigma}_u^2\to\sigma_{u0}^2$.

3.9.2 Proof of Lemma 3.1

The characteristic function of Z_{T_1,T_2} is, letting ϕ_U denote the characteristic function of a random variable U,

$$\begin{aligned}
\phi_{Z_{T_1,T_2}}(s) &= E\left[\exp\left\{is\left(\sqrt{\frac{T_1}{T_1+T_2}}\xi_{T_1}+\sqrt{\frac{T_2}{T_1+T_2}}\zeta_{T_2}\right)\right\}\right] \\
&= E\left\{\exp\left(is\sqrt{\frac{T_1}{T_1+T_2}}\xi_{T_1}\right)\right\}E\left\{\exp\left(is\sqrt{\frac{T_2}{T_1+T_2}}\zeta_{T_2}\right)\right\} \\
&= \phi_{\xi_{T_1}}\left(s\sqrt{\frac{T_1}{T_1+T_2}}\right)\phi_{\zeta_{T_2}}\left(s\sqrt{\frac{T_2}{T_1+T_2}}\right) \\
&= \left\{\phi_{\xi_{T_1}}\left(s\sqrt{\frac{T_1}{T_1+T_2}}\right)-\exp\left(-\frac{1}{2}s^2\frac{T_1}{T_1+T_2}\right)+\exp\left(-\frac{1}{2}s^2\frac{T_1}{T_1+T_2}\right)\right\} \\
&\quad\times\left\{\phi_{\zeta_{T_2}}\left(s\sqrt{\frac{T_2}{T_1+T_2}}\right)-\exp\left(-\frac{1}{2}s^2\frac{T_2}{T_1+T_2}\right)+\exp\left(-\frac{1}{2}s^2\frac{T_2}{T_1+T_2}\right)\right\} \\
&= \left\{\phi_{\xi_{T_1}}\left(s\sqrt{\frac{T_1}{T_1+T_2}}\right)-\exp\left(-\frac{1}{2}s^2\frac{T_1}{T_1+T_2}\right)\right\} \\
&\quad\times\left\{\phi_{\zeta_{T_2}}\left(s\sqrt{\frac{T_2}{T_1+T_2}}\right)-\exp\left(-\frac{1}{2}s^2\frac{T_2}{T_1+T_2}\right)\right\} \\
&\quad+\exp\left(-\frac{1}{2}s^2\frac{T_1}{T_1+T_2}\right)\left\{\phi_{\zeta_{T_2}}\left(s\sqrt{\frac{T_2}{T_1+T_2}}\right)-\exp\left(-\frac{1}{2}s^2\frac{T_2}{T_1+T_2}\right)\right\} \\
&\quad+\exp\left(-\frac{1}{2}s^2\frac{T_2}{T_1+T_2}\right)\left\{\phi_{\xi_{T_1}}\left(s\sqrt{\frac{T_1}{T_1+T_2}}\right)-\exp\left(-\frac{1}{2}s^2\frac{T_1}{T_1+T_2}\right)\right\} \\
&\quad+\exp\left(-\frac{1}{2}s^2\right).
\end{aligned}$$

Now,

$$\begin{aligned}
\left|\phi_{\xi_{T_1}}\left(s\sqrt{\frac{T_1}{T_1+T_2}}\right)-\exp\left(-\frac{1}{2}s^2\frac{T_1}{T_1+T_2}\right)\right| &\leqslant \sup_{|t|<|s|}\left|\phi_{\xi_{T_1}}(t)-\exp\left(-\frac{1}{2}t^2\right)\right| \\
&\to 0,
\end{aligned}$$

as $T_1\to\infty$, as sequences of characteristic functions converge uniformly in any closed interval (see, for example, Lukacs, 1970, Corollary 1 to Theorem 3.6.1). Similarly,

$$\left|\phi_{\zeta_{T_2}}\left(s\sqrt{\frac{T_2}{T_1+T_2}}\right)-\exp\left(-\frac{1}{2}s^2\frac{T_2}{T_1+T_2}\right)\right|\to 0$$

as $T_2\to\infty$. Hence

$$\phi_{Z_{T_1,T_2}}(s)\to\exp\left(-\frac{1}{2}s^2\right)$$

and so the distribution of Z_{T_1,T_2} converges to the standard normal.

3.9.3 Proof of Theorem 3.2

From the mean value theorem, letting θ_j denote the jth component of θ, we have

$$0 = \frac{\partial l_0\left(\widehat{\theta}\right)}{\partial \theta_j} = \frac{\partial l_0\left(\theta_0\right)}{\partial \theta_j} + \frac{\partial^2 l_0\left(\widetilde{\theta}_j\right)}{\partial \theta_j \partial \theta'}\left(\widehat{\theta} - \theta_0\right),$$

where $\widetilde{\theta}_j$ lies on the line segment between θ_0 and $\widehat{\theta}$. From Theorem 3.1, $\widetilde{\theta}_j \to \theta_0$ as $T_1, T_2 \to \infty$ for all j, and so $\widehat{\theta} - \theta_0$ has the same asymptotic distribution as

$$-\left\{\frac{\partial^2 l_0\left(\theta_0\right)}{\partial \theta \partial \theta'}\right\}^{-1}\left\{\frac{\partial l_0\left(\theta_0\right)}{\partial \theta}\right\}. \tag{3.3}$$

The first derivatives of $l_0\left(\theta\right)$ are

$$\begin{aligned}
\frac{\partial l_0\left(\theta\right)}{\partial \beta} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{t=p}^{T_1-1} b_\beta\left(z\right) X_t \begin{bmatrix} X_{t-1} \\ \vdots \\ X_{t-p} \end{bmatrix} - \frac{1}{\sigma_u^2}\sum_{t=p}^{T_2-1} b_\beta\left(z\right) Y_t \begin{bmatrix} Y_{t-1} \\ \vdots \\ Y_{t-p} \end{bmatrix}, \\
\frac{\partial l_0\left(\theta\right)}{\partial \sigma_\varepsilon^2} &= -\frac{T_1}{2\sigma_\varepsilon^2} + \frac{1}{2\left(\sigma_\varepsilon^2\right)^2}\sum_{t=p}^{T_1-1}\left\{b_\beta\left(z\right) X_t\right\}^2, \\
\frac{\partial l_0\left(\theta\right)}{\partial \sigma_u^2} &= -\frac{T_2}{2\sigma_u^2} + \frac{1}{2\left(\sigma_u^2\right)^2}\sum_{t=p}^{T_2-1}\left\{b_\beta\left(z\right) Y_t\right\}^2.
\end{aligned}$$

The second derivatives of $l_0\left(\theta\right)$ are

$$\begin{aligned}
\frac{\partial^2 l_0\left(\theta\right)}{\partial \beta \partial \beta'} &= -\frac{T_1}{\sigma_\varepsilon^2} C_X - \frac{T_2}{\sigma_u^2} C_Y, \\
\frac{\partial^2 l_0\left(\theta\right)}{\left(\partial \sigma_\varepsilon^2\right)^2} &= \frac{T_1}{2\left(\sigma_\varepsilon^2\right)^2} - \frac{1}{\left(\sigma_\varepsilon^2\right)^3}\sum_{t=p}^{T_1-1}\left\{b_\beta\left(z\right) X_t\right\}^2, \\
\frac{\partial^2 l_0\left(\theta\right)}{\left(\partial \sigma_u^2\right)^2} &= \frac{T_2}{2\left(\sigma_u^2\right)^2} - \frac{1}{\left(\sigma_u^2\right)^3}\sum_{t=p}^{T_2-1}\left\{b_\beta\left(z\right) Y_t\right\}^2, \\
\frac{\partial^2 l_0\left(\theta\right)}{\partial \beta \partial \sigma_\varepsilon^2} &= \frac{1}{\left(\sigma_\varepsilon^2\right)^2}\sum_{t=p}^{T_1-1} b_\beta\left(z\right) X_t \begin{bmatrix} X_{t-1} \\ \vdots \\ X_{t-p} \end{bmatrix}, \\
\frac{\partial^2 l_0\left(\theta\right)}{\partial \beta \partial \sigma_u^2} &= \frac{1}{\left(\sigma_u^2\right)^2}\sum_{t=p}^{T_2-1} b_\beta\left(z\right) Y_t \begin{bmatrix} Y_{t-1} \\ \vdots \\ Y_{t-p} \end{bmatrix}, \\
\frac{\partial^2 l_0\left(\theta\right)}{\partial \sigma_\varepsilon^2 \partial \sigma_u^2} &= 0.
\end{aligned} \tag{3.4}$$

Consider

$$
\begin{aligned}
&a'\left(T_1+T_2\right)^{-1/2}\frac{\partial l_0\left(\theta_0\right)}{\partial\beta}+b_1T_1^{-1/2}\frac{\partial l_0\left(\theta_0\right)}{\partial\sigma_\varepsilon^2}+b_2T_2^{-1/2}\frac{\partial l_0\left(\theta_0\right)}{\partial\sigma_u^2}\\
&=\frac{T_1^{1/2}}{\left(T_1+T_2\right)^{1/2}}a'Z_{1,T_1}+\frac{T_2^{1/2}}{\left(T_1+T_2\right)^{1/2}}a'Z_{2,T_2}+b_1Z_{3,T_1}+b_2Z_{4,T_2},
\end{aligned}
$$

where $a=\left[\begin{array}{ccc} a_1 & \cdots & a_p \end{array}\right]'$, b_1 and b_2 are constants, and

$$
Z_{1,T_1,j}=-T_1^{-1/2}\frac{1}{\sigma_{\varepsilon0}^2}\sum_{t=p}^{T_1-1}\varepsilon_tX_{t-j},\qquad Z_{2,T_2,j}=-T_2^{-1/2}\frac{1}{\sigma_{u0}^2}\sum_{t=p}^{T_2-1}u_tY_{t-j},
$$

$$
Z_{3,T_1}=T_1^{-1/2}\frac{1}{2\left(\sigma_{\varepsilon0}^2\right)^2}\sum_{t=p}^{T_1-1}\left(\varepsilon_t^2-\sigma_{\varepsilon0}^2\right)\qquad Z_{4,T_2}=T_2^{-1/2}\frac{1}{2\left(\sigma_u^2\right)^2}\sum_{t=p}^{T_2-1}\left(u_t^2-\sigma_{u0}^2\right).
$$

Let

$$
V_t=-\frac{1}{\sigma_{\varepsilon0}^2}\varepsilon_t\sum_{j=1}^{p}a_jX_{t-j}+\frac{b_1}{2\left(\sigma_{\varepsilon0}^2\right)^2}\left(\varepsilon_t^2-\sigma_{\varepsilon0}^2\right).
$$

Then

$$
\begin{aligned}
E\left(V_t\mid\mathcal{F}_{t-1}\right)&=-\frac{1}{\sigma_{\varepsilon0}^2}E\left(\sum_{j=1}^{p}a_jX_{t-j}\mid\mathcal{F}_{t-1}\right)E\left(\varepsilon_t\mid\mathcal{F}_{t-1}\right)+\frac{b_1}{2\left(\sigma_{\varepsilon0}^2\right)^2}\left\{E\left(\varepsilon_t^2\mid\mathcal{F}_{t-1}\right)-\sigma_{\varepsilon0}^2\right\}\\
&=0
\end{aligned}
$$

and

$$
\begin{aligned}
E\left(V_t^2\right)&=E\left\{E\left(V_t^2\mid\mathcal{F}_{t-1}\right)\right\}\\
&=\frac{1}{\left(\sigma_{\varepsilon0}^2\right)^2}E\left[\left(\sum_{j=1}^{p}a_jX_{t-j}\right)^2E\left(\varepsilon_t^2\mid\mathcal{F}_{t-1}\right)\right]\\
&\quad-\frac{b_1}{\left(\sigma_{\varepsilon0}^2\right)^3}E\left\{\sum_{j=1}^{p}a_jX_{t-j}E\left(\varepsilon_t^3\mid\mathcal{F}_{t-1}\right)\right\}\\
&\quad+\frac{b_1^2}{4\left(\sigma_{\varepsilon0}^2\right)^4}E\left\{E\left(\varepsilon_t^4\mid\mathcal{F}_{t-1}\right)-2\sigma_{\varepsilon0}^2E\left(\varepsilon_t^2\mid\mathcal{F}_{t-1}\right)+\left(\sigma_{\varepsilon0}^2\right)^2\right\}\\
&=\frac{1}{\sigma_{\varepsilon0}^2}E\left(\sum_{j=1}^{p}a_j\sum_{k=1}^{p}a_kX_{t-j}X_{t-k}\right)+\frac{b_1^2}{4\left(\sigma_{\varepsilon0}^2\right)^4}\left\{E\left(\varepsilon_t^4\right)-\left(\sigma_{\varepsilon0}^2\right)^2\right\}\\
&=a'\Omega a+\frac{b_1^2}{4}\upsilon,
\end{aligned}
$$

where

$$
\upsilon=\frac{1}{\left(\sigma_{\varepsilon0}^2\right)^4}E\left(\varepsilon_t^4\right)-\frac{1}{\left(\sigma_{\varepsilon0}^2\right)^2}.
$$

Thus, by the martingale central limit theorem, $a'Z_{1,T_1} + b_1 Z_{3,T_1} = T_1^{-1/2}\sum_{t=p}^{T_1-1} V_t$ is asymptotically normal with variance $a'\Omega a + b_1^2\upsilon/4$. Similarly, $a'Z_{2,T_2} + b_2 Z_{4,T_2}$ is asymptotically normal with variance $a'\Omega a + b_2^2\eta/4$, where

$$\eta = \frac{1}{\left(\sigma_{u0}^2\right)^4} E\left(u_t^4\right) - \frac{1}{\left(\sigma_{u0}^2\right)^2}.$$

Note that if $\{\varepsilon_t\}$ and $\{u_t\}$ are Gaussian, $\upsilon = 2\left(\sigma_{\varepsilon 0}^2\right)^{-2}$ and $\eta = 2\left(\sigma_{u0}^2\right)^{-2}$. Since $a'Z_{1,T_1} + b_1 Z_{3,T_1}$ and $a'Z_{2,T_2} + b_2 Z_{4,T_2}$ are independent, then Z_{1,T_1}, Z_{2,T_2}, Z_{3,T_1} and Z_{4,T_2} are jointly asymptotically normal. Also, from Lemma 3.1,

$$\frac{T_1^{1/2}}{\left(T_1+T_2\right)^{1/2}} a'Z_{1,T_1} + \frac{T_2^{1/2}}{\left(T_1+T_2\right)^{1/2}} a'Z_{2,T_2}$$

is asymptotically normal with mean zero and variance $a'\Omega a$ as $T_1, T_2 \to \infty$. Thus

$$a'\left(T_1+T_2\right)^{-1/2}\frac{\partial l\left(\theta_0\right)}{\partial\beta} + b_1 T_1^{-1/2}\frac{\partial l\left(\theta_0\right)}{\partial\sigma_\varepsilon^2} + b_2 T_2^{-1/2}\frac{\partial l\left(\theta_0\right)}{\partial\sigma_u^2}$$

is asymptotically normal with mean zero and variance

$$\begin{bmatrix} a' & b_1 & b_2 \end{bmatrix} \jmath \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix},$$

where

$$\jmath = \begin{bmatrix} \Omega & 0 & 0 \\ 0 & \frac{\upsilon}{4} & 0 \\ 0 & 0 & \frac{\eta}{4} \end{bmatrix}.$$

Hence

$$\begin{bmatrix} \left(T_1+T_2\right)^{-1/2}\frac{\partial l_0(\theta_0)}{\partial\beta} \\ T_1^{-1/2}\frac{\partial l_0(\theta_0)}{\partial\sigma_\varepsilon^2} \\ T_2^{-1/2}\frac{\partial l_0(\theta_0)}{\partial\sigma_u^2} \end{bmatrix}$$

is asymptotically normal with mean zero and covariance matrix $\jmath$. Now, from (3.3) and (3.4), $\begin{bmatrix} \left(\widehat{\beta}-\beta_0\right)' & \widehat{\sigma}_\varepsilon^2 - \sigma_{\varepsilon 0}^2 & \widehat{\sigma}_u^2 - \sigma_{u0}^2 \end{bmatrix}'$ has the same asymptotic distribution as

$$-\begin{bmatrix} \frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\beta'} & \frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\sigma_\varepsilon^2} & \frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\sigma_u^2} \\ \frac{\partial^2 l_0(\theta_0)}{\partial\sigma_\varepsilon^2\partial\beta'} & \frac{\partial^2 l_0(\theta_0)}{(\partial\sigma_\varepsilon^2)^2} & 0 \\ \frac{\partial^2 l_0(\theta_0)}{\partial\sigma_u^2\partial\beta'} & 0 & \frac{\partial^2 l_0(\theta_0)}{(\partial\sigma_u^2)^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial l(\theta_0)}{\partial\beta} \\ \frac{\partial l(\theta_0)}{\partial\sigma_\varepsilon^2} \\ \frac{\partial l(\theta_0)}{\partial\sigma_\varepsilon^2} \end{bmatrix}.$$

Let

$$M_{T_1,T_2} = \begin{bmatrix} \left(T_1+T_2\right)^{-1/2} I_p & 0 & 0 \\ 0 & T_1^{-1/2} & 0 \\ 0 & 0 & T_2^{-1/2} \end{bmatrix}.$$

Then

$$M_{T_1,T_2}^{-1} = \begin{bmatrix} (T_1+T_2)^{1/2}\, I_p & 0 & 0 \\ 0 & T_1^{1/2} & 0 \\ 0 & 0 & T_2^{1/2} \end{bmatrix}$$

and so

$$M_{T_1,T_2}^{-1} \begin{bmatrix} \left(\widehat{\beta} - \beta_0\right)' & \widehat{\sigma}_\varepsilon^2 - \sigma_{\varepsilon 0}^2 & \widehat{\sigma}_u^2 - \sigma_{u0}^2 \end{bmatrix}'$$

has the same asymptotic distribution as

$$\begin{aligned} & - M_{T_1,T_2}^{-1} \begin{bmatrix} \frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\beta'} & \frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\sigma_\varepsilon^2} & \frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\sigma_u^2} \\ \frac{\partial^2 l_0(\theta_0)}{\partial\sigma_\varepsilon^2\partial\beta'} & \frac{\partial^2 l_0(\theta_0)}{(\partial\sigma_\varepsilon^2)^2} & 0 \\ \frac{\partial^2 l_0(\theta_0)}{\partial\sigma_u^2\partial\beta'} & 0 & \frac{\partial^2 l_0(\theta_0)}{(\partial\sigma_u^2)^2} \end{bmatrix}^{-1} M_{T_1,T_2}^{-1} M_{T_1,T_2} \begin{bmatrix} \frac{\partial l_0(\theta_0)}{\partial\beta} \\ \frac{\partial l_0(\theta_0)}{\partial\sigma_\varepsilon^2} \\ \frac{\partial l_0(\theta_0)}{\partial\sigma_\varepsilon^2} \end{bmatrix} \\ = & - \left(M_{T_1,T_2} \begin{bmatrix} \frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\beta'} & \frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\sigma_\varepsilon^2} & \frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\sigma_u^2} \\ \frac{\partial^2 l_0(\theta_0)}{\partial\sigma_\varepsilon^2\partial\beta'} & \frac{\partial^2 l_0(\theta_0)}{(\partial\sigma_\varepsilon^2)^2} & 0 \\ \frac{\partial^2 l_0(\theta_0)}{\partial\sigma_u^2\partial\beta'} & 0 & \frac{\partial^2 l_0(\theta_0)}{(\partial\sigma_u^2)^2} \end{bmatrix} M_{T_1,T_2} \right)^{-1} M_{T_1,T_2} \begin{bmatrix} \frac{\partial l_0(\theta_0)}{\partial\beta} \\ \frac{\partial l_0(\theta_0)}{\partial\sigma_\varepsilon^2} \\ \frac{\partial l_0(\theta_0)}{\partial\sigma_\varepsilon^2} \end{bmatrix}. \end{aligned}$$

That is,

$$\begin{bmatrix} (T_1+T_2)^{1/2}\left(\widehat{\beta} - \beta_0\right) \\ T_1^{1/2}\left(\widehat{\sigma}_\varepsilon^2 - \sigma_{\varepsilon 0}^2\right) \\ T_2^{1/2}\left(\widehat{\sigma}_u^2 - \sigma_{u0}^2\right) \end{bmatrix}$$

has the same asymptotic distribution as

$$\begin{aligned} & - \begin{bmatrix} (T_1+T_2)^{-1}\,\frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\beta'} & T_1^{-1/2}(T_1+T_2)^{-1/2}\,\frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\sigma_\varepsilon^2} & T_2^{-1/2}(T_1+T_2)^{-1/2}\,\frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\sigma_u^2} \\ T_1^{-1/2}(T_1+T_2)^{-1/2}\,\frac{\partial^2 l_0(\theta_0)}{\partial\sigma_\varepsilon^2\partial\beta'} & T_1^{-1}\frac{\partial^2 l_0(\theta_0)}{(\partial\sigma_\varepsilon^2)^2} & 0 \\ T_2^{-1/2}(T_1+T_2)^{-1/2}\,\frac{\partial^2 l_0(\theta_0)}{\partial\sigma_u^2\partial\beta'} & 0 & T_2^{-1}\frac{\partial^2 l_0(\theta_0)}{(\partial\sigma_u^2)^2} \end{bmatrix}^{-1} \\ & \times \begin{bmatrix} (T_1+T_2)^{-1/2}\,\frac{\partial l(\theta_0)}{\partial\beta} \\ T_1^{-1/2}\frac{\partial l(\theta_0)}{\partial\sigma_\varepsilon^2} \\ T_2^{-1/2}\frac{\partial l(\theta_0)}{\partial\sigma_\varepsilon^2} \end{bmatrix}. \end{aligned}$$

But

$$T_1^{-1}\frac{\partial^2 l_0(\theta_0)}{(\partial\sigma_\varepsilon^2)^2} \to -\frac{1}{2\left(\sigma_\varepsilon^2\right)^2} \qquad \text{and} \qquad T_2^{-1}\frac{\partial^2 l_0(\theta_0)}{(\partial\sigma_u^2)^2} \to -\frac{1}{2\left(\sigma_u^2\right)^2}.$$

Also,

$$\begin{aligned} (T_1+T_2)^{-1}\frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\beta'} + \Omega &= (T_1+T_2)^{-1}\left\{-\frac{T_1}{\sigma_{\varepsilon 0}^2}\left(C_X - \sigma_{\varepsilon 0}^2\Omega\right) - \frac{T_2}{\sigma_{u0}^2}\left(C_Y - \sigma_{u0}^2\Omega\right)\right\} \\ &\to 0, \end{aligned}$$

and so

$$(T_1+T_2)^{-1}\frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\beta'} \to -\Omega.$$

In addition,

$$\frac{\partial^2 l_0(\theta_0)}{\partial\beta\partial\sigma_\varepsilon^2} = \frac{1}{\left(\sigma_{\varepsilon 0}^2\right)^2}\sum_{t=p}^{T_1-1}\varepsilon_t\begin{bmatrix} X_{t-1} \\ \vdots \\ X_{t-p}\end{bmatrix}$$

and, for all $j = 1, \ldots, p$,

$$\begin{aligned} T_1^{-1/2}(T_1+T_2)^{-1/2}\left|\frac{\partial^2 l_0(\theta_0)}{\partial\beta_j\partial\sigma_\varepsilon^2}\right| &\leqslant T_1^{-1}\left|\frac{\partial^2 l_0(\theta_0)}{\partial\beta_j\partial\sigma_\varepsilon^2}\right| \\ &= \frac{1}{\left(\sigma_{\varepsilon 0}^2\right)^2}\left|T_1^{-1}\sum_{t=p}^{T_1-1}\varepsilon_t X_{t-j}\right| \\ &\to 0. \end{aligned}$$

Similarly, for all $j = 1, \ldots, p$,

$$T_2^{-1/2}(T_1+T_2)^{-1/2}\left|\frac{\partial^2 l_0(\theta_0)}{\partial\beta_j\partial\sigma_u^2}\right| \to 0.$$

That is

$$M_{T_1,T_2}\frac{\partial l_0(\theta_0)}{\partial\theta\partial\theta'}M_{T_1,T_2} \to -J,$$

where

$$J = \begin{bmatrix} \Omega & 0 & 0 \\ 0 & \frac{1}{2(\sigma_\varepsilon^2)^2} & 0 \\ 0 & 0 & \frac{1}{2(\sigma_u^2)^2}\end{bmatrix}.$$

Thus,

$$\begin{bmatrix} (T_1+T_2)^{1/2}\left(\widehat{\beta}-\beta_0\right) \\ T_1^{1/2}\left(\widehat{\sigma}_\varepsilon^2-\sigma_{\varepsilon 0}^2\right) \\ T_2^{1/2}\left(\widehat{\sigma}_u^2-\sigma_{u0}^2\right)\end{bmatrix}$$

is asymptotically normal with mean zero and covariance matrix Σ, where

$$\Sigma = J^{-1}\jmath J^{-1} = \begin{bmatrix} \Omega^{-1} & 0 & 0 \\ 0 & \left(\sigma_{\varepsilon 0}^2\right)^4\upsilon & 0 \\ 0 & 0 & \left(\sigma_{u0}^2\right)^4\eta\end{bmatrix}.$$

If $\{\varepsilon_t\}$ and $\{u_t\}$ are Gaussian, then

$$\Sigma = \begin{bmatrix} \Omega^{-1} & 0 & 0 \\ 0 & 2\left(\sigma_{\varepsilon 0}^2\right)^2 & 0 \\ 0 & 0 & 2\left(\sigma_{u0}^2\right)^2\end{bmatrix}.$$

3.9.4 Proof of Theorem 3.3

Let $\theta = \begin{bmatrix} \theta_1' & \sigma_\varepsilon^2 & \sigma_u^2 & \theta_2' \end{bmatrix}'$, where $\theta_1 = \beta_X$ and $\theta_2 = \beta_Y - \beta_X$. The hypothesis test is then

$$H_0 : \theta_2 = 0$$

$$H_A : \theta_2 \neq 0.$$

The test statistic is

$$\Lambda = 2\left\{\sup_\theta l(\theta) - \sup_{\theta_2 = 0} l(\theta)\right\},$$

where

$$l(\theta) = l_X\left(\beta_X, \sigma_\varepsilon^2\right) + l_Y\left(\beta_Y, \sigma_u^2\right).$$

Let $\theta_* = \begin{bmatrix} \theta_1' & \sigma_\varepsilon^2 & \sigma_u^2 \end{bmatrix}'$ and denote the true value of θ under H_0 by $\theta_0 = \begin{bmatrix} \theta_{*0}' & 0 \end{bmatrix}'$. The estimators under H_0 and H_A, denoted $\widehat{\theta}_0$ and $\widehat{\theta}_A$, respectively, satisfy

$$0 = \frac{\partial l\left(\widehat{\theta}_A\right)}{\partial \theta} \qquad \text{and} \qquad 0 = \frac{\partial l\left(\widehat{\theta}_0\right)}{\partial \theta_*},$$

where $\widehat{\theta}_0 = \begin{bmatrix} \widehat{\theta}_{*0}' & 0 \end{bmatrix}'$. From the mean value theorem, letting θ_j, θ_{*j} and θ_{Aj} denote the jth components of θ, θ_* and θ_A, respectively,

$$0 = \frac{\partial l\left(\widehat{\theta}_A\right)}{\partial \theta_j} = \frac{\partial l(\theta_0)}{\partial \theta_j} + \frac{\partial^2 l\left(\widetilde{\theta}_{Aj}\right)}{\partial \theta_j \partial \theta'}\left(\widehat{\theta}_A - \theta_0\right) \tag{3.5}$$

and

$$0 = \frac{\partial l\left(\widehat{\theta}_0\right)}{\partial \theta_{*j}} = \frac{\partial l(\theta_0)}{\partial \theta_{*j}} + \frac{\partial^2 l\left(\widetilde{\theta}_{0j}\right)}{\partial \theta_{*j} \partial \theta_*'}\left(\widehat{\theta}_{*0} - \theta_{*0}\right), \tag{3.6}$$

where $\widetilde{\theta}_{Aj}$ is a point on the line segment between θ_0 and $\widehat{\theta}_A$, and $\widetilde{\theta}_{0j}$ is a point on the line segment between θ_0 and $\widehat{\theta}_0$. Since $\beta_X = \theta_1$ and $\beta_Y = \theta_1 + \theta_2$, the first derivatives of $l(\theta)$ with respect to θ_1 and θ_2 at θ_0 are

$$\frac{\partial l(\theta_0)}{\partial \theta_1} = \frac{\partial l(\theta_0)}{\partial \beta_X} + \frac{\partial l(\theta_0)}{\partial \beta_Y} = -T_1^{1/2} z_1 - T_2^{1/2} z_2$$

and

$$\frac{\partial l(\theta_0)}{\partial \theta_2} = \frac{\partial l(\theta_0)}{\partial \beta_Y} = -T_2^{1/2} z_2,$$

where

$$z_1' = \frac{T_1^{-1/2}}{\sigma_\varepsilon^2} \sum_{t=p}^{T_1 - 1} \varepsilon_t \begin{bmatrix} X_{t-1} & \cdots & X_{t-p} \end{bmatrix} \qquad \text{and} \qquad z_2' = \frac{T_2^{-1/2}}{\sigma_u^2} \sum_{t=p}^{T_2 - 1} u_t \begin{bmatrix} Y_{t-1} & \cdots & Y_{t-p} \end{bmatrix}.$$

Therefore

$$\frac{\partial l(\theta_0)}{\partial \theta} = -N_{T_1, T_2} Z,$$

where

$$N_{T_1,T_2} = \begin{bmatrix} T_1^{1/2} I_p & 0 & T_2^{1/2} I_p \\ 0 & I_2 & 0 \\ 0 & 0 & T_2^{1/2} I_p \end{bmatrix},$$

$$Z = \begin{bmatrix} z_1' & a' & z_2' \end{bmatrix}'$$

and

$$a = \begin{bmatrix} \frac{\partial l(\theta_0)}{\partial \sigma_\varepsilon^2} & \frac{\partial l(\theta_0)}{\partial \sigma_u^2} \end{bmatrix}'.$$

The second derivatives of $l(\theta)$ with respect to θ_1 and θ_2 at θ_0 are

$$\frac{\partial^2 l(\theta_0)}{\partial \theta_1 \partial \theta_1'} = \frac{\partial^2 l(\theta_0)}{\partial \beta_X \partial \beta_X'} + \frac{\partial^2 l(\theta_0)}{\partial \beta_X \partial \beta_Y'} + \frac{\partial^2 l(\theta_0)}{\partial \beta_Y \partial \beta_X'} + \frac{\partial^2 l(\theta_0)}{\partial \beta_Y \partial \beta_Y'} = -\frac{T_1}{\sigma_\varepsilon^2} C_X - \frac{T_2}{\sigma_u^2} C_Y,$$

$$\frac{\partial^2 l(\theta_0)}{\partial \theta_2 \partial \theta_2'} = \frac{\partial^2 l(\theta_0)}{\partial \beta_Y \partial \beta_Y'} = -\frac{T_2}{\sigma_u^2} C_Y,$$

$$\frac{\partial^2 l(\theta_0)}{\partial \theta_1 \partial \theta_2'} = \frac{\partial^2 l(\theta_0)}{\partial \beta_X \partial \beta_Y'} + \frac{\partial^2 l(\theta_0)}{\partial \beta_Y \partial \beta_Y'} = -\frac{T_2}{\sigma_u^2} C_Y,$$

$$\frac{\partial^2 l(\theta_0)}{\partial \theta_2 \partial \theta_1'} = \frac{\partial^2 l(\theta_0)}{\partial \beta_Y \partial \beta_X'} + \frac{\partial^2 l(\theta_0)}{\partial \beta_Y \partial \beta_Y'} = -\frac{T_2}{\sigma_u^2} C_Y.$$

Under H_0, using a second order Taylor expansion of $l\left(\widehat{\theta}_A\right)$ around θ_0 and since $\widehat{\theta}_A \to \theta_0$, $l\left(\widehat{\theta}_A\right) - l(\theta_0)$ has the same asymptotic properties as

$$\frac{\partial l(\theta_0)}{\partial \theta'}\left(\widehat{\theta}_A - \theta_0\right) + \frac{1}{2}\left(\widehat{\theta}_A - \theta_0\right)' \frac{\partial^2 l(\theta_0)}{\partial \theta \partial \theta'}\left(\widehat{\theta}_A - \theta_0\right),$$

which, because of (3.5), is equal to

$$\begin{aligned} -\frac{1}{2} Z' N_{T_1,T_2}' \left\{\frac{\partial^2 l(\theta_0)}{\partial \theta \partial \theta'}\right\}^{-1} N_{T_1,T_2} Z &= -\frac{1}{2} Z' \left\{N_{T_1,T_2}^{-1} \frac{\partial^2 l(\theta_0)}{\partial \theta \partial \theta'} \left(N_{T_1,T_2}'\right)^{-1}\right\}^{-1} Z \\ &= \frac{1}{2} Z' \begin{bmatrix} \frac{1}{\sigma_\varepsilon^2} C_X & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & \frac{1}{\sigma_u^2} C_Y \end{bmatrix}^{-1} Z, \end{aligned}$$

where

$$A = \begin{bmatrix} \frac{\partial^2 l(\theta)}{\left(\partial \sigma_\varepsilon^2\right)^2} & 0 \\ 0 & \frac{\partial^2 l(\theta)}{\left(\partial \sigma_u^2\right)^2} \end{bmatrix}.$$

Similarly, from (3.6), $l\left(\widehat{\theta}_0\right) - l(\theta_0)$ has the same asymptotic properties under H_0 as

$$l(\theta_0) + \frac{1}{2} Z' \begin{bmatrix} \frac{1}{\sigma_\varepsilon^2} C_X & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & 0 \end{bmatrix}^{-1} Z.$$

Thus Λ has the same asymptotic distribution under H_0 as

$$z_2'\left(\frac{1}{\sigma_u^2}C_Y\right)^{-1}z_2.$$

But z_2 is asymptotically normal with mean zero and covariance matrix Ω and

$$\frac{1}{\sigma_u^2}C_Y \to \Omega.$$

Thus, under H_0, Λ asymptotically has the χ^2 distribution with p degrees of freedom since $\dim\Omega = p$.

3.9.5 The Asymptotic Distribution of the Range Statistic Under H_0

Suppose that the X_t's are i.i.d. and Gaussian, then the random variables $I_{T,X}(\omega_j)/\sigma_\varepsilon^2$, $j = 1,\ldots,q = \lfloor (T-1)/2 \rfloor$, are asymptotically independent and follow the χ^2 distribution with 2 degrees of freedom. Thus, letting $c = \sigma_\varepsilon^2/\sigma_u^2$,

$$J(\omega_j) = \frac{I_{T,X}(\omega_j)}{cI_{T,Y}(\omega_j)}, \qquad j = 1,\ldots,q,$$

asymptotically follow the $F_{2,2}$ distribution. Let $Z_j = \log J(\omega_j)$ and $Z_{(k)}$ be the kth smallest of the Z_j, $j = 1,\ldots,q$. Then

$$P(Z_j \leqslant z) = P\left(J(\omega_j) \leqslant e^z\right) \to \frac{e^z}{1+e^z} = \left(1+e^{-z}\right)^{-1}$$

as $T \to \infty$. Let $U = Z_{(1)} + \log q$ and $V = Z_{(q)} - \log q$. Then

$$P(V \leqslant v) = \{P(Z_j \leqslant \log q + v)\}^q = \left(1+q^{-1}e^{-v}\right)^{-q} \to e^{-e^{-v}}$$

as $T \to \infty$. It follows that, as $T \to \infty$,

$$P(U \leqslant u) \to 1 - e^{-e^u}.$$

Since

$$\begin{aligned} P(U > u, V \leqslant v) &= P(-\log q + u \leqslant Z_j \leqslant \log q + v,\ 1 \leqslant j \leqslant q) \\ &= \{P(Z_j \leqslant \log q + v) - P(Z_j \leqslant -\log q + u)\}^q \\ &\to e^{-e^{-v}}e^{-e^u} \end{aligned}$$

as $T \to \infty$,

$$\begin{aligned} F_{U,V}(u,v) &= P(U \leqslant u, V \leqslant v) \\ &= P(V \leqslant v) - P(U > u, V \leqslant v) \\ &\to e^{-e^{-v}} - e^{-e^{-v}}e^{-e^u} \\ &= e^{-e^{-v}}\left(1 - e^{-e^u}\right), \end{aligned}$$

as $T \to \infty$, which is the cdf of independent random variables U and V. Hence $W_T = Z_{(q)} - Z_{(1)} - 2\log q$ has limiting cdf

$$\begin{aligned} F_W(w) &= P(V - U \leqslant w) \\ &= \int_{-\infty}^{\infty} P(V \leqslant u + w) f_U(u)\, du \\ &= \int_{-\infty}^{\infty} e^{-e^{-w}e^{-u}} e^u e^{-e^u} du. \end{aligned}$$

Let $e^{-w} = \alpha^2$ and $\alpha s = e^u$. Then $e^u du = \alpha ds$ and

$$F_W(w) = \alpha \int_0^{\infty} e^{-\alpha\left(s+s^{-1}\right)} ds.$$

Now put $s = e^z$. Then $ds = e^z dz$ and we have

$$\begin{aligned} F_W(\omega) &= \alpha \int_{-\infty}^{\infty} e^{-2\alpha\cosh z} e^z dz \\ &= 2\alpha \int_0^{\infty} e^{-2\alpha\cosh z} \cosh z dz \\ &= 2\alpha K_1(2\alpha), \end{aligned}$$

from Abramowitz and Stegun (1965), Section 9.6.24. Thus

$$F_W(w) = 2e^{-\omega/2} K_1\left(2e^{-\omega/2}\right).$$

4

ARMA Spectral Discrimination

4.1 Introduction

In Chapter 3 we developed a parametric test to discriminate between time series on the basis of their spectral shape. The test was based on fitting fixed order autoregressions to the time series and using a pseudo-likelihood ratio procedure. It was shown that fitting long-order autoregressions in this way produced a test which performed well when discriminating between time series which have been generated by processes which are not purely autoregressive, for example, which are from a moving average process. A natural extension is to develop a test based on fitting autoregressive-moving average (ARMA) models. If the true orders of the underlying processes are known, then a test based on fitting ARMA models would be expected to have higher power than one based on fitting long-order autoregressions.

Let $\{X_t\}$ and $\{Y_t\}$ be univariate, stationary stochastic processes, assumed to have zero means. Given samples of sizes T_1 and T_2, respectively, we fit the ARMA models

$$X_t + \beta_{X,1} X_{t-1} + \cdots + \beta_{X,p_X} X_{t-p_X} = \varepsilon_t + \alpha_{X,1}\varepsilon_{t-1} + \cdots \alpha_{X,q_X}\varepsilon_{t-q_X} \tag{4.1}$$

and

$$Y_t + \beta_{Y,1} Y_{t-1} + \cdots + \beta_{Y,p_Y} Y_{t-p_X} = u_t + \alpha_{Y,1} u_{t-1} + \cdots \alpha_{Y,q_Y} u_{t-q_Y}. \tag{4.2}$$

We assume that the innovation processes, $\{\varepsilon_t\}$ and $\{u_t\}$, have zero mean with $E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2$ and $E\left(u_t^2\right) = \sigma_u^2$.

The null hypothesis is that the spectral densities of the two processes have the same shape, that is, that their ratio is constant. If $\{X_t\}$ and $\{Y_t\}$ satisfy (4.1) and (4.2), respectively, then their spectral densities are

$$f_X\left(\omega\right) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{\left|1 + \sum_{k=1}^{q_X} \alpha_{X,k} e^{-ik\omega}\right|^2}{\left|1 + \sum_{j=1}^{p_X} \beta_{X,j} e^{-ij\omega}\right|^2} \qquad \text{and} \qquad f_Y\left(\omega\right) = \frac{\sigma_u^2}{2\pi} \frac{\left|1 + \sum_{k=0}^{q_Y} \alpha_{Y,k} e^{-ik\omega}\right|^2}{\left|1 + \sum_{j=0}^{p_Y} \beta_{Y,j} e^{-ij\omega}\right|^2}.$$

Thus, $f_X\left(\omega\right)$ and $f_Y\left(\omega\right)$ will have the same shape if and only if the autoregressive parameters are equal and the moving average parameters are equal. The null hypothesis is then

$$H_0 : \beta_{X,j} = \beta_{Y,j},\ \forall j,\ \alpha_{X,k} = \alpha_{Y,k},\ \forall k,$$

and the alternative hypothesis, H_A, is its complement. In order to estimate the parameters under H_0, we therefore require a procedure for fitting ARMA models with common autoregressive and moving average parameters to two time series with potentially different innovation variances and sample sizes.

Fitting ARMA models to two or more time series with common parameters has been previously studied by Bowden and Clarke (2012), who proposed the interleaving method. This method combines n time series of length T, denoted $\{X_{k,t}\}$, $k = 1, \ldots, n$, to create a single series given by

$$X_{1,0}, \ldots, X_{m,0}, X_{1,1}, \ldots, X_{m,1}, \ldots, X_{1,T-1}, \ldots, X_{m,T-1}.$$

An ARMA model is then fitted to the new series using standard procedures, subject to the constraint that the autoregressive and moving average parameters are only non-zero for lags which are multiples of n. The technique was used to model maximum daily temperatures for a given week in the year by considering measurements over a sixty six year period. Bowden and Clarke (2017) extended the interleaving method to the multivariate case. The method assumes that the innovation variances are the same and that the sample sizes are equal. If the sample sizes are not equal, the shorter series' are zero-padded to be the length of the longest.

The procedure that we develop is motivated by the Hannan–Rissanen procedure for estimating ARMA parameters (Hannan and Rissanen, 1982; Hannan and Kavalieris, 1984a). This procedure is based on minimising the least squares function and can incorporate order estimation using an information criterion.

In this chapter we show how the pseudo-likelihood ratio procedure used in Chapter 3 can be adapted to the case of fitting ARMA models. We detail the Hannan–Rissanen procedure which can be used to maximise the pseudo-likelihoods under H_A. We then show how to extend the procedure for fitting ARMA models with common parameters to two time series in order to maximise the pseudo-likelihood under H_0. The results of simulation studies are presented that compare the new test to that based on fitting fixed order autoregressions. It is shown that fitting ARMA models results in a more powerful test when the orders are known. However, when the orders are not known and need to be estimated, which will be the case in practice, the fixed order autoregressive test performs better than the one based on fitting ARMA models.

4.2 The Likelihood Ratio Test

The conditional Gaussian log-likelihoods of $\{X_t\}$ and $\{Y_t\}$ are

$$-\frac{T_1}{2}\log\left(2\pi\sigma_\varepsilon^2\right)-\frac{1}{2\sigma_\varepsilon^2}\left[\sum_{t=\max(p_X,q_X)}^{T_1-1}\left\{\frac{b_{\beta_X}(z)}{a_{\alpha_X}(z)}X_t\right\}^2\right]$$

and

$$-\frac{T_2}{2}\log\left(2\pi\sigma_u^2\right)-\frac{1}{2\sigma_u^2}\left[\sum_{t=\max(p_Y,q_Y)}^{T_2-1}\left\{\frac{b_{\beta_Y}(z)}{a_{\alpha_Y}(z)}Y_t\right\}^2\right],$$

respectively. Under H_A, these can be maximised independently using, for example, the Hannan–Rissanen procedure. This procedure is described in Section 4.3. The maximised conditional Gaussian log-likelihood under H_A is the sum of the two maximised individual ones, which is

$$\widehat{l}_A=-\frac{T_1+T_2}{2}\left\{1+\log(2\pi)\right\}-\frac{T_1}{2}\log\widehat{\sigma}_{\varepsilon;A}^2-\frac{T_2}{2}\log\widehat{\sigma}_{u;A}^2,$$

where $\widehat{\sigma}_{\varepsilon;A}^2$ and $\widehat{\sigma}_{u;A}^2$ are the estimators of σ_ε^2 and σ_u^2, respectively.

Under H_0, $\{X_t\}$ and $\{Y_t\}$ have the same autoregressive and moving average parameters and orders. Let $\beta=\begin{bmatrix}\beta_1 & \cdots & \beta_p\end{bmatrix}'$ and $\alpha=\begin{bmatrix}\alpha_1 & \cdots & \alpha_q\end{bmatrix}'$ be the common model parameters and let p and q be the common orders. The conditional Gaussian log-likelihood is

$$l_0\left(\beta,\alpha,\sigma_\varepsilon^2,\sigma_u^2\right)=-\frac{T_1}{2}\log\left(2\pi\sigma_\varepsilon^2\right)-\frac{T_2}{2}\log\left(2\pi\sigma_u^2\right)-\frac{1}{2\sigma_\varepsilon^2}\left[\sum_{t=\max(p,q)}^{T_1-1}\left\{\frac{b_\beta(z)}{a_\alpha(z)}X_t\right\}^2\right]$$
$$-\frac{1}{2\sigma_u^2}\left[\sum_{t=\max(p,q)}^{T_2-1}\left\{\frac{b_\beta(z)}{a_\alpha(z)}Y_t\right\}^2\right].$$

In Section 4.4 we show how to maximise this using an approach based on the Hannan–Rissanen procedure.

4.3 The Hannan–Rissanen Procedure

The Hannan–Rissanen procedure was first proposed by Hannan and Rissanen (1982) and later modified by Hannan and Kavalieris (1984a) to estimate the parameters of the model

$$X_t + \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q},$$

which we can write as

$$b_\beta (z)\, X_t = a_\alpha (z)\, \varepsilon_t.$$

The method has three stages. In the first stage, the unobserved innovation process, $\{\varepsilon_t\}$, is estimated by fitting an autoregression to $\{X_t\}$ of order m, where m is chosen using AIC (see Section 3.3.4). Denoting the estimators of the autoregressive parameters by $\widetilde{\beta}_{0,1}, \ldots, \widetilde{\beta}_{0,m}$, $\{\varepsilon_t\}$ is estimated by

$$\widetilde{\varepsilon}_t = X_t + \sum_{j=1}^{m} \widetilde{\beta}_{0,j} X_{t-j},$$

$t = 0, \ldots, T-1$, where T is the sample size and $X_t = 0$ for $t < 0$. In the second stage, initial estimates of p autoregressive and q moving average parameters are obtained by regressing $-X_t$ on

$$X_{t-1}, \ldots, X_{t-p}, -\widetilde{\varepsilon}_{t-1}, \ldots, -\widetilde{\varepsilon}_{t-q},$$

$t = 0, \ldots, T-1$. These initial estimates are then used in the third stage to maximise the conditional Gaussian log-likelihood, or equivalently, the least squares function. The method of doing this proposed by Hannan and Rissanen (1982) comes from applying the Newton–Raphson method to the least squares function as follows. Let

$$S = \sum_{t=\max(p,q)}^{T-1} \left\{ \frac{b_\beta (z)}{a_\alpha (z)} X_t \right\}^2 .$$

In what follows, a summation with a subscript t denotes summing from $t = \max(p, q)$ to the maximum value indicated. Let

$$\beta = \begin{bmatrix} \beta_1 & \cdots & \beta_p \end{bmatrix}', \qquad \alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_q \end{bmatrix}' \qquad \text{and} \qquad \theta = \begin{bmatrix} \beta' & \alpha' \end{bmatrix}'.$$

Given a current estimate of θ, denoted $\widetilde{\theta}$, a new estimate is obtained by

$$\widetilde{\theta} - \left(\frac{\partial^2 S}{\partial \widetilde{\theta} \partial \widetilde{\theta}'} \right)^{-1} \frac{\partial S}{\partial \widetilde{\theta}}. \tag{4.3}$$

Let

$$\eta_t = \frac{1}{a_\alpha (z)} X_t = \frac{1}{b_\beta (z)} \varepsilon_t$$

and

$$\xi_t = \frac{b_\beta(z)}{a_\alpha^2(z)} X_t = \frac{1}{a_\alpha(z)} \varepsilon_t,$$

$t = 0, \ldots, T-1$, where $X_t = 0$ and $\varepsilon_t = 0$ for $t < 0$. The first and second derivatives of S with respect to β and α are given by

$$\begin{aligned}
\frac{\partial S}{\partial \beta_j} &= 2 \sum_t^{T-1} \varepsilon_t \eta_{t-j}, \\
\frac{\partial S}{\partial \alpha_j} &= -2 \sum_t^{T-1} \varepsilon_t \xi_{t-j}, \\
\frac{\partial^2 S}{\partial \beta_j \partial \beta_k} &= 2 \sum_t^{T-1} \eta_{t-k} \eta_{t-j},
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 S}{\partial \beta_j \partial \alpha_k} &= -2 \sum_t^{T-1} \xi_{t-k} \eta_{t-j} - 2 \sum_t^{T-1} \frac{1}{a_\alpha^2(z)} \varepsilon_t X_{t-j-k}, \\
\frac{\partial^2 S}{\partial \alpha_j \partial \alpha_k} &= 2 \sum_t^{T-1} \xi_{t-k} \xi_{t-j} + 4 \sum_t^{T-1} \frac{b_\beta(z)}{a_\alpha^3(z)} \varepsilon_t X_{t-j-k}.
\end{aligned}$$

But, by the strong law of large numbers,

$$T^{-1} \sum_t^{T-1} \frac{1}{a_\alpha^2(z)} \varepsilon_t X_{t-j-k} \to \frac{1}{a_\alpha^2(z)} E(\varepsilon_t X_{t-j-k}) = 0,$$

for all $j, k = 1, \ldots, \max(p, q)$. Note that convergence here means convergence in the almost sure sense, and will do so in what follows. Similarly,

$$T^{-1} \sum_t^{T-1} \frac{b_\beta(z)}{a_\alpha^3(z)} \varepsilon_t X_{t-j-k} \to 0,$$

for all $j, k = 1, \ldots, \max(p, q)$. Thus

$$\frac{\partial^2 S}{\partial \beta_j \partial \alpha_k} = -2 \sum_t^{T-1} \xi_{t-k} \eta_{t-j} + o(T)$$

and

$$\frac{\partial^2 S}{\partial \alpha_j \partial \alpha_k} = 2 \sum_t^{T-1} \xi_{t-k} \xi_{t-j} + o(T),$$

where $o(\cdot)$ is in the almost sure sense. Let

$$H_t = \begin{bmatrix} \widetilde{\eta}_{t-1} & \cdots & \widetilde{\eta}_{t-p} \end{bmatrix}' \qquad \text{and} \qquad \Xi_t = \begin{bmatrix} \widetilde{\xi}_{t-1} & \cdots & \widetilde{\xi}_{t-p} \end{bmatrix}',$$

where $\tilde{}$ is used to indicate that the quantities have been computed using $\widetilde{\theta}$ in place of θ. Then (4.3) is asymptotically equivalent to

$$\widetilde{\theta} - \begin{bmatrix} \sum_t^{T-1} H_t H_t' & -\sum_t^{T-1} H_t \Xi_t' \\ -\sum_t^{T-1} \Xi_t H_t' & \sum_t^{T-1} \Xi_t \Xi_t' \end{bmatrix}^{-1} \begin{bmatrix} \sum_t^{T-1} H_t \widetilde{\varepsilon}_t \\ -\sum_t^{T-1} \Xi_t \widetilde{\varepsilon}_t \end{bmatrix},$$

which is equal to

$$\begin{bmatrix} \sum_t^{T-1} H_t H_t' & -\sum_t^{T-1} H_t \Xi_t' \\ -\sum_t^{T-1} \Xi_t H_t' & \sum_t^{T-1} \Xi_t \Xi_t' \end{bmatrix}^{-1} \begin{bmatrix} \sum_t^{T-1} H_t \left(H_t' \widetilde{\beta} - \Xi_t' \widetilde{\alpha} - \widetilde{\varepsilon}_t \right) \\ -\sum_t^{T-1} \Xi_t \left(H_t' \widetilde{\beta} - \Xi_t' \widetilde{\alpha} - \widetilde{\varepsilon}_t \right) \end{bmatrix}. \tag{4.4}$$

Now

$$\begin{aligned} H_t' \widetilde{\beta} - \Xi_t' \widetilde{\alpha} - \widetilde{\varepsilon}_t &= \sum_{j=1}^{p} \widetilde{\beta}_j \widetilde{\eta}_{t-j} - \sum_{j=1}^{q} \widetilde{\alpha}_j \xi_{t-j} - \widetilde{\varepsilon}_t \\ &= -\widetilde{\eta}_t + \widetilde{\xi}_t - \widetilde{\varepsilon}_t. \end{aligned}$$

Computing (4.4) is therefore equivalent to regressing $-\widetilde{\eta}_t + \widetilde{\xi}_t - \widetilde{\varepsilon}_t$ on $\begin{bmatrix} H_t' & -\Xi_t' \end{bmatrix}$. This regression gives the updated estimate of θ and the procedure repeats until convergence.

Denoting the autoregressive parameter estimators by $\widehat{\beta}_1, \ldots, \widehat{\beta}_p$ and the moving average parameter estimators by $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_q$, the variance of ε_t is estimated by

$$\widehat{\sigma}_{p,q}^2 = T^{-1} \sum_t^{T-1} \left\{ \frac{b_{\widehat{\beta}}(z)}{a_{\widehat{\alpha}}(z)} X_t \right\}^2.$$

The orders p and q may be estimated by running the procedure over all $p = 0, \ldots, P$ and $q = 0, \ldots, Q$, where P and Q are assumed to be greater than the true orders, then using an information criterion, for example BIC. That is, the estimates of p and q are chosen as those that minimise, for example,

$$T \log \widehat{\sigma}_{p,q}^2 + (p+q) \log T.$$

4.4 Parameter Estimation Under the Null Hypothesis

As we did in Chapter 3, we reparametrise the conditional Gaussian log-likelihood and then use a profile likelihood approach to maximise it. Let $\sigma_\varepsilon^2 = \lambda \sigma_u^2$. Then $l_0\left(\beta, \alpha, \sigma_\varepsilon^2, \sigma_u^2\right)$ can be rewritten as

$$-\frac{T_1 + T_2}{2} \log\left(2\pi\sigma_\varepsilon^2\right) + \frac{T_2}{2} \log \lambda - \frac{1}{2\sigma_\varepsilon^2} \left[\sum_t^{T_1-1} \left\{ \frac{b_\beta(z)}{a_\alpha(z)} X_t \right\}^2 + \lambda \sum_t^{T_1-1} \left\{ \frac{b_\beta(z)}{a_\alpha(z)} Y_t \right\}^2 \right]. \tag{4.5}$$

In order to maximise this with respect to α and β, for fixed λ, we minimise

$$S = \sum_t^{T_1-1} \left\{ \frac{b_\beta(z)}{a_\alpha(z)} X_t \right\}^2 + \lambda \sum_t^{T_2-1} \left\{ \frac{b_\beta(z)}{a_\alpha(z)} Y_t \right\}^2$$

using the Newton–Raphson method. Let $\theta = \begin{bmatrix} \beta' & \alpha' \end{bmatrix}'$. Given a current estimate of θ, denoted $\widetilde{\theta}$, a new estimate is obtained by

$$\widetilde{\theta} - \left(\frac{\partial^2 S}{\partial \widetilde{\theta} \partial \widetilde{\theta}'} \right)^{-1} \frac{\partial S}{\partial \widetilde{\theta}}. \tag{4.6}$$

Let

$$\eta_{X,t} = \frac{1}{a_\alpha(z)} X_t = \frac{1}{b_\beta(z)} \varepsilon_t, \qquad \text{and} \qquad \eta_{Y,t} = \frac{1}{a_\alpha(z)} Y_t = \frac{1}{b_\beta(z)} u_t,$$

$t = 0, \ldots, T_1 - 1$, and let

$$\xi_{X,t} = \frac{b_\beta(z)}{a_\alpha^2(z)} X_t = \frac{1}{a_\alpha(z)} \varepsilon_t, \qquad \text{and} \qquad \xi_{Y,t} = \frac{b_\beta(z)}{a_\alpha^2(z)} Y_t = \frac{1}{a_\alpha(z)} u_t,$$

$t = 0, \ldots, T_1 - 2$. Also let

$$H_{X,t} = \begin{bmatrix} \widetilde{\eta}_{X,t-1} \\ \vdots \\ \widetilde{\eta}_{X,t-p} \end{bmatrix}, \; H_{Y,t} = \begin{bmatrix} \widetilde{\eta}_{Y,t-1} \\ \vdots \\ \widetilde{\eta}_{Y,t-p} \end{bmatrix}, \; \Xi_{X,t} = \begin{bmatrix} \widetilde{\xi}_{X,t-1} \\ \vdots \\ \widetilde{\xi}_{X,t-q} \end{bmatrix} \text{ and } \Xi_{Y;t} = \begin{bmatrix} \widetilde{\xi}_{Y;t-1} \\ \vdots \\ \widetilde{\xi}_{Y;t-q} \end{bmatrix},$$

where, again, $\tilde{}$ is used to indicate that the quantities have been computed using $\widetilde{\theta}$ in place of θ. Following the same calculations as in the previous section, (4.6) is asymptotically equivalent to

$$\left(\begin{bmatrix} \sum_t^{T_1-1} H_{X,t} H'_{X,t} & -\sum_t^{T_1-1} H_{X,t} \Xi'_{X,t} \\ -\sum_t^{T_1-1} \Xi_{X,t} H'_{X,t} & \sum_t^{T_1-1} \Xi_{X,t} \Xi'_{X,t} \end{bmatrix} + \lambda \begin{bmatrix} \sum_t^{T_2-1} H_{Y,t} H'_{Y,t} & -\sum_t^{T_2-1} H_{Y,t} \Xi'_{Y,t} \\ -\sum_t^{T_2-1} \Xi_{Y,t} H'_{Y,t} & \sum_t^{T_2-1} \Xi_{Y,t} \Xi'_{Y,t} \end{bmatrix} \right)^{-1}$$
$$\times \left(\begin{bmatrix} \sum_t^{T_1-1} H_{X,t} \left(H'_{X,t} \widetilde{\beta} - \Xi'_{X,t} \widetilde{\alpha} - \widetilde{\varepsilon}_t \right) \\ -\sum_t^{T_1-1} \Xi_{X,t} \left(H'_{X,t} \widetilde{\beta} - \Xi'_{X,t} \widetilde{\alpha} - \widetilde{\varepsilon}_t \right) \end{bmatrix} + \lambda \begin{bmatrix} \sum_t^{T_2-1} H_{Y,t} \left(H'_{Y,t} \widetilde{\beta} - \Xi'_{Y,t} \widetilde{\alpha} - \widetilde{u}_t \right) \\ -\sum_t^{T_2-1} \Xi_{Y,t} \left(H'_{Y,t} \widetilde{\beta} - \Xi'_{Y,t} \widetilde{\alpha} - \widetilde{u}_t \right) \end{bmatrix} \right). \tag{4.7}$$

Computing this gives the updated estimate of θ and the process repeats until convergence. Of course, λ above needs to be updated. This will be discussed later.

In order to obtain an initial estimate of θ, we can follow a similar process to the first two stages of the Hannan–Rissanen procedure, utilising the methods developed in Chapter 3 for fitting common autoregressive parameters to two time series. We begin by fitting autoregressions of order m to $\{X_t\}$ and $\{Y_t\}$ with the same autoregressive parameters, denoted by $\widetilde{\beta}_{0,1}, \ldots, \widetilde{\beta}_{0,m}$, using AIC to estimate m (see Section 3.3). Put

$$\widetilde{\varepsilon}_t = X_t + \sum_{j=1}^{m} \widetilde{\beta}_{0,j} X_{t-j},$$

$t = 0, \ldots, T_1 - 1$, letting $X_t = 0$ for $t < 0$, and

$$\widetilde{u}_t = Y_t + \sum_{j=1}^{m} \widetilde{\beta}_{0,j} Y_{t-j},$$

$t = 0, \ldots, T_2 - 1$, letting $Y_t = 0$ for $t < 0$. We then obtain an initial estimate of θ from (4.7) with $H_{X,t}$, $H_{Y,t}$, $\Xi_{X,t}$ and $\Xi_{Y,t}$ computed with

$$\widetilde{\eta}_{X,t} = X_t \qquad \text{and} \qquad \widetilde{\xi}_{X,t} = \widetilde{\varepsilon}_t,$$

$t = 0, \dots, T_1 - 1$, and

$$\widetilde{\eta}_{Y,t} = Y_t \qquad \text{and} \qquad \widetilde{\xi}_{Y,t} = \widetilde{u}_t,$$

$t = 0, \dots, T_2 - 1$.

Denoting the common autoregressive parameter estimators by $\widehat{\beta}_1, \dots, \widehat{\beta}_p$ and the common moving average parameter estimators by $\widehat{\alpha}_1, \dots, \widehat{\alpha}_q$, (4.5) is maximised with respect to σ_ε^2, for fixed λ, by

$$\widetilde{\sigma}^2_{\varepsilon;p,q}(\lambda) = (T_1 + T_2)^{-1} \left[\sum_t^{T_1-1} \left\{ \frac{b_{\widehat{\beta}}(z)}{a_{\widehat{\alpha}}(z)} X_t \right\}^2 + \lambda \sum_t^{T_2-1} \left\{ \frac{b_{\widehat{\beta}}(z)}{a_{\widehat{\alpha}}(z)} Y_t \right\}^2 \right].$$

The maximised conditional Gaussian log-likelihood, for fixed λ, is therefore

$$\widetilde{l}_0(\lambda) = -\frac{T_1 + T_2}{2} \{1 + \log(2\pi)\} - \frac{T_1 + T_2}{2} \log \left\{ \widetilde{\sigma}^2_{\varepsilon;p,q}(\lambda) \right\} + \frac{T_2}{2} \log \lambda.$$

By maximising $\widetilde{l}_0(\lambda)$ with respect to λ, we obtain estimators of all the parameters. The procedure above will therefore need to be iterated in order to update λ. One method for updating λ is as follows. Given a current estimator of λ, denoted by $\widetilde{\lambda}$, estimate the model parameters as above and then re-estimate λ by

$$\frac{T_2 \widetilde{\sigma}^2_{\varepsilon;p,q}\left(\widetilde{\lambda}\right)}{\sum_t^{T_2-1} \left\{ \frac{b_{\widehat{\beta}}(z)}{a_{\widehat{\alpha}}(z)} Y_t \right\}^2}.$$

Use this new estimate of λ to re-estimate the model parameters and repeat the whole procedure until convergence. An initial estimate of λ can be obtained from the estimates under H_A, that is by letting $\widetilde{\lambda} = \widehat{\sigma}^2_{\varepsilon;A} / \widehat{\sigma}^2_{u;A}$.

To estimate p and q we make use of an information criterion. Let $\widehat{\lambda}_{p,q}$ be the value of λ which maximises $\widetilde{l}_0(\lambda)$ for a given p and q. Then let

$$\phi(p, q) = (T_1 + T_2) \log \left\{ \widetilde{\sigma}^2_{\varepsilon;p,q}\left(\widehat{\lambda}_{p,q}\right) \right\} - T_2 \log \widehat{\lambda}_{p,q} + (p + q) \log (T_1 + T_2).$$

The estimators of p and q, denoted $\widehat{p}$ and $\widehat{q}$, respectively, are the minimisers of $\phi(p, q)$ over all $p = 0, \dots, P$ and $q = 0, \dots, Q$, where P and Q are assumed to be greater than the true orders. The estimators of σ_ε^2 and σ_u^2 under H_0 are then

$$\widehat{\sigma}^2_{\varepsilon;0} = \widetilde{\sigma}^2_{\varepsilon;\widehat{p},\widehat{q}}\left(\widehat{\lambda}_{\widehat{p},\widehat{q}}\right) \qquad \text{and} \qquad \widehat{\sigma}^2_{u;0} = \widehat{\sigma}^2_{\varepsilon;0} / \widehat{\lambda}_{\widehat{p},\widehat{q}},$$

respectively.

4.5 The Test Statistic

The test statistic is

$$\Lambda = T_1 \log\left(\frac{\widehat{\sigma}^2_{\varepsilon;0}}{\widehat{\sigma}^2_{\varepsilon;A}}\right) + T_2 \log\left(\frac{\widehat{\sigma}^2_{u;0}}{\widehat{\sigma}^2_{u;A}}\right).$$

If p and q are known, H_0 is rejected at significance level α when Λ is greater than the $100\,(1-\alpha)$th percentile of the χ^2 distribution with $p+q$ degrees of freedom. If p and q are unknown, H_0 is rejected when Λ is greater than the $100\,(1-\alpha)$th percentile of the χ^2 distribution with $\widehat{p}+\widehat{q}$ degrees of freedom.

4.6 Simulations

In this section we show the results of simulations which compare the test based on fitting ARMA models developed in this chapter with the method based on fitting fixed order autoregressions which was developed in Chapter 3. Pairs of time series were simulated from the AR(1) processes

$$X_t + 0.5X_{t-1} = \varepsilon_t \qquad \text{and} \qquad Y_t + \beta Y_{t-1} = u_t,$$

the MA(1) processes

$$X_t = \varepsilon_t + 0.5\varepsilon_{t-1} \qquad \text{and} \qquad Y_t = u_t + \beta u_{t-1},$$

the ARMA$(1,1)$ processes

$$X_t + 0.5X_{t-1} = \varepsilon_t + \alpha\varepsilon_{t-1} \qquad \text{and} \qquad Y_t + \beta Y_{t-1} = u_t + \alpha u_{t-1},$$

and the ARMA$(2,2)$ processes

$$X_t + 0.5X_{t-1} + \beta X_{t-2} = \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha\varepsilon_{t-2} \qquad \text{and} \qquad Y_t + 0.5Y_{t-1} + \beta Y_{t-2} = u_t + \alpha u_{t-1} + \alpha u_{t-2}.$$

The floating parameter β was varied from $0.01, 0.02, \ldots, 0.99$. For the ARMA processes, the simulations were run with both $\alpha = 0.6$ and $\alpha = 0.8$. In each case, $\{\varepsilon_t\}$ and $\{u_t\}$ were simulated from normal distributions with mean zero and variances 1 and 4, respectively. The AR(1) and MA(1) processes are the same as those used in the power simulations in Section 3.5 and the null spectral densities of $\{X_t\}$ are described in Figure 3.1. The null spectral densities of $\{X_t\}$ for the ARMA processes are described in Figure 4.1.

For each of the six processes and for each level of β, three tests were applied with 10,000 replications with sample sizes $T_1 = 100$ and $T_2 = 125$, and 10,000 replications with sample sizes $T_1 = 1,000$ and $T_2 = 1,250$. The first test applied was that developed in this chapter based on fitting ARMA models where p and q were known. The second test applied also

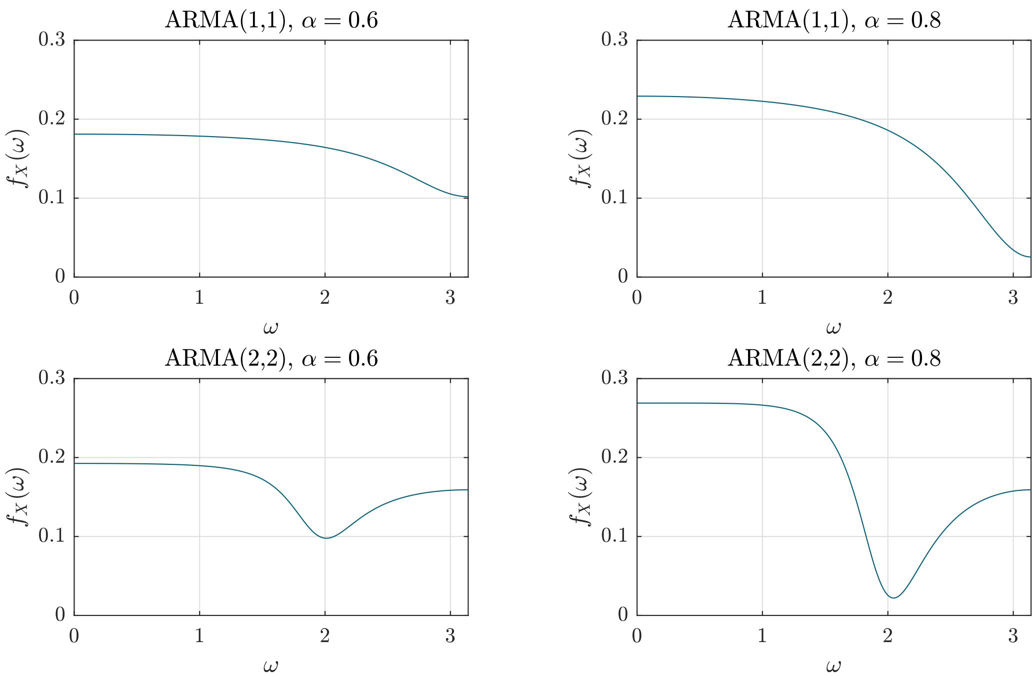

Figure 4.1: Spectral densities of $\{X_t\}$ for the ARMA(1,1) and ARMA(2,2) processes described in Section 4.6.

fitted ARMA models but the orders were unknown and were estimated using the information criterion given in Section 4.4. The maximum orders were set to one greater than the true values. The third test applied was that developed in Chapter 3 based on fitting fixed order autoregressions. The autoregressive orders were $\lfloor \log(1,000) \rfloor^{1.1} = 5$ for the short series and $\lfloor \log(1,000) \rfloor^{1.1} = 8$ for the long series. The empirical powers are shown in Figures 4.2 and 4.3.

For the short series, the ARMA method with known orders and the fixed order autoregressive method performed well, with the former having higher power, as expected. When the orders were estimated the ARMA method did not perform well, with Type I error rates between 0.23 and 0.85.

For the long series, when the time series were from the AR(1) and MA(1) processes, the ARMA method with estimated orders had almost identical power to the ARMA method with known orders, although the Type I error rate was a little over 0.05 when the orders were estimated. Both methods had higher power than the fixed order autoregressive method. When the time series were simulated from the ARMA$(1,1)$ and ARMA$(2,2)$ processes with $\alpha = 0.6$, the ARMA method with estimated orders did not perform well, with Type I error rates of 0.71 and 0.82. It performed better when $\alpha = 0.8$, although the Type I error rates were still higher than 0.05. The ARMA method with known orders and the fixed order autoregressive method performed similarly well, with the former having slightly higher power.

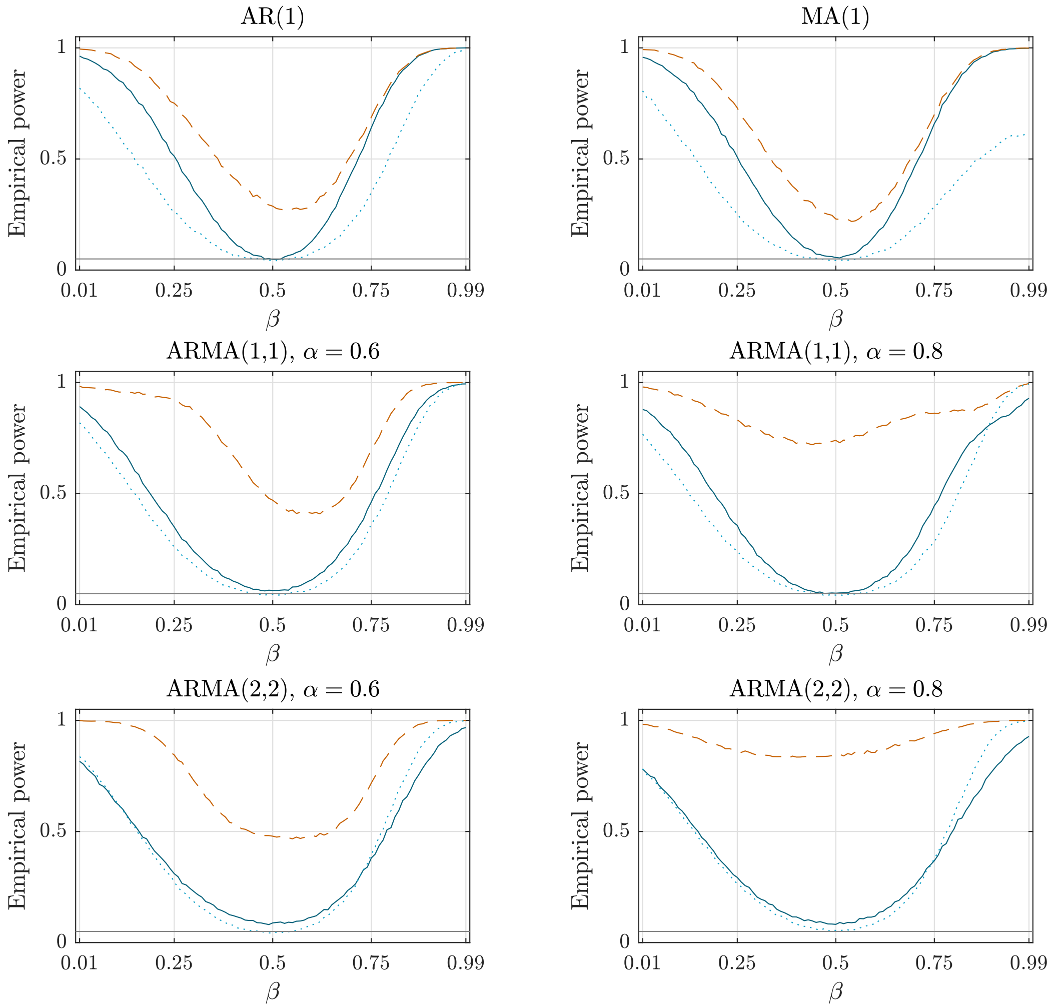

Figure 4.2: Empirical power of the likelihood ratio test using ARMA(p, q) models with the orders known (solid), ARMA(p, q) models with the orders estimated (dashes) and the fixed order autoregressive method (small dashes), for $T_1 = 100$ and $T_2 = 125$. The lowest horizontal line in each plot indicates the significance level of 0.05.

4.7 Discussion

The simulations have shown that if ARMA models are fitted with the true orders, then the pseudo-likelihood ratio procedure will result in a test which is more powerful than fitting fixed order autoregressions. This is expected, since the methods will more closely approximate the true spectral densities. In practice, however, the ARMA orders will not be known and will need to be estimated. The simulations suggest that this method will not always perform well. As in the previous chapter, simulations have demonstrated that the fixed order autoregressive method will work well even when the time series are not from autoregressive processes.

In light of this, the fixed order autoregressive method is a better choice for time series discrimination than the method which fits ARMA models in practice, since the true orders will not be known. For this reason we will not pursue the ARMA methods any further in the

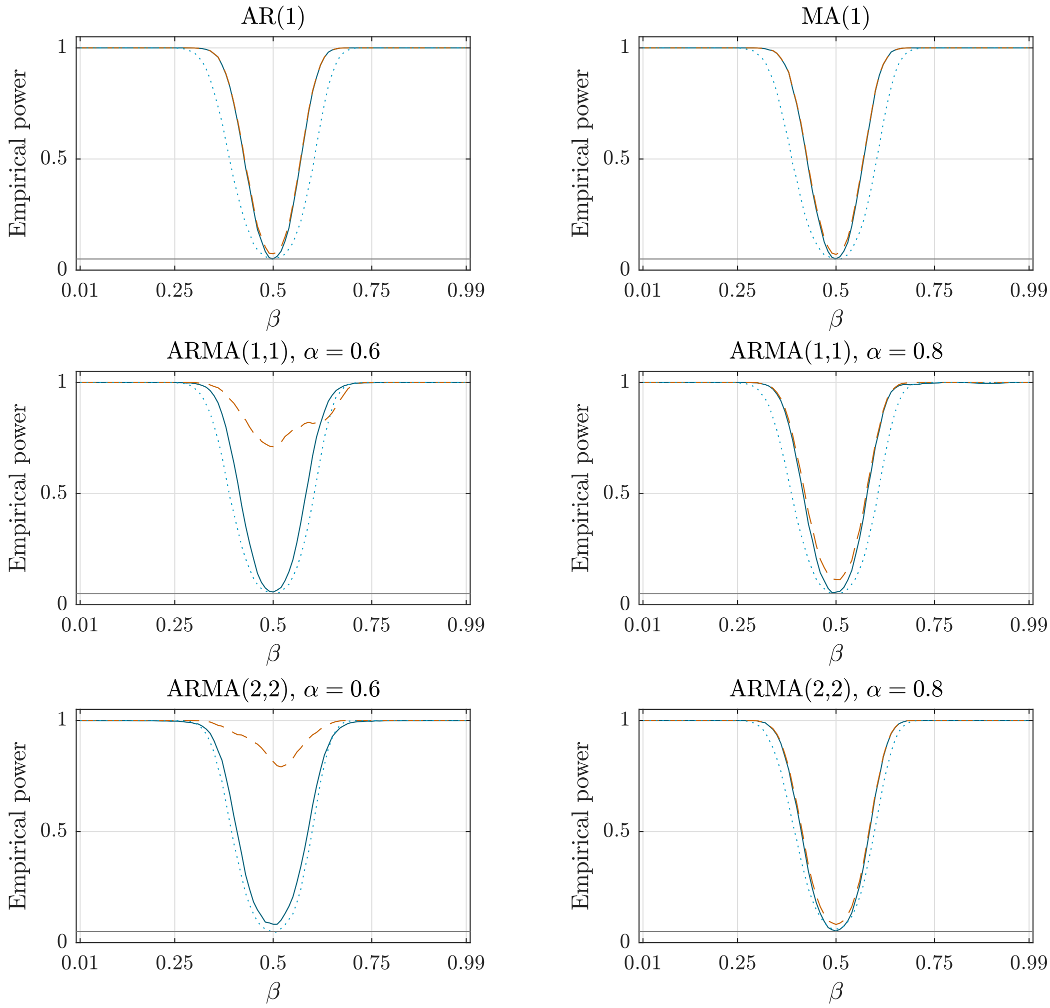

Figure 4.3: Empirical power of the likelihood ratio test using ARMA(p, q) models with the orders known (solid), ARMA(p, q) models with the orders estimated (dashes) and the fixed order autoregressive method (small dashes), for $T_1 = 1,000$ and $T_2 = 1,250$. The lowest horizontal line in each plot indicates the significance level of 0.05.

discrimination context when considering, for example, multivariate time series. However, the technique that has been developed in this chapter for fitting common ARMA models to two time series may have applications beyond the parametric test for time series discrimination. For example, Bowden and Clarke (2012) have discussed the use of modelling measurements of daily maximum temperatures in a given period over several years to forecast electricity demand. An area for future research therefore will be to extend the procedure developed in this chapter to, say, the case of more than two time series, and to study the properties of the parameter estimators.

5

Comparing Multivariate Time Series

5.1 Introduction

While there is an extensive literature on comparing univariate time series from stationary processes, there has been much less work for the multivariate case. Most of the existing methods test the null hypothesis that the spectral densities of two independent stationary processes are the same. For this null hypothesis, Bassily et al. (2009), Lund et al. (2009) and Ravishanker et al. (2010) have developed nonparametric methods based on comparing smoothed periodograms at the Fourier frequencies. Tugnait (2016) considered the complex case using a similar approach. Lund et al. (2009) also proposed a test based on comparing sample autocovariances. Kakizawa et al. (1998) compares several time series using disparity measures between smoothed periodograms for the purposes of clustering and classification.

An alternative null hypothesis, suggested by Ravishanker et al. (2010), is that the spectral densities differ in scale but still share the same second order dynamics. That is, that the spectral densities of each of the corresponding components of the two processes have the same shape. This is a multivariate generalisation of the null hypothesis considered in Chapters 3 and 4.

Another null hypothesis of interest is that two time series are from vector autoregressions with the same autoregressive parameters. A test of Maharaj (1999) fits vector autoregressions to two time series using an information criterion to select the autoregressive orders, and considers the differences between the independent parameter estimates for each process.

The above suggests a number of null hypotheses related to comparing multivariate time series. We can develop tests for these by generalising the parametric approach of Chapter 3 to the multivariate case.

Let $\{X_t\}$ and $\{Y_t\}$ be d-dimensional stationary stochastic processes, assumed to have zero mean. We fit a d-dimensional vector autoregression of order p_X to $\{X_t\}$ and of order p_Y to $\{Y_t\}$. That is, we fit the models

$$X_t + \beta_{X,1} X_{t-1} + \cdots + \beta_{X,p_X} X_{t-p_X} = \varepsilon_t \tag{5.1}$$

and

$$Y_t + \beta_{Y,1} Y_{t-1} + \cdots + \beta_{Y,p_Y} Y_{t-p_Y} = u_t, \tag{5.2}$$

where $\beta_{X,j}$, $j = 1, \ldots, p_X$, and $\beta_{Y,j}$, $j = 1, \ldots, p_Y$, are $d \times d$ and $\{\varepsilon_t\}$ and $\{u_t\}$ are independent d-dimensional innovation processes. We make the usual assumptions that $\{\varepsilon_t\}$ and $\{u_t\}$ are sequences of martingale differences, that is that

$$E\left(\varepsilon_t \mid \mathcal{F}_{t-1}\right) = E\left(u_t \mid \mathcal{G}_{t-1}\right) = 0,$$

and also that

$$E\left(\varepsilon_t \varepsilon_t' \mid \mathcal{F}_{t-1}\right) = \Sigma_\varepsilon,\ E\left(u_t u_t' \mid \mathcal{G}_{t-1}\right) = \Sigma_u,$$

where $\mathcal{F}_t$ and $\mathcal{G}_t$ are the σ-fields generated by $\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$ and $\{u_t, u_{t-1}, \ldots\}$, respectively. The spectral densities of $\{X_t\}$ and $\{Y_t\}$ are

$$f_X\left(\omega\right) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma_X\left(j\right) e^{-ij\omega} \qquad \text{and} \qquad f_Y\left(\omega\right) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma_Y\left(j\right) e^{-ij\omega}, \tag{5.3}$$

respectively, where $\Gamma_X\left(j\right) = E\left(X_t X_{t+j}'\right)$ and $\Gamma_Y\left(j\right) = E\left(Y_t Y_{t+j}'\right)$. If $\{X_t\}$ and $\{Y_t\}$ satisfy (5.1) and (5.2), then their spectral densities are

$$f_X\left(\omega\right) = \frac{1}{2\pi} \left(I_d + \sum_{j=1}^{p_X} \beta_{X,j} e^{-i\omega j}\right)^{-1} \Sigma_\varepsilon \left\{\left(I_d + \sum_{j=1}^{p_X} \beta_{X,j} e^{-i\omega j}\right)^{-1}\right\}^*$$

and

$$f_Y\left(\omega\right) = \frac{1}{2\pi} \left(I_d + \sum_{j=1}^{p_Y} \beta_{Y,j} e^{-i\omega j}\right)^{-1} \Sigma_u \left\{\left(I_d + \sum_{j=1}^{p_Y} \beta_{Y,j} e^{-i\omega j}\right)^{-1}\right\}^*,$$

respectively, where $*$ denotes the complex conjugate transpose (Reinsel, 1993, Section 2.3).

We do not want to assume that $\{X_t\}$ and $\{Y_t\}$ truly are autoregressive but instead use long-order autoregressions to approximate more general processes. Following the approach of Chapter 3, we let $p_X = p_Y = p$ and derive test statistics, parameter estimators and their asymptotic properties for fixed p. When applying the test procedures in practice we then let $p = \lfloor (\log T_{\min})^c \rfloor$, where $T_{\min} = \min(T_1, T_2)$ and $c > 1$ (see Section 3.6).

The first null hypothesis we consider is that the autoregressive parameters of $\{X_t\}$ and $\{Y_t\}$ are equal, that is

$$H_0^{(1)} : \beta_{X,j} = \beta_{Y,j} \; \forall j.$$

Under $H_0^{(1)}$, the innovation processes can have different covariance matrices, and so the spectral densities of $\{X_t\}$ and $\{Y_t\}$ are not necessarily the same.

The second null hypothesis we consider is that $f_X(\omega) = \lambda f_Y(\omega)$ for some positive constant λ. This is equivalent to

$$H_0^{(2)} : \beta_{X,j} = \beta_{Y,j} \; \forall j, \; \Sigma_\varepsilon = \lambda \Sigma_u.$$

Under $H_0^{(2)}$, the spectral densities of each of the corresponding components of $\{X_t\}$ and $\{Y_t\}$ differ only by a common scale. A special case is when $\lambda = 1$, which is when the spectral densities are equal.

The third null hypothesis we consider is that $f_X(\omega) = \Lambda f_Y(\omega) \Lambda$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$ and $\lambda_1, \ldots, \lambda_d$ are positive constants. This is the null hypothesis suggested by Ravishanker et al. (2010). Its interpretation is that the spectral density of each component of $\{X_t\}$ has the same shape as the corresponding component of $\{Y_t\}$. If this is the case then, from (5.3),

$$\Gamma_Y(j) = E\left(Y_t Y_{t+j}'\right) = E\left(\Lambda X_t X_{t+j}' \Lambda\right) = \Lambda \Gamma_X(j) \Lambda,$$

and it follows that this null hypothesis is equivalent to

$$H_0^{(3)} : \beta_{X,j} = \Lambda \beta_{Y,j} \Lambda^{-1} \; \forall j, \; \Sigma_\varepsilon = \Lambda \Sigma_u \Lambda.$$

In each case, the alternative hypothesis, H_A, is the complement of the null hypothesis. We shall introduce three more null hypotheses when we extend the tests to compare more than two time series.

In this chapter we show how to use the pseudo-likelihood ratio procedure for the multivariate case to derive tests for each of the hypotheses given above. We present methods for estimating the parameters under each null hypothesis and establish asymptotic theory for the estimators. We also show how to extend the tests to compare more than two time series. The results of simulation studies are presented which demonstrate the behaviour of the test statistics under the null hypotheses.

5.2 The Likelihood Ratio Procedure

Let

$$X + \beta_X Z_X = \varepsilon$$

and

$$Y + \beta_Y Z_Y = u,$$

where

$$\begin{aligned}
X &= \left[\begin{array}{ccc} X_p & \cdots & X_{T_1-1} \end{array}\right], & Y &= \left[\begin{array}{ccc} Y_p & \cdots & Y_{T_1-1} \end{array}\right], \\
Z_X &= \left[\begin{array}{ccc} Z_{X,p} & \cdots & Z_{X,T_1-1} \end{array}\right], & Z_Y &= \left[\begin{array}{ccc} Z_{Y,p} & \cdots & Z_{Y,T_2-1} \end{array}\right], \\
Z_{X,t} &= \left[\begin{array}{ccc} X'_{t-1} & \cdots & X'_{t-p} \end{array}\right]', & Z_{Y,t} &= \left[\begin{array}{ccc} Y'_{t-1} & \cdots & Y'_{t-p} \end{array}\right]', \\
\varepsilon &= \left[\begin{array}{ccc} \varepsilon_p & \cdots & \varepsilon_{T_1-1} \end{array}\right], & u &= \left[\begin{array}{ccc} u_p & \cdots & u_{T_2-1} \end{array}\right], \\
\beta_X &= \left[\begin{array}{ccc} \beta_{X,1} & \cdots & \beta_{X,p} \end{array}\right], & \beta_Y &= \left[\begin{array}{ccc} \beta_{Y,1} & \cdots & \beta_{Y,p} \end{array}\right].
\end{aligned}$$

The conditional Gaussian log-likelihoods for $\{X_t\}$ and $\{Y_t\}$ are

$$l_X(\beta_X, \Sigma_\varepsilon) = -\frac{T_1 d}{2}\log(2\pi) - \frac{T_1}{2}\log|\Sigma_\varepsilon| - \frac{T_1}{2}\operatorname{tr}\left\{\Sigma_\varepsilon^{-1} s_{X,T_1}(\beta_X)\right\}$$

and

$$l_Y(\beta_Y, \Sigma_u) = -\frac{T_2 d}{2}\log(2\pi) - \frac{T_2}{2}\log|\Sigma_u| - \frac{T_2}{2}\operatorname{tr}\left\{\Sigma_u^{-1} s_{Y,T_2}(\beta_Y)\right\},$$

where

$$s_{X,T_1}(\beta) = T_1^{-1}(X + \beta Z_X)(X + \beta Z_X)',$$

$$s_{Y,T_2}(\beta) = T_2^{-1}(Y + \beta Z_Y)(Y + \beta Z_Y)'$$

and $|\cdot|$ denotes the determinant. Under the alternative hypotheses $\{X_t\}$ and $\{Y_t\}$ are independent vector autoregressive processes of order p, denoted VAR(p), and their likelihoods can be maximised separately. The maximum Gaussian likelihood, or least squares, estimators for the parameters of VAR(p) processes, and their properties, are well known (see, for example, Hannan, 1970). The estimators of β_X and β_Y are

$$\widehat{\beta}_X = -X Z_X' \left(Z_X Z_X'\right)^{-1} \qquad \text{and} \qquad \widehat{\beta}_Y = -Y Z_Y' \left(Z_Y Z_Y'\right)^{-1},$$

respectively, and the estimators of Σ_ε and Σ_u are

$$\widehat{\Sigma}_{\varepsilon;A} = s_{X,T_1}\left(\widehat{\beta}_X\right) \qquad \text{and} \qquad \widehat{\Sigma}_{u;A} = s_{Y,T_2}\left(\widehat{\beta}_Y\right),$$

respectively. Alternatively, the parameters can be estimated using the Whittle recursion (Whittle, 1963), which, asymptotically, is equivalent to least squares estimation. The Whittle recursion makes use of the properties of block Toeplitz matrices, and is computationally fast since it involves only one $d \times d$ matrix inversion at each step. The recursion is summarised in the Appendix. The maximised conditional Gaussian log-likelihood under the alternative hypotheses is the sum of the individual ones, which is

$$\widehat{l}_A = -\frac{(T_1 + T_2)\, d}{2} \{1 + \log(2\pi)\} - \frac{T_1}{2} \log \left| \widehat{\Sigma}_{\varepsilon;A} \right| - \frac{T_2}{2} \log \left| \widehat{\Sigma}_{u;A} \right| .$$

Under the null hypotheses, the maximised conditional Gaussian log-likelihood is again the sum of the individual ones maximised over the relevant parameter subspace. That is, letting $\widehat{l}_0^{(r)}$ be the maximised log-likelihood under $H_0^{(r)}$, $r = 1, 2, 3$,

$$\widehat{l}_0^{(r)} = \max_{H_0^{(r)}} \{ l_X(\beta_X, \Sigma_\varepsilon) + l_Y(\beta_Y, \Sigma_u) \} .$$

We show in the following section how to compute the parameter estimators which maximise $\widehat{l}_0^{(r)}$.

5.3 Parameter Estimation Under the Null Hypotheses

5.3.1 Parameter Estimation Under $H_0^{(1)}$

Under $H_0^{(1)}$ $\beta_X = \beta_Y$. Letting $\beta = \begin{bmatrix} \beta_1 & \cdots & \beta_p \end{bmatrix}'$ be the common autoregressive parameters, the conditional Gaussian log-likelihood is

$$\begin{aligned} l_0^{(1)}(\beta, \Sigma_\varepsilon, \Sigma_u) = &-\frac{(T_1 + T_2)\, d}{2} \log(2\pi) - \frac{T_1}{2} \log |\Sigma_\varepsilon| - \frac{T_2}{2} \log |\Sigma_u| \\ &- \frac{T_1}{2} \operatorname{tr} \left\{ \Sigma_\varepsilon^{-1} s_{X,T_1}(\beta) \right\} - \frac{T_2}{2} \operatorname{tr} \left\{ \Sigma_u^{-1} s_{Y,T_2}(\beta) \right\} . \end{aligned}$$

This is maximised with respect to β when

$$\Sigma_\varepsilon^{-1} \beta Z_X Z_X' + \Sigma_u^{-1} \beta Z_Y Z_Y' = -\left(\Sigma_\varepsilon^{-1} X Z_X' + \Sigma_u^{-1} Y Z_Y' \right) ,$$

and maximised with respect to Σ_ε and Σ_u when $\Sigma_\varepsilon = s_{X,T_1}(\beta)$ and $\Sigma_u = s_{Y,T_2}(\beta)$, respectively. The estimates can therefore be computed iteratively as follows. Given current estimates of Σ_u and Σ_u, denoted $\widehat{\Sigma}_{\varepsilon;1}$ and $\widehat{\Sigma}_{u;1}$, estimate β by $\widehat{\beta}$ where

$$\operatorname{vec}\left(\widehat{\beta}\right) = -\left\{ \left(Z_X Z_X' \otimes \widehat{\Sigma}_{\varepsilon;1}^{-1} \right) + \left(Z_Y Z_Y' \otimes \widehat{\Sigma}_{u;1}^{-1} \right) \right\}^{-1} \left\{ \operatorname{vec}\left(\widehat{\Sigma}_{\varepsilon;1}^{-1} X Z_X' \right) + \operatorname{vec}\left(\widehat{\Sigma}_{u;1}^{-1} Y Z_Y' \right) \right\} .$$

Then re-estimate Σ_ε and Σ_u by

$$\widehat{\Sigma}_{\varepsilon;1} = s_{X,T_1}\left(\widehat{\beta}\right) \qquad \text{and} \qquad \widehat{\Sigma}_{u;1} = s_{Y,T_2}\left(\widehat{\beta}\right) ,$$

respectively. Use these new estimates to update $\widehat{\beta}$ and repeat the process until convergence. For initial estimates of Σ_u and Σ_u, we use $\widehat{\Sigma}_{\varepsilon;A}$ and $\widehat{\Sigma}_{u;A}$, respectively.

5.3.2 Parameter Estimation Under $H_0^{(2)}$

Let $\beta_X = \beta_Y = \beta$ as before. Since $\Sigma_\varepsilon = \lambda\Sigma_u$, where λ is some positive constant, the conditional Gaussian log-likelihood is

$$\begin{aligned} l_0^{(2)}(\beta, \Sigma_\varepsilon, \lambda) = &-\frac{(T_1+T_2)\,d}{2}\log(2\pi) - \frac{T_1+T_2}{2}\log|\Sigma_\varepsilon| + \frac{T_2 d}{2}\log\lambda \\ &- \frac{T_1}{2}\operatorname{tr}\left\{\Sigma_\varepsilon^{-1} s_{X,T_1}(\beta)\right\} - \lambda\frac{T_2}{2}\operatorname{tr}\left\{\Sigma_\varepsilon^{-1} s_{Y,T_2}(\beta)\right\}. \end{aligned}$$

We maximise this using a profile likelihood approach. For a given λ, $l_0^{(2)}(\beta, \Sigma_\varepsilon, \lambda)$ is maximised with respect to β by

$$\widehat{\beta}_\lambda = -\left(XZ_X' + \lambda Y Z_Y'\right)\left(Z_X Z_X' + \lambda Z_Y Z_Y'\right)^{-1}. \tag{5.4}$$

Then $l_0^{(2)}\left(\widehat{\beta}_\lambda, \Sigma_\varepsilon, \lambda\right)$ is maximised with respect to Σ_ε by

$$\widetilde{\Sigma}_{\varepsilon;\lambda} = (T_1+T_2)^{-1}\left\{T_1 s_{X,T_1}\left(\widehat{\beta}_\lambda\right) + \lambda T_2 s_{Y,T_2}\left(\widehat{\beta}_\lambda\right)\right\}, \tag{5.5}$$

and we thus obtain the profile log-likelihood

$$\begin{aligned} \widetilde{l}_0^{(2)}(\lambda) &= l_0^{(2)}\left(\widehat{\beta}_\lambda, \widetilde{\Sigma}_{\varepsilon;\lambda}, \lambda\right) \\ &= -\frac{(T_1+T_2)\,d}{2}\{1+\log(2\pi)\} - \frac{T_1+T_2}{2}\log\left|\widetilde{\Sigma}_{\varepsilon;\lambda}\right| + \frac{T_2 d}{2}\log\lambda. \end{aligned}$$

Let $\widehat{\lambda}$ be the maximiser of $\widetilde{l}_0^{(2)}(\lambda)$. The parameter estimators of β, Σ_ε and Σ_u are then

$$\widehat{\beta} = \widehat{\beta}_{\widehat{\lambda}}, \qquad \widehat{\Sigma}_{\varepsilon;2} = \widetilde{\Sigma}_{\varepsilon;\widehat{\lambda}} \qquad \text{and} \qquad \widehat{\Sigma}_{u;2} = \widehat{\Sigma}_{\varepsilon;2}/\widehat{\lambda},$$

respectively.

Any optimisation procedure can be used to maximise $\widetilde{l}_0^{(2)}(\lambda)$. The derivative of $l_0^{(2)}(\beta, \Sigma_\varepsilon, \lambda)$ with respect to λ is

$$\frac{T_2 d}{\lambda} - \frac{T_2}{2}\operatorname{tr}\left\{\Sigma_\varepsilon^{-1} s_{Y,T_2}(\beta)\right\},$$

which equals zero when

$$\lambda = d/\operatorname{tr}\left\{\Sigma_\varepsilon^{-1} s_{Y,T_2}(\beta)\right\}.$$

This suggests the following iterative procedure. Given a current estimate of λ, denoted by $\widetilde{\lambda}$, compute $\widehat{\beta}_{\widetilde{\lambda}}$ and $\widetilde{\Sigma}_{\varepsilon;\widetilde{\lambda}}$. Then re-estimate λ by

$$d/\operatorname{tr}\left\{\widetilde{\Sigma}_{\varepsilon;\widetilde{\lambda}}^{-1} s_{Y,T_2}\left(\widehat{\beta}_{\widetilde{\lambda}}\right)\right\}$$

and repeat the process until convergence. For an initial estimate of λ, we use

$$d/\operatorname{tr}\left(\widehat{\Sigma}_{\varepsilon;A}^{-1}\widehat{\Sigma}_{u;A}\right).$$

A special case of $H_0^{(2)}$ is when $\lambda = 1$. That is, the null hypothesis is that $f_X(\omega) = f_Y(\omega)$. This simplifies the procedure since λ no longer needs to be estimated. Therefore β and Σ_ε can be estimated using (5.4) and (5.5) with $\lambda = 1$, and no iteration is required.

5.3.3 Parameter Estimation Under $H_0^{(3)}$

Let $\beta_X = \beta$, then $\beta_Y = \Lambda^{-1}\beta\,(I_p \otimes \Lambda)$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$ and $\lambda_1, \ldots, \lambda_d$ are positive constants. Since $\Sigma_\varepsilon = \Lambda\Sigma_u\Lambda$, the conditional Gaussian log-likelihood is

$$\begin{aligned} l_0^{(3)}(\beta, \Sigma_\varepsilon, \Lambda) = &-\frac{(T_1+T_2)\,d}{2}\log(2\pi) - \frac{T_1+T_2}{2}\log|\Sigma_\varepsilon| + T_2\log\Lambda \\ &- \frac{T_1}{2}\operatorname{tr}\left\{\Sigma_\varepsilon^{-1}s_{X,T_1}(\beta)\right\} - \frac{T_2}{2}\operatorname{tr}\left[\Sigma_\varepsilon^{-1}\Lambda s_{Y,T_2}\left\{\Lambda^{-1}\beta\,(I_p\otimes\Lambda)\right\}\Lambda\right]. \end{aligned}$$

As before, we use a profile likelihood approach to maximise this. For a given Λ, $l_0^{(3)}(\beta, \Sigma_\varepsilon, \Lambda)$ is maximised with respect to β by

$$\widehat{\beta}_\Lambda = -\left\{XZ_X' + \Lambda YZ_Y'(I_p\otimes\Lambda)\right\}\left\{Z_XZ_X' + (I_p\otimes\Lambda)\,Z_YZ_Y'(I_p\otimes\Lambda)\right\}^{-1}.$$

Then $l_0^{(3)}\left(\widehat{\beta}_\Lambda, \Sigma_\varepsilon, \Lambda\right)$ is maximised with respect to Σ_ε by

$$\widetilde{\Sigma}_{\varepsilon;\Lambda} = (T_1+T_2)^{-1}\left[T_1 s_{X,T_1}\left(\widehat{\beta}_\Lambda\right) + T_2\Lambda s_{Y,T_2}\left\{\Lambda^{-1}\widehat{\beta}_\Lambda\,(I_p\otimes\Lambda)\right\}\Lambda\right]$$

and the profile log-likelihood is

$$\begin{aligned} \widetilde{l}_0^{(3)}(\Lambda) &= l_0^{(3)}\left(\widehat{\beta}_\Lambda, \widetilde{\Sigma}_{\varepsilon;\Lambda}, \Lambda\right) \\ &= -\frac{(T_1+T_2)\,d}{2}\left\{1+\log(2\pi)\right\} - \frac{T_1+T_2}{2}\log\left|\widetilde{\Sigma}_{\varepsilon;\Lambda}\right| + T_2\log|\Lambda|. \end{aligned}$$

Let $\widehat{\Lambda}$ be the maximiser of $\widetilde{l}_0^{(3)}(\Lambda)$. The parameter estimators of β, Σ_ε and Σ_u are then

$$\widehat{\beta} = \widehat{\beta}_{\widehat{\Lambda}}, \qquad \widehat{\Sigma}_{\varepsilon;3} = \widetilde{\Sigma}_{\varepsilon;\widehat{\Lambda}} \qquad \text{and} \qquad \widehat{\Sigma}_{u;3} = \widehat{\Lambda}^{-1}\widehat{\Sigma}_{\varepsilon;3}\widehat{\Lambda}^{-1}.$$

In practice, since we cannot obtain a closed form solution for Λ given β and Σ_ε, we make use of computer search algorithms to maximise $\widetilde{l}_0^{(3)}(\Lambda)$. For an initial estimate of Λ we let

$$\widehat{\lambda}_j = \sqrt{\frac{\sigma_{\varepsilon;A}(j,j)}{\sigma_{u;A}(j,j)}}, \qquad j = 1, \ldots, d,$$

where $\sigma_{\varepsilon;A}(i,j)$ and $\sigma_{u;A}(i,j)$ are the (i,j)th elements of $\widehat{\Sigma}_{\varepsilon;A}$ and $\widehat{\Sigma}_{u;A}$, respectively.

5.3.4 Parameter Estimation Using the Whittle Recursion

Under $H_0^{(2)}$ and $H_0^{(3)}$, the parameters can be estimated using the Whittle recursion by replacing the statistics XZ_X', YZ_Y', Z_XZ_X' and Z_YZ_Y' with their block Toeplitz versions. Let

$$\widehat{\Gamma}_X(j) = T_1^{-1}\sum_{t=j}^{T_1-1} X_{t-j}X_t' \qquad \text{and} \qquad \widehat{\Gamma}_Y(j) = T_2^{-1}\sum_{t=j}^{T_2-1} Y_{t-j}Y_t'.$$

To estimate the parameters under $H_0^{(2)}$, for a given λ, let

$$\widehat{\Gamma}_\lambda(j) = (T_1+T_2)^{-1}\left\{T_1\widehat{\Gamma}_X(j) + \lambda T_2\widehat{\Gamma}_Y(j)\right\}.$$

Then let $\widehat{\Gamma}_\lambda$ be the $dp \times dp$ block matrix with (i,j)th block $\widehat{\Gamma}_\lambda(i-j)$, $i \geqslant j$, and $\widehat{\Gamma}'_\lambda(j-i)$, $i < j$. Also let $\widehat{\gamma}_\lambda$ be the $d \times dp$ block matrix with $(1,j)$th block $\widehat{\Gamma}'_\lambda(j)$. The estimators of β and Σ_ε are then

$$\widehat{\beta}_\lambda = -\widehat{\gamma}_\lambda \widehat{\Gamma}_\lambda^{-1}$$

and

$$\widetilde{\Sigma}_{\varepsilon;\lambda} = \widehat{\Gamma}_\lambda(0) - \widehat{\gamma}_\lambda \widehat{\Gamma}_\lambda^{-1} \widehat{\gamma}'_\lambda,$$

which can be computed using the Whittle recursion.

To estimate the parameters under $H_0^{(3)}$, for a given Λ, let

$$\widehat{\Gamma}_\Lambda(j) = (T_1+T_2)^{-1}\left\{T_1\widehat{\Gamma}_X(j) + T_2\Lambda\widehat{\Gamma}_Y(j)\Lambda\right\}.$$

Then let $\widehat{\Gamma}_\Lambda$ be the $dp \times dp$ block matrix with (i,j)th block $\widehat{\Gamma}_\Lambda(i-j)$, $i \geqslant j$, and $\widehat{\Gamma}'_\Lambda(j-i)$, $i < j$. Also let $\widehat{\gamma}_\Lambda$ be the $d \times dp$ block matrix with $(1,j)$th block $\widehat{\Gamma}'_\Lambda(j)$. The estimators of β and Σ_ε are then

$$\widehat{\beta}_\Lambda = -\widehat{\gamma}_\Lambda \widehat{\Gamma}_\Lambda^{-1}$$

and

$$\widetilde{\Sigma}_{\varepsilon;\Lambda} = \widehat{\Gamma}_\Lambda(0) - \widehat{\gamma}_\Lambda \widehat{\Gamma}_\Lambda^{-1} \widehat{\gamma}'_\Lambda,$$

which can be computed using the Whittle recursion.

5.4 The Test Statistics

The maximised conditional Gaussian log-likelihood under $H_0^{(r)}$, $r = 1,2,3$, is

$$\widehat{l}_0^{(r)} = -\frac{(T_1+T_2)\,d}{2}\left\{1+\log(2\pi)\right\} - \frac{T_1}{2}\log\left|\widehat{\Sigma}_{\varepsilon;r}\right| - \frac{T_2}{2}\log\left|\widehat{\Sigma}_{u;r}\right|.$$

Thus, the test statistic for $H_0^{(r)}$ is

$$\begin{aligned}\theta^{(r)} &= 2\left(\widehat{l}_A - \widehat{l}_0^{(r)}\right)\\ &= T_1\log\left(\frac{\left|\widehat{\Sigma}_{\varepsilon;r}\right|}{\left|\widehat{\Sigma}_{\varepsilon;A}\right|}\right) + T_2\log\left(\frac{\left|\widehat{\Sigma}_{u;r}\right|}{\left|\widehat{\Sigma}_{u;A}\right|}\right).\end{aligned}$$

The null hypothesis is rejected at significance level α when $\theta^{(r)}$ is greater than the $100(1-\alpha)$th percentile of the χ^2 distribution with $v^{(r)}$ degrees of freedom, where

$$\begin{aligned}v^{(1)} &= d^2p,\\ v^{(2)} &= d^2p + d(d+1)/2 - 1,\\ v^{(3)} &= d^2p + d(d+1)/2 - d.\end{aligned}$$

5.5 Asymptotic Properties of the Estimators Under the Null Hypotheses

In this section we establish the strong consistency of the estimators under the null hypotheses, and also establish the central limit theorem for the estimators of the autoregressive parameters. In order to prove the central limit theorem under $H_0^{(1)}$ we need to make the assumption that $T_2 = \kappa T_1$ for some constant κ. That is, that the sample sizes increase at the same rate. This assumption is not needed for the other theorems. The proofs of the theorems are in the Appendix.

Let Γ_X and Γ_Y be the $dp \times dp$ matrices with (i,j)th block given by $\Gamma_X(i-j)$ and $\Gamma_Y(i-j)$, respectively. Note that

$$T_1^{-1} Z_X Z_X' \to \Gamma_X \qquad \text{and} \qquad T_2^{-1} Z_Y Z_Y' \to \Gamma_Y$$

almost surely as $T_1 \to \infty$ and $T_2 \to \infty$, respectively. In the theorems below and their proofs, a parameter written with a 0 in the subscript will denote the true value of that parameter.

Theorem 5.1 *Under $H_0^{(1)}$, $\widehat{\beta} \to \beta_0$, $\widehat{\Sigma}_{\varepsilon;1} \to \Sigma_{\varepsilon 0}$ and $\widehat{\Sigma}_{u;1} \to \Sigma_{u0}$ almost surely as $T_1, T_2 \to \infty$.*

Theorem 5.2 *Under $H_0^{(2)}$, $\widehat{\lambda} \to \lambda_0$, $\widehat{\beta}_{\widehat{\lambda}} \to \beta_0$ and $\widetilde{\Sigma}_{\varepsilon;\widehat{\lambda}} \to \Sigma_{\varepsilon 0}$ almost surely as $T_1, T_2 \to \infty$.*

Theorem 5.3 *Under $H_0^{(3)}$, $\widehat{\Lambda} \to \Lambda_0$, $\widehat{\beta}_{\widehat{\Lambda}} \to \beta_0$ and $\widetilde{\Sigma}_{\varepsilon;\widehat{\Lambda}} \to \Sigma_{\varepsilon 0}$ almost surely as $T_1, T_2 \to \infty$.*

In order to prove the central limit theorems we will make use of the following lemma. The proof of the lemma is in the Appendix.

Lemma 5.1 *The distribution of $T_1^{-1/2} \operatorname{vec}(\varepsilon Z_X')$ converges to the normal distribution with mean zero and covariance matrix $\Gamma_X \otimes \Sigma_\varepsilon$ as $T_1 \to \infty$. The distribution of $T_2^{-1/2} \operatorname{vec}(u Z_Y')$ converges to the normal distribution with mean zero and covariance matrix $\Gamma_Y \otimes \Sigma_u$ as $T_2 \to \infty$.*

Theorem 5.4 *Let $T_2 = \kappa T_1$ for some constant κ. Then, under $H_0^{(1)}$, the distribution of $T_1^{1/2} \operatorname{vec}\left(\widehat{\beta} - \beta_0\right)$ converges to the normal distribution with mean zero and covariance matrix*

$$\left\{ \left(\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1} \right) + \kappa \left(\Gamma_Y \otimes \Sigma_{u0}^{-1} \right) \right\}^{-1},$$

as $T_1 \to \infty$.

Theorem 5.5 *Under $H_0^{(2)}$, the distribution of $(T_1+T_2)^{1/2}\operatorname{vec}\left(\widehat{\beta}-\beta_0\right)$ converges to the normal distribution with mean zero and covariance matrix*

$$\Gamma_X^{-1}\otimes\Sigma_{\varepsilon 0}=\Gamma_Y^{-1}\otimes\Sigma_{u0},$$

as $T_1,T_2\to\infty$.

Theorem 5.6 *Under $H_0^{(3)}$, the distribution of $(T_1+T_2)^{1/2}\operatorname{vec}\left(\widehat{\beta}-\beta_0\right)$ converges to the normal distribution with mean zero and covariance matrix*

$$\Gamma_X^{-1}\otimes\Sigma_{\varepsilon 0}=\Gamma_Y^{-1}\otimes\Sigma_{u0},$$

as $T_1,T_2\to\infty$.

5.6 Comparing More Than Two Time Series

In this section we show how to extend the above procedures to the case of comparing more than two time series. Let $\{X_{k,t}\}$ be a d-dimensional stationary process. If we have n such processes, we fit VAR(p) models to samples of size T_k from $\{X_{k,t}\}$, $k=1,\ldots,n$. That is, for $k=1,\ldots,n$,

$$X_{k,t}+\beta_{k,1}X_{k,t-1}+\cdots+\beta_{k,p}X_{k,t-p}=\varepsilon_{k,t},$$

where $\beta_{k,j}$, $j=1,\ldots,p$, $k=1,\ldots,n$, are $d\times d$ and $\{\varepsilon_{k,t}\}$, $k=1,\ldots,n$, are independent d-dimensional innovation processes. We make the usual assumptions that the $\{\varepsilon_{k,t}\}$ are sequences of martingale differences and that

$$E\left(\varepsilon_{k,t}\varepsilon_{k,t}'\mid\mathcal{F}_{k,t-1}\right)=\Sigma_k,$$

where $\mathcal{F}_{k,t}$ is the σ-field generated by $\{\varepsilon_{k,t},\varepsilon_{k,t-1},\ldots\}$. We redefine the three null hypotheses as follows.

$$H_0^{(1)}:\beta_{1,j}=\cdots=\beta_{n,j}\ \forall j,$$

$$H_0^{(2)}:\beta_{1,j}=\cdots=\beta_{n,j}\ \forall j,\ \Sigma_1=\lambda_2\Sigma_2=\cdots=\lambda_n\Sigma_n,$$

where $\lambda_2,\ldots\lambda_n$ are positive constants, and

$$H_0^{(3)}:\beta_{1,j}=\Lambda_2\beta_{2,j}\Lambda_2^{-1}=\cdots=\Lambda_n\beta_{n,j}\Lambda_n^{-1}\ \forall j,\ \Sigma_1=\Lambda_2\Sigma_2\Lambda_2=\cdots=\Lambda_n\Sigma_n\Lambda_n,$$

where $\Lambda_k=\operatorname{diag}\left(\lambda_{k,1},\ldots,\lambda_{k,d}\right)$ and $\lambda_{k,1},\ldots,\lambda_{k,d}$, $k=2,\ldots n$, are positive constants. In each case, the alternative hypothesis is the complement of the null hypothesis. In what follows, the test statistics will be derived for fixed p. In practice we let $p=\lfloor\log\left(T_{\min}\right)^c\rfloor$, where $T_{\min}=\min\left(T_1,\ldots T_n\right)$ and $c>1$.

Let

$$X_k + \beta_k Z_k = \varepsilon_k,$$

where

$$X_k = \left[\begin{array}{ccc} X_{k,p} & \cdots & X_{k,T_k-1} \end{array}\right],$$

$$Z_k = \left[\begin{array}{ccc} Z_{k,p} & \cdots & Z_{k,T_k-1} \end{array}\right],$$

$$Z_{k,t} = \left[\begin{array}{ccc} X'_{k,t-1} & \cdots & X'_{k,t-p} \end{array}\right]',$$

$$\varepsilon_k = \left[\begin{array}{ccc} \varepsilon_{k,p} & \cdots & \varepsilon_{k,T_1-1} \end{array}\right],$$

$$\beta_k = \left[\begin{array}{ccc} \beta_{k,1} & \cdots & \beta_{k,p} \end{array}\right].$$

Under the alternative hypotheses, the estimators of β_k and Σ_k are

$$\widehat{\beta}_k = -X_k Z'_k \left(Z_k Z'_k\right)^{-1}$$

and

$$\widehat{\Sigma}_{k;A} = T_k^{-1} \left(X_k + \beta Z_k\right) \left(X_k + \beta Z_k\right)',$$

$k = 1, \ldots, n$, and the maximised conditional Gaussian log-likelihood is

$$\widehat{l}_A = -\frac{d}{2} \left\{1 + \log\left(2\pi\right)\right\} \sum_{j=1}^{n} T_j - \sum_{j=1}^{n} \frac{T_j}{2} \log \left|\widehat{\Sigma}_{j;A}\right|.$$

In order to maximise the conditional Gaussian log-likelihood under $H_0^{(1)}$, we extend the procedure of Section 5.3.1 as follows. Given current estimates of Σ_k, denoted $\widehat{\Sigma}_{k;1}$, $k = 1, \ldots, n$, estimate β by $\widehat{\beta}$ where

$$\operatorname{vec}\left(\widehat{\beta}\right) = -\left\{\sum_{j=1}^{n} \left(Z_j Z'_j \otimes \widehat{\Sigma}_{j;1}^{-1}\right)\right\}^{-1} \left\{\sum_{j=1}^{n} \operatorname{vec}\left(\widehat{\Sigma}_{j;1}^{-1} X_j Z'_j\right)\right\}.$$

Then re-estimate Σ_k by

$$\widehat{\Sigma}_{k;1} = T_k^{-1} \left(X_k + \widehat{\beta} Z_k\right) \left(X_k + \widehat{\beta} Z_k\right)',$$

$k = 1, \ldots, n$, and repeat the process until convergence.

Under $H_0^{(2)}$, the profile log-likelihood is

$$\widetilde{l}_0^{(2)}\left(\lambda\right) = -\frac{d}{2} \left\{1 + \log\left(2\pi\right)\right\} \sum_{j=1}^{n} T_j - \frac{1}{2} \left(\sum_{j=1}^{n} T_j\right) \log \left|\widetilde{\Sigma}_\lambda\right| + \frac{d}{2} \sum_{j=2}^{n} T_j \log \lambda_j,$$

where

$$\lambda = \left[\begin{array}{ccc} \lambda_2 & \cdots & \lambda_n \end{array}\right]',$$

$$\widetilde{\Sigma}_{\lambda} = \left(\sum_{j=1}^{n} T_j\right)^{-1} \left\{\sum_{j=1}^{n} \lambda_j \left(X_j + \widehat{\beta}_{\lambda} Z_j\right) \left(X_j + \widehat{\beta}_{\lambda} Z_j\right)'\right\},$$

$$\widehat{\beta}_{\lambda} = -\left(\sum_{j=1}^{n} \lambda_j X_j Z_j'\right) \left(\sum_{j=1}^{n} \lambda_j Z_j Z_j'\right)^{-1}$$

and $\lambda_1 = 1$. Letting $\widehat{\lambda} = \left[\begin{array}{ccc} \widehat{\lambda}_2 & \cdots & \widehat{\lambda}_n \end{array}\right]$ be the maximiser of $\widetilde{l}_0^{(2)}(\lambda)$, the parameter estimators of β and Σ_k, $k = 1, \ldots, n$, are

$$\widehat{\beta} = \widehat{\beta}_{\widehat{\lambda}}, \qquad \widehat{\Sigma}_{1;2} = \widetilde{\Sigma}_{\widehat{\lambda}} \qquad \text{and} \qquad \widehat{\Sigma}_{k;2} = \widehat{\Sigma}_{1;2}/\widehat{\lambda}_k,\ k \geqslant 2.$$

We can use a similar iterative procedure to maximise $\widetilde{l}_0^{(2)}(\lambda)$ as that given in Section 5.3.2. Given a current estimate of λ, denoted by $\widetilde{\lambda}$, compute $\widehat{\beta}_{\widetilde{\lambda}}$ and $\widetilde{\Sigma}_{\widetilde{\lambda}}$. Then re-estimate λ_k, $k = 2, \ldots, n$, by

$$\widehat{\lambda}_k = d/\operatorname{tr}\left[\widetilde{\Sigma}_{\widetilde{\lambda}}^{-1}\left\{T_k^{-1}\left(X_k + \widehat{\beta}_{\widetilde{\lambda}} Z_k\right)\left(X_k + \widehat{\beta}_{\widetilde{\lambda}} Z_k\right)'\right\}\right]$$

and repeat the process until convergence. Initial estimates of λ_k can be obtained from

$$d/\operatorname{tr}\left(\widehat{\Sigma}_{1;A}^{-1}\widehat{\Sigma}_{k;A}\right),$$

$k = 1, \ldots, n$.

Under $H_0^{(3)}$, the profile log-likelihood is

$$\widetilde{l}_0^{(3)}(\Lambda) = -\frac{d}{2}\left\{1 + \log(2\pi)\right\}\sum_{j=1}^{n} T_j - \frac{1}{2}\left(\sum_{j=1}^{n} T_j\right)\log\left|\widetilde{\Sigma}_{\Lambda}\right| + \sum_{j=2}^{n} T_j \log\left|\Lambda_j\right|,$$

where

$$\Lambda = \left[\begin{array}{ccc} \Lambda_2 & \cdots & \Lambda_n \end{array}\right],$$

$$\widetilde{\Sigma}_{\Lambda} = \left(\sum_{j=1}^{n} T_j\right)^{-1}\left[\sum_{j=1}^{n} \Lambda_j \left\{X_j + \Lambda_j^{-1}\widehat{\beta}_{\Lambda}\left(I_p \otimes \Lambda_j\right) Z_k\right\}\left\{X_j + \Lambda_j^{-1}\widehat{\beta}_{\Lambda}\left(I_p \otimes \Lambda_j\right) Z_k\right\}' \Lambda_j\right],$$

$$\widehat{\beta}_{\Lambda} = -\left\{\sum_{j=1}^{n} \Lambda_j X_j Z_j' \left(I_p \otimes \Lambda_j\right)\right\}\left\{\sum_{j=1}^{n} \left(I_p \otimes \Lambda_j\right) Z_j Z_j' \left(I_p \otimes \Lambda_j\right)\right\}^{-1}$$

and $\Lambda_1 = I_d$. Letting $\widehat{\Lambda} = \left[\begin{array}{ccc} \widehat{\Lambda}_2 & \cdots & \widehat{\Lambda}_n \end{array}\right]$ be the maximiser of $\widetilde{l}_0^{(3)}(\Lambda)$, the parameter estimators of β and Σ_k, $k = 1, \ldots, n$, are

$$\widehat{\beta} = \widehat{\beta}_{\widehat{\Lambda}}, \qquad \widehat{\Sigma}_{1;3} = \widetilde{\Sigma}_{\widehat{\Lambda}} \qquad \text{and} \qquad \widehat{\Sigma}_{k;3} = \widehat{\Lambda}_k^{-1}\widehat{\Sigma}_{1;3}\widehat{\Lambda}_k^{-1},\ k \geqslant 2.$$

As before, we make use of computer search algorithms to maximise $\widetilde{l}_0(\Lambda)$ in practice, with initial estimates of Λ_k, $k = 2, \ldots, n$, given by

$$\widehat{\lambda}_{k,j} = \sqrt{\frac{\sigma_{1;A}(j,j)}{\sigma_{k;A}(j,j)}},$$

$j = 1, \ldots, d$, where $\sigma_{k;A}(i, j)$ is the (i, j)th element of $\widehat{\Sigma}_{k;A}$.

The test statistic under $H_0^{(r)}$, $r = 1, 2, 3$, is

$$\theta^{(r)} = \sum_{j=1}^{n} T_j \log\left(\frac{\left|\widehat{\Sigma}_{j;r}\right|}{\left|\widehat{\Sigma}_{j;A}\right|}\right).$$

The null hypothesis is rejected at significance level α when $\theta^{(r)}$ is greater than the $100(1-\alpha)$th percentile of the χ^2 distribution with $v^{(r)}$ degrees of freedom, where

$$\begin{aligned} v^{(1)} &= (n-1)\, d^2 p, \\ v^{(2)} &= (n-1)\, d^2 p + (n-1)\, d\,(d+1)/2 - (n-1), \\ v^{(3)} &= (n-1)\, d^2 p + (n-1)\, d\,(d+1)/2 - (n-1)\, d. \end{aligned}$$

5.7 Simulations

In order to examine the behaviour of the test statistics under the null hypotheses when comparing two time series, that is when $n = 2$, the tests were applied to pairs of time series which were simulated from either the VAR(1) processes

$$X_t + \beta_1 X_{t-1} = \varepsilon_t \qquad \text{and} \qquad Y_t + \beta_1^* Y_{t-1} = u_t,$$

the VAR(2) processes

$$X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} = \varepsilon_t \qquad \text{and} \qquad Y_t + \beta_1^* Y_{t-1} + \beta_2^* Y_{t-2} = u_t,$$

the VMA(1) processes

$$X_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} \qquad \text{and} \qquad Y = u_t + \alpha_1^* u_{t-1},$$

the VMA(2) processes

$$X_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} \qquad \text{and} \qquad Y = u_t + \alpha_1^* u_{t-1} + \alpha_2^* u_{t-1},$$

or the VARMA$(1, 1)$ processes

$$X_t + \beta_1 X_{t-1} = \varepsilon_t + \alpha_1 \varepsilon_{t-1} \qquad \text{and} \qquad Y_t + \beta_1^* Y_{t-1} = u_t + \alpha_1^* u_{t-1}.$$

Note that VMA(q) denotes a vector moving average process of order q and VARMA(p, q) denotes a vector autoregressive-moving average process of orders p and q. Simulations were run for $d = 2$ and $d = 3$. In all cases, $\{\varepsilon_t\}$ was simulated from the normal distribution with covariance matrix $\Sigma_\varepsilon = I_d$. In the simulations run under $H_0^{(1)}$ and $H_0^{(2)}$, $\beta_1^* = \beta_1$,

$\beta_2^* = \beta_2$, $\alpha_1^* = \alpha_1$, $\alpha_2^* = \alpha_2$ and $\{u_t\}$ was simulated from the normal distribution with mean zero and covariance matrix $\Sigma_u = 2\Sigma_\varepsilon$. For the simulations run under $H_0^{(3)}$, $\beta_1^* = \Lambda_2^{-1}\beta_1\Lambda_2$, $\beta_2^* = \Lambda_2^{-1}\beta_2\Lambda_2$, $\alpha_1^* = \Lambda_2^{-1}\alpha_1\Lambda_2$, $\alpha_2^* = \Lambda_2^{-1}\alpha_2\Lambda_2$ and $\{u_t\}$ was simulated from the normal distribution with mean zero and covariance matrix $\Sigma_u = \Lambda_2^{-1}\Sigma_\varepsilon\Lambda_2^{-1}$. The parameters when $d = 2$ were

$$\beta_1 = \begin{bmatrix} 0.7 & 0.3 \\ -0.3 & 0.7 \end{bmatrix}, \qquad \beta_2 = \begin{bmatrix} 0.3 & 0.1 \\ -0.2 & 0.3 \end{bmatrix},$$

$$\alpha_1 = \begin{bmatrix} 0.8 & 0.1 \\ -0.1 & 0.8 \end{bmatrix}, \qquad \alpha_2 = \begin{bmatrix} 0.2 & 0.2 \\ -0.1 & 0.2 \end{bmatrix} \qquad \text{and} \qquad \Lambda_2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 1.5 \end{bmatrix}.$$

The parameters when $d = 3$ were

$$\beta_1 = \begin{bmatrix} 0.7 & 0.3 & 0.1 \\ -0.3 & 0.7 & 0.2 \\ 0.1 & 0.05 & 0.7 \end{bmatrix}, \qquad \beta_2 = \begin{bmatrix} 0.3 & 0.1 & 0.1 \\ -0.2 & 0.3 & 0.05 \\ 0.1 & 0.05 & 0.5 \end{bmatrix},$$

$$\alpha_1 = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ -0.1 & 0.8 & 0.05 \\ 0.1 & 0.05 & 0.4 \end{bmatrix}, \qquad \alpha_2 = \begin{bmatrix} 0.2 & 0.2 & 0.05 \\ -0.2 & 0.2 & 0 \\ 0.05 & 0 & 0.2 \end{bmatrix}$$

and

$$\Lambda_2 = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Figure 5.1 shows the spectral densities for each component of the $\{X_t\}$ processes for the $d = 2$ case. Also shown is their coherency, which is

$$\frac{|f_{X,12}(\omega)|^2}{f_{X,11}(\omega)\, f_{X,22}(\omega)},$$

where $f_{X,ij}(\omega)$ is the (i,j)th element of $f_X(\omega)$. Figure 5.2 shows the spectral densities for each component of the $\{X_t\}$ processes for the $d = 3$ case.

The simulations were run with sample sizes of $T_1 = 1,000$ and $T_2 = 1,250$, and also $T_1 = 2,000$ and $T_2 = 2,500$. The tests were applied by fitting vector autoregressions of order 7, which is the integer component of both $\log(1,000)^{1.01}$ and $\log(2,000)^{1.01}$. When computing the test statistic under $H_0^{(3)}$, the fminsearch function in MATLAB was used.

Tables 5.1–5.6 give the means and variances of the resulting test statistics as well as the Type I error rates, that is, the proportion of times the null hypothesis was rejected at the 5% significance level. The means and variances are mostly close to the theoretical means and variances of the χ^2 distribution with the indicated degrees of freedom. For the smaller

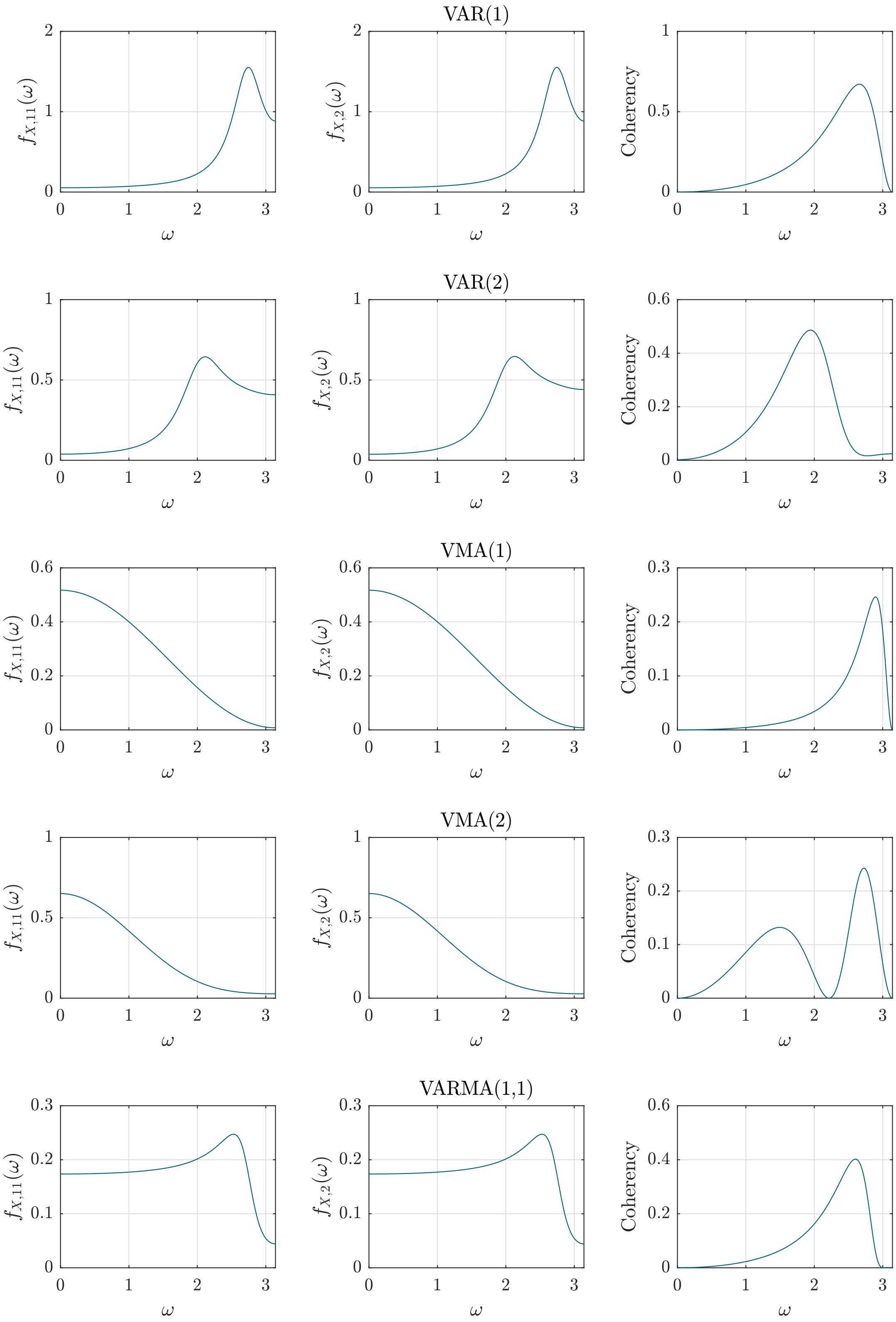

Figure 5.1: The spectral densities of each component, as well as their coherency, for each 2-dimensional $\{X_t\}$ process described in Section 5.7.

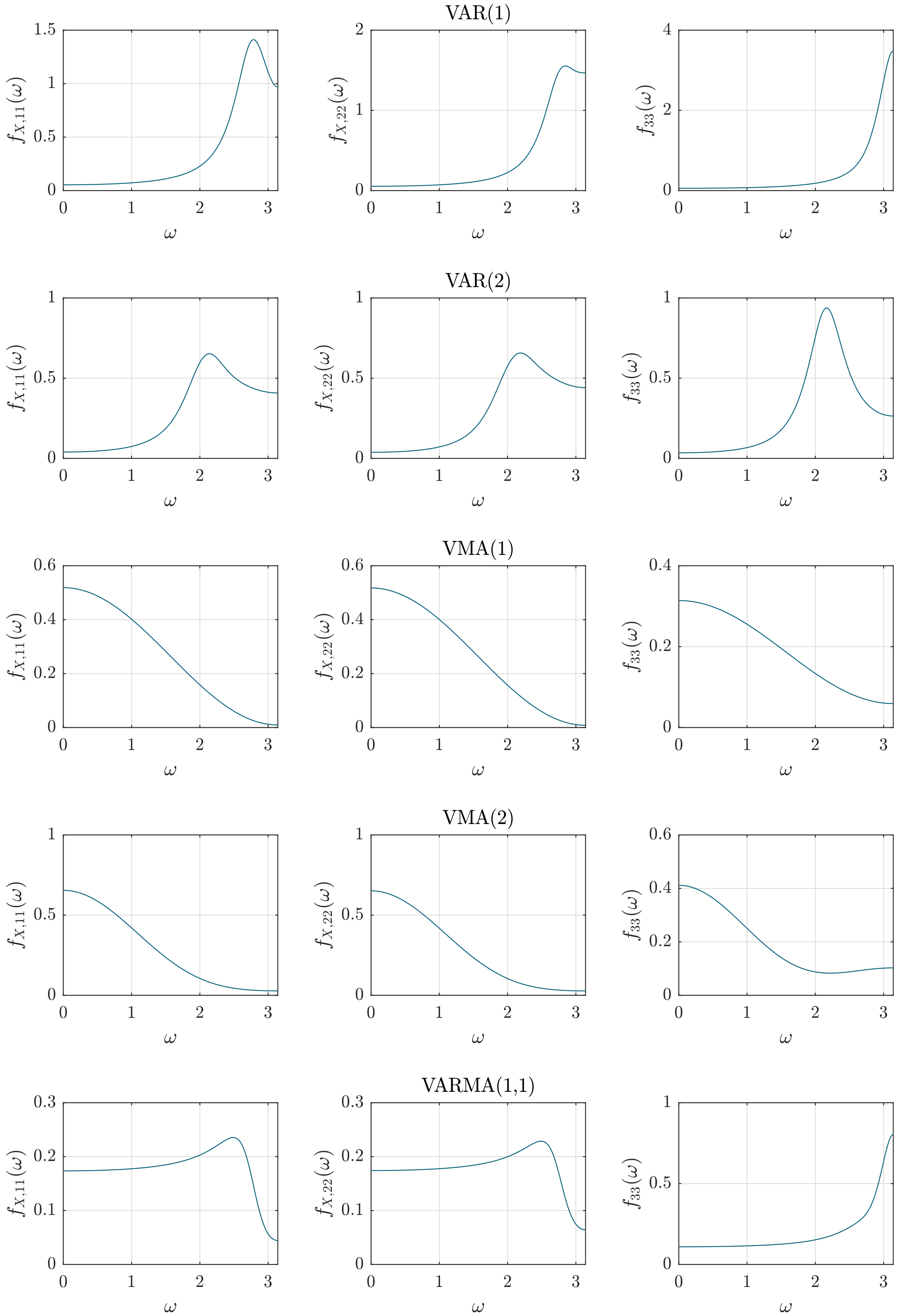

Figure 5.2: The spectral densities of each component for each 3-dimensional $\{X_t\}$ process described in Section 5.7.

sample sizes, the Type I error rates are mostly close to 0.06. For the larger sample sizes the Type I errors are mostly close to 0.05.

To examine the behaviour of the test statistics under the null hypotheses when comparing more than two time series, a third time series, $\{Z_t\}$, was simulated along with $\{X_t\}$ and $\{Y_t\}$ from either the VAR(1) process

$$Z_t + \beta_1^{**} Z_{t-1} = w_t,$$

the VAR(2) process

$$Z_t + \beta_1^{**} Z_{t-1} + \beta_2^{**} Z_{t-2} = w_t,$$

the VMA(1) process

$$Z_t = w_t + \alpha_1^{**} w_{t-1},$$

the VMA(2) process

$$Z_t = w_t + \alpha_1^{**} w_{t-1} + \alpha_2^{**} w_{t-2},$$

or the VARMA$(1, 1)$ process

$$Z_t + \beta_1^{**} Z_{t-1} = w_t + \alpha_1^{**} w_{t-1}.$$

For the simulations run under $H_0^{(1)}$ and $H_0^{(2)}$, $\beta_1^{**} = \beta_1^* = \beta_1$, $\beta_2^{**} = \beta_2^* = \beta_2$, $\alpha_1^{**} = \alpha_1^* = \alpha_1$, $\alpha_2^{**} = \alpha_2^* = \alpha_2$ and $\{w_t\}$ was simulated from the normal distribution with mean zero and covariance matrix $\Sigma_w = 0.5\Sigma_\varepsilon$. For the simulations run under $H_0^{(3)}$, the autoregressive and moving average parameters were given by $\beta_1 = \Lambda_2 \beta_1^* \Lambda_2^{-1} = \Lambda_3 \beta_1^{**} \Lambda_3^{-1}$, $\beta_2 = \Lambda_2 \beta_2^* \Lambda_2^{-1} = \Lambda_3 \beta_2^{**} \Lambda_3^{-1}$, $\alpha_1 = \Lambda_2 \alpha_1^* \Lambda_2^{-1} = \Lambda_3 \alpha_1^{**} \Lambda_3^{-1}$ and $\alpha_2 = \Lambda_2 \alpha_2^* \Lambda_2^{-1} = \Lambda_3 \alpha_2^{**} \Lambda_3^{-1}$, and $\{w_t\}$ was simulated from the normal distribution with mean zero and covariance matrix given by $\Sigma_\varepsilon = \Lambda_2 \Sigma_u \Lambda_2 = \Lambda_3 \Sigma_w \Lambda_3$, where

$$\Lambda_3 = \begin{bmatrix} 0.75 & 0 \\ 0 & 1.25 \end{bmatrix}$$

when $d = 2$ and

$$\Lambda_3 = \begin{bmatrix} 0.75 & 0 & 0 \\ 0 & 1.25 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

when $d = 3$.

Tables 5.7–5.12 give the means, variances and Type I error rates of the resulting test statistics. The means and variances are mostly close to their theoretical values, although they are a little higher for the $d = 3$ simulations, particularly for the shorter time series. For the shorter time series, the Type I error rates are around 0.06 for the $d = 2$ cases and around 0.07 for the $d = 3$ cases. For the longer time series, the Type I error rates are closer to 0.05.

Table 5.1: Summary of simulations under $H_0^{(1)}$ when $n = 2$ and $d = 2$.

		$T_1 = 1,000$			$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	28	28.312	56.394	0.055	28.192	57.454	0.057
VAR(2)	28	28.226	54.386	0.051	28.189	57.770	0.054
VMA(1)	28	28.547	58.515	0.061	28.156	56.892	0.053
VMA(2)	28	28.542	58.478	0.056	28.326	57.302	0.056
VARMA(1, 1)	28	28.306	56.609	0.053	28.159	57.049	0.053

Table 5.2: Summary of simulations under $H_0^{(2)}$ when $n = 2$ and $d = 2$.

		$T_1 = 1,000$			$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	30	30.437	61.471	0.056	30.281	61.465	0.056
VAR(2)	30	30.605	61.867	0.058	30.236	61.586	0.053
VMA(1)	30	30.459	62.303	0.060	30.289	62.788	0.057
VMA(2)	30	30.573	63.783	0.063	30.266	60.743	0.053
VARMA(1, 1)	30	30.217	63.549	0.055	30.130	59.270	0.051

Table 5.3: Summary of simulations under $H_0^{(3)}$ when $n = 2$ and $d = 2$.

		$T_1 = 1,000$			$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	29	29.452	59.250	0.058	29.250	58.319	0.051
VAR(2)	29	29.475	60.035	0.057	29.184	57.873	0.052
VMA(1)	29	29.415	59.114	0.059	29.197	59.852	0.056
VMA(2)	29	29.594	59.548	0.059	29.200	56.510	0.051
VARMA(1, 1)	29	29.419	59.355	0.056	29.168	58.293	0.052

Table 5.4: Summary of simulations under $H_0^{(1)}$ when $n = 2$ and $d = 3$.

	$T_1 = 1,000$				$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	63	64.357	130.537	0.066	63.592	127.580	0.055
VAR(2)	63	64.091	132.386	0.065	63.694	126.771	0.055
VMA(1)	63	64.452	132.225	0.065	63.808	127.054	0.055
VMA(2)	63	64.482	131.225	0.063	63.637	126.864	0.055
VARMA(1, 1)	63	64.171	131.726	0.062	63.525	130.989	0.057

Table 5.5: Summary of simulations under $H_0^{(2)}$ when $n = 2$ and $d = 3$.

	$T_1 = 1,000$				$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	68	69.285	138.706	0.063	68.762	139.288	0.056
VAR(2)	68	69.531	143.669	0.070	68.747	140.467	0.058
VMA(1)	68	69.595	143.638	0.070	68.569	142.632	0.056
VMA(2)	68	69.624	143.532	0.065	68.902	143.231	0.060
VARMA(1, 1)	68	69.445	146.348	0.073	68.771	141.682	0.060

Table 5.6: Summary of simulations under $H_0^{(3)}$ when $n = 2$ and $d = 3$.

	$T_1 = 1,000$				$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	66	67.308	138.384	0.066	66.780	137.056	0.059
VAR(2)	66	67.382	138.232	0.068	66.614	132.128	0.056
VMA(1)	66	67.605	139.483	0.073	66.522	131.531	0.051
VMA(2)	66	67.550	136.227	0.068	66.778	136.942	0.060
VARMA(1, 1)	66	67.320	136.891	0.067	66.583	138.284	0.057

Table 5.7: Summary of simulations under $H_0^{(1)}$ when $n = 3$ and $d = 2$.

	$T_1 = 1,000$				$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	56	56.909	114.701	0.059	56.359	114.595	0.056
VAR(2)	56	56.683	112.586	0.056	56.475	113.747	0.054
VMA(1)	56	56.728	117.763	0.062	56.158	113.059	0.053
VMA(2)	56	57.116	117.236	0.064	56.569	110.778	0.056
VARMA(1, 1)	56	56.428	113.487	0.053	56.359	114.700	0.056

Table 5.8: Summary of simulations under $H_0^{(2)}$ when $n = 3$ and $d = 2$.

	$T_1 = 1,000$				$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	60	60.918	123.845	0.060	60.252	120.459	0.052
VAR(2)	60	60.985	123.060	0.061	60.286	116.923	0.050
VMA(1)	60	60.707	123.230	0.058	60.237	124.625	0.057
VMA(2)	60	61.263	123.257	0.065	60.733	121.923	0.060
VARMA(1, 1)	60	60.692	123.090	0.057	60.154	119.101	0.051

Table 5.9: Summary of simulations under $H_0^{(3)}$ when $n = 3$ and $d = 2$.

	$T_1 = 1,000$				$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	58	58.820	119.354	0.060	58.507	116.387	0.054
VAR(2)	58	58.662	117.137	0.057	58.608	115.720	0.056
VMA(1)	58	58.690	122.232	0.059	58.377	121.126	0.057
VMA(2)	58	59.065	124.436	0.068	58.535	121.191	0.058
VARMA(1, 1)	58	58.670	118.422	0.058	58.132	117.195	0.052

Table 5.10: Summary of simulations under $H_0^{(1)}$ when $n = 3$ and $d = 3$.

		$T_1 = 1,000$			$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	126	128.466	258.064	0.071	127.169	262.934	0.058
VAR(2)	126	128.484	260.588	0.067	127.304	254.056	0.058
VMA(1)	126	128.245	260.805	0.067	127.212	254.869	0.057
VMA(2)	126	128.479	267.426	0.072	127.253	254.975	0.056
VARMA(1, 1)	126	128.162	267.925	0.068	126.908	260.099	0.057

Table 5.11: Summary of simulations under $H_0^{(2)}$ when $n = 3$ and $d = 3$.

		$T_1 = 1,000$			$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	136	138.673	278.539	0.071	137.303	277.509	0.058
VAR(2)	136	138.647	286.885	0.070	137.181	278.882	0.060
VMA(1)	136	138.763	284.278	0.071	137.316	279.728	0.057
VMA(2)	136	138.641	292.403	0.073	137.696	276.100	0.058
VARMA(1, 1)	136	138.256	285.143	0.069	136.855	282.930	0.060

Table 5.12: Summary of simulations under $H_0^{(3)}$ when $n = 3$ and $d = 3$.

		$T_1 = 1,000$			$T_1 = 2,000$		
	df	Mean	Var	Type I	Mean	Var	Type I
VAR(1)	132	134.340	274.708	0.067	133.416	272.047	0.061
VAR(2)	132	134.495	269.531	0.068	133.451	257.342	0.055
VMA(1)	132	134.345	274.657	0.069	133.509	277.072	0.062
VMA(2)	132	134.589	278.575	0.071	133.381	268.399	0.059
VARMA(1, 1)	132	134.399	278.490	0.068	132.982	270.539	0.057

5.A Appendix

In what follows, where convergence is indicated, it will mean convergence in the almost sure sense, unless otherwise indicated. Where order notation is used, it will also mean orders in the almost sure sense.

5.A.1 The Whittle Recursion

Consider the vector autoregression of order k,

$$X_t + \sum_{j=1}^{k} \beta_{p,j} X_{t-j} = \varepsilon_t,$$

where $E\left(\varepsilon_t \varepsilon_t'\right) = G_k$. The Whittle recursion, originally due to Whittle (1963), computes $\beta_{k,1}, \ldots, \beta_{k,k}$ and G_k, for successive values of k, given the autocovariances $\Gamma(j) = E\left(X_t X_{t+j}'\right)$. The version given below includes modifications from Quinn (1980).

Let $G_0 = \overline{G}_0 = \Gamma(0)$. Then, for $k = 0, \ldots,$

$$\Delta_{k+1} = \Gamma(-k-1) + \sum_{j=1}^{k} \beta_{k,j} \Gamma(j-k-1)$$

$$\overline{\Delta}_{k+1} = \Gamma(k+1) + \sum_{j=1}^{k} \overline{\beta}_{k,j} \Gamma(k+1-j)$$

$$\beta_{k+1,k+1} = -\Delta_{k+1} \overline{G}_k^{-1}$$

$$\overline{\beta}_{k+1,k+1} = -\overline{\Delta}_{k+1} G_k^{-1}$$

$$\beta_{k+1,j} = \beta_{k,j} + \beta_{k+1,k+1} \overline{\beta}_{k,k+1-j},\ j = 1, \ldots, k,$$

$$\overline{\beta}_{k+1,j} = \overline{\beta}_{k,j} + \overline{\beta}_{k+1,k+1} \beta_{k,k+1-j},\ j = 1, \ldots, k,$$

$$G_{k+1} = \left(I_d - \beta_{k+1,k+1} \overline{\beta}_{k+1,k+1}\right) G_k$$

$$\overline{G}_{k+1} = \left(I_d - \overline{\beta}_{k+1,k+1} \beta_{k+1,k+1}\right) \overline{G}_k.$$

Parameter estimators are obtained by replacing the $\Gamma(j)$ with the sample autocovariances, $\widehat{\Gamma}(j)$, noting that $\Gamma(-j) = \Gamma'(j)$.

5.A.2 Proof of Theorem 5.1

For a given β, $l_0^{(1)}(\beta, \Sigma_\varepsilon, \Sigma_u)$ is maximised by $\widehat{\Sigma}_\varepsilon(\beta) = s_{X,T_1}(\beta)$ and $\widehat{\Sigma}_u(\beta) = s_{Y,T_2}(\beta)$. Let

$$\begin{aligned}\widetilde{l}_0^{(1)}(\beta) &= l_0^{(2)}\left\{\beta, \widehat{\Sigma}_\varepsilon(\beta), \widehat{\Sigma}_u(\beta)\right\} \\ &= -\frac{T_1+T_2}{2}\left\{1 + d\log(2\pi)\right\} - \frac{T_1}{2}\log\left|\widehat{\Sigma}_\varepsilon(\beta)\right| - \frac{T_2}{2}\log\left|\widehat{\Sigma}_u(\beta)\right|.\end{aligned}$$

Then

$$\widetilde{l}_0^{(1)}(\beta) - \widetilde{l}_0^{(1)}(\beta_0) = -\frac{T_1}{2}\log\left(\frac{\left|\widehat{\Sigma}_\varepsilon(\beta)\right|}{\left|\widehat{\Sigma}_\varepsilon(\beta)_0\right|}\right) - \frac{T_2}{2}\log\left(\frac{\left|\widehat{\Sigma}_u(\beta)\right|}{\left|\widehat{\Sigma}_u(\beta)_0\right|}\right).$$

Now,

$$\begin{aligned}X + \beta Z_X &= X + \beta_0 Z_X + (\beta - \beta_0) Z_X \\ &= \varepsilon + (\beta - \beta_0) Z_X,\end{aligned}$$

and so

$$\widehat{\Sigma}_\varepsilon(\beta) = T_1^{-1}\left\{\varepsilon\varepsilon' + \varepsilon Z_X'(\beta-\beta_0)' + (\beta-\beta_0) Z_X\varepsilon' + (\beta-\beta_0) Z_X Z_X (\beta-\beta_0)'\right\}.$$

But, as $T_1 \to \infty$,

$$T_1^{-1}\varepsilon\varepsilon' \to E\left(\varepsilon_t\varepsilon_t'\right) = E\left\{E\left(\varepsilon_t\varepsilon_t' \mid \mathcal{F}_{t-1}\right)\right\} = \Sigma_{\varepsilon 0} \tag{5.6}$$

and

$$T_1^{-1} Z_X \varepsilon' \to 0, \tag{5.7}$$

since

$$T_1^{-1}\sum_{t=p}^{T_1-1} X_{t-j}\varepsilon_t' \to E\left(X_{t-j}\varepsilon_t'\right) = E\left\{X_{t-j}E\left(\varepsilon_t' \mid \mathcal{F}_{t-1}\right)\right\} = 0.$$

Thus, as $T_1 \to \infty$,

$$\widehat{\Sigma}_\varepsilon(\beta) \to \Sigma_{\varepsilon 0} + (\beta-\beta_0)\Gamma_X(\beta-\beta_0)', \tag{5.8}$$

and so

$$\frac{\left|\widehat{\Sigma}_\varepsilon(\beta)\right|}{\left|\widehat{\Sigma}_\varepsilon(\beta)_0\right|} \to \left|I_d + \Sigma_{\varepsilon 0}^{-1/2}(\beta-\beta_0)\Gamma_X(\beta-\beta_0)'\Sigma_{\varepsilon 0}^{-1/2}\right|.$$

Similarly,

$$\frac{\left|\widehat{\Sigma}_u(\beta)\right|}{\left|\widehat{\Sigma}_u(\beta)_0\right|} \to \left|I_d + \Sigma_{u 0}^{-1/2}(\beta-\beta_0)\Gamma_Y(\beta-\beta_0)'\Sigma_{u 0}^{-1/2}\right|.$$

Now,

$$\left|I_d + \Sigma_{\varepsilon 0}^{-1/2}(\beta-\beta_0)\Gamma_X(\beta-\beta_0)'\Sigma_{\varepsilon 0}^{-1/2}\right| = \prod_{j=1}^{d}(1+e_j),$$

where the e_j are the eigenvalues of

$$\Sigma_{\varepsilon 0}^{-1/2}\left(\beta-\beta_0\right)\Gamma_X\left(\beta-\beta_0\right)'\Sigma_{\varepsilon 0}^{-1/2}, \tag{5.9}$$

which are non-negative since (5.9) is non-negative definite. However, the e_j are all zero if and only if $\left(\beta-\beta_0\right)=0$. The same holds for

$$\left|I_d+\Sigma_{u0}^{-1/2}\left(\beta-\beta_0\right)\Gamma_Y\left(\beta-\beta_0\right)'\Sigma_{u0}^{-1/2}\right|$$

and so, for any $\delta>0$,

$$\sup_{\beta;\|\beta-\beta_0\|>\delta}\left\{\widetilde{l}_0^{(1)}\left(\beta\right)-\widetilde{l}_0^{(1)}\left(\beta_0\right)\right\}$$

diverges to $-\infty$ as $T_1,T_2\to\infty$, where $\|\cdot\|$ is any norm. It follows from Theorem 2.5 that $\widehat{\beta}\to\beta_0$. It then follows from (5.8) that $\widehat{\Sigma}_{\varepsilon;1}\to\Sigma_{\varepsilon 0}$ and, similarly, $\widehat{\Sigma}_{u;1}\to\Sigma_{u0}$.

5.A.3 Proof of Theorem 5.2

Consider

$$\widetilde{l}_0^{(2)}\left(\lambda\right)-\widetilde{l}_0^{(2)}\left(\lambda_0\right)=-\frac{T_1+T_2}{2}\log\left(\frac{\left|\widetilde{\Sigma}_{\varepsilon;\lambda}\right|}{\left|\widetilde{\Sigma}_{\varepsilon;\lambda_0}\right|}\right)-\frac{T_2 d}{2}\log\left(\frac{\lambda_0}{\lambda}\right),$$

where

$$\widetilde{\Sigma}_{\varepsilon;\lambda}=\left(T_1+T_2\right)^{-1}\left\{T_1 s_{X,T_1}\left(\beta\right)+\lambda T_2 s_{Y,T_2}\left(\beta\right)\right\}.$$

Now,

$$\begin{aligned}T_1 s_{X,T_1}\left(\beta\right)&=\left(\varepsilon-\beta_0 Z_X+\beta Z_X\right)\left(\varepsilon-\beta_0 Z_X+\beta Z_X\right)'\\&=\varepsilon\varepsilon'+\varepsilon Z_X'\left(\beta-\beta_0\right)'+\left(\beta-\beta_0\right)Z_X\varepsilon'+\left(\beta-\beta_0\right)Z_X Z_X'\left(\beta-\beta_0\right)'\end{aligned}$$

and, similarly,

$$T_2 s_{Y,T_2}\left(\beta\right)=uu'+uZ_Y'\left(\beta-\beta_0\right)'+\left(\beta-\beta_0\right)Z_Y u'+\left(\beta-\beta_0\right)Z_Y Z_Y'\left(\beta-\beta_0\right)'.$$

Also,

$$\begin{aligned}\widehat{\beta}_\lambda&=-\left\{\left(\varepsilon-\beta_0 Z_X\right)Z_X'+\lambda\left(u-\beta_0 Z_Y\right)Z_Y'\right\}\left(Z_X Z_X'+\lambda Z_Y Z_Y'\right)^{-1}\\&=\beta_0-\left(\varepsilon Z_X'+\lambda u Z_Y'\right)\left(Z_X Z_X'+\lambda Z_Y Z_Y'\right)^{-1},\end{aligned} \tag{5.10}$$

and so

$$\left(T_1+T_2\right)\widetilde{\Sigma}_{\varepsilon;\lambda}=\varepsilon\varepsilon'+uu'-\left(\varepsilon Z_X'+\lambda u Z_Y'\right)\left(Z_X Z_X'+\lambda Z_Y Z_Y'\right)^{-1}\left(Z_X\varepsilon'+\lambda Z_Y u'\right). \tag{5.11}$$

Therefore

$$
\begin{aligned}
&\widetilde{l}_0^{(2)}(\lambda) - \widetilde{l}_0^{(2)}(\lambda_0) \\
&= -\frac{T_1+T_2}{2}\log\left(\frac{\left|\varepsilon\varepsilon' + uu' - (\varepsilon Z_X' + \lambda u Z_Y')(Z_X Z_X' + \lambda Z_Y Z_Y')^{-1}(Z_X\varepsilon' + \lambda Z_Y u')\right|}{\left|\varepsilon\varepsilon' + uu' - \left(\varepsilon Z_X' + \lambda_0 u Z_Y'\right)\left(Z_X Z_X' + \lambda_0 Z_Y Z_Y'\right)^{-1}(Z_X\varepsilon' + \lambda_0 Z_Y u')\right|}\right) \\
&\quad - \frac{T_2 d}{2}\log\left(\frac{\lambda_0}{\lambda}\right).
\end{aligned}
$$

From (5.6),

$$
T_1^{-1}\varepsilon\varepsilon' \to \Sigma_{\varepsilon 0} \tag{5.12}
$$

and, similarly,

$$
T_2^{-1}uu' \to \Sigma_{u0} = \frac{1}{\lambda_0}\Sigma_{\varepsilon 0}. \tag{5.13}
$$

From (5.7),

$$
T_1^{-1}\varepsilon Z_X' \to 0 \tag{5.14}
$$

and, similarly,

$$
T_2^{-1}uZ_Y' \to 0. \tag{5.15}
$$

Also,

$$
T_1^{-1}Z_X Z_X' \to \Gamma_X \qquad \text{and} \qquad T_2^{-1}Z_Y Z_Y' \to \Gamma_Y = \frac{1}{\lambda_0}\Gamma_X. \tag{5.16}
$$

Thus

$$
\begin{aligned}
&\frac{\left|\varepsilon\varepsilon' + uu' - (\varepsilon Z_X' + \lambda u Z_Y')(Z_X Z_X' + \lambda Z_Y Z_Y')^{-1}(Z_X\varepsilon' + \lambda Z_Y u')\right|}{\left|\varepsilon\varepsilon' + uu' - \left(\varepsilon Z_X' + \lambda_0 u Z_Y'\right)\left(Z_X Z_X' + \lambda_0 Z_Y Z_Y'\right)^{-1}(Z_X\varepsilon' + \lambda_0 Z_Y u')\right|} \\
&= \frac{\left|T_1\Sigma_{\varepsilon 0} + \frac{\lambda}{\lambda_0}T_2\Sigma_{\varepsilon 0} + o(T_1) + o(T_2)\right|}{|T_1\Sigma_{\varepsilon 0} + T_2\Sigma_{\varepsilon 0} + o(T_1) + o(T_2)|} \\
&= \frac{\left|\left(T_1 + \frac{\lambda}{\lambda_0}T_2\right)I_d + o(T_1) + o(T_2)\right|}{|(T_1+T_2)I_d + o(T_1) + o(T_2)|} \\
&= \left(\frac{T_1 + \frac{\lambda}{\lambda_0}T_2}{T_1+T_2}\right)^d \{1 + o(1)\}.
\end{aligned}
$$

Hence

$$
\widetilde{l}_0^{(2)}(\lambda) - \widetilde{l}_0^{(2)}(\lambda_0) = -\frac{(T_1+T_2)\,d}{2}\log\left(\frac{T_1 + \frac{\lambda}{\lambda_0}T_2}{T_1+T_2}\right) - \frac{T_2 d}{2}\log\left(\frac{\lambda_0}{\lambda}\right) + o(T_1+T_2).
$$

Let

$$
f(x) = -(T_1+T_2)\log\left(\frac{T_1 + xT_2}{T_1+T_2}\right) + T_2\log x.
$$

Then

$$\frac{d}{dx}f(x) = -(T_1+T_2)\frac{T_2}{T_1+xT_2}+\frac{T_2}{x}$$
$$= \frac{T_1T_2}{x(T_1+xT_2)}(1-x),$$

and so $f(x)$ is evidently maximised when $x=1$. Consequently,

$$\lim_{T_1,T_2\to\infty}\sup_{\lambda}\left\{\frac{\widetilde{l}_0^{(2)}(\lambda)-\widetilde{l}_0^{(2)}(\lambda_0)}{T_1+T_2}\right\}\leqslant 0$$

with equality if and only if $\lambda=\lambda_0$ and so, for any $\delta>0$,

$$\lim_{T_1,T_2\to\infty}\sup_{\lambda;|\lambda-\lambda_0|\geqslant\delta}\left\{\frac{\widetilde{l}_0^{(2)}(\lambda)-\widetilde{l}_0^{(2)}(\lambda_0)}{T_1+T_2}\right\}<0.$$

It follows from Theorem 2.5 that $\widehat{\lambda}\to\lambda_0$. It also follows from (5.10), (5.12), (5.13), (5.14), (5.15) and (5.16) that $\widehat{\beta}_{\widehat{\lambda}}\to\beta_0$ and, furthermore, from (5.11) that $\widehat{\Sigma}_{\varepsilon;\widehat{\lambda}}\to\Sigma_{\varepsilon 0}$.

5.A.4 Proof of Theorem 5.3

Consider

$$\widetilde{l}_0^{(3)}(\Lambda)-\widetilde{l}_0^{(3)}(\Lambda_0) = \frac{T_1+T_2}{2}\log\frac{\left|\widetilde{\Sigma}_{\varepsilon;\Lambda_0}\right|}{\left|\widetilde{\Sigma}_{\varepsilon;\Lambda}\right|}-T_2\log\frac{|\Lambda_0|}{|\Lambda|}$$
$$= -\frac{T_1+T_2}{2}\log\left(\frac{\left|\widetilde{\Sigma}_{\varepsilon;\Lambda}\right|}{\left|\widetilde{\Sigma}_{\varepsilon;\Lambda_0}\right|}\right)+T_2\log|\Omega|,$$

where

$$\widetilde{\Sigma}_{\varepsilon;\Lambda}=(T_1+T_2)^{-1}\left[T_1 s_{X,T_1}\left(\widehat{\beta}_\Lambda\right)+T_2\Lambda s_{Y,T_2}\left\{\Lambda^{-1}\widehat{\beta}_\Lambda(I_p\otimes\Lambda)\right\}\Lambda\right]$$

and $\Omega=\Lambda\Lambda_0^{-1}$. Now,

$$\begin{aligned}
\widehat{\beta} &= -\left\{\varepsilon Z_X'-\beta_0 Z_X Z_X'+\Lambda u Z_Y'(I_p\otimes\Lambda)-\Lambda\Lambda_0^{-1}\beta_0(I\otimes\Lambda_0)Z_Y Z_Y'(I_p\otimes\Lambda)\right\}\\
&\quad\times\left\{Z_X Z_X'+(I_p\otimes\Lambda)Z_Y Z_Y'(I_p\otimes\Lambda)\right\}^{-1}\\
&= -\left[\varepsilon Z_X'+\Lambda u Z_Y'(I_p\otimes\Lambda)-\beta_0\left\{Z_X Z_X'-(I_p\otimes\Lambda)Z_Y Z_Y'(I_p\otimes\Lambda)\right\}\right.\\
&\quad\left.+\beta_0(I_p\otimes\Lambda)Z_Y Z_Y'(I_p\otimes\Lambda)-\Omega\beta_0(I\otimes\Lambda_0)Z_Y Z_Y'(I_p\otimes\Lambda)\right]\\
&\quad\times\left\{Z_X Z_X'+(I_p\otimes\Lambda)Z_Y Z_Y'(I_p\otimes\Lambda)\right\}^{-1}\\
&= \beta_0-\left\{\varepsilon Z_X'+\Lambda u Z_Y'(I_p\otimes\Lambda)\right\}\left\{Z_X Z_X'+(I_p\otimes\Lambda)Z_Y Z_Y'(I_p\otimes\Lambda)\right\}^{-1}\\
&\quad-\left[\left\{\beta_0(I_p\otimes\Lambda)Z_Y Z_Y'(I_p\otimes\Lambda)\right\}-\Omega\beta_0(I\otimes\Lambda_0)Z_Y Z_Y'(I_p\otimes\Lambda)\right]\\
&\quad\times\left\{Z_X Z_X'+(I_p\otimes\Lambda)Z_Y Z_Y'(I_p\otimes\Lambda)\right\}^{-1}. \qquad (5.17)
\end{aligned}$$

Also,

$$\begin{aligned}T_1 s_{X,T_1}\left(\widehat{\beta}_\Lambda\right) &= \{\varepsilon + (\beta - \beta_0) Z_X\}\{\varepsilon + (\beta - \beta_0) Z_X\}' \\ &= \varepsilon\varepsilon' + \varepsilon Z_X' (\beta - \beta_0)' + (\beta - \beta_0) Z_X \varepsilon' + (\beta - \beta_0) Z_X Z_X' (\beta - \beta_0)' \\ &= T_1 \Sigma_{\varepsilon 0} + T_1 (\beta - \beta_0) \Gamma_X (\beta - \beta_0)' + o(T_1)\end{aligned}$$

and, similarly,

$$\begin{aligned}T_2 \Lambda s_{Y,T_2}\left\{\Lambda^{-1}\beta (I_p \otimes \Lambda)\right\}\Lambda = \Omega T_2 \Sigma_{u0}\Omega &+ (\beta - \beta_0)(I_p \otimes \Omega) T_2 \Gamma_X (I_p \otimes \Omega)(\beta - \beta_0)' \\ &+ \left\{\Omega\beta_0 - (I_p \otimes \Omega)\beta_0'\right\} T_2 \Gamma_X \left\{\Omega\beta_0 - (I_p \otimes \Omega)\beta_0'\right\}' \\ &+ (\beta - \beta_0)(I_p \otimes \Omega) T_2 \Gamma_X \{\beta_0 (I_p \otimes \Omega) - \Omega\beta_0\}' \\ &+ \{\beta_0 (I_p \otimes \Omega) - \Omega\beta_0\} T_2 \Gamma_X (I_p \otimes \Omega)(\beta - \beta_0)' + o(T_2),\end{aligned}$$

since, from (5.6) and (5.7),

$$T_1^{-1}\varepsilon\varepsilon' \to \Sigma_{\varepsilon 0} \qquad \text{and} \qquad T_1^{-1}\varepsilon Z_X' \to 0 \tag{5.18}$$

and, similarly,

$$T_2^{-1} u u' \to \Sigma_{u0} = \Lambda_0^{-1}\Sigma_{\varepsilon 0}\Lambda_0^{-1} \qquad \text{and} \qquad T_2^{-1} u Z_Y' \to 0, \tag{5.19}$$

and, in addition,

$$T_1^{-1} Z_X Z_X' \to \Gamma_X \qquad \text{and} \qquad T_2^{-1} Z_Y Z_Y' \to \Gamma_Y = \left(I_p \otimes \Lambda_0^{-1}\right)\Gamma_X \left(I_p \otimes \Lambda_0^{-1}\right). \tag{5.20}$$

Thus

$$\begin{aligned}&(T_1 + T_2)\widetilde{\Sigma}_{\varepsilon;\Lambda} \\ &= T_1 \Sigma_{\varepsilon 0} + \Omega T_2 \Sigma_{u0}\Omega \\ &\quad + \left(\widehat{\beta} - \beta_0\right) T_1 \Gamma_X \left(\widehat{\beta} - \beta_0\right)' + \left(\widehat{\beta} - \beta_0\right)(I_p \otimes \Omega) T_2 \Gamma_X (I_p \otimes \Omega)\left(\widehat{\beta} - \beta_0\right)' \\ &\quad + \left\{\Omega\beta_0 - (I_p \otimes \Omega)\beta_0'\right\} T_2 \Gamma_X \left\{\Omega\beta_0 - (I_p \otimes \Omega)\beta_0'\right\} \\ &\quad + \left(\widehat{\beta} - \beta_0\right)(I_p \otimes \Omega) T_2 \Gamma_X \{\beta_0 (I_p \otimes \Omega) - \Omega\beta_0\}' \\ &\quad + \{\beta_0 (I_p \otimes \Omega) - \Omega\beta_0\} T_2 \Gamma_X (I_p \otimes \Omega)\left(\widehat{\beta} - \beta_0\right)' + o(T_1 + T_2) \\ &= T_1 \Sigma_{\varepsilon 0} + \Omega T_2 \Sigma_{u0}\Omega + z T_2 \Gamma_X z' - z (T_2 \Gamma_X)^{1/2} Z \left(I_{dp} + Z'Z\right)^{-1} Z' (T_2 \Gamma_X)^{1/2} z' \\ &\quad + o(T_1 + T_2),\end{aligned}$$

where

$$z = \beta_0 (I_p \otimes \Omega) - \Omega\beta_0$$

and

$$Z = (T_2\Gamma_X)^{1/2}\,(I_p \otimes \Omega)\,(T_1\Gamma_X)^{1/2}\,.$$

But

$$\begin{aligned}
&T_2\Gamma_X - (T_2\Gamma_X)^{1/2}\,Z\left(I_{dp} + Z'Z\right)^{-1} Z'\,(T_2\Gamma_X)^{1/2} \\
&= T_2\Gamma_X + (T_2\Gamma_X)^{1/2}\left(I_{dp} + ZZ'\right)^{-1}(T_2\Gamma_X)^{1/2} - T_2\Gamma_X \\
&= (T_2\Gamma_X)^{1/2}\left(I_{dp} + ZZ'\right)^{-1}(T_2\Gamma_X)^{1/2}\,,
\end{aligned}$$

since

$$Z\left(I_{dp} + Z'Z\right)^{-1} Z' = I_{dp} - \left(I_{dp} + ZZ'\right)^{-1},$$

and so

$$\begin{aligned}
&(T_1 + T_2)\,\widetilde{\Sigma}_{\varepsilon;\Lambda} \\
&= T_1\Sigma_{\varepsilon 0} + \Omega T_2\Sigma_{u0}\Omega + z\,(T_2\Gamma_X)^{1/2}\left(I_{dp} + ZZ'\right)^{-1}(T_2\Gamma_X)^{1/2}\,z' + o\,(T_1 + T_2)\,. \qquad (5.21)
\end{aligned}$$

Therefore

$$\begin{aligned}
\left|\widetilde{\Sigma}_{\varepsilon;\Lambda}\right| &\geqslant \left|\widetilde{\Sigma}\right| + \left|z\,(T_2\Gamma_X)^{1/2}\left(I_{dp} + ZZ'\right)^{-1}(T_2\Gamma_X)^{1/2}\,z'\right| + o\,(1) \\
&\geqslant \left|\widetilde{\Sigma}\right|,
\end{aligned}$$

where

$$\widetilde{\Sigma} = \frac{T_1\Sigma_{\varepsilon 0} + \Omega T_2\Sigma_{\varepsilon 0}\Omega}{T_1 + T_2},$$

since $z\,(T_2\Gamma_X)^{1/2}\,(I_{dp} + ZZ')^{-1}\,(T_2\Gamma_X)^{1/2}\,z'$ is non-negative definite. Let

$$\Sigma_{\varepsilon 0} = PAP'$$

be the Jordan canonical form of $\Sigma_{\varepsilon 0}$. Then

$$\begin{aligned}
\left|\widetilde{\Sigma}\right| &= \left|\frac{T_1A + T_2P'\Omega PAP'\Omega P}{T_1 + T_2}\right| \\
&= |A|\left|\frac{T_2I_d + T_2A^{-1/2}P'\Omega PAP'\Omega PA^{-1/2}}{T_1 + T_2}\right| \\
&= |A|\left|\frac{T_1I_d + T_2QQ'}{T_1 + T_2}\right|,
\end{aligned}$$

where

$$Q = A^{-1/2}P'\Omega PA^{1/2}.$$

Note that $|QQ'| = 2|\Omega|$ and so

$$\begin{aligned}
&(T_1+T_2)\log\left(\frac{\left|\widetilde{\Sigma}_{\varepsilon;\Lambda}\right|}{\left|\widetilde{\Sigma}_{\varepsilon;\Lambda_0}\right|}\right) - 2T_2\log|\Omega| \\
&\geqslant (T_1+T_2)\log\left|\frac{T_1 I_d + T_2 QQ'}{T_1+T_2}\right| - 2T_2\log\left|QQ'\right| + o\left(T_1+T_2\right) \\
&= \sum_{j=1}^{d}\left\{(T_1+T_2)\log\left(\frac{T_1+T_2\mu_j}{T_1+T_2}\right) - T_2\log\mu_j\right\} + o\left(T_1+T_2\right), \qquad (5.22)
\end{aligned}$$

where the μ_j are the eigenvalues of QQ'. But

$$\begin{aligned}
\frac{\partial}{\partial\mu_j}\left\{(T_1+T_2)\log\left(\frac{T_1+T_2\mu_j}{T_1+T_2}\right) - T_2\log\mu_j\right\} &= (T_1+T_2)\frac{T_2}{T_1+T_2\mu_j} - \frac{T_2}{\mu_j} \\
&= \frac{T_2}{\mu_j\left(T_1+T_2\mu_j\right)}\left\{(T_1+T_2)\mu_j - (T_1+T_2\mu_j)\right\} \\
&= \frac{T_1T_2}{\mu_j\left(T_1+T_2\mu_j\right)}\left(\mu_j - 1\right). \qquad (5.23)
\end{aligned}$$

The derivatives of

$$(T_1+T_2)\log\left(\frac{\left|\widetilde{\Sigma}_{\varepsilon;\Lambda}\right|}{\left|\widetilde{\Sigma}_{\varepsilon;\Lambda_0}\right|}\right) - 2T_2\log|\Omega|$$

are thus all 0 if and only if all of the μ_j are 1, that is when $QQ' = I_d$. But then

$$P'\Omega PAP'\Omega P = A,$$

that is,

$$\Omega\Sigma_{\varepsilon 0}\Omega = \Sigma_{\varepsilon 0},$$

which can only occur when $\Omega = I_d$ which is when $\Lambda = \Lambda_0$. However, from (5.23), there is only one turning point, which is clearly a minimum since (5.23) is negative when $\mu_j < 1$ and positive when $\mu_j > 1$. Moreover, (5.22) is 0 at this turning point. Thus for any $\delta > 0$,

$$\lim_{T_1,T_2\to\infty}\ \sup_{\max_j|\lambda_j-\lambda_{0j}|>\delta}\left\{\widetilde{l}_0^{(3)}(\Lambda) - \widetilde{l}_0^{(3)}(\Lambda_0)\right\} < 0.$$

It follows from Theorem 2.5 that $\widehat{\Lambda} \to \Lambda_0$ as $T_1, T_2 \to \infty$. It then follows from (5.17), (5.18), (5.19) and (5.20) that $\widehat{\beta} \to \beta_0$ and, furthermore, from (5.21) that $\widetilde{\Sigma}_{\varepsilon;\Lambda} \to \Sigma_{\varepsilon 0}$ as $T_1, T_2 \to \infty$.

5.A.5 Proof of Lemma 5.1

Let

$$W_{T_1} = a'\operatorname{vec}\left(\varepsilon Z_X'\right) = a'\sum_{t=p}^{T_1-1} V_t,$$

where $a = \begin{bmatrix} a_1 & \cdots & a_{d^2p}\end{bmatrix}'$, $a_1, \ldots, a_{d^2p}$ are constants and $V_t = (Z_{X,t} \otimes \varepsilon_t)$. Now,

$$E\left(V_t \mid \mathcal{F}_{t-1}\right) = Z_{X,t} \otimes E\left(\varepsilon_t \mid \mathcal{F}_{t-1}\right) = 0$$

and

$$E\left(V_tV_t'\right) = E\left\{E\left(V_tV_t' \mid \mathcal{F}_{t-1}\right)\right\} = \left(\Gamma_X \otimes \Sigma_{\varepsilon 0}\right).$$

Thus, by the martingale central limit theorem, $T_1^{-1/2}W_{T_1}$ is asymptotically normal with mean zero and covariance matrix $a'\left(\Gamma_X \otimes \Sigma_{\varepsilon 0}\right)a$ as $T_1 \to \infty$. The distribution of $T_1^{-1/2}\,\mathrm{vec}\left(\varepsilon Z_X'\right)$ therefore converges to the normal distribution with mean zero and covariance matrix $\Gamma_X \otimes \Sigma_{\varepsilon 0}$ as $T_1 \to \infty$. Similarly, the distribution of $T_2^{-1/2}\,\mathrm{vec}\left(uZ_Y'\right)$ converges to the normal distribution with mean zero and covariance matrix $\Gamma_Y \otimes \Sigma_{u0}$ as $T_2 \to \infty$.

5.A.6 Proof of Theorem 5.4

Let $B = \mathrm{vec}\,\beta$, $A_\varepsilon = \mathrm{vec}\left(\Sigma_\varepsilon^{-1}\right)$, $A_u = \mathrm{vec}\left(\Sigma_u^{-1}\right)$ and $\theta = \begin{bmatrix} B' & A_\varepsilon' & A_u' \end{bmatrix}'$. From the mean value theorem, and since $\widehat{\theta} \to \theta_0$, $\left(\widehat{\theta} - \theta_0\right)$ has the same asymptotic distribution as

$$-\left\{\frac{\partial^2 l_0\left(\theta_0\right)}{\partial\theta\partial\theta'}\right\}^{-1}\frac{\partial l_0\left(\theta_0\right)}{\partial\theta}.$$

Now, if C and D are $n \times m$ and $n \times n$ matrices, respectively, then

$$\mathrm{tr}\left(C'DC\right) = \left(\mathrm{vec}\,C\right)'\left(I_m \otimes D\right)\mathrm{vec}\,C.$$

Thus

$$\begin{aligned}
&T_1\,\mathrm{tr}\left\{\Sigma_\varepsilon^{-1}s_{X,T_1}\left(\beta\right)\right\}\\
&= \mathrm{tr}\left\{\left(X + \beta Z_X\right)'\Sigma_\varepsilon^{-1}\left(X + \beta Z_X\right)\right\}\\
&= \left\{\mathrm{vec}\left(X + \beta Z_X\right)\right\}'\left(I_{T_1} \otimes \Sigma_\varepsilon^{-1}\right)\mathrm{vec}\left(X + \beta Z_X\right)\\
&= \left(\mathrm{vec}\,X\right)'\left(I_{T_1} \otimes \Sigma_\varepsilon^{-1}\right)\mathrm{vec}\,X + 2B'\left(Z_X \otimes \Sigma_\varepsilon^{-1}\right)\mathrm{vec}\,X + B'\left(Z_XZ_X' \otimes \Sigma_\varepsilon^{-1}\right)B
\end{aligned}$$

and so

$$\frac{\partial}{\partial B}T_1\,\mathrm{tr}\left\{\Sigma_\varepsilon^{-1}s_{X,T_1}\left(\beta\right)\right\} = 2\left(Z_X \otimes \Sigma_\varepsilon^{-1}\right)\mathrm{vec}\,X + 2\left(Z_XZ_X' \otimes \Sigma_\varepsilon^{-1}\right)B.$$

Similarly,

$$\frac{\partial}{\partial B}T_2\,\mathrm{tr}\left\{\Sigma_u^{-1}s_{Y,T_2}\left(\beta\right)\right\} = 2\left(Z_Y \otimes \Sigma_u^{-1}\right)\mathrm{vec}\,Y + 2\left(Z_YZ_Y' \otimes \Sigma_u^{-1}\right)B.$$

The first derivative of $l_0\left(\theta\right)$ with respect to B is therefore

$$\begin{aligned}
\frac{\partial l_0\left(\theta\right)}{\partial B} &= -\left(Z_X \otimes \Sigma_\varepsilon^{-1}\right)\mathrm{vec}\,X - \left(Z_XZ_X' \otimes \Sigma_\varepsilon^{-1}\right)B - \left(Z_Y \otimes \Sigma_u^{-1}\right)\mathrm{vec}\,Y - \left(Z_YZ_Y' \otimes \Sigma_u^{-1}\right)B\\
&= -\left\{Z_X\left(X' + Z_X'\beta'\right) \otimes I_d\right\}A_\varepsilon - \left\{Z_Y\left(Y' + Z_Y'\beta'\right) \otimes I_d\right\}A_u.
\end{aligned}$$

Thus

$$\frac{\partial^2 l_0\left(\theta\right)}{\partial B\partial A_\varepsilon'} = -\left\{\left(X + \beta Z_X\right)Z_X' \otimes I_d\right\}$$

and so

$$T_1^{-1}\frac{\partial^2 l_0(\theta_0)}{\partial B \partial A_\varepsilon'} = -\left(T_1^{-1}\varepsilon Z_X' \otimes I_d\right) \to 0$$

from (5.7). Similarly,

$$T_1^{-1}\frac{\partial^2 l_0(\theta_0)}{\partial B \partial A_u'} = -\left(\kappa T_2^{-1} u Z_Y' \otimes I_d\right) \to 0.$$

It follows that $T_1^{1/2}\left(\widehat{B}-B_0\right)$ has the same asymptotic distribution as

$$-\left\{T_1^{-1}\frac{\partial^2 l_0(\theta_0)}{\partial B \partial B'}\right\}^{-1}\left\{T_1^{-1/2}\frac{\partial l_0(\theta_0)}{\partial B}\right\}.$$

The second derivative of $l_0(\theta)$ with respect to B is

$$\frac{\partial^2 l_0(\theta)}{\partial B \partial B'} = -\left(Z_X Z_X' \otimes \Sigma_\varepsilon^{-1}\right) - \left(Z_Y Z_Y' \otimes \Sigma_u^{-1}\right)$$

and so

$$-T_1^{-1}\frac{\partial^2 l_0(\theta_0)}{\partial B \partial B'} \to \left(\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1}\right) + \kappa\left(\Gamma_Y \otimes \Sigma_{u0}^{-1}\right).$$

Also,

$$\begin{aligned}\frac{\partial l_0(\theta_0)}{\partial B} &= -\left(Z_X \varepsilon' \otimes I_d\right) A_{\varepsilon 0} - \left(Z_Y u' \otimes I_d\right) A_{u0} \\ &= -\left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1}\right)\operatorname{vec}\left(\varepsilon Z_X'\right) - \left(I_{dp} \otimes \Sigma_{u0}^{-1}\right)\operatorname{vec}\left(u Z_Y'\right).\end{aligned}$$

From Lemma 5.1, the distribution of

$$T_1^{-1/2}\left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1}\right)\operatorname{vec}\left(\varepsilon Z_X'\right)$$

converges to the normal distribution with mean zero and covariance matrix $\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1}$ as $T_1 \to \infty$. Similarly, the distribution of

$$T_1^{-1/2}\left(I_{dp} \otimes \Sigma_{u0}^{-1}\right)\operatorname{vec}\left(u Z_Y'\right)$$

converges to the normal distribution with mean zero and covariance matrix $\kappa\left(\Gamma_Y \otimes \Sigma_{u0}^{-1}\right)$ as $T_1 \to \infty$. The distribution of vec $\left(\widehat{\beta}-\beta_0\right)$ therefore converges to the normal distribution with mean zero and covariance matrix

$$\left\{\left(\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1}\right) + \kappa\left(\Gamma_Y \otimes \Sigma_{u0}^{-1}\right)\right\}^{-1}$$

as $T_1 \to \infty$.

5.A.7 Proof of Theorem 5.5

Let $B = \operatorname{vec}\beta$, $A_\varepsilon = \operatorname{vec}\left(\Sigma_\varepsilon^{-1}\right)$ and $\theta = \begin{bmatrix} B' & A_\varepsilon' & \lambda \end{bmatrix}'$. From the mean value theorem, and since $\widehat{\theta} \to \theta_0$, $\left(\widehat{\theta} - \theta_0\right)$ has the same asymptotic distribution as

$$-\left\{\frac{\partial^2 l_0\left(\theta_0\right)}{\partial\theta\partial\theta'}\right\}^{-1}\frac{\partial l_0\left(\theta_0\right)}{\partial\theta}.$$

The first derivative of $l_0\left(\theta\right)$ with respect to B is

$$\begin{aligned}\frac{\partial l_0\left(\theta\right)}{\partial B} &= -\left(Z_X \otimes \Sigma_\varepsilon^{-1}\right)\operatorname{vec}X - \left(Z_X Z_X' \otimes \Sigma_\varepsilon^{-1}\right)B - \lambda\left(Z_Y \otimes \Sigma_\varepsilon^{-1}\right)\operatorname{vec}Y \\ &\quad - \lambda\left(Z_Y Z_Y' \otimes \Sigma_\varepsilon^{-1}\right)B \\ &= -\left\{Z_X\left(X' + Z_X'\beta\right) \otimes I_d\right\}A_\varepsilon - \lambda\left\{Z_Y\left(Y' + Z_Y'\beta'\right) \otimes I_d\right\}A_\varepsilon.\end{aligned}$$

Thus

$$\frac{\partial^2 l_0\left(\theta\right)}{\partial B\partial A_\varepsilon'} = -\left\{\left(X + \beta Z_X\right)Z_X' \otimes I_d\right\} - \lambda\left\{\left(Y + \beta Z_Y\right)Z_Y' \otimes I_d\right\}$$

and

$$\frac{\partial^2 l_0\left(\theta\right)}{\partial B\partial\lambda} = -\left\{Z_Y\left(Y' + Z_Y'\beta\right) \otimes I_d\right\}A_\varepsilon.$$

Therefore

$$-\left(T_1 + T_2\right)^{-1}\frac{\partial^2 l_0\left(\theta_0\right)}{\partial B\partial A_\varepsilon'} = \left(T_1 + T_2\right)^{-1}\left\{\left(\varepsilon Z_X' \otimes I_d\right) + \lambda\left(uZ_Y' \otimes I_d\right)\right\} \to 0$$

and

$$\begin{aligned}-T_2^{-1}\frac{\partial^2 l_0\left(\theta_0\right)}{\partial B\partial\lambda} &= T_2^{-1}\left(Z_Y u' \otimes I_d\right)A_{\varepsilon 0} \\ &= T_2^{-1}\left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1}\right)\operatorname{vec}\left(uZ_Y'\right) \\ &\to 0.\end{aligned}$$

It follows that $\left(T_1 + T_2\right)^{1/2}\left(\widehat{B} - B_0\right)$ has the same asymptotic distribution as

$$-\left\{\left(T_1 + T_2\right)^{-1}\frac{\partial^2 l_0\left(\theta_0\right)}{\partial B\partial B'}\right\}^{-1}\left\{\left(T_1 + T_2\right)^{-1/2}\frac{\partial l_0\left(\theta_0\right)}{\partial B}\right\}.$$

The second derivative of $l_0\left(\theta\right)$ with respect to B is

$$\frac{\partial^2 l_0\left(\theta\right)}{\partial B\partial B'} = -\left(Z_X Z_X' \otimes \Sigma_\varepsilon^{-1}\right) - \lambda\left(Z_Y Z_Y' \otimes \Sigma_\varepsilon^{-1}\right).$$

Now,

$$\begin{aligned}\Gamma_X\left(0\right) &= E\left(X_0 X_0'\right) \\ &= E\left\{\left(\varepsilon_0 - \beta_0 Z_{X,0}\right)\left(\varepsilon_0 - \beta_0 Z_{X,0}\right)'\right\} \\ &= E\left(\varepsilon_0\varepsilon_0' - \varepsilon_0 Z_{X,0}'\beta_0' - \beta_0 Z_{X,0}\varepsilon_0' + \beta Z_{X,0}Z_{X,0}'\beta'\right) \\ &= \Sigma_{\varepsilon 0} + \beta_0\Gamma_X\beta_0'\end{aligned}$$

and, similarly,

$$\Gamma_Y(0) = \Sigma_{u0} + \beta_0 \Gamma_Y \beta_0'.$$

Since $\Sigma_{\varepsilon 0} = \lambda_0 \Sigma_{u0}$,

$$\Gamma_X(0) - \beta_0 \Gamma_X \beta_0' = \lambda_0 \Gamma_Y(0) - \lambda_0 \beta_0 \Gamma_Y \beta_0'.$$

Thus $\Gamma_X = \lambda_0 \Gamma_Y$ and so

$$-(T_1 + T_2)^{-1} \frac{\partial^2 l_0(\theta_0)}{\partial B \partial B'} \to \left(\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1}\right).$$

Also,

$$\begin{aligned} \frac{\partial l_0(\theta_0)}{\partial B} &= -\left(Z_X \varepsilon' \otimes I_d\right) A_{\varepsilon 0} - \lambda_0 \left(Z_Y u' \otimes I_d\right) A_{\varepsilon 0} \\ &= -\left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1}\right) \operatorname{vec}\left(\varepsilon Z_X'\right) - \lambda_0 \left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1}\right) \operatorname{vec}\left(u Z_Y'\right). \end{aligned}$$

From Lemma 5.1, the distribution of

$$T_1^{-1/2} \left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1}\right) \operatorname{vec}\left(\varepsilon Z_X'\right)$$

converges to the normal distribution with mean zero and covariance matrix $\left(\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1}\right)$ as $T_1 \to \infty$. Similarly, the distribution of

$$T_2^{-1/2} \lambda_0 \left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1}\right) \operatorname{vec}\left(u Z_Y'\right)$$

converges to the normal distribution with mean zero and covariance matrix $\left(\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1}\right)$ as $T_2 \to \infty$. Let $a = \left[\begin{array}{ccc} a_1 & \cdots & a_{d^2 p} \end{array}\right]$ for constants $a_1, \ldots, a_{d^2 p}$. Then, from Lemma 3.1, the distribution of

$$(T_1 + T_2)^{-1/2} a \left[\left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1}\right) \operatorname{vec}\left(\varepsilon Z_X'\right) + \lambda_0 \left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1}\right) \operatorname{vec}\left(u Z_Y'\right)\right]$$

converges to the normal distribution with mean zero and variance $a' \left(\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1}\right) a$. The distribution of $(T_1 + T_2)^{1/2} \operatorname{vec}\left(\widehat{\beta} - \beta_0\right)$ therefore converges to the normal distribution with mean zero and covariance matrix $\Gamma_X^{-1} \otimes \Sigma_{\varepsilon 0}$ as $T_1, T_2 \to \infty$.

5.A.8 Proof of Theorem 5.6

Let $B = \operatorname{vec} \beta$, $A_\varepsilon = \operatorname{vec}\left(\Sigma_\varepsilon^{-1}\right)$, $L = \operatorname{vec} \Lambda$ and $\theta = \left[\begin{array}{ccc} B' & A_\varepsilon' & L \end{array}\right]'$. From the mean value theorem, and since $\widehat{\theta} \to \theta_0$, $\left(\widehat{\theta} - \theta_0\right)$ has the same asymptotic distribution as

$$-\left\{\frac{\partial^2 l_0(\theta_0)}{\partial \theta \partial \theta'}\right\}^{-1} \frac{\partial l_0(\theta)}{\partial \theta}.$$

The first derivative of $l_0(\theta)$ with respect to B is

$$\begin{aligned}\frac{\partial l_0(\theta)}{\partial B} &= -\left(Z_X \otimes \Sigma_\varepsilon^{-1}\right)\operatorname{vec} X - \left(Z_X Z_X' \otimes \Sigma_\varepsilon^{-1}\right) B - \left\{\left(I_p \otimes \Lambda\right) Z_Y \otimes \Sigma_\varepsilon^{-1}\Lambda\right\}\operatorname{vec} Y \\ &\quad - \left\{\left(I_p \otimes \Lambda\right) Z_Y Z_Y' \left(I_p \otimes \Lambda\right) \otimes \Sigma_\varepsilon^{-1}\right\} B \\ &= -\left\{Z_X\left(X' + Z_X'\beta'\right) \otimes I_d\right\} A_\varepsilon - \left[\left(I_p \otimes \Lambda\right) Z_Y \left\{Y'\Lambda + Z_Y'\left(I_p \otimes \Lambda\right)\beta'\right\} \otimes I_d\right] A_\varepsilon.\end{aligned}$$

As before,

$$-\left(T_1 + T_2\right)^{-1} \frac{\partial^2 l_0(\theta_0)}{\partial B \partial A_\varepsilon'} \to 0.$$

It can also be shown that

$$-\left(T_1 + T_2\right)^{-1} \frac{\partial^2 l_0(\theta_0)}{\partial B \partial L'} \to 0.$$

It follows that $(T_1 + T_2)^{1/2}\left(\widehat{B} - B_0\right)$ has the same asymptotic distribution as

$$-\left\{\left(T_1 + T_2\right)^{-1} \frac{\partial^2 l_0(\theta_0)}{\partial B \partial B'}\right\}^{-1}\left\{\left(T_1 + T_2\right)^{-1/2} \frac{\partial l_0(\theta_0)}{\partial B}\right\}.$$

The second derivative of $l_0(\theta)$ with respect to B is

$$\frac{\partial^2 l_0(\theta)}{\partial B \partial B'} = -\left(Z_X Z_X' \otimes \Sigma_\varepsilon^{-1}\right) - \left\{\left(I_p \otimes \Lambda\right) Z_Y Z_Y' \left(I_p \otimes \Lambda\right) \otimes \Sigma_\varepsilon^{-1}\right\}.$$

Now,

$$\Gamma_X(0) = \Sigma_{\varepsilon 0} + \beta \Gamma_X \beta'$$

and

$$\Gamma_Y(0) = \Sigma_u + \Lambda_0^{-1}\beta_0\left(I_p \otimes \Lambda_0\right)\Gamma_Y\left(I_p \otimes \Lambda_0\right)\beta_0'\Lambda_0^{-1}.$$

Since $\Sigma_{\varepsilon 0} = \Lambda \Sigma_{u0} \Lambda$,

$$\Gamma_X(0) - \beta\Gamma_X\beta' = \Lambda_0 \Gamma_Y(0)\Lambda_0 - \beta_0\left(I_p \otimes \Lambda_0\right)\Gamma_Y\left(I_p \otimes \Lambda_0\right)\beta_0'.$$

Thus $\Gamma_X = \left(I_p \otimes \Lambda_0\right)\Gamma_Y\left(I_p \otimes \Lambda_0\right)$ and so

$$-\left(T_1 + T_2\right)^{-1} \frac{\partial^2 l_0(\theta_0)}{\partial B \partial B'} \to \left(\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1}\right).$$

Also,

$$\begin{aligned}\frac{\partial l_0(\theta_0)}{\partial B} &= -\left(Z_X \varepsilon' \otimes I_d\right) A_{\varepsilon 0} - \left\{\left(I_p \otimes \Lambda_0\right) Z_Y u' \Lambda_0 \otimes I_d\right\} A_{\varepsilon 0} \\ &= -\left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1}\right)\operatorname{vec}\left(\varepsilon Z_X'\right) - \left\{\left(I_p \otimes \Lambda_0\right) \otimes \Sigma_{\varepsilon 0}^{-1}\Lambda_0\right\}\operatorname{vec}\left(u Z_Y'\right).\end{aligned}$$

From Lemma 5.1, the distribution of

$$T_1^{-1/2}\left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1}\right)\operatorname{vec}\left(\varepsilon Z_X'\right)$$

converges to the normal distribution with mean zero and covariance matrix $\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1}$ as $T_1 \to \infty$. Similarly, the distribution of

$$T_2^{-1/2} \left\{ (I_p \otimes \Lambda_0) \otimes \Sigma_{\varepsilon 0}^{-1} \Lambda_0 \right\} \operatorname{vec} \left(u Z_Y' \right)$$

converges to the normal distribution with mean zero and covariance matrix $\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1}$ as $T_2 \to \infty$. Let $a = \begin{bmatrix} a_1 & \cdots & a_{d^2p} \end{bmatrix}$ for constants $a_1, \ldots, a_{d^2p}$. Then, from Lemma 3.1, the distribution of

$$(T_1 + T_2)^{-1/2} \left[\left(I_{dp} \otimes \Sigma_{\varepsilon 0}^{-1} \right) \operatorname{vec} \left(\varepsilon Z_X' \right) + \left\{ (I_p \otimes \Lambda_0) \otimes \Sigma_{\varepsilon 0}^{-1} \Lambda_0 \right\} \operatorname{vec} \left(u Z_Y' \right) \right]$$

converges to the normal distribution with mean zero and covariance matrix $a' \left(\Gamma_X \otimes \Sigma_{\varepsilon 0}^{-1} \right) a$. The distribution of $(T_1 + T_2)^{1/2} \operatorname{vec} \left(\widehat{\beta} - \beta_0 \right)$ therefore converges to the normal distribution with mean zero and covariance matrix $\Gamma_X^{-1} \otimes \Sigma_{\varepsilon 0}$ as $T_1, T_2 \to \infty$.

6

The Estimation of Frequency in the Multichannel Sinusoidal Model

6.1 Introduction

An important class of stationary processes, arising in many applications, consist of those which contain periodic components. An example of a periodic process is the vibration through the air produced by a musical note. The soundwave will oscillate at a fixed, regular, interval, known as the period. The number of oscillations that occur for each unit of time is the frequency. In practice, a recording of the note will not oscillate exactly at this fixed frequency since the recording will be subject to, for example, background noise or variation in the recording device. The time series can therefore be considered to be generated by a process which consists of both a deterministic periodic component, which we refer to as the signal, and a stationary stochastic component, which we refer to as the noise. The stationary component also reflects the fact that some phenomena which we model in this way are only approximately periodic. It is often of interest to estimate the fixed frequency, or frequencies, of the periodic components given the noisy signal.

Typically, periodic processes are modelled as the sum of sinusoids and a stationary stochastic noise process, that is,

$$X_t = \mu + \sum_{j=1}^{f} \rho_j \cos(\omega_j t + \phi_j) + \varepsilon_t, \tag{6.1}$$

where ρ_j is the amplitude and ϕ_j is the phase of the jth sinusoid, μ is a constant mean term and $\omega \in (0, \pi)$. These are called mixed spectra models. The noise process, $\{\varepsilon_t\}$, is assumed to be stationary with a smooth spectral density, and can be modelled as, for example, an autoregression. It is often convenient to reparametrise (6.1) as

$$X_t = \mu + \sum_{j=1}^{f} \{\alpha_j \cos(\omega_j t) + \beta_j \sin(\omega_j t)\} + \varepsilon_t, \tag{6.2}$$

where $\alpha_j = \rho_j \cos \phi_j$ and $\beta_j = -\rho_j \sin \phi_j$.

There is an extensive literature on the estimation of the parameters in (6.1), or equivalently in (6.2), when $\{X_t\}$ is univariate (see, for example, Quinn and Hannan, 2001). There has, however, been little work on the case where $\{X_t\}$ is a vector process. In the multivariate case, the model we consider is the same as (6.2), but where α_j and β_j, $j = 1, \ldots, f$, are $d \times 1$ and $\{\varepsilon_t\}$ is d-dimensional. Its interpretation is that we have a periodic vector process where each component of the deterministic part is made up of sinusoids with common frequencies but with possibly different amplitudes and phases. It has been referred to as the multichannel sinusoidal model (Sakai, 1993).

The estimation of frequency in the multichannel model was considered by Sakai (1993) for the case where the true frequencies were Fourier frequencies. It was also assumed that the noise was Gaussian and white. In practice, the true frequencies are unlikely to be exactly Fourier frequencies and the noise may be non-Gaussian and/or coloured. In this chapter we develop procedures for estimating the frequencies in this more general setting.

The chapter begins with a brief overview of univariate frequency estimation which will motivate the methods used in the multivariate case. We then develop procedures for estimating a single frequency in the multichannel model and establish the asymptotic properties of the estimators. We also discuss how to use the procedures to estimate more than one frequency. Results of simulation studies are presented which demonstrate the performance of the estimation procedures in practice.

6.2 Univariate Frequency Estimation

6.2.1 Estimating a Single Frequency

We begin by considering the case where there is a single frequency, that is where $f = 1$, and $\{\varepsilon_t\}$ is Gaussian and white. That is, we consider the model

$$X_t = \mu + \alpha \cos(\omega t) + \beta \sin(\omega t) + \varepsilon_t, \tag{6.3}$$

where ε_t is normal with mean zero and variance σ^2, $t = 0, \ldots, T-1$, and T is the sample size. Gaussianity is assumed initially so that the maximum likelihood and least squares techniques are the same. Later on, the assumption of Gaussianity will be dropped.

The log-likelihood is

$$-\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\{X - M_T(\omega)\theta\}'\{X - M_T(\omega)\theta\}, \tag{6.4}$$

where

$$X = \begin{bmatrix} X_0 & \cdots & X_{T-1} \end{bmatrix}',$$
$$\theta = \begin{bmatrix} \mu & \alpha & \beta \end{bmatrix}'$$

and $M_T(\omega)$ is the $T \times 3$ matrix with $(t+1)$th row

$$\begin{bmatrix} 1 & \cos(\omega t) & \sin(\omega t) \end{bmatrix},$$

$t = 0, \ldots, T-1$. For fixed ω, (6.4) is maximised with respect to θ by

$$\left\{M_T'(\omega) M_T(\omega)\right\}^{-1}\left\{M_T'(\omega) X\right\}.$$

But

$$M_T'(\omega) M_T(\omega) = T \operatorname{diag}\left(1, \frac{1}{2}, \frac{1}{2}\right) + O(1) \tag{6.5}$$

since, for $\omega \neq 0, \pi$,

$$\sum_{t=0}^{T-1} \cos^2(\omega t) = \frac{T}{2} + O(1), \qquad \sum_{t=0}^{T-1} \sin^2(\omega t) = \frac{T}{2} + O(1), \tag{6.6}$$

and

$$\sum_{t=0}^{T-1} \cos(\omega t) = O(1), \qquad \sum_{t=0}^{T-1} \sin(\omega t) = O(1), \qquad \sum_{t=0}^{T-1} \cos(\omega t)\sin(\omega t) = O(1). \tag{6.7}$$

Hence the maximum likelihood estimator of θ has the same asymptotic properties as

$$\widehat{\theta}_T(\omega) = \begin{bmatrix} \overline{X} & 2T^{-1}\sum_{t=0}^{T-1} X_t \cos(\omega t) & 2T^{-1}\sum_{t=0}^{T-1} X_t \sin(\omega t) \end{bmatrix}', \tag{6.8}$$

where $\overline{X} = T^{-1}\sum_{t=0}^{T-1} X_t$. Maximum likelihood estimation of ω therefore consists of minimising

$$X'X - 2\widehat{\theta}_T'(\omega)\left\{M_T'(\omega)X\right\} + \widehat{\theta}_T'(\omega)\left\{M_T'(\omega)M_T(\omega)\right\}\widehat{\theta}_T(\omega) \tag{6.9}$$

with respect to ω. From (6.5) and (6.8), (6.9) is equal to

$$\sum_{t=0}^{T-1}\left(X_t - \overline{X}\right)^2 - 2T^{-1}\left\{\sum_{t=0}^{T-1} X_t\cos(\omega t)\right\}^2 - 2T^{-1}\left\{\sum_{t=0}^{T-1} X_t\sin(\omega t)\right\}^2 + o(1).$$

Thus, the maximum likelihood estimator of ω is asymptotically equivalent to the maximiser of

$$\begin{aligned} I_{T,X}(\omega) &= 2T^{-1}\left\{\sum_{t=0}^{T-1} X_t\cos(\omega t)\right\}^2 + 2T^{-1}\left\{\sum_{t=0}^{T-1} X_t\sin(\omega t)\right\}^2 \\ &= 2T^{-1}\left|\sum_{t=0}^{T-1} X_t e^{-i\omega t}\right|^2, \end{aligned}$$

which is the periodogram.

Let $\widehat{\omega}$ be the maximiser of $I_{T,X}(\omega)$ and ω_0 be the true frequency. Walker (1971) showed that $T(\widehat{\omega} - \omega_0) \to 0$ almost surely and that the distribution of $T^{3/2}(\widehat{\omega} - \omega_0)$ converges to the normal distribution with mean zero and variance $24\sigma^2/\rho^2$, where $\rho^2 = \alpha^2 + \beta^2$. Hannan (1973b) showed that if $\{\varepsilon_t\}$ is not white, but coloured, with spectral density $f_\varepsilon(\omega)$, and not necessarily Gaussian, $T(\widehat{\omega} - \omega_0) \to 0$ almost surely and the distribution of $T^{3/2}(\widehat{\omega} - \omega_0)$ converges to the normal distribution with mean zero and variance $48\pi f_\varepsilon(\omega_0)/\rho^2$.

6.2.2 Estimating More Than One Frequency

We now consider the case where $f > 1$. If $\{\varepsilon_t\}$ is Gaussian and white, the same calculations as above show that maximum likelihood estimation of $\omega_1, \ldots, \omega_f$ is asymptotically equivalent to minimising

$$\sum_{t=0}^{T-1}\left(X_t - \overline{X}\right)^2 - \sum_{j=1}^{f} I_{T,X}(\omega_j). \tag{6.10}$$

This is the same as finding local maxima for $I_{T,X}(\omega)$. This cannot be done in practice, as associated with each local maximum are "sidelobes", which give rise to other local maxima close to these. Providing none of the frequencies are too close together, they can be estimated one at a time, with the sinusoid corresponding to each estimated frequency removed from the time series by regression before estimating the next. The procedure for estimating the f frequencies is given in Algorithm 6.1 (see Quinn and Hannan, 2001).

Now suppose that $\{\varepsilon_t\}$ is coloured with spectral density $f_\varepsilon(\omega)$ and let Γ_ε be the $T \times T$ matrix with (i,j)th element $\Gamma_\varepsilon(|i-j|)$, where

$$\Gamma_\varepsilon(j) = E(\varepsilon_t\varepsilon_{t+j}),$$

Algorithm 6.1 Estimating f frequencies in the univariate sinusoidal model

1. Let $j = 1$.
2. Maximise the periodogram of $\{X_t\}$ in order to obtain $\widehat{\omega}_j$, where $\widehat{\omega}_j$ is the estimator of the frequency with the jth largest amplitude.
3. Estimate α_j and β_j by

$$\widehat{\alpha}_j = 2T^{-1} \sum_{t=0}^{T-1} X_t \cos(\widehat{\omega}_j t) \qquad \text{and} \qquad \widehat{\beta}_j = 2T^{-1} \sum_{t=0}^{T-1} X_t \sin(\widehat{\omega}_j t),$$

   respectively.
4. Put

$$X_t = X_t - \widehat{\alpha}_j \cos(\widehat{\omega}_j t) - \widehat{\beta}_j \sin(\widehat{\omega}_j t).$$

5. Let $j = j + 1$.
6. Repeat steps 2-5 until f frequencies have been estimated.

---

which is the jth autocovariance. The Gaussian log-likelihood is then

$$-\frac{T}{2}\log(2\pi) - \frac{1}{2}\log|\Gamma_\varepsilon| - \frac{1}{2}\{X - M_{T,f}(\omega)\theta_f\}' \Gamma_\varepsilon^{-1} \{X - M_{T,f}(\omega)\theta_f\}, \tag{6.11}$$

where

$$\theta_f = \begin{bmatrix} \mu & \alpha_1 & \cdots & \alpha_f & \beta_1 & \cdots & \beta_f \end{bmatrix}'$$

and $M_{T,f}(\omega)$ is the $T \times (2f+1)$ matrix with $(t+1)$th row

$$\begin{bmatrix} 1 & \cos(\omega_1 t) & \cdots & \cos(\omega_f t) & \sin(\omega_1 t) & \cdots & \sin(\omega_f t) \end{bmatrix},$$

$t = 0, \ldots, T-1$. As shown in Quinn and Hannan (2001), the maximiser of (6.11) is asymptotically equivalent to the maximiser of

$$\sum_{j=1}^{f} \frac{I_{T,X}(\omega_j)}{4\pi f_\varepsilon(\omega_j)}.$$

It follows that even if the spectral density is known, or unknown but estimated in some consistent way, we do not get better estimates, asymptotically, than we do by simply minimising (6.10). We can therefore use Algorithm 6.1 to estimate frequencies without assuming that the noise is Gaussian and white. The exception to this is the case where the frequencies are harmonics of a fundamental frequency, that is, when

$$X_t = \mu + \sum_{j=1}^{f} \{\alpha_j \cos(j\omega t) + \beta_j \sin(j\omega t)\} + \varepsilon_t. \tag{6.12}$$

Quinn and Thomson (1991) showed that ω in (6.12) can be estimated by maximising

$$\sum_{j=1}^{f} \frac{I_{T,X}(j\omega)}{\widehat{f}_{\varepsilon}(j\omega)},$$

where $\widehat{f}_{\varepsilon}(\omega)$ is a consistent estimator of $f_{\varepsilon}(\omega)$. Such an estimator may be obtained, for example, from a smoothed periodogram or by fitting a long-order autoregression.

We finally mention, for completeness, the case where two frequencies are close together. The procedures described above may not work in this case, since they may not discriminate between a local maximum of the periodogram due to one of the frequencies and a sidelobe of the other. We will not consider this case here but refer the reader to Hannan and Quinn (1989) for more details.

6.2.3 Maximising the Periodogram

In practice, the periodogram is very difficult to maximise. However, it can be easily evaluated at the set of Fourier frequencies using the fast Fourier transform algorithm. The maximiser of the periodogram over the Fourier frequencies will give an estimator for ω which is accurate to $O\left(T^{-1}\right)$ which is less than the desired accuracy of $O\left(T^{-3/2}\right)$. It is often used, however, as an initial estimate in some other estimation technique. For example, it may be used to initialise the Gauss–Newton algorithm to maximise the Gaussian white log-likelihood. Using the Gauss–Newton algorithm in this way is not guaranteed to converge to the true frequency (see, for example, Rice and Rosenblatt, 1988). However, Quinn et al. (2008) have shown that it will converge if the initial estimate is computed using the periodogram of the time series zero-padded to four times its length. An alternative approach to maximising the periodogram is to use the Quinn–Fernandes technique (Quinn and Fernandes, 1991). This technique is computationally fast and will result in an estimator which has the same central limit theorem as the periodogram maximiser. Both approaches are detailed below.

Maximising the Gaussian White Log-Likelihood Using the Gauss–Newton Algorithm

From the asymptotic results in Section 6.2.1, we can maximise the Gaussian white log-likelihood by minimising

$$R_T(\omega) = E_T'(\omega)\, E_T(\omega),$$

where

$$E_T(\omega) = X - M_T(\omega)\, \widehat{\theta}(\omega).$$

The derivative of $R_T(\omega)$ is

$$2E_T'(\omega)\frac{d}{d\omega}E_T(\omega)$$

and so, given a current estimate of ω, denoted $\widetilde{\omega}$, the Gauss–Newton algorithm updates the estimate by

$$\widetilde{\omega} - \frac{E_T'(\widetilde{\omega})\frac{d}{d\omega}E_T(\widetilde{\omega})}{\frac{d}{d\omega}E_T'(\widetilde{\omega})\frac{d}{d\omega}E_T(\widetilde{\omega})} \tag{6.13}$$

and repeats until convergence. The derivative of $E_T(\omega)$ is

$$\begin{aligned}&-\left\{\frac{d}{d\omega}M_T(\omega)\right\}\widehat{\theta}_T(\omega)\\&+M_T(\omega)\left\{M_T'(\omega)M_T(\omega)\right\}\left[\left\{\frac{d}{d\omega}M_T'(\omega)M_T(\omega)\right\}\widehat{\theta}_T(\omega)-\left\{\frac{d}{d\omega}M_T'(\omega)\right\}X\right].\end{aligned} \tag{6.14}$$

Even using asymptotic results, this can be onerous to compute. However, the algorithm can be computed efficiently, and the required quantities calculated exactly, by reparametrising (6.3) as

$$X_t = \mu^* + \alpha^*\left[\cos\{\omega(t-\nu)\} - \bar{c}(\omega)\right] + \beta^*\sin\{\omega(t-\nu)\} + \varepsilon_t,$$

where $\nu = (T-1)/2$,

$$\bar{c}_T(\omega) = T^{-1}\sum_{t=0}^{T-1}\cos\{\omega(t-\nu)\} \qquad \text{and} \qquad \mu^* = \mu + \alpha^*\bar{c}_T(\omega).$$

Let

$$E_T^*(\omega) = X - M_T^*(\omega)\widehat{\theta}_T^*(\omega),$$

where

$$\widehat{\theta}_T^*(\omega) = \left\{M_T^{*\prime}(\omega)M_T^*(\omega)\right\}^{-1}\left\{M_T^{*\prime}(\omega)X\right\}$$

and $M_T^*(\omega)$ is the $T\times 3$ matrix with $(t+1)$th row

$$\left[\begin{array}{ccc}1 & \cos\{\omega(t-\nu)\}-\bar{c}_T(\omega) & \sin\{\omega(t-\nu)\}\end{array}\right], \tag{6.15}$$

$t = 0,\ldots,T-1$. Under the reparametrisation, $M_T^{*\prime}(\omega)M_T^*(\omega)$, its inverse and its derivative can easily be computed exactly since

$$\sum_{t=0}^{T-1}\cos\{2\omega(t-\nu)\}$$

is the real part of

$$\sum_{t=0}^{T-1}e^{2i\omega(t-\nu)} = e^{-2i\omega\nu}\frac{e^{2i\omega T}-1}{e^{2i\omega}-1} = \frac{e^{i\omega T}-e^{-i\omega T}}{e^{i\omega}-e^{-i\omega}} = \frac{\sin(\omega T)}{\sin(\omega)}.$$

Letting $s_T(\omega) = \sin(\omega T) / \sin(\omega)$,

$$M_T^{*\prime}(\omega) M_T^*(\omega) = \begin{bmatrix} T & 0 & 0 \\ 0 & D_{T1}(\omega) & 0 \\ 0 & 0 & D_{T2}(\omega) \end{bmatrix},$$

where

$$D_{T1}(\omega) = \sum_{t=0}^{T-1} [\cos\{\omega(t-1)\} - \bar{c}(\omega)]^2 = \frac{T}{2} + \frac{1}{2} s_T(\omega) - T^{-1} s_T^2(\omega/2)$$

and

$$D_{T2}(\omega) = \sum_{t=0}^{T-1} \sin^2\{\omega(t-1)\} = \frac{T}{2} - \frac{1}{2} s_T(\omega).$$

The derivatives of $s_T(\omega)$ and $s_T^2(\omega)$ are

$$\frac{d}{d\omega} s_T(\omega) = \frac{T\cos(\omega T)}{\sin(\omega)} - \frac{\cos(\omega)\sin(\omega T)}{\sin^2(\omega)}$$

and

$$\frac{d}{d\omega} s_T^2(\omega) = 2\left\{\frac{d}{d\omega} s_T(\omega)\right\} s_T(\omega).$$

Thus

$$\frac{d}{d\omega}\left\{M_T^{*\prime}(\omega) M_T^*(\omega)\right\} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{d}{d\omega} D_{T1}(\omega) & 0 \\ 0 & 0 & \frac{d}{d\omega} D_{T2}(\omega) \end{bmatrix},$$

where

$$\frac{d}{d\omega} D_{T1}(\omega) = \frac{1}{2}\frac{d}{d\omega} s_T(\omega) - T^{-1}\left\{\frac{d}{d\omega} s_T(\omega/2)\right\} s_T(\omega/2)$$

and

$$\frac{d}{d\omega} D_{T2}(\omega) = -\frac{1}{2}\frac{d}{d\omega} s_T(\omega).$$

Finally, $dM_T(\omega)/d\omega$ is the $T \times 3$ matrix with $(t+1)$th row

$$\begin{bmatrix} 0 & -(t-\nu)\sin\{\omega(t-\nu)\} - (2T)^{-1}\frac{d}{d\omega} s_T(\omega/2) & (t-\nu)\cos\{\omega(t-\nu)\} \end{bmatrix},$$

$t = 0, \ldots, T-1$. The algorithm can therefore be performed by replacing $E_T(\omega)$ and $dE_T(\omega)/d\omega$ in (6.13) by $E_T^*(\omega)$ and $dE_T^*(\omega)/d\omega$, respectively, which are given by replacing $\widehat{\theta}_T(\omega)$ and $M_T(\omega)$ with $\widehat{\theta}_T^*(\omega)$ and $M_T^*(\omega)$, respectively.

The Quinn–Fernandes Technique

The Quinn–Fernandes technique is motivated by the fact that, assuming $\mu = 0$ and with $\{\varepsilon_t\}$ not necessarily white, (6.3) can be rewritten as

$$X_t - 2\cos(\omega) X_{t-1} + X_{t-2} = \varepsilon_t - 2\cos(\omega)\varepsilon_{t-1} + \varepsilon_{t-2},$$

which is the ARMA$(2,2)$-like model

$$X_t - bX_{t-1} + X_{t-2} = \varepsilon_t - a\varepsilon_{t-1} + \varepsilon_{t-2}, \tag{6.16}$$

where $a = b = 2\cos\omega$. Letting $\{\xi_t\}$ be the process given by

$$\xi_t = X_t + a\xi_{t-1} - \xi_{t-2},$$

where $\xi_{-1} = \xi_{-2} = 0$, (6.16) becomes

$$\xi_t + \xi_{t-2} = b\xi_{t-1} + \varepsilon_t.$$

Thus, given an estimate of a, b can be estimated by regressing $\xi_t + \xi_{t-2}$ on ξ_{t-1}. That is, by

$$\frac{\sum_{t=0}^{T_1-1} (\xi_t + \xi_{t-2})\, \xi_{t-1}}{\sum_{t=0}^{T_1-1} \xi_{t-1}^2} = a + \frac{\sum_{t=0}^{T-1} X_t \xi_{t-1}}{\sum_{t=0}^{T-1} \xi_{t-1}^2}. \tag{6.17}$$

This estimate of b can be used to re-estimate a which can in turn be used to re-compute $\{\xi_t\}$. An accelerated version of the algorithm multiplies the final term in (6.17) by two, increasing the rate of convergence. The full procedure is given in Algorithm 6.2. Note that since we are assuming that the process has zero mean, the data should be first mean corrected.

Algorithm 6.2 The Quinn–Fernandes Technique

1. Put $\widehat{a} = 2\cos\widehat{\omega}$, where $\widehat{\omega}$ is an initial estimate of ω.
2. For $t = 0, \ldots, T-1$, let
$$\xi_t = X_t + \widehat{a}\xi_{t-1} - \xi_{t-2},$$
where $\xi_{-1} = \xi_{-2} = 0$.
3. Let $\widehat{b} = \widehat{a} + \nu$, where
$$\nu = 2\frac{\sum_{t=0}^{T-1} X_t \xi_{t-1}}{\sum_{t=0}^{T-1} \xi_{t-1}^2},$$
and put $\widehat{a} = \widehat{b}$.
4. Repeat steps 2 and 3 until $|\nu|$ converges to 0.
5. Put $\widehat{\omega} = \cos^{-1}(\widehat{a}/2)$.

Quinn and Fernandes (1991) showed that the estimator of ω obtained by Algorithm 6.2 is strongly consistent and follows the same central limit theorem as the periodogram maximiser. Furthermore, provided the initial estimator of ω is accurate to $O\left(T^{-1/2}\right)$, steps 2 and 3 need to be repeated only once in order for the algorithm to be accurate to $O\left(T^{-3/2}\right)$. Such an initial estimator will be obtained from the maximiser of the periodogram of $\{X_t\}$ evaluated at the Fourier frequencies.

6.2.4 Estimating the Number of Frequencies

Until now we have assumed that the number of frequencies, f, is known. However, in practice this will not be the case. Quinn (1989), Wang (1993) and Kavalieris and Hannan (1994) have proposed information criteria to estimate the number of frequencies. The most general result is that of Kavalieris and Hannan (1994) which uses the information criterion

$$\phi(f,p) = T\log\left(\widehat{\sigma}^2_{f,p}\right) + (p+5f)\log T,$$

where $\widehat{\sigma}^2_{f,p}$ is the estimator of the residual variance obtained by fitting an autoregression of order p to the time series given by removing f sinusoids from $\{X_t\}$ by regression. The $5f$ term reflects the fact that, for each additional frequency estimated, $\widehat{\sigma}^2_{f,p}$ involves the estimation of an additional two parameters, α_f and β_f, which have asymptotic standard error of $O\left(T^{-1/2}\right)$ and of one additional parameter, ω_f, which has asymptotic standard error of $O\left(T^{-3/2}\right)$. For each $f = 0,\ldots,F$, where F is assumed to be greater than the true number of frequencies, the autoregressive order is estimated by $\widehat{p}$, and then the number of frequencies is estimated by the minimiser of $\phi(f,\widehat{p})$.

6.3 Multichannel Frequency Estimation

6.3.1 Estimating a Single Frequency

As in the univariate case, we begin with the case where $f = 1$. That is, we consider the model

$$X_t = \mu + \alpha\cos(\omega t) + \beta\sin(\omega t) + \varepsilon_t, \tag{6.18}$$

where μ, α, β, X_t and ε_t are $d\times 1$. We will motivate the estimation techniques by maximising the log-likelihood as though ε_t is d-dimensional multivariate normal with mean zero and covariance matrix Σ, $t = 0,\ldots,T-1$. The log-likelihood is then

$$-\frac{Td}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma| - \frac{1}{2}\operatorname{tr}\left[\left\{X - \theta M_T'(\omega)\right\}'\Sigma^{-1}\left\{X - \theta M_T'(\omega)\right\}\right], \tag{6.19}$$

where

$$X = \begin{bmatrix} X_0 & \cdots & X_{T-1}\end{bmatrix},$$
$$\theta = \begin{bmatrix} \mu & \alpha & \beta\end{bmatrix}$$

and $M_T(\omega)$ is defined in the same way as in the univariate case. Note that X is now a $d\times T$ matrix whose columns are the different samples. For fixed ω, the maximiser of (6.19) with respect to θ is

$$\left\{XM_T(\omega)\right\}\left\{M_T'(\omega)M_T(\omega)\right\}^{-1},$$

which, because of (6.6) and (6.7), has the same asymptotic properties as

$$\widehat{\theta}_T(\omega) = \left[\begin{array}{ccc} \overline{X} & 2T^{-1}\sum_{t=0}^{T-1}\cos(\omega t)X_t & 2T^{-1}\sum_{t=0}^{T-1}\sin(\omega t)X_t \end{array}\right].$$

The maximiser of (6.19) with respect to Σ therefore has the same asymptotic properties as

$$\widehat{\Sigma}_T(\omega) = T^{-1}\left\{\sum_{t=0}^{T-1}\left(X_t - \overline{X}\right)\left(X_t - \overline{X}\right)' - \left[\begin{array}{ccc} I_{11}(\omega) & \cdots & I_{1d}(\omega) \\ \vdots & \ddots & \vdots \\ I_{d1}(\omega) & \cdots & I_{dd}(\omega) \end{array}\right]\right\},$$

where

$$\begin{aligned} I_{jk}(\omega) &= 2T^{-1}\left[\left\{\sum_{t=0}^{T-1}\cos(\omega t)X_{t,j}\right\}\left\{\sum_{t=0}^{T-1}\cos(\omega t)X_{t,k}\right\}\right. \\ &\quad \left. + \left\{\sum_{t=0}^{T-1}\sin(\omega t)X_{t,j}\right\}\left\{\sum_{t=0}^{T-1}\sin(\omega t)X_{t,k}\right\}\right] \\ &= 2T^{-1}\,\mathrm{Re}\left[\left\{\sum_{t=0}^{T-1}e^{-i\omega t}X_{t,j}\right\}\left\{\sum_{t=0}^{T-1}e^{i\omega t}X_{t,k}\right\}\right] \end{aligned}$$

and $X_{t,j}$ is the jth component of X_t. Substituting $\widehat{\theta}_T(\omega)$ and $\widehat{\Sigma}_T(\omega)$ into (6.19), we have

$$-\frac{Td}{2}\left\{1 + \log(2\pi)\right\} - \frac{T}{2}\log\left|\widehat{\Sigma}_T(\omega)\right|.$$

The maximum likelihood estimator of ω is thus found by minimising $\left|\widehat{\Sigma}_T(\omega)\right|$. In order to minimise this, and to derive the asymptotic properties of the minimiser, the following lemma will be useful. The proof of the lemma is in the Appendix.

Lemma 6.1 *If A and B are d-dimensional vectors then*

$$\left|I_d - AA' - BB'\right| = 1 - \left(A'A + B'B\right) + \left(A'A\right)\left(B'B\right) - \left(A'B\right)^2$$

and

$$\left|I_d + AA' + BB'\right| = 1 + \left(A'A + B'B\right) + \left(A'A\right)\left(B'B\right) - \left(A'B\right)^2.$$

Now, letting

$$V_T = \sum_{t=0}^{T-1}\left(X_t - \overline{X}\right)\left(X_t - \overline{X}\right)',$$

$$C_T(\omega) = \sqrt{2}T^{-1/2}\sum_{t=0}^{T-1}\cos(\omega t)X_t$$

and

$$S_T(\omega) = \sqrt{2}T^{-1/2}\sum_{t=0}^{T-1}\sin(\omega t)X_t,$$

$\widehat{\Sigma}_T(\omega)$ can be rewritten as

$$\widehat{\Sigma}_T(\omega) = T^{-1}\left\{V_T - C_T(\omega)\,C_T'(\omega) - S_T(\omega)\,S_T'(\omega)\right\}.$$

The determinant of $\widehat{\Sigma}_T(\omega)$ is therefore

$$\begin{aligned}
&T^{-d}\left|V_T - C_T(\omega)\,C_T'(\omega) - S_T(\omega)\,S_T'(\omega)\right| \\
&= T^{-d}\left|V_T^{1/2}\right|\left|I_d - V_T^{-1/2}C_T(\omega)\,C_T'(\omega)\,V_T^{-1/2} - V_T^{-1/2}S_T(\omega)\,S_T'(\omega)\,V_T^{-1/2}\right|\left|V_T^{1/2}\right|.
\end{aligned}$$

From Lemma 6.1, putting

$$A = V_T^{-1/2}C_T(\omega) \qquad \text{and} \qquad B = V_T^{-1/2}S_T(\omega),$$

we have

$$\begin{aligned}
&\left|I_d - V_T^{-1/2}C_T(\omega)\,C_T'(\omega)\,V_T^{-1/2} - V_T^{-1/2}S_T(\omega)\,S_T'(\omega)\,V_T^{-1/2}\right| \\
&= 1 - C_T'(\omega)\,V_T^{-1}C_T(\omega) - S_T'(\omega)\,V_T^{-1}S_T(\omega) + \left\{C_T'(\omega)\,V_T^{-1}C_T(\omega)\right\}\left\{S_T'(\omega)\,V_T^{-1}S_T(\omega)\right\} \\
&\quad - \left\{C_T'(\omega)\,V_T^{-1}S_T(\omega)\right\}^2.
\end{aligned}$$

In order to compute the maximum likelihood estimator of ω we must therefore maximise

$$\begin{aligned}
J_T(\omega) = {}& C_T'(\omega)\,V_T^{-1}C_T(\omega) + S_T'(\omega)\,V_T^{-1}S_T(\omega) - \left\{C_T'(\omega)\,V_T^{-1}C_T(\omega)\right\}\left\{S_T'(\omega)\,V_T^{-1}S_T(\omega)\right\} \\
&+ \left\{C_T'(\omega)\,V_T^{-1}S_T(\omega)\right\}^2.
\end{aligned}$$

Suppose now that $\{\varepsilon_t\}$ is not white, but coloured, with spectral density $f_\varepsilon(\omega)$. Let

$$\widehat{\omega} = \arg\max_{\omega} J_T(\omega)$$

and denote by ω_0, Σ_0 and $\theta_0 = \begin{bmatrix} \mu_0 & \alpha_0 & \beta_0 \end{bmatrix}$ the true values of ω, Σ and θ, respectively. Theorem 6.1 shows that $T(\widehat{\omega} - \omega_0)$ converges almost surely to 0 and Theorem 6.2 establishes the central limit theorem. The proofs of the theorems are in the Appendix.

Theorem 6.1 *$T(\widehat{\omega} - \omega_0) \to 0$ almost surely as $T \to \infty$.*

Theorem 6.2 *The distribution of $T^{3/2}(\widehat{\omega} - \omega_0)$ converges to the normal distribution with mean zero and variance*

$$48\pi\frac{\alpha_0'\Sigma_0^{-1}f_\varepsilon(\omega_0)\,\Sigma_0^{-1}\alpha_0 + \beta_0'\Sigma_0^{-1}f_\varepsilon(\omega_0)\,\Sigma_0^{-1}\beta_0}{\left(\alpha_0'\Sigma_0^{-1}\alpha_0 + \beta_0'\Sigma_0^{-1}\beta_0\right)^2}$$

as $T \to \infty$.

The above suggests that rather than maximising $J_T(\omega)$, we might maximise the simpler function

$$\begin{aligned}\widetilde{J}_{T,\Omega}(\omega) &= C_T'(\omega)\,\Omega C_T(\omega) + S_T'(\omega)\,\Omega S_T(\omega)\\ &= F_T^*(\omega)\,\Omega F_T(\omega),\end{aligned}$$

where

$$\begin{aligned}F_T(\omega) &= C_T(\omega) - iS_T(\omega)\\ &= \sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} e^{-i\omega t}X_t\end{aligned}$$

and Ω is a suitable positive definite symmetric matrix. Let

$$\widetilde{\omega} = \arg\max_{\omega} \widetilde{J}_{T,\Omega}(\omega).$$

Theorem 6.3 shows that $T(\widetilde{\omega} - \omega_0)$ converges almost surely to 0 and Theorem 6.4 establishes the central limit theorem.

Theorem 6.3 *$T(\widetilde{\omega} - \omega_0) \to 0$ almost surely as $T \to \infty$.*

Theorem 6.4 *The distribution of $T^{3/2}(\widetilde{\omega} - \omega_0)$ converges to the normal distribution with mean zero and variance*

$$48\pi\frac{\alpha_0'\Omega f_\varepsilon(\omega_0)\,\Omega\alpha_0 + \beta_0'\Omega f_\varepsilon(\omega_0)\,\Omega\beta_0}{\left(\alpha_0'\Omega\alpha_0 + \beta_0'\Omega\beta_0\right)^2} \tag{6.20}$$

as $T \to \infty$.

Letting $\eta_0 = \alpha_0 - i\beta_0$, (6.20) can be rewritten as

$$48\pi\frac{\eta_0^*\Omega f_\varepsilon(\omega_0)\,\Omega\eta_0}{\left(\eta_0^*\Omega\eta_0\right)^2}.$$

An application of the Kantorovich matrix inequality (see Gentle, 2017, Section 8.4.3) shows that

$$\frac{\eta_0^*\Omega f_\varepsilon(\omega_0)\,\Omega\eta_0}{\left(\eta_0^*\Omega\eta_0\right)^2} \geqslant \frac{1}{\eta_0^* f_\varepsilon^{-1}(\omega_0)\,\eta_0} \tag{6.21}$$

with equality if and only if $\Omega = cf_\varepsilon^{-1}(\omega_0)$ for some constant c. To see this, consider the Cauchy–Schwarz inequality,

$$(x'x)(y'y) \geqslant (x'y)^2$$

with equality if and only if $x = cy$, for vectors x and y and any constant c. Putting $x = Q^{-1/2}z$ and $y = Q^{1/2}z$, where Q is positive definite, we have

$$\left(z'Q^{-1}z\right)\left(z'Qz\right) \geqslant \left(z'z\right)^2$$

with equality if and only if

$$Q^{-1/2}z = cQ^{1/2}z,$$

that is, if and only if $Q = cI_d$. Now, putting $z = \Omega^{1/2}\eta_0$ and $Q = \Omega^{-1/2} f_\varepsilon^{-1}(\omega_0)\,\Omega^{-1/2}$, we obtain (6.21) and have equality if and only if

$$\Omega^{-1/2} f_\varepsilon^{-1}(\omega_0)\,\Omega^{-1/2} = cI_d,$$

that is, if and only if $\Omega = cf_\varepsilon^{-1}(\omega_0)$, for some scalar constant c.

An efficient estimator of ω can therefore be obtained by maximising $\widetilde{J}_{T,\Omega}(\omega)$ with Ω equal to $f_\varepsilon^{-1}(\omega_0)$. In practice, $f_\varepsilon(\omega_0)$ will not be known, and will need to be estimated. Note that the multivariate case is thus very different from the univariate, where estimation of the spectral density is not needed.

6.3.2 Autoregressive Approximation

Since $\{\varepsilon_t\}$ is stationary, it can be modelled by a long-order autoregression. We thus fit

$$X_t = \mu + \alpha\cos(\omega t) + \beta\sin(\omega t) + \varepsilon_t,$$

where

$$\sum_{j=0}^{p} \delta_j \varepsilon_{t-j} = u_t,$$

$\delta_0 = I_d$, δ_j, $j = 1, \ldots, p$, are $d \times d$ and $\{u_t\}$ is d-dimensional. We assume that $\{u_t\}$ is a sequence of martingale differences and that

$$E\left(u_t u_t' \mid \mathcal{F}_{t-1}\right) = G,$$

where $\mathcal{F}_t$ is the σ-field generated by $\{u_t, u_{t-1}, \ldots\}$. Note that

$$f_\varepsilon(\omega) = \frac{1}{2\pi}\left(I_d + \sum_{j=1}^{p}\delta_j e^{-ij\omega}\right)^{-1} G \left\{\left(I_d + \sum_{j=1}^{p}\delta_j e^{-ij\omega}\right)^{*}\right\}^{-1}.$$

The autoregressive order, p, will in general be unknown and may be estimated using an information criterion (see Section 3.3.4). Alternatively, the order could be fixed at $\lfloor(\log T)^c\rfloor$, $c > 1$ (see Section 3.6). In what follows, we will assume that p is known, and then introduce its estimation into the resulting algorithms.

Let

$$D = \operatorname{vec}\delta = \operatorname{vec}\left[\begin{array}{ccc}\delta_1 & \cdots & \delta_p\end{array}\right]$$

and $\widetilde{G} = \operatorname{vec} G$. The Gaussian log-likelihood is

$$l(\Theta) = -\frac{Td}{2}\log(2\pi) - \frac{T}{2}\log|G| - \frac{1}{2}\sum_{t=0}^{T-1} U_t' G^{-1} U_t,$$

where

$$\Theta = \left[\begin{array}{cccccc} \mu' & D' & \widetilde{G}' & \alpha' & \beta' & \omega \end{array} \right]',$$

$$U_t = d_\delta (z) \{X_t - \mu - \alpha \cos(\omega t) - \beta \sin(\omega t)\}$$

and $d_\delta (z) = I_d + \sum_{j=1}^{p} \delta_j z^j$. Let

$$\widehat{\omega} = \arg\max_{\omega} l(\Theta)$$

and let ω_0 be the true value of ω. Theorem 6.5 establishes the central limit theorem for $\widehat{\omega}$. The proof of the theorem is in the Appendix.

Theorem 6.5 *The distribution of $T^{3/2}(\widehat{\omega} - \omega_0)$ converges to the normal distribution with mean zero and variance*

$$\frac{48\pi}{\eta_0^* f_\varepsilon^{-1}(\omega_0)\,\eta_0}$$

as $T \to \infty$.

In light of this, an efficient estimator of ω may be obtained by fitting a long-order autoregression to the noise process $\{\varepsilon_t\}$ and using the inverse of the estimated spectral density in place of Ω in maximising $\widetilde{J}_{T,\Omega}(\omega)$.

6.3.3 A Procedure for Multichannel Frequency Estimation

As we have seen, an efficient estimator of ω in (6.18) is obtained by maximising $\widetilde{J}_{T,\Omega}(\omega)$ with $\Omega = \widehat{f}_\varepsilon^{-1}(\omega)$, where $\widehat{f}_\varepsilon(\omega)$ is an estimator of $f_\varepsilon(\omega)$ which can be obtained by fitting a long-order autoregression to $\{\varepsilon_t\}$. However, in order to estimate $f_\varepsilon(\omega)$ we first need an estimate of ω. We can therefore estimate ω in two stages. We begin by maximising $\widetilde{J}_{T,\Omega}(\omega)$ with Ω equal to a suitable symmetric positive definite matrix, for example I_d. Although the resulting estimator will not be efficient, it is of the correct order of efficiency. The sinusoid at the estimated frequency is then removed by regression and $f_\varepsilon(\omega)$ is estimated by fitting a long-order autoregression to the residuals. The inverse of the estimated spectral density is then used in place of Ω when maximising $\widetilde{J}_{T,\Omega}(\omega)$ in the second stage. The full procedure is given in Algorithm 6.3.

It remains to maximise $\widetilde{J}_{T,\Omega}(\omega)$ for a given Ω. This can be done, for example, using the Gauss–Newton algorithm. Given a current estimate of ω, denoted $\widetilde{\omega}$, the Gauss–Newton algorithm updates the estimate by

$$\widetilde{\omega} + \frac{\operatorname{Re}\left\{F_T^*(\widetilde{\omega})\,\Omega \frac{\partial}{\partial\omega} F_T(\widetilde{\omega})\right\}}{\frac{\partial}{\partial\omega} F_T^*(\widetilde{\omega})\,\Omega \frac{\partial}{\partial\omega} F_T(\widetilde{\omega})},$$

Algorithm 6.3 Estimating a single frequency in the multichannel sinusoidal model

1. Put $\Omega = I_d$ and let
$$\widetilde{\omega} = \arg\max_{\omega} \widetilde{J}_{T,\Omega}(\omega).$$
2. Fit an autoregression of order p to
$$X_t - \overline{X} - \widehat{\alpha}\cos(\omega t) - \widehat{\beta}\sin(\omega t),$$
where
$$\widehat{\alpha} = 2T^{-1}\sum_{t=0}^{T-1}\cos(\widetilde{\omega}t)X_t, \qquad \widehat{\beta} = 2T^{-1}\sum_{t=0}^{T-1}\sin(\widetilde{\omega}t)X_t$$
and p is either estimated using an information criterion or fixed at $\left\lfloor (\log T)^{1+c}\right\rfloor$, $c > 0$. Denote the autoregressive parameter estimates by $\widehat{\delta}_1, \ldots, \widehat{\delta}_p$ and the residual covariance matrix estimate by $\widehat{G}$.
3. Put
$$\Omega = 2\pi\left(I_d + \sum_{j=1}^{p}\widehat{\delta}_j e^{-ij\widetilde{\omega}}\right)^{*}\widehat{G}^{-1}\left(I_d + \sum_{j=1}^{p}\widehat{\delta}_j e^{-ij\widetilde{\omega}}\right).$$
4. Let
$$\widetilde{\omega} = \arg\max_{\omega} \widetilde{J}_{T,\Omega}(\omega).$$

---

where
$$\frac{\partial}{\partial\omega}F_T(\omega) = -i\sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} te^{-i\omega t}X_t.$$

An alternative method for maximising $\widetilde{J}_{T,\Omega}(\omega)$ is to use a multivariate version of the Quinn–Fernandes technique which we introduce below.

As discussed in Section 6.2.3, the estimator obtained using the Gauss–Newton algorithm will converge to the true value in the univariate case if the initial value is computed by maximising the periodogram at the Fourier frequencies with the time series first zero-padded to four times its length. Whether this result applies to the multivariate case remains an open problem, however simulations suggest that a zero-padding factor of four is appropriate here. For example, for the simulation study presented in Section 6.4, zero-padding the time series to four times their length generally produced the same results as zero-padding them to eight times their length. For now, therefore, we will compute initial values in practice using a zero-padding factor of four. Confirming the appropriate zero-padding factor theoretically will be left for future work.

6.3.4 The Multivariate Quinn–Fernandes Technique

As in the univariate case, and assuming that $\mu = 0$, we can rewrite (6.18) as

$$X_t - 2\cos(\omega) X_{t-1} + X_{t-2} = \varepsilon_t - 2\cos(\omega)\varepsilon_{t-1} + \varepsilon_{t-2}.$$

We therefore fit the VARMA(p,q)-like model

$$X_t - bX_{t-1} + X_{t-2} = \varepsilon_t - a\varepsilon_{t-1} + \varepsilon_{t-2}$$

subject to the constraint that $a = b$. Let

$$\xi_t = X_t + a\xi_{t-1} - \xi_{t-2},$$

where $\xi_{-1} = \xi_{-2} = 0$. We then wish to estimate b in

$$\xi_t + \xi_{t-2} = b\xi_{t-1} + \varepsilon_t.$$

We can estimate b by, for example,

$$\frac{\sum_{t=0}^{T-1} \xi'_{t-1}\Omega(\xi_t + \xi_{t-2})}{\sum_{t=0}^{T-1} \xi'_{t-1}\Omega\xi_{t-1}} = a + \frac{\sum_{t=0}^{T-1} \xi'_{t-1}\Omega X_t}{\sum_{t=0}^{T-1} \xi'_{t-1}\Omega\xi_{t-1}}. \tag{6.22}$$

Thus, given an estimate of a, b can be estimated using (6.22). This estimate of b can be used to re-estimate a and the procedure can be repeated until convergence. As in the univariate case, an accelerated version of the algorithm multiplies the second term in (6.22) by two, which increases the rate of convergence. The full procedure is given in Algorithm 6.4.

Algorithm 6.4 The multivariate Quinn–Fernandes technique for a given Ω

1. Put $\widehat{a} = 2\cos(\widehat{\omega})$, where $\widehat{\omega}$ is an initial estimate of ω.
2. For $t = 0, \ldots, T-1$, let
$$\xi_t = X_t + \widehat{a}\xi_{t-1} - \xi_{t-2},$$
where $\xi_{-1} = \xi_{-2} = 0$.
3. Replace $\widehat{a}$ by $\widehat{a} + \nu$, where
$$\nu = 2\frac{\sum_{t=0}^{T-1} \xi'_{t-1}\Omega X_t}{\sum_{t=0}^{T-1} \xi'_{t-1}\Omega\xi_{t-1}}.$$
4. Repeat steps 2 and 3 until $|\nu|$ converges to 0.
5. Put $\widehat{\omega} = \cos^{-1}(\widehat{a}/2)$.

Let a_j be the jth iterate of a and $\omega_j = \cos^{-1}(a_j/2)$. Then

$$a_{j+1} - a_j = 2\sin\omega_j h_T(\omega_j),$$

where

$$h_T(\omega) = \frac{\sin\omega \sum_{t=0}^{T-1} \xi'_{t-1}\Omega X_t}{\sin^2\omega \sum_{t=0}^{T-1} \xi'_{t-1}\Omega \xi_{t-1}}.$$

Now, $a_j = 2\cos\omega_j$ and so

$$\begin{aligned} a_{j+1} - a_j &= 2\cos\omega_{j+1} - 2\cos\omega_j \\ &= -2\sin\omega_j(\omega_{j+1} - \omega_j) + O\left\{(\omega_{j+1}-\omega_j)^2\right\}. \end{aligned}$$

Thus, provided the algorithm begins with an initial estimate which is $o\left(T^{-1/2}\right)$, the $(j+1)$th iterate of ω is asymptotically equivalent to

$$\omega_j - h_T(\omega_j).$$

An alternative version of Algorithm 6.4 therefore replaces step 3 with updating $\widehat{\omega}$ by

$$\widehat{\omega} - \frac{\sum_{t=0}^{T-1} \xi'_{t-1}\Omega X_t}{\sum_{t=0}^{T-1} \xi'_{t-1}\Omega \xi_{t-1}}$$

then letting $\widehat{a} = 2\cos\widehat{\omega}$. Step 5 is then not required.

Theorem 6.6 shows that there is a unique point, $\widehat{\omega}_T$, such that $T^{\nu}(\widehat{\omega}_T - \omega_0) \to 0$ almost surely for all $\nu < 3/2$. Theorem 6.7 shows how many iterates of the algorithm are needed to converge to the unique point and, in particular, that if the initial estimator is $O\left(T^{-1}\right)$, only two iterates are needed. Theorem 6.8 shows that the estimator has the same central limit theorem as the maximiser of $\widetilde{J}_{T,\Omega}(\omega)$. The proofs of the theorems are in the Appendix.

Theorem 6.6 *Let $A_T(\nu) = \{\omega; |\omega - \omega_0| < cT^{-\nu}\}$, for fixed constants $c > 0$ and $1 < \nu < 3/2$. Then there exists a unique $\widehat{\omega}_T \in A_T(\nu)$ such that $h_T(\widehat{\omega}_T) = 0$ almost surely as $T \to \infty$. Thus there is a unique solution to $h_T(\widehat{\omega}_T) = 0$ for which $T^{\nu}(\widehat{\omega}_T - \omega_0) \to 0$ almost surely as $T \to \infty$, for all $\nu < 3/2$.*

Theorem 6.7 *Let $\omega_1 \in A_T(\nu)$ and $\omega_{j+1} = \omega_j - h_T(\omega_j)$. If $1 < \nu < 3/2$, then*

$$\omega_{j+1} - \widehat{\omega}_T = (\omega_j - \widehat{\omega}_T)\, O\left\{T^{-1/2}(\log T)^{1/2}\right\},$$

while if $1/2 < \nu \leqslant 1$,

$$\omega_{j+1} - \widehat{\omega}_T = (\omega_j - \widehat{\omega}_T)\, O\left\{T^{-1/2-\nu}(\log T)^{1/2}\right\} + O\left\{T^{-1/2-2\nu}(\log T)^{1/2}\right\}.$$

Also,

$$\widehat{\omega}_T - \omega_0 = O\left\{T^{-3/2}(\log T)^{1/2}\right\}$$

and

$$\omega_j - \widehat{\omega}_T = o\left(T^{-3/2}\right)$$

for

$$j \geqslant \lfloor 3 - \log(2\nu - 1)/\log 2 \rfloor.$$

Theorem 6.8 *The distribution of $T^{3/2}\left(\widehat{\omega}_T - \omega_0\right)$ converges to the normal distribution with mean zero and variance*

$$48\pi \frac{\alpha_0' \Omega f_\varepsilon\left(\omega_0\right) \Omega \alpha_0 + \beta_0' \Omega f_\varepsilon\left(\omega_0\right) \Omega \beta_0}{\left(\alpha_0' \Omega \alpha_0 + \beta_0' \Omega \beta_0\right)^2}$$

as $T \to \infty$.

We can therefore estimate ω in (6.18) by using the procedure described in Section 6.3.3 with Algorithm 6.4 replacing the maximisation of $\widetilde{J}_{T,\Omega}\left(\omega\right)$. For an initial estimate, we can use the maximiser of $\widetilde{J}_{T,\Omega}\left(\omega\right)$ over the Fourier frequencies.

6.3.5 Estimating More Than One Frequency

When there is more than one frequency, that is, when $f > 1$, the Gaussian white log-likelihood is

$$-\frac{Td}{2}\log\left(2\pi\right) - \frac{T}{2}\log\left|\Sigma\right| - \frac{1}{2}\operatorname{tr}\left[\left\{X - \theta_f M_{T,f}'\left(\omega\right)\right\}' \Sigma^{-1}\left\{X - \theta_f M_{T,f}'\left(\omega\right)\right\}\right],$$

where

$$\theta = \left[\begin{array}{ccccccc} \mu & \alpha_1 & \cdots & \alpha_f & \beta_1 & \cdots & \beta_f \end{array}\right].$$

This is maximised when

$$\left|V_T - \sum_{j=1}^{f}\left\{C_T\left(\omega_j\right) C_T'\left(\omega_j\right) + S_T\left(\omega_j\right) S_T'\left(\omega_j\right)\right\}\right|$$

is minimised. Thus, just as in the univariate case, the frequencies can be estimated one at a time, providing none of them are too close together, using a procedure similar to that given in Section 6.2.2. The estimation is performed in two stages. In the first, Ω is set to I_d. In the second, Ω is set to the inverse of the estimated spectral density of the residuals after removing the f sinusoids. The full procedure is given in Algorithm 6.5.

To estimate the number of frequencies we can use an information criterion along the lines of that proposed by Kavalieris and Hannan (1994) for the univariate case (see Section 6.2.4). We estimate f using

$$\phi\left(f,p\right) = T\log\left|\widehat{G}_{f,p}\right| + \left\{dp^2 + \left(2d+3\right)f\right\}\log T,$$

where $\widehat{G}_{f,p}$ is the estimator of Σ assuming there are f frequencies and the autoregressive order is p. For each $f = 0,\ldots,F$, where F is assumed to be greater than the true number of frequencies, the autoregressive order is estimated by $\widehat{p}_f$ and then the estimator of f is the minimiser of $\phi\left(f,\widehat{p}_f\right)$.

Algorithm 6.5 Estimating f frequencies in the multichannel sinusoidal model

1. Put $\Omega_j = I_d$, $j = 1, \ldots, f$.
2. Let $j = 1$.
3. Let
$$\widetilde{\omega}_j = \arg\max_{\omega} \widetilde{J}_{T,\Omega_j}(\omega)$$
and
$$X_t = X_t - \widetilde{a}\cos(\widetilde{\omega}_j t) - \widetilde{\beta}\sin(\widetilde{\omega}_j t),$$
where
$$\widetilde{a} = 2T^{-1}\sum_{t=0}^{T-1}\cos(\widetilde{\omega}_j t) X_t \qquad \text{and} \qquad \widetilde{\beta}_j = 2T^{-1}\sum_{t=0}^{T-1}\sin(\widetilde{\omega}_j t) X_t.$$
4. Let $j = j + 1$.
5. Repeat steps 2–4 until f frequencies have been estimated.
6. Fit an autoregression of order p to $\{X_t\}$ where p is either estimated using an information criterion or fixed at $\lfloor(\log T)^c\rfloor$, $c > 1$. Denote the autoregressive parameter estimates by $\widehat{\delta}_1, \ldots, \widehat{\delta}_{\widehat{p}}$ and the residual covariance matrix estimate by $\widehat{G}$.
7. For $j = 1, \ldots, f$, put
$$\Omega_j = 2\pi\left(I_d + \sum_{j=1}^{p}\widehat{\delta}_j e^{-ij\widetilde{\omega}_j}\right)^* \widehat{G}^{-1}\left(I_d + \sum_{j=1}^{p}\widehat{\delta}_j e^{-ij\widetilde{\omega}_j}\right).$$
8. Repeat steps 2–5 once.

6.4 Simulations

The frequency estimation procedure described in Section 6.3.3 was applied to sets of time series with 10,000 replications simulated from the model

$$X_t = \mu + \alpha\cos(\omega t) + \beta\sin(\omega t) + \varepsilon_t,$$

where

$$\mu = 0, \qquad \alpha = \left[\begin{array}{cc} 1/\sqrt{2} & 1/\sqrt{2} \end{array}\right]' \qquad \text{and} \qquad \beta = \left[\begin{array}{cc} -1/\sqrt{2} & -1/\sqrt{2} \end{array}\right]'.$$

Note that the amplitude of the sinusoid is 1. The sample sizes were $T = 100$, 250, 500 and 1,000. The noise process was either generated from white noise, where $\varepsilon_t = u_t$, the autoregressive model

$$\varepsilon_t + \delta_1\varepsilon_{t-1} = u_t$$

with

$$\delta_1 = \begin{bmatrix} 0.7 & 0.3 \\ -0.3 & 0.7 \end{bmatrix},$$

or the moving average model

$$\varepsilon_t = u_t + \delta_2 u_{t-1}$$

with

$$\delta_2 = \begin{bmatrix} 0.8 & 0.1 \\ -0.1 & 0.8 \end{bmatrix}.$$

The residuals, $\{u_t\}$, were simulated from the multivariate normal distribution with mean zero and covariance matrix G, where G was equal to gI_2, $g = 0.1, 0.2, \ldots 4$. Both the Gauss–Newton algorithm and Quinn–Fernandes technique were used to maximise $\widetilde{J}_{T,\Omega}(\omega)$ with a tolerance for convergence of 1.0^{-6}. That is, the algorithms were stopped when successive iterates were within 1.0^{-6} of each other. The autoregressive orders were estimated using BIC.

The AR(1) and MA(1) parts of Figure 5.1 show the component spectral densities and coherency for the two processes used here for the case where $g = 1$. The simulations were run for both $\omega = \pi/5 + \pi/(4T)$ and $\omega = 4\pi/5 - \pi/(4T)$. When the true frequency is high, the autoregressive noise causes a spike in $\widetilde{J}_{T,\Omega}(\omega)$ close to the true frequency, making it more difficult to estimate it particularly with a low signal to noise ratio. The same occurs when the true frequency is low and the noise is from the moving average process. The addition or subtraction of $\pi/(4T)$ ensures the true frequency falls in the middle of two Fourier frequencies, which is a worst case scenario.

Figures 6.1–6.10 show the resulting mean estimates of ω and the logarithm of their mean squared errors (MSE). The plots show that the estimation procedure works well, with the mean estimates of ω close to the true values and their MSEs close to the theoretical variances, up to a point at which the signal-to-noise ratio becomes too small. That is, when g becomes large. This is known as the threshold effect (Quinn and Kootsookos, 1994) and, when it occurs, can be clearly seen in the log(MSE) plots. For example, when $T = 100$ and $\{\varepsilon_t\}$ is white noise, Figure 6.2 shows the threshold effect occurring around $g = 1.8$. It occurs sooner in the AR(1) and MA(1) cases than for the white noise cases. After the threshold effect occurs, the mean estimates of ω move toward the peak of the noise spectral density. In all cases, there was no threshold effect when $T = 1,000$ up to $g = 4$. Although the two methods used produced very similar results, there appears to be a small bias when the Quinn–Fernandes technique is used.

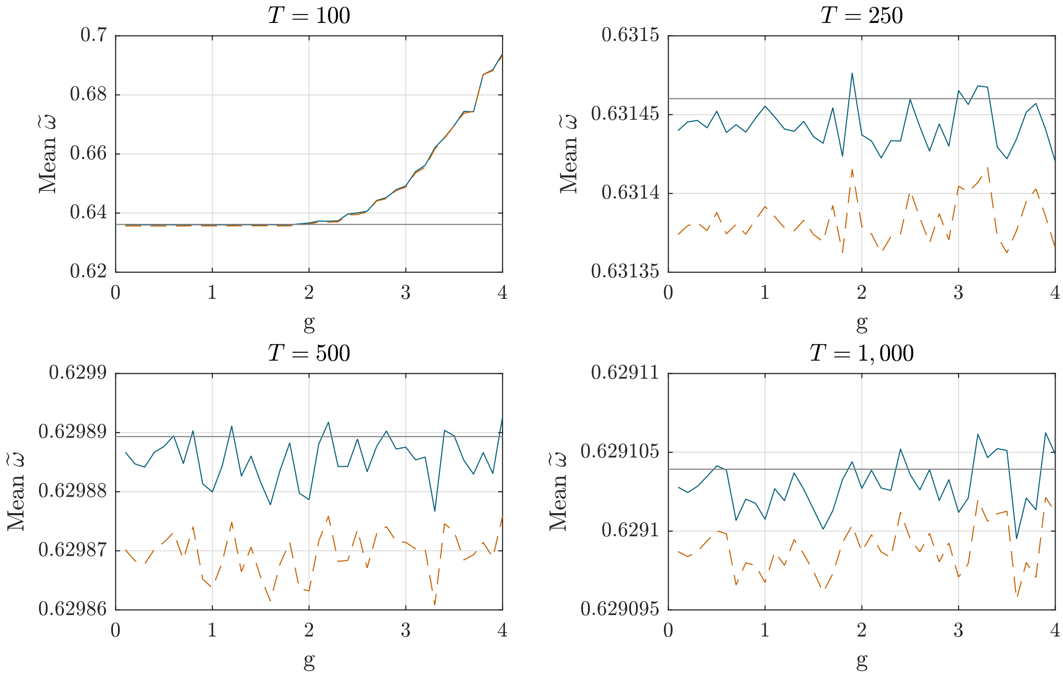

Figure 6.1: Mean $\widetilde{\omega}$ when $\{\varepsilon_t\}$ is white noise and $\omega_0 = \frac{\pi}{5} + \frac{\pi}{4T}$, using the Gauss–Newton (solid) and Quinn–Fernandes (dashes) methods. The true frequency is indicated by the grey line.

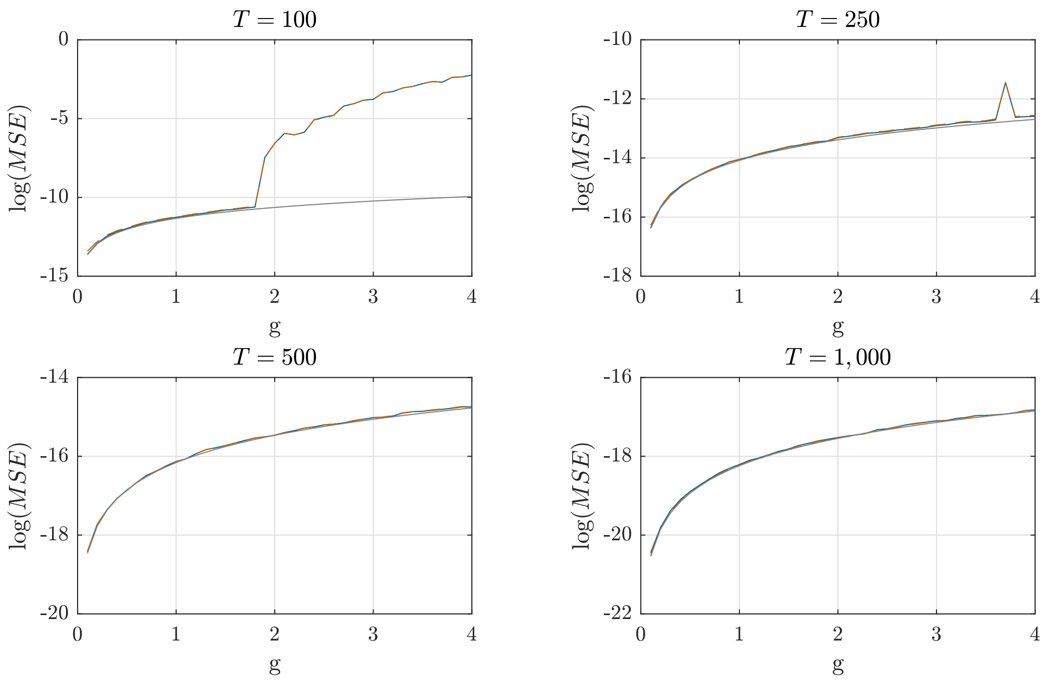

Figure 6.2: log(MSE) for $\widetilde{\omega}$ when $\{\varepsilon_t\}$ is white noise and $\omega_0 = \frac{\pi}{5} + \frac{\pi}{4T}$, using the Gauss–Newton (solid) and Quinn–Fernandes (dashes) methods. The theoretical variance is indicated by the grey line.

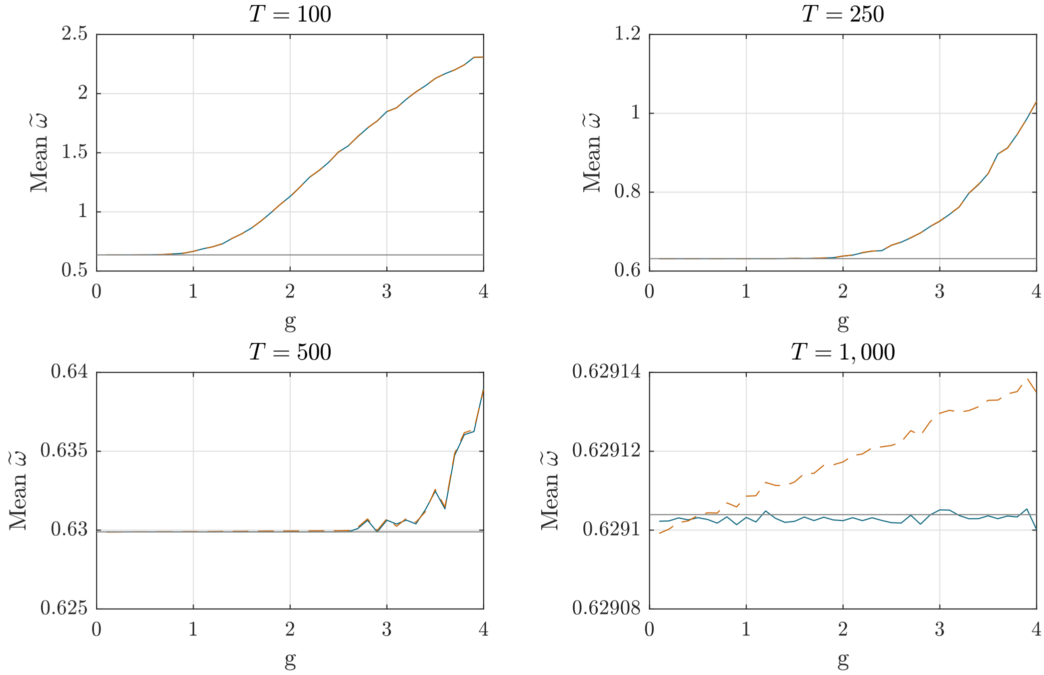

Figure 6.3: Mean $\widetilde{\omega}$ when $\{\varepsilon_t\}$ is an AR(1) and $\omega_0 = \frac{\pi}{5} + \frac{\pi}{4T}$, using the Gauss–Newton (solid) and Quinn–Fernandes (dashes) methods. The true frequency is indicated by the grey line.

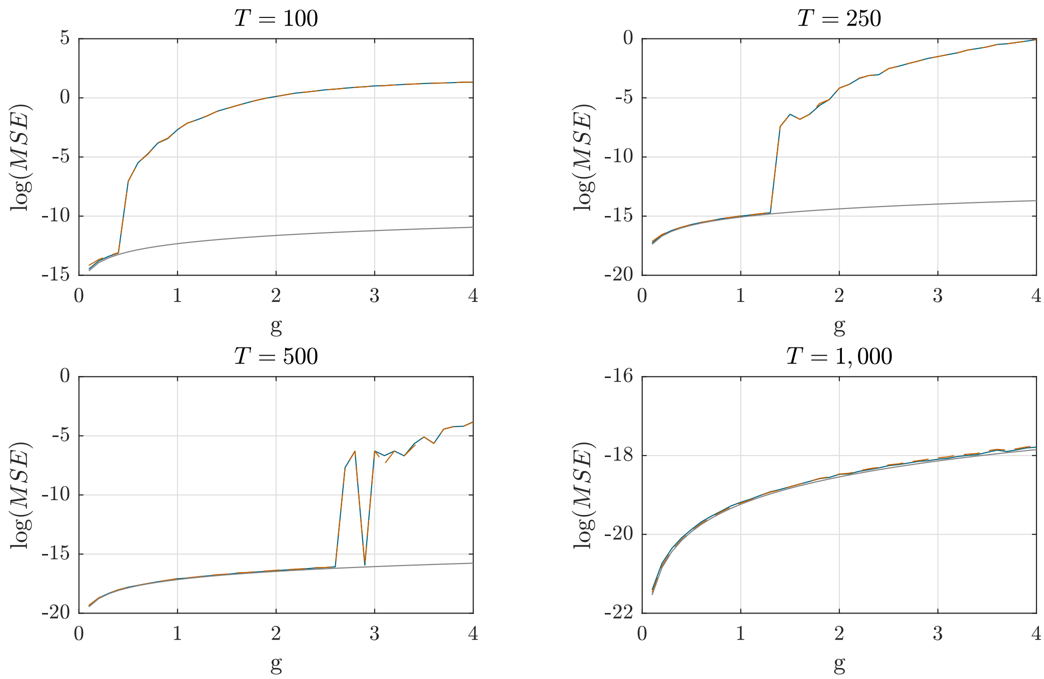

Figure 6.4: log(MSE) for $\widetilde{\omega}$ when $\{\varepsilon_t\}$ is an AR(1) and $\omega_0 = \frac{\pi}{5} + \frac{\pi}{4T}$, using the Gauss–Newton (solid) and Quinn–Fernandes (dashes) methods. The theoretical variance is indicated by the grey line.

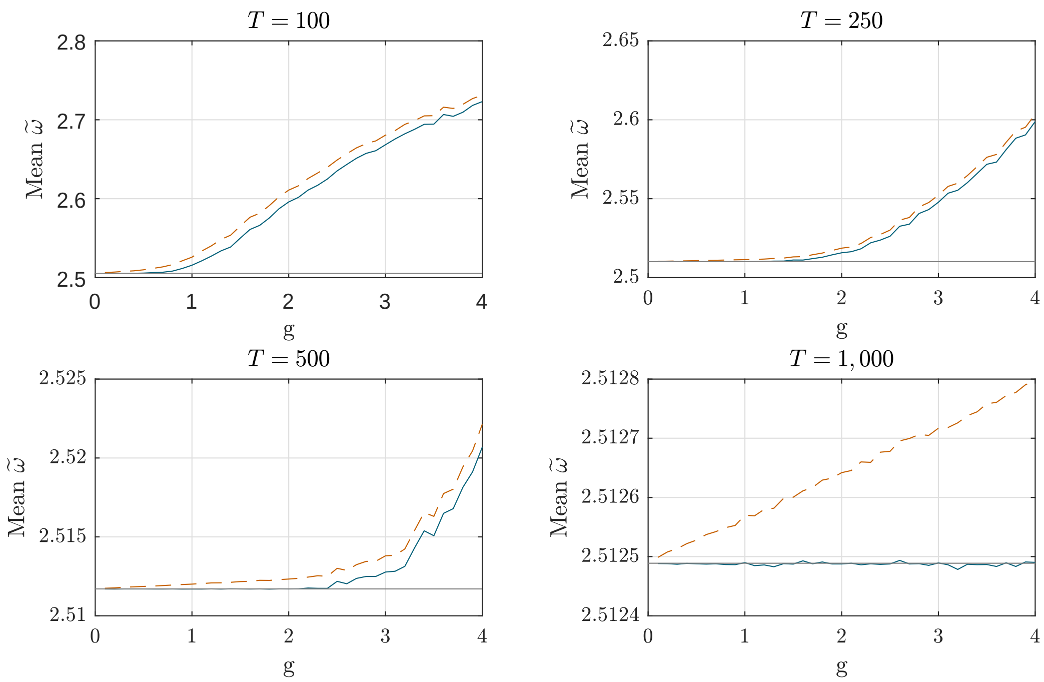

Figure 6.5: Mean $\widetilde{\omega}$ when $\{\varepsilon_t\}$ is an AR(1) and $\omega_0 = \frac{4\pi}{5} - \frac{\pi}{4T}$, using the Gauss–Newton (solid) and Quinn–Fernandes (dashes) methods. The true frequency is indicated by the grey line.

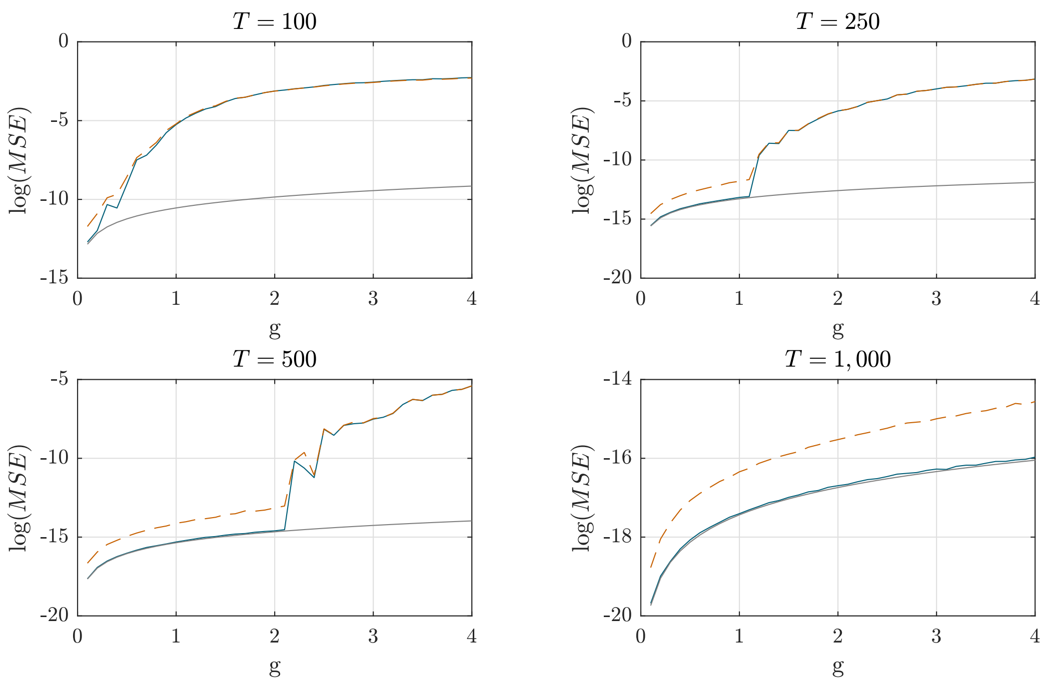

Figure 6.6: log(MSE) for $\widetilde{\omega}$ when $\{\varepsilon_t\}$ is an AR(1) and $\omega_0 = \frac{4\pi}{5} - \frac{\pi}{4T}$, using the Gauss–Newton (solid) and Quinn–Fernandes (dashes) methods. The theoretical variance is indicated by the grey line.

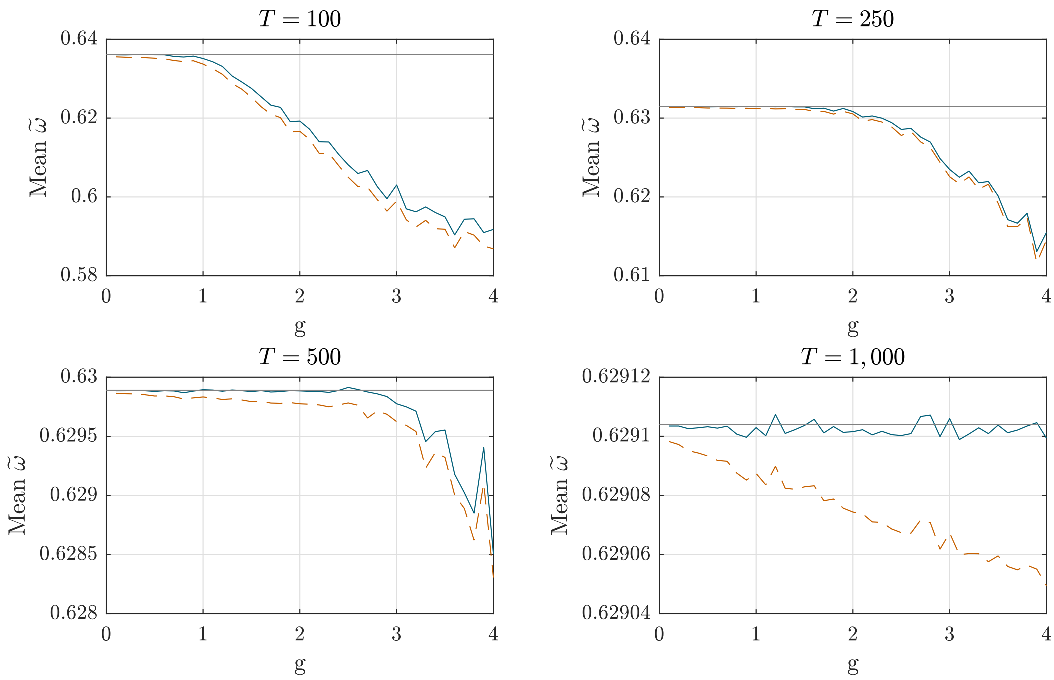

Figure 6.7: Mean $\widetilde{\omega}$ when $\{\varepsilon_t\}$ is an MA(1) and $\omega_0 = \frac{\pi}{5} + \frac{\pi}{4T}$, using the Gauss–Newton (solid) and Quinn–Fernandes (dashes) methods. The true frequency is indicated by the grey line.

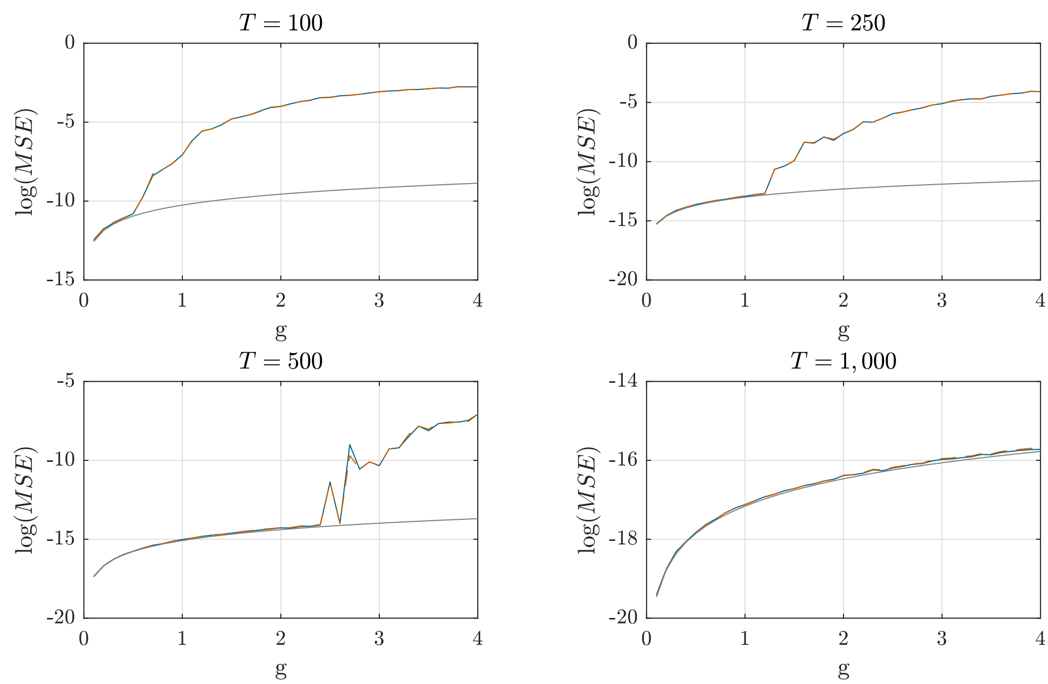

Figure 6.8: log(MSE) for $\widetilde{\omega}$ when $\{\varepsilon_t\}$ is an MA(1) and $\omega_0 = \frac{\pi}{5} + \frac{\pi}{4T}$, using the Gauss–Newton (solid) and Quinn–Fernandes (dashes) methods. The theoretical variance is indicated by the grey line.

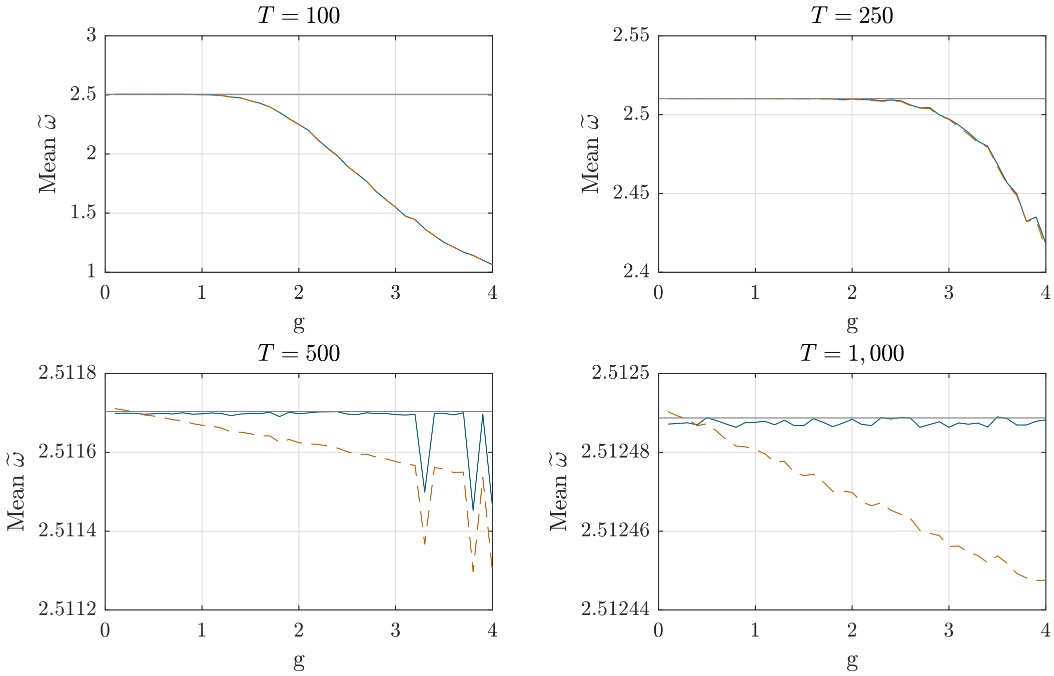

Figure 6.9: Mean $\widetilde{\omega}$ when $\{\varepsilon_t\}$ is an MA(1) and $\omega_0 = \frac{4\pi}{5} - \frac{\pi}{4T}$, using the Gauss–Newton (solid) and Quinn–Fernandes (dashes) methods. The true frequency is indicated by the grey line.

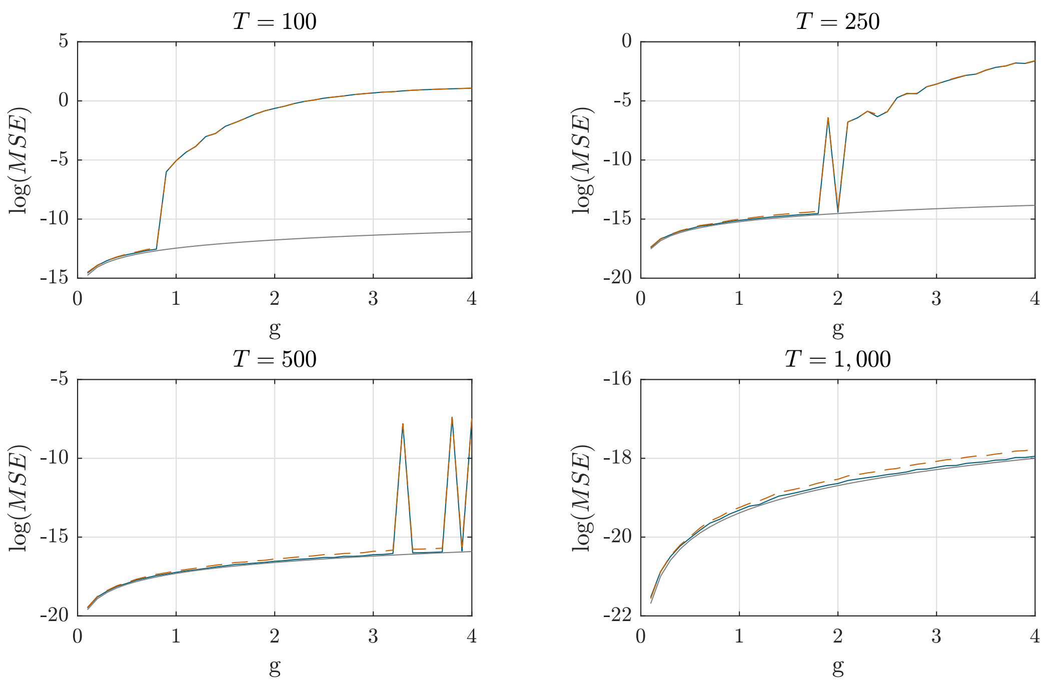

Figure 6.10: log(MSE) for $\widetilde{\omega}$ when $\{\varepsilon_t\}$ is an MA(1) and $\omega_0 = \frac{4\pi}{5} - \frac{\pi}{4T}$, using the Gauss–Newton (solid) and Quinn–Fernandes (dashes) methods. The theoretical variance is indicated by the grey line.

6.A Appendix

In what follows, where convergence is indicated, it will mean convergence in the almost sure sense, unless otherwise stated. Where order notation is used, it will also mean orders in the almost sure sense.

6.A.1 Proof of Lemma 6.1

Letting $V = \begin{bmatrix} A & B \end{bmatrix}$,

$$\left|I_d - AA' - BB'\right| = \left|I_d - VV'\right|.$$

Now, for any matrices C, D, E and F of appropriate dimension,

$$\left|\begin{bmatrix} C & D \\ E & F \end{bmatrix}\right| = |F| \left|C - DF^{-1}C\right| = |C| \left|F - EC^{-1}D\right|.$$

Thus

$$\left|\begin{bmatrix} I_d & V \\ V' & I_2 \end{bmatrix}\right| = \left|I_d - VV'\right| = \left|I_2 - V'V\right|.$$

But

$$\begin{aligned}\left|I_2 - V'V\right| &= \left|\begin{bmatrix} 1 - A'A & -A'B \\ -B'A & 1 - B'B \end{bmatrix}\right| \\ &= 1 - \left(A'A + B'B\right) + \left(A'A\right)\left(B'B\right) - \left(A'B\right)^2,\end{aligned}$$

which is the first part of the lemma. Similarly,

$$\left|\begin{bmatrix} I_d & V \\ -V' & I_2 \end{bmatrix}\right| = \left|I_d + VV'\right| = \left|I_2 + V'V\right|$$

and the second part of the lemma follows.

6.A.2 Proof of Theorem 6.1

Consider

$$\begin{aligned}F_T(\omega) &= C_T(\omega) - iS_T(\omega) \\ &= \sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} X_t e^{-i\omega t}.\end{aligned}$$

Let $W_T(\omega) = \sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} e^{-i\omega t}\varepsilon_t$, $\widetilde{\theta}_0 = \left[\begin{array}{cc} \alpha_0 & \beta_0 \end{array}\right]$ and $m_t(\omega) = \left[\begin{array}{cc} \cos(\omega t) & \sin(\omega t) \end{array}\right]'$. Then

$$
\begin{aligned}
F_T(\omega) &= W_T(\omega) + \sqrt{2}T^{-1/2}\mu_0 \sum_{t=0}^{T-1} e^{-i\omega t} + \sqrt{2}T^{-1/2}\widetilde{\theta}_0 \sum_{t=0}^{T-1} m_t(\omega_0) e^{-i\omega t} \\
&= W_T(\omega) + \frac{1}{\sqrt{2}}T^{-1/2}\widetilde{\theta}_0 \sum_{t=0}^{T-1} \left[\begin{array}{c} e^{i\omega_0 t} \\ \frac{1}{i}e^{i\omega_0 t} \end{array}\right] e^{-i\omega t} + O(1) \\
&= W_T(\omega) + \frac{1}{\sqrt{2}}T^{-1/2}\widetilde{\theta}_0 \sum_{t=0}^{T-1} \left[\begin{array}{c} e^{-i(\omega-\omega_0)t} \\ \frac{1}{i}e^{-i(\omega-\omega_0)t} \end{array}\right] + O(1) \\
&= W_T(\omega) + \frac{1}{\sqrt{2}}T^{1/2}\eta_0 h_T(\omega-\omega_0) + O(1),
\end{aligned}
$$

where $\eta_0 = \alpha_0 - i\beta_0$ and

$$
\begin{aligned}
h_T(x) &= T^{-1}\sum_{t=0}^{T-1} e^{-ixt} \\
&= T^{-1}\frac{e^{-ixT}-1}{e^{-ix}-1} \\
&= T^{-1}e^{-ix(T-1)/2}\frac{\sin(xT/2)}{\sin(x/2)},
\end{aligned}
$$

noting that $h_T(0) = 1$. Consider the case where $\omega = \omega_0 + a/T$ for some $a \geqslant 0$. Then

$$
T^{-1}F_T(\omega_0 + a/T)F_T^*(\omega_0 + a/T) = \frac{1}{2}\widetilde{\theta}_0\widetilde{\theta}_0'\left|h_T(a/T)\right|^2 + O\left(T^{-1}\log T\right),
$$

since $\eta_0\eta_0^* = \widetilde{\theta}_0\widetilde{\theta}_0'$ and $W_T(\omega) = O\left\{(\log T)^{1/2}\right\}$. Now,

$$
\begin{aligned}
\left|h_T(a/T)\right|^2 &= \left|e^{-ix(T-1)/2}\frac{\sin(a/2)}{T\sin(a/2T)}\right|^2 \\
&\to \frac{\sin^2(a/2)}{(a/2)^2},
\end{aligned}
$$

as $T \to \infty$, and so

$$
T^{-1}F_T(\omega_0 + a/T)F_T^*(\omega_0 + a/T) \to \frac{1}{2}\widetilde{\theta}_0\widetilde{\theta}_0'\frac{\sin^2(a/2)}{(a/2)^2}
$$

as $T \to \infty$. Also,

$$
\overline{X} = \mu_0 + T^{-1}\widetilde{\theta}_0\sum_{t=0}^{T-1} m_t(\omega) + T^{-1}\sum_{t=0}^{T-1}\varepsilon_t,
$$

and so

$$
X_t - \overline{X} = \widetilde{\theta}_0 m_t(\omega) + \varepsilon_t + O\left\{T^{-1}(\log\log T)^{1/2}\right\}
$$

since

$$
T^{-1}\sum_{t=0}^{T-1}\varepsilon_t = O\left\{T^{-1}(\log\log T)^{1/2}\right\},
$$

from Theorem 2.2. Thus

$$\begin{aligned} T^{-1}V_T &= T^{-1}\sum_{t=0}^{T-1}\varepsilon_t\varepsilon_t' + T^{-1}\widetilde{\theta}_0\left\{\sum_{t=0}^{T-1}m_t(\omega)\,m_t'(\omega)\right\}\widetilde{\theta}_0' + O\left\{T^{-1/2}(\log\log T)^{1/2}\right\} \\ &\to \Sigma_0 + \frac{1}{2}\widetilde{\theta}_0\widetilde{\theta}_0' \end{aligned} \tag{6.23}$$

as $T\to\infty$. Now,

$$\begin{aligned} J_T(\omega) &= 1 - \left|I_d - V_T^{-1/2}C_T(\omega)\,C_T'(\omega)\,V_T^{-1/2} - V_T^{-1/2}S_T(\omega)\,S_T'(\omega)\,V_T^{-1/2}\right| \\ &= 1 - \left|V_T^{-1}\right|\left|V_T - F(\omega)\,F^*(\omega)\right|. \end{aligned}$$

Therefore

$$J_T(\omega_0 + a/T) \to 1 - \left|\Sigma_0 + \frac{1}{2}\widetilde{\theta}_0\widetilde{\theta}_0'\right|^{-1}\left|\Sigma_0 + \frac{1}{2}\widetilde{\theta}_0\widetilde{\theta}_0' - \frac{1}{2}\widetilde{\theta}_0\widetilde{\theta}_0'\frac{\sin^2(a/2)}{(a/2)^2}\right|.$$

Let

$$\begin{bmatrix} A & B \end{bmatrix} = \frac{1}{\sqrt{2}}\Sigma_0^{-1/2}\widetilde{\theta}_0,$$

then

$$\begin{bmatrix} A' \\ B' \end{bmatrix}\begin{bmatrix} A & B \end{bmatrix} = \frac{1}{2}\widetilde{\theta}_0'\Sigma_0^{-1}\widetilde{\theta}_0,$$

and so

$$A'A + B'B = \frac{1}{2}\operatorname{tr}\left(\widetilde{\theta}_0'\Sigma_0^{-1}\widetilde{\theta}_0\right)$$

and

$$(A'A)\,(B'B) - (A'B)^2 = \frac{1}{4}\left|\widetilde{\theta}_0'\Sigma_0^{-1}\widetilde{\theta}_0\right|.$$

From Lemma 6.1,

$$\begin{aligned} \left|\Sigma_0 + \frac{1}{2}\widetilde{\theta}_0\widetilde{\theta}_0'\right| &= |\Sigma_0|\left|I_d + \frac{1}{2}\Sigma_0^{-1/2}\widetilde{\theta}_0\widetilde{\theta}_0'\Sigma_0^{-1/2}\right| \\ &= |\Sigma_0|\left\{1 + \operatorname{tr}\left(\widetilde{\theta}_0'\Sigma_0^{-1}\widetilde{\theta}_0\right) + \frac{1}{4}\left|\widetilde{\theta}_0'\Sigma_0^{-1}\widetilde{\theta}_0\right|\right\} \end{aligned}$$

and

$$\left|\Sigma_0 + \frac{1}{2}\widetilde{\theta}_0\widetilde{\theta}_0' - \frac{1}{2}\widetilde{\theta}_0\widetilde{\theta}_0'\frac{\sin^2(a/2)}{(a/2)^2}\right| = |\Sigma_0|\left\{1 + \frac{c}{2}\operatorname{tr}\left(\widetilde{\theta}_0'\Sigma_0^{-1}\widetilde{\theta}_0\right) + \frac{c^2}{4}\left|\widetilde{\theta}_0'\Sigma_0^{-1}\widetilde{\theta}_0\right|\right\},$$

where

$$0 \leqslant c = 1 - \frac{\sin^2(a/2)}{(a/2)^2} \leqslant 1.$$

Thus

$$J_T(\omega_0 + a/T) \to 1 - \frac{1 + \frac{c}{2}\operatorname{tr}\left(\widetilde{\theta}_0'\Sigma_0^{-1}\widetilde{\theta}_0\right) + \frac{c^2}{4}\left|\widetilde{\theta}_0'\Sigma_0^{-1}\widetilde{\theta}_0\right|}{1 + \operatorname{tr}\left(\widetilde{\theta}_0'\Sigma_0^{-1}\widetilde{\theta}_0\right) + \frac{1}{4}\left|\widetilde{\theta}_0'\Sigma_0^{-1}\widetilde{\theta}_0\right|}$$

as $T \to \infty$, and so the almost sure limit of $J_T(\omega_0 + a/T)$ is 1 if and only if $a = 0$. That is, if $\kappa > 0$,

$$\liminf_{T\to\infty} \inf_{|\omega-\omega_0|>\kappa/T} \{J_T(\omega_0) - J_T(\omega)\} > 0$$

and it follows from Theorem 2.5 that $T(\widehat{\omega} - \omega_0) \to 0$ as $T \to \infty$.

6.A.3 Proof of Theorem 6.2

Let

$$K_T(\omega) = \log \left| \widehat{\Sigma}_T(\omega) \right|.$$

From the mean value theorem,

$$0 = \frac{d}{d\omega} K_T(\widehat{\omega}) = \frac{d}{d\omega} K_T(\omega_0) + \frac{d^2}{d\omega^2} K_T(\omega^*)(\widehat{\omega} - \omega_0),$$

where ω^* is a point on the line segment between ω_0 and $\widehat{\omega}$. Since $T(\widehat{\omega} - \omega_0) \to 0$, it follows that $T^{3/2}(\widehat{\omega} - \omega_0)$ has the same asymptotic distribution as

$$-\frac{T^{-1/2}\frac{d}{d\omega}K_T(\omega_0)}{T^{-2}\frac{d^2}{d\omega^2}K_T(\omega_0)}.$$

The first and second derivatives of $K_T(\omega)$ are

$$\frac{d}{d\omega} K_T(\omega) = \operatorname{tr}\left\{ \widehat{\Sigma}_T^{-1}(\omega) \frac{d}{d\omega} \widehat{\Sigma}_T(\omega) \right\}$$

and

$$\frac{d^2}{d\omega^2} K_T(\omega) = \operatorname{tr}\left[-\widehat{\Sigma}_T^{-1}(\omega) \left\{ \frac{d}{d\omega}\widehat{\Sigma}_T(\omega) \right\} \widehat{\Sigma}_T^{-1}(\omega) \left\{ \frac{d}{d\omega}\widehat{\Sigma}_T(\omega) \right\} + \widehat{\Sigma}_T^{-1}(\omega) \left\{ \frac{d^2}{d\omega^2}\widehat{\Sigma}_T(\omega) \right\} \right].$$

The first and second derivatives of $\widehat{\Sigma}_T(\omega)$ are

$$\begin{aligned} \frac{d}{d\omega}\widehat{\Sigma}_T(\omega) = &-T^{-1} \begin{bmatrix} \frac{d}{d\omega}C_T(\omega) & \frac{d}{d\omega}S_T(\omega) \end{bmatrix} \begin{bmatrix} C'_T(\omega) \\ S'_T(\omega) \end{bmatrix} \\ &- T^{-1} \begin{bmatrix} C_T(\omega) & S_T(\omega) \end{bmatrix} \begin{bmatrix} \frac{d}{d\omega}C'_T(\omega) \\ \frac{d}{d\omega}S'_T(\omega) \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} \frac{d^2}{d\omega^2}\widehat{\Sigma}_T(\omega) = &-T^{-1} \begin{bmatrix} \frac{d^2}{d\omega^2}C_T(\omega) & \frac{d^2}{d\omega^2}S_T(\omega) \end{bmatrix} \begin{bmatrix} C'_T(\omega) \\ S'_T(\omega) \end{bmatrix} \\ &- T^{-1} \begin{bmatrix} C_T(\omega) & S_T(\omega) \end{bmatrix} \begin{bmatrix} \frac{d^2}{d\omega^2}C'_T(\omega) \\ \frac{d^2}{d\omega^2}S'_T(\omega) \end{bmatrix} \\ &- 2T^{-1} \begin{bmatrix} \frac{d}{d\omega}C_T(\omega) & \frac{d}{d\omega}S_T(\omega) \end{bmatrix} \begin{bmatrix} \frac{d}{d\omega}C'_T(\omega) \\ \frac{d}{d\omega}S'_T(\omega) \end{bmatrix}, \end{aligned}$$

where the first and second derivatives of $C_T(\omega)$ are

$$\frac{d}{d\omega}C_T(\omega) = -\sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} t\sin(\omega t)\, X_t \quad \text{and} \quad \frac{d^2}{d\omega^2}C_T(\omega) = -\sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} t^2\cos(\omega t)\, X_t,$$

and the first and second derivatives of $S_T(\omega)$ are

$$\frac{d}{d\omega}S_T(\omega) = \sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} t\cos(\omega t)\, X_t, \quad \text{and} \quad \frac{d^2}{d\omega^2}S_T(\omega) = -\sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} t^2\sin(\omega t)\, X_t.$$

Let

$$Y_T(\omega) = \sqrt{2}T^{-1/2}\sum_{t=0}^{T-1}\cos(\omega t)\,\varepsilon_t, \qquad Z_T(\omega) = \sqrt{2}T^{-1/2}\sum_{t=0}^{T-1}\sin(\omega t)\,\varepsilon_t,$$

$$Y_{T1}(\omega) = -\sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} t\sin(\omega t)\,\varepsilon_t, \qquad Z_{T1}(\omega) = \sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} t\cos(\omega t)\,\varepsilon_t,$$

$$Y_{T2}(\omega) = \sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} t^2\cos(\omega t)\,\varepsilon_t, \qquad Z_{T2}(\omega) = -\sqrt{2}T^{-1/2}\sum_{t=0}^{T-1} t^2\sin(\omega t)\,\varepsilon_t.$$

Then, evaluating $C_T(\omega)$, $S_T(\omega)$ and their derivatives at the true parameter values,

$$\begin{aligned}
C_T(\omega_0) &= \sqrt{2}T^{-1/2}\theta_0 \begin{bmatrix} \sum_{t=0}^{T-1}\cos(\omega_0 t) \\ \sum_{t=0}^{T-1}\cos^2(\omega_0 t) \\ \sum_{t=0}^{T-1}\cos(\omega_0 t)\sin(\omega_0 t) \end{bmatrix} + Y_T(\omega_0) \\
&= \frac{1}{\sqrt{2}}T^{1/2}\alpha_0 + Y_T(\omega_0) + O\left(T^{-1/2}\right),
\end{aligned}$$

$$\begin{aligned}
S_T(\omega_0) &= \sqrt{2}T^{-1/2}\theta_0 \begin{bmatrix} \sum_{t=0}^{T-1}\sin(\omega_0 t) \\ \sum_{t=0}^{T-1}\cos(\omega_0 t)\sin(\omega_0 t) \\ \sum_{t=0}^{T-1}\sin^2(\omega_0 t) \end{bmatrix} + Z_T(\omega_0) \\
&= \frac{1}{\sqrt{2}}T^{1/2}\beta_0 + Z_T(\omega_0) + O\left(T^{-1/2}\right),
\end{aligned}$$

$$\begin{aligned}
\frac{d}{d\omega}C_T(\omega_0) &= -\sqrt{2}T^{-1/2}\theta_0 \begin{bmatrix} \sum_{t=0}^{T-1} t\sin(\omega_0 t) \\ \sum_{t=0}^{T-1} t\sin(\omega_0 t)\cos(\omega_0 t) \\ \sum_{t=0}^{T-1} t\sin^2(\omega_0 t) \end{bmatrix} + Y_{T1}(\omega_0) \\
&= -\frac{1}{2\sqrt{2}}T^{3/2}\beta_0 + Y_{T1}(\omega_0) + O\left(T^{1/2}\right),
\end{aligned}$$

$$\begin{aligned}
\frac{d}{d\omega}S_T(\omega_0) &= \sqrt{2}T^{-1/2}\theta_0 \begin{bmatrix} \sum_{t=0}^{T-1} t\cos(\omega_0 t) \\ \sum_{t=0}^{T-1} t\cos^2(\omega_0 t) \\ \sum_{t=0}^{T-1} t\cos(\omega_0 t)\sin(\omega_0 t) \end{bmatrix} + Z_{T1}(\omega_0) \\
&= \frac{1}{2\sqrt{2}}T^{3/2}\alpha_0 + Z_{T1}(\omega_0) + O\left(T^{1/2}\right),
\end{aligned}$$

$$\begin{aligned}
\frac{d^2}{d\omega^2} C_T(\omega_0) &= -\sqrt{2} T^{-1/2} \theta_0 \begin{bmatrix} \sum_{t=0}^{T-1} t^2 \cos(\omega_0 t) \\ \sum_{t=0}^{T-1} t^2 \cos^2(\omega_0 t) \\ \sum_{t=0}^{T-1} t^2 \cos(\omega_0 t) \sin(\omega_0 t) \end{bmatrix} + Y_{T2}(\omega_0) \\
&= -\frac{1}{3\sqrt{2}} T^{5/2} \alpha_0 + Y_{T2}(\omega_0) + O\left(T^{3/2}\right),
\end{aligned}$$

$$\begin{aligned}
\frac{d^2}{d\omega^2} S_T(\omega_0) &= -\sqrt{2} T^{-1/2} \theta_0 \begin{bmatrix} \sum_{t=0}^{T-1} t^2 \sin(\omega_0 t) \\ \sum_{t=0}^{T-1} t^2 \sin(\omega_0 t) \cos(\omega_0 t) \\ \sum_{t=0}^{T-1} t^2 \sin^2(\omega_0 t) \end{bmatrix} + Z_{T2}(\omega_0) \\
&= -\frac{1}{3\sqrt{2}} T^{5/2} \beta_0 + Z_{T2}(\omega_0) + O\left(T^{3/2}\right).
\end{aligned}$$

Thus

$$\widehat{\Sigma}_T(\omega_0) = \Sigma_0 + O\left\{T^{-1/2} (\log\log T)^{1/2}\right\},$$

$$\begin{aligned}
\frac{d}{d\omega} \widehat{\Sigma}_T(\omega_0) &= \frac{1}{2\sqrt{2}} T^{1/2} \beta_0 Y_T'(\omega_0) - \frac{1}{\sqrt{2}} T^{-1/2} Y_{T1}(\omega_0) \alpha_0' - \frac{1}{2\sqrt{2}} T^{1/2} \alpha_0 Z_T'(\omega_0) \\
&\quad - \frac{1}{\sqrt{2}} T^{-1/2} Z_{T1}(\omega_0) \beta_0' - \frac{1}{\sqrt{2}} T^{-1/2} \alpha_0 Y_{T1}'(\omega_0) + \frac{1}{2\sqrt{2}} T^{1/2} Y_T(\omega_0) \beta_0' \\
&\quad - \frac{1}{\sqrt{2}} T^{-1/2} \beta_0 Z_{T1}'(\omega_0) - \frac{1}{2\sqrt{2}} T^{1/2} Z_T(\omega_0) \alpha_0' + O(\log\log T) \\
&= O\left\{T^{1/2} (\log\log T)^{1/2}\right\},
\end{aligned}$$

$$\begin{aligned}
\frac{d^2}{d\omega^2} \widehat{\Sigma}_T(\omega_0) &= \left(\frac{1}{3} - \frac{1}{4}\right) T^2 \left(\alpha_0 \alpha_0' + \beta_0 \beta_0'\right) + O\left\{T^{3/2} (\log\log T)^{1/2}\right\} \\
&= \frac{1}{12} T^2 \left(\alpha_0 \alpha_0' + \beta_0 \beta_0'\right) + O\left\{T^{3/2} (\log\log T)^{1/2}\right\},
\end{aligned}$$

from (6.23) and since

$$\begin{aligned}
Y_T(\omega_0) &= O\left\{(\log\log T)^{1/2}\right\}, & Z_T(\omega_0) &= O\left\{(\log\log T)^{1/2}\right\}, \\
Y_{T1}(\omega_0) &= O\left\{T (\log\log T)^{1/2}\right\}, & Z_{T1}(\omega_0) &= O\left\{T (\log\log T)^{1/2}\right\}, \\
Y_{T2}(\omega_0) &= O\left\{T^2 (\log\log T)^{1/2}\right\}, & Z_{T2}(\omega_0) &= O\left\{T^2 (\log\log T)^{1/2}\right\}
\end{aligned}$$

(Hannan and Mackisack, 1986). Therefore

$$\begin{aligned}
T^{-2} \frac{d^2}{d\omega^2} K_T(\omega_0) &= \frac{1}{12} \operatorname{tr}\left[\left\{\Sigma_0 + O\left(T^{-1/2} (\log\log T)^{1/2}\right)\right\}^{-1} \left(\alpha_0 \alpha_0' + \beta_0 \beta_0'\right)\right] \\
&\quad + O\left\{T^{-1/2} (\log\log T)^{1/2}\right\} \\
&\to \frac{1}{12} \left(\alpha_0' \Sigma_0^{-1} \alpha_0 + \beta_0' \Sigma_0^{-1} \beta_0\right).
\end{aligned}$$

Also, $T^{-1/2}\frac{d}{d\omega}K_T(\omega_0)$ has the same asymptotic distribution as

$$\frac{1}{2\sqrt{2}}\beta_0'\Sigma_0^{-1}Y_T(\omega_0)-\frac{1}{\sqrt{2}}T^{-1}\alpha_0'\Sigma_0^{-1}Y_{T1}(\omega_0)-\frac{1}{2\sqrt{2}}\alpha_0'\Sigma_0^{-1}Z_T(\omega_0)-\frac{1}{\sqrt{2}}T^{-1}\beta_0'\Sigma_0^{-1}Z_{T1}(\omega_0)$$
$$-\frac{1}{\sqrt{2}}T^{-1}\alpha_0'\Sigma_0^{-1}Y_{T1}(\omega_0)+\frac{1}{2\sqrt{2}}\beta_0'\Sigma_0^{-1}Y_T(\omega_0)-\frac{1}{\sqrt{2}}T^{-1}\beta_0'\Sigma_0^{-1}Z_{T1}(\omega_0)-\frac{1}{2\sqrt{2}}\alpha_0'\Sigma_0^{-1}Z_T(\omega_0)$$
$$=-\frac{1}{\sqrt{2}}\alpha_0'\Sigma_0^{-1}\left\{2T^{-1}Y_{T1}(\omega_0)+Z_T(\omega_0)\right\}-\frac{1}{\sqrt{2}}\beta_0'\Sigma_0^{-1}\left\{2T^{-1}Z_{T1}(\omega_0)-Y_T(\omega_0)\right\}.$$

Now,

$$2T^{-1}Y_{T1}(\omega_0)+Z_T(\omega_0)=\sqrt{2}T^{-1/2}\sum_{t=0}^{T-1}\left(1-\frac{2t}{T}\right)\sin(\omega_0 t)\,\varepsilon_t, \tag{6.24}$$

and

$$2T^{-1}Z_{T1}(\omega_0)-Y_T(\omega_0)=-\sqrt{2}T^{-1/2}\sum_{t=0}^{T-1}\left(1-\frac{2t}{T}\right)\cos(\omega_0 t)\,\varepsilon_t. \tag{6.25}$$

These are both asymptotically normal with mean zero and covariance matrix $2\pi f_\varepsilon(\omega_0)/3$. To see this, let $\zeta_t=c'\varepsilon_t$, where c is a $d\times 1$ vector of constants, and consider

$$y=-T^{-1/2}\sum_{t=0}^{T-1}e^{i\omega_0 t}\zeta_t \qquad \text{and} \qquad z=-T^{-3/2}\sum_{t=0}^{T-1}te^{i\omega_0 t}\zeta_t.$$

From Theorem 2.6, both the real and imaginary components of $\begin{bmatrix} y & z \end{bmatrix}'$ are asymptotically normal with mean zero and covariance matrix

$$\pi f_\zeta(\omega_0)\begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix},$$

where

$$\begin{aligned} f_\zeta(\omega) &= \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_\zeta(j)\,e^{-ij\omega} \\ &= c'\left\{\frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_\varepsilon(j)\,e^{-ij\omega}\right\}c \\ &= c' f_\varepsilon(\omega)\,c \end{aligned}$$

and $\gamma_\zeta(j)$ and $\gamma_\varepsilon(j)$ are the autocovariance functions of $\{\zeta_t\}$ and $\{\varepsilon_t\}$, respectively. Thus the real and imaginary components of $-\sqrt{2}y+2\sqrt{2}z$ are both asymptotically normal with mean zero and variance

$$\begin{aligned} \pi f_\zeta(\omega_0)\begin{bmatrix} -\sqrt{2} & 2\sqrt{2} \end{bmatrix}\begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}\begin{bmatrix} -\sqrt{2} \\ 2\sqrt{2} \end{bmatrix} &= \frac{2\pi}{3}f_\zeta(\omega_0) \\ &= c'\left\{\frac{2\pi}{3}f_\varepsilon(\omega)\right\}c. \end{aligned}$$

The asymptotic distributions of (6.24) and (6.25) follow by applying the Cramér-Wold device. Therefore $T^{-1/2}\frac{d}{d\omega}K(\omega_0)$ is asymptotically normal with mean zero and variance

$$\begin{aligned}
&\frac{1}{2}\frac{2\pi}{3}\alpha_0'\Sigma_0^{-1}f_\varepsilon(\omega_0)\Sigma^{-1}\alpha_0+\frac{1}{2}\frac{2\pi}{3}\beta_0'\Sigma_0^{-1}f_\varepsilon(\omega_0)\Sigma^{-1}\beta_0\\
&=\frac{\pi}{3}\left(\alpha_0'\Sigma_0^{-1}f_\varepsilon(\omega_0)\Sigma^{-1}\alpha_0+\beta_0'\Sigma_0^{-1}f_\varepsilon(\omega_0)\Sigma_0^{-1}\beta_0\right).
\end{aligned}$$

It follows that $T^{3/2}(\widehat{\omega}-\omega_0)$ is asymptotically normal with mean zero and variance

$$\begin{aligned}
&\frac{\frac{\pi}{3}\left(\alpha_0'\Sigma_0^{-1}f_\varepsilon(\omega_0)\Sigma^{-1}\alpha_0+\beta_0'\Sigma_0^{-1}f_\varepsilon(\omega_0)\Sigma_0^{-1}\beta_0\right)}{\left\{\frac{1}{12}\left(\alpha_0'\Sigma_0^{-1}\alpha_0+\beta_0'\Sigma_0^{-1}\beta_0\right)\right\}^2}\\
&=48\pi\frac{\alpha_0'\Sigma_0^{-1}f_\varepsilon(\omega_0)\Sigma_0^{-1}\alpha_0+\beta_0'\Sigma_0^{-1}f_\varepsilon(\omega_0)\Sigma_0^{-1}\beta_0}{\left(\alpha_0'\Sigma_0^{-1}\alpha_0+\beta_0'\Sigma_0^{-1}\beta_0\right)^2}.
\end{aligned}$$

6.A.4 Proof of Theorem 6.3

From the proof of Theorem 6.1,

$$\begin{aligned}
T^{-1}\widetilde{J}_{T,\Omega}(\omega_0+a/T)&=T^{-1}F^*(\omega_0+a/T)\,\Omega F(\omega_0+a/T)\\
&\to\frac{1}{2}\widetilde{\theta}_0'\Omega\widetilde{\theta}_0\frac{\sin^2(a/2)}{(a/2)^2},\qquad\text{as } T\to\infty\\
&\leqslant\frac{1}{2}\widetilde{\theta}_0'\Omega\widetilde{\theta}_0,
\end{aligned}$$

with equality if and only if $a=0$. Thus

$$T^{-1}\left\{\widetilde{J}_T(\omega_0)-\widetilde{J}_T(\omega_0+a/T)\right\}\to\frac{1}{2}\widetilde{\theta}_0'\Omega\widetilde{\theta}_0\left\{1-\frac{\sin^2(a/2)}{(a/2)^2}\right\}$$

as $T\to\infty$ and, for $\kappa>0$,

$$\lim_{T\to\infty}T^{-1}\left\{\widetilde{J}_T(\omega_0)-\widetilde{J}_T(\omega_0+\kappa/T)\right\}>0.$$

It follows from Theorem 2.5 that $T(\widetilde{\omega}-\omega_0)\to 0$ as $T\to\infty$.

6.A.5 Proof of Theorem 6.4

From the mean value theorem,

$$0=\frac{d}{d\omega}\widetilde{J}_{T,\Omega}(\widetilde{\omega})=\frac{d}{d\omega}\widetilde{J}_{T,\Omega}(\omega_0)+\frac{d^2}{d\omega^2}\widetilde{J}_{T,\Omega}(\omega^*)(\widetilde{\omega}-\omega_0)$$

where ω^* is a point on the line segment between ω_0 and $\widetilde{\omega}$. Since $T(\widetilde{\omega}-\omega_0)\to 0$, it follows that $T^{3/2}(\widetilde{\omega}-\omega_0)$ has the same asymptotic distribution as

$$-\frac{T^{-3/2}\frac{d}{d\omega}\widetilde{J}_{T,\Omega}(\omega_0)}{T^{-3}\frac{d^2}{d\omega^2}\widetilde{J}_{T,\Omega}(\omega_0)}.$$

The first and second derivatives of $\widetilde{J}_{T,\Omega}(\omega)$ are

$$\frac{d}{d\omega}\widetilde{J}_{T,\Omega}(\omega)=2\begin{bmatrix}\frac{d}{d\omega}C_T'(\omega) & \frac{d}{d\omega}S_T'(\omega)\end{bmatrix}\Omega\begin{bmatrix}C_T(\omega)\\ S_T(\omega)\end{bmatrix}$$

and

$$\begin{aligned}\frac{d^2}{d\omega^2}\widetilde{J}_{T,\Omega}(\omega)=&\,2\begin{bmatrix}\frac{d^2}{d\omega^2}C_T'(\omega) & \frac{d^2}{d\omega^2}S_T'(\omega)\end{bmatrix}\Omega\begin{bmatrix}C_T(\omega)\\ S_T(\omega)\end{bmatrix}\\ &+2\begin{bmatrix}\frac{d}{d\omega}C_T'(\omega) & \frac{d}{d\omega}S_T'(\omega)\end{bmatrix}\Omega\begin{bmatrix}\frac{d}{d\omega}C_T(\omega)\\ \frac{d}{d\omega}S_T(\omega)\end{bmatrix}.\end{aligned}$$

Thus, using the results of the proof of Theorem 6.2,

$$\frac{d^2}{d\omega^2}\widetilde{J}_{T,\Omega}(\omega_0)=-\frac{1}{12}T^3\left(\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0\right)+O\left\{T^{5/2}(\log\log T)^{1/2}\right\}$$

and so

$$-T^{-3}\frac{d^2}{d\omega^2}\widetilde{J}_{T,\Omega}(\omega_0)\to\frac{1}{12}\left(\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0\right)$$

as $T\to\infty$. Also

$$\begin{aligned}\frac{d}{d\omega}\widetilde{J}_{T,\Omega}(\omega_0)=&-\frac{1}{\sqrt{2}}T^{3/2}\beta_0'\Omega Y_T(\omega_0)+\sqrt{2}T^{1/2}\alpha_0'\Omega Y_{T1}(\omega_0)+\frac{1}{\sqrt{2}}T^{3/2}\alpha_0'\Omega Z_T(\omega_0)\\ &+\sqrt{2}T^{1/2}\beta_0'\Omega Z_{T1}(\omega_0)+O(T\log\log T)\end{aligned}$$

and so $T^{-3/2}\frac{d}{d\omega}\widetilde{J}_T(\omega_0)$ has the same asymptotic distribution as

$$\frac{1}{\sqrt{2}}\alpha_0'\Omega\left\{2T^{-1}Y_{T1}(\omega_0)+Z_T(\omega_0)\right\}+\frac{1}{\sqrt{2}}\beta_0'\Omega\left\{2T^{-1}Z_{T1}(\omega_0)-Y_T(\omega_0)\right\},$$

which, as shown in the proof of Theorem 6.2, is asymptotically normal with mean zero and variance

$$\frac{\pi}{3}\left\{\alpha_0'\Omega f_\varepsilon(\omega_0)\Omega\alpha_0+\beta_0'\Omega f_\varepsilon(\omega_0)\Omega\beta_0\right\}.$$

Therefore $T^{3/2}(\widetilde{\omega}-\omega_0)$ is asymptotically normal with mean zero and variance

$$48\pi\frac{\alpha_0'\Omega f_\varepsilon(\omega_0)\Omega\alpha_0+\beta_0'\Omega f_\varepsilon(\omega_0)\Omega\beta_0}{\left(\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0\right)^2}.$$

6.A.6 Proof of Theorem 6.5

Let Θ_0 denote the true value of Θ and let

$$\mathcal{I}=\lim_{T\to\infty}-N_T^{-1}\frac{\partial^2 l(\Theta_0)}{\partial\Theta_0\partial\Theta_0'}N_T^{-1},$$

where

$$N_T = \begin{bmatrix} T^{1/2} I_{3d+(p+1)d^2} & 0 \\ 0 & T^{3/2} \end{bmatrix}.$$

The first derivatives of $l(\Theta)$ with respect to Θ are

$$\frac{\partial l(\Theta)}{\partial \mu} = -\sum_{t=0}^{T-1} \frac{\partial U_t}{\partial \mu} G^{-1} U_t,$$

$$\frac{\partial l(\Theta)}{\partial D} = -\sum_{t=0}^{T-1} \frac{\partial U_t}{\partial D'} G^{-1} U_t,$$

$$\frac{\partial l(\Theta)}{\partial \widetilde{G}} = -\left(G^{-1} \otimes G^{-1}\right) \operatorname{vec}\left\{\sum_{t=0}^{T-1} U_t U_t'\right\},$$

$$\frac{\partial l(\Theta)}{\partial \alpha} = -\sum_{t=0}^{T-1} \frac{\partial U_t}{\partial \alpha'} G^{-1} U_t,$$

$$\frac{\partial l(\Theta)}{\partial \beta} = -\sum_{t=0}^{T-1} \frac{\partial U_t}{\partial \beta'} G^{-1} U_t,$$

$$\frac{\partial l(\Theta)}{\partial \omega} = -\sum_{t=0}^{T-1} \frac{\partial U_t'}{\partial \omega} G^{-1} U_t.$$

Letting

$$E_t = X_t - \mu - \alpha \cos(\omega t) - \beta \sin(\omega t)$$

and

$$Z_t = \begin{bmatrix} E_{t-1}' & \cdots & E_{t-p}' \end{bmatrix}',$$

we have

$$U_t = E_t + \delta Z_t = E_t + \left(Z_t' \otimes I_d\right) D$$

and so the first derivatives of U_t are

$$\frac{\partial U_t(\Theta)}{\partial \mu} = 1,$$

$$\frac{\partial U_t(\Theta)}{\partial D'} = Z_t' \otimes I_d,$$

$$\frac{\partial U_t(\Theta)}{\partial \alpha'} = -\sum_{j=0}^{p} \delta_j \cos\{\omega(t-j)\},$$

$$\frac{\partial U_t(\Theta)}{\partial \beta'} = -\sum_{j=0}^{p} \delta_j \sin\{\omega(t-j)\},$$

$$\frac{\partial U_t(\Theta)}{\partial \omega} = \sum_{j=0}^{p} \delta_j \left[\alpha(t-j)\sin\{\omega(t-j)\} - \beta(t-j)\cos\{\omega(t-j)\}\right].$$

The second derivatives of $l(\Theta)$ with respect to Θ are then

$$\frac{\partial^2 l(\Theta)}{\partial \mu^2} = -TG^{-1},$$

$$\begin{aligned}\frac{\partial^2 l(\Theta)}{\partial \mu \partial \alpha} &= -G^{-1}\sum_{t=0}^{T-1}\frac{\partial U_t}{\partial \alpha}\\ &= G^{-1}\sum_{t=0}^{T-1}\sum_{j=0}^{p}\delta_j \cos\{\omega(t-j)\},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l(\Theta)}{\partial \mu \partial \beta} &= -G^{-1}\sum_{t=0}^{T-1}\frac{\partial U_t}{\partial \beta}\\ &= G^{-1}\sum_{t=0}^{T-1}\sum_{j=0}^{p}\delta_j \sin\{\omega(t-j)\},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l(\Theta)}{\partial \mu \partial D'} &= -G^{-1}\sum_{t=0}^{T-1}\frac{\partial U_t}{\partial \widetilde{\delta}'}\\ &= -G^{-1}\sum_{t=0}^{T-1}\left(Z_t' \otimes I_d\right),\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l(\Theta)}{\partial \mu \partial \omega} &= -G^{-1}\sum_{t=0}^{T-1}\frac{\partial U_t}{\partial \omega}\\ &= -G^{-1}\sum_{t=0}^{T-1}\sum_{j=0}^{p}\delta_j\left[\alpha(t-j)\sin\{\omega(t-j)\} - \beta(t-j)\cos\{\omega(t-j)\}\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l(\Theta)}{\partial D \partial D'} &= -\sum_{t=0}^{T-1}\frac{\partial^2 U_t}{\partial \widetilde{\delta}\partial \widetilde{\delta}'}G^{-1}U_t(\Theta) - \sum_{t=0}^{T-1}\frac{\partial U_t}{\partial \widetilde{\delta}}G^{-1}\frac{\partial U_t}{\partial \widetilde{\delta}'}\\ &\sim -\sum_{t=0}^{T-1}(Z_t \otimes I_d)\,G^{-1}\left(Z_t' \otimes I_d\right),\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l(\Theta)}{\partial D \delta \alpha'} &= -\sum_{t=0}^{T-1}\frac{\partial^2 U_t}{\partial \widetilde{\delta}\partial \alpha'}G^{-1}U_t - \sum_{t=0}^{T-1}\frac{\partial U_t'}{\partial \widetilde{\delta}}G^{-1}\frac{\partial U_t}{\partial \alpha'}\\ &\sim -\sum_{t=0}^{T-1}(Z_t \otimes I_d)\,G^{-1}\left\{-\sum_{j=0}^{p}\delta_j\cos\{\omega(t-j)\}\right\},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l(\Theta)}{\partial D \partial \beta'} &= -\sum_{t=0}^{T-1}\frac{\partial^2 U_t}{\partial \widetilde{\delta}\partial \beta'}G^{-1}U_t - \sum_{t=0}^{T-1}\frac{\partial U_t}{\partial \widetilde{\delta}}G^{-1}\frac{\partial U_t}{\partial \beta'}\\ &\sim -\sum_{t=0}^{T-1}(Z_t \otimes I_d)\,G^{-1}\left\{-\sum_{j=0}^{p}\delta_j\sin\{\omega(t-j)\}\right\},\end{aligned}$$

$$\frac{\partial^2 l(\Theta)}{\partial D \partial \omega} = -\sum_{t=0}^{T-1} \frac{\partial^2 U_t}{\partial \widetilde{\delta} \partial \omega} G^{-1} U_t - \sum_{t=0}^{T-1} \frac{\partial U_t}{\partial \widetilde{\delta}} G^{-1} \frac{\partial U_t}{\partial \omega}$$
$$\sim \sum_{t=0}^{T-1} (Z_t \otimes I_d) G^{-1} \sum_{j=0}^{p} \delta_j [\alpha (t-j) \sin \{\omega (t-j)\} - \beta (t-j) \cos \{\omega (t-j)\}],$$

$$\frac{\partial^2 l(\Theta)}{\partial \widetilde{G} \partial \mu} = -\left(G^{-1} \otimes G^{-1}\right) \operatorname{vec}\left(2 \sum_{t=0}^{T-1} \frac{\partial U_t}{\partial \mu} U_t'\right)$$
$$= -2\left(G^{-1} \otimes G^{-1}\right) \sum_{t=0}^{T-1} U_t,$$

$$\frac{\partial^2 l(\Theta)}{\partial \widetilde{G} \partial D} = -\left(G^{-1} \otimes G^{-1}\right) \operatorname{vec}\left(2 \sum_{t=0}^{T-1} \frac{\partial U_t}{\partial D'} U_t'\right)$$
$$= -\left(G^{-1} \otimes G^{-1}\right) \operatorname{vec}\left\{2 \sum_{t=0}^{T-1} \left(Z_t' \otimes I_d\right) U_t'\right\},$$

$$\frac{\partial^2 l(\Theta)}{\partial \widetilde{G} \partial \alpha'} = -\left(G^{-1} \otimes G^{-1}\right) \operatorname{vec}\left(2 \sum_{t=0}^{T-1} \frac{\partial U_t}{\partial \alpha'} U_t'\right)$$
$$= \left(G^{-1} \otimes G^{-1}\right) \operatorname{vec}\left(2 \sum_{t=0}^{T-1} \sum_{j=0}^{p} \delta_j \cos \{\omega (t-j)\} U_t'\right),$$

$$\frac{\partial^2 l(\Theta)}{\partial \widetilde{G} \partial \beta'} = -\left(G^{-1} \otimes G^{-1}\right) \operatorname{vec}\left(2 \sum_{t=0}^{T-1} \frac{\partial U_t}{\partial \beta'} U_t'\right)$$
$$= \left(G^{-1} \otimes G^{-1}\right) \operatorname{vec}\left(2 \sum_{t=0}^{T-1} \sum_{j=0}^{p} \delta_j \sin \{\omega (t-j)\} U_t'\right),$$

$$\frac{\partial^2 l(\Theta)}{\partial \widetilde{G} \partial \omega} = -\left(G^{-1} \otimes G^{-1}\right) \operatorname{vec}\left(2 \sum_{t=0}^{T-1} \frac{\partial U_t}{\partial \omega} U_t'\right)$$
$$= -\left(G^{-1} \otimes G^{-1}\right) \operatorname{vec}\left(2 \sum_{t=0}^{T-1} \sum_{j=0}^{p} \delta_j [\alpha (t-j) \sin \{\omega (t-j)\} - \beta (t-j) \cos \{\omega (t-j)\}] U_t'\right)$$

$$\frac{\partial^2 l(\Theta)}{\partial \alpha \partial \alpha'} = -\sum_{t=0}^{T-1} \frac{\partial^2 U_t}{\partial \alpha \partial \alpha'} G^{-1} U_t - \sum_{t=0}^{T-1} \frac{\partial U_t}{\partial \alpha} G^{-1} \frac{\partial U_t}{\partial \alpha'}$$
$$\sim -\sum_{j,k=0}^{p} \delta_j' G^{-1} \delta_k \sum_{t=0}^{T-1} \cos \{\omega (t-j)\} \cos \{\omega (t-k)\},$$

$$\frac{\partial^2 l(\Theta)}{\partial \beta \partial \beta'} = -\sum_{t=0}^{T-1} \frac{\partial^2 U_t}{\partial \beta \partial \beta'} G^{-1} U_t - \sum_{t=0}^{T-1} \frac{\partial U_t}{\partial \beta} G^{-1} \frac{\partial U_t}{\partial \beta'}$$
$$\sim -\sum_{j,k=0}^{p} \delta_j' G^{-1} \delta_k \sum_{t=0}^{T-1} \sin \{\omega (t-j)\} \sin \{\omega (t-k)\},$$

$$\frac{\partial^2 l(\Theta)}{\partial\alpha\partial\beta'} = -\sum_{t=0}^{T-1}\frac{\partial^2 U_t}{\partial\alpha\partial\beta'}G^{-1}U_t - \sum_{t=0}^{T-1}\frac{\partial U_t}{\partial\alpha}G^{-1}\frac{\partial U_t}{\partial\beta'}$$
$$\sim -\sum_{j,k=0}^{p}\delta_j' G^{-1}\delta_k\sum_{t=0}^{T-1}\cos\{\omega(t-j)\}\sin\{\omega(t-k)\},$$

$$\begin{aligned}
\frac{\partial^2 l(\Theta)}{\partial\alpha\partial\omega} &= -\sum_{t=0}^{T-1}\frac{\partial^2 U_t}{\partial\alpha\partial\omega}G^{-1}U_t - \sum_{t=0}^{T-1}\frac{\partial U_t}{\partial\alpha}G^{-1}\frac{\partial U_t}{\partial\omega}\\
&= -\sum_{j=0}^{p}\delta_j' G^{-1}\sum_{t=0}^{T-1}(t-j)\sin\{\omega(t-j)\}U_t\\
&\quad+\sum_{j,k=0}^{p}\delta_j' G^{-1}\delta_k\\
&\quad\times\sum_{t=0}^{T-1}\cos\{\omega(t-j)\}\left[\alpha(t-k)\sin\{\omega(t-k)\}-\beta(t-k)\cos\{\omega(t-k)\}\right],
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l(\Theta)}{\partial\beta\partial\omega} &= -\sum_{t=0}^{T-1}\frac{\partial^2 U_t}{\partial\beta\partial\omega}G^{-1}U_t - \sum_{t=0}^{T-1}\frac{\partial U_t}{\partial\beta}G^{-1}\frac{\partial U_t}{\partial\omega}\\
&= \sum_{j=0}^{p}\delta_j' G^{-1}\sum_{t=0}^{T-1}(t-j)\cos\{\omega(t-j)\}U_t\\
&\quad+\sum_{j,k=0}^{p}\delta_j' G^{-1}\delta_k\\
&\quad\times\sum_{t=0}^{T-1}\sin\{\omega(t-j)\}\left[\alpha(t-k)\sin\{\omega(t-k)\}-\beta(t-k)\cos\{\omega(t-k)\}\right],
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l(\Theta)}{\partial\omega^2} &= -\sum_{t=0}^{T-1}\frac{\partial^2 U_t}{\partial\omega^2}G^{-1}U_t - \sum_{t=0}^{T-1}\frac{\partial U_t'}{\partial\omega}G^{-1}\frac{\partial U_t}{\partial\omega}\\
&= \sum_{t=0}^{T-1}\sum_{j=0}^{p}\delta_j'\left[\alpha'(t-j)^2\cos\{\omega(t-j)\}+\beta'(t-j)^2\sin\{\omega(t-j)\}\right]G^{-1}U_t\\
&\quad-\sum_{t=0}^{T-1}\sum_{j,k=0}^{p}\left[\alpha'(t-j)\sin\{\omega(t-j)\}-\beta'(t-j)\cos\{\omega(t-j)\}\right]\\
&\quad\times\delta_j' G^{-1}\delta_k\left[\alpha(t-k)\sin\{\omega(t-k)\}-\beta(t-k)\cos\{\omega(t-k)\}\right],
\end{aligned}$$

where ~ denotes the term of the highest order. Now, $E\{U_t(\Theta_0)\} = E(u_t) = 0$, and it follows that the α, β and ω components of $\mathcal{I}$ are given by

$$\jmath = \frac{1}{2}\begin{bmatrix} \Delta_1 & \Delta_2 & \frac{1}{2}(-\Delta_2\alpha_0+\Delta_1\beta_0)\\ -\Delta_2 & \Delta_1 & \frac{1}{2}(-\Delta_2\beta_0-\Delta_1\alpha)\\ \frac{1}{2}(\alpha_0'\Delta_2+\beta_0'\Delta_1) & \frac{1}{2}(\beta_0'\Delta_2-\alpha_0'\Delta_1) & \frac{1}{3}(\alpha_0'\Delta_1\alpha_0+\beta_0'\Delta_1\beta_0-\beta_0'\Delta_2\alpha_0+\alpha_0'\Delta_2\beta_0)\end{bmatrix},$$

where

$$\Delta_1 = \delta_R' G_0^{-1} \delta_R + \delta_I' G_0^{-1} \delta_I, \qquad \Delta_2 = -\delta_I' G_0^{-1} \delta_R + \delta_R' G_0^{-1} \delta_I,$$

$$\delta_R = \mathrm{Re}\left(I_d + \sum_{j=1}^{p} \delta_{0j} e^{-ij\omega} \right), \qquad \delta_I = \mathrm{Im}\left(I_d + \sum_{j=1}^{p} \delta_{0j} e^{-ij\omega} \right).$$

The non-diagonal blocks of the μ, D and $\widetilde{G}$ components of $\mathcal{I}$ are zero. Let

$$\tau = \begin{bmatrix} \Delta_1 & \Delta_2 \\ -\Delta_2 & \Delta_1 \end{bmatrix} \qquad \text{and} \qquad \varphi = \begin{bmatrix} \beta_0 \\ -\alpha_0 \end{bmatrix}.$$

Then, since τ is symmetric,

$$\jmath = \frac{1}{2} \begin{bmatrix} \tau & \frac{1}{2}\tau\varphi \\ \frac{1}{2}\varphi'\tau & \frac{1}{3}\varphi'\tau\varphi \end{bmatrix}$$

and so, from the matrix inversion lemma,

$$\jmath^{-1} = 2 \begin{bmatrix} \tau^{-1} + \frac{1}{4}\tau^{-1}(\tau\varphi) H (\varphi'\tau)\tau^{-1} & -\frac{1}{2}\tau^{-1}(\tau\varphi) H \\ -\frac{1}{2} H (\varphi'\tau)\tau^{-1} & H \end{bmatrix},$$

where

$$H^{-1} = \left\{ \frac{1}{3}\varphi'\tau\varphi - \frac{1}{4}(\varphi'\tau)\tau^{-1}(\tau\varphi) \right\} = \frac{1}{12}\varphi'\tau\varphi.$$

That is,

$$\jmath^{-1} = \begin{bmatrix} 2\tau^{-1} + \frac{6}{\theta'\tau\theta}\varphi\varphi' & -\frac{12}{\varphi'\tau\varphi}\varphi \\ -\frac{12}{\varphi'\tau\varphi}\varphi' & \frac{24}{\varphi'\tau\varphi} \end{bmatrix}.$$

The first derivatives of $l(\Theta)$ with respect to α, β and ω at Θ_0 are

$$\frac{\partial l(\Theta_0)}{\partial \alpha} = \sum_{t=0}^{T-1} \left[\sum_{j=0}^{p} \delta_{0j} \cos\{\omega_0 (t-j)\} \right] G_0^{-1} u_t,$$

$$\frac{\partial l(\Theta_0)}{\partial \beta} = \sum_{t=0}^{T-1} \left[\sum_{j=0}^{p} \delta_{0j} \sin\{\omega_0 (t-j)\} \right] G_0^{-1} u_t,$$

$$\frac{\partial l(\Theta_0)}{\partial \omega} = -\sum_{t=0}^{T-1} \sum_{j=0}^{p} \delta_{0j} \left[\alpha_0 (t-j) \sin\{\omega_0 (t-j)\} - \beta_0 (t-j) \cos\{\omega_0 (t-j)\} \right] G_0^{-1} u_t.$$

Thus, since $E(u_t u_t') = G_0$,

$$E\left\{ \frac{\partial l(\Theta_0)}{\partial \widetilde{\theta}} \frac{\partial l(\Theta_0)}{\partial \widetilde{\theta}'} \right\} = \jmath.$$

It follows from the martingale central limit theorem that $T^{3/2}(\widehat{\omega} - \omega_0)$ is asymptotically normal with mean zero and variance $24/\varphi'\tau\varphi$. But

$$\begin{aligned} \varphi'\tau\varphi &= \beta_0'\Delta_1\beta_0 + \alpha_0'\Delta_2\beta_0 - \beta_0'\Delta_2\alpha_0 + \alpha_0'\Delta_1\alpha_0 \\ &= \frac{1}{2\pi}\eta_0^* f_\varepsilon^{-1}(\omega_0)\eta_0, \end{aligned}$$

since

$$f_{\varepsilon}(\omega) = \frac{1}{2\pi}\left(I_d + \sum_{j=1}^{p}\delta_j e^{-ij\omega}\right)^{-1} G\left\{\left(I_d + \sum_{j=1}^{p}\delta_j e^{-ij\omega}\right)^{*}\right\}^{-1}.$$

6.A.7 Proof of Theorem 6.6

Dieudonné's fixed point theorem (Dieudonné, 1960, Section 10.1) says that there exists a unique point $\widehat{\omega}_T \in A_T(\nu)$ such that $h(\widehat{\omega}_T) = 0$ provided the following two conditions are met as $T \to \infty$.

Condition 6.1 *There exists k, $0 \leqslant k < 1$, such that if $\omega, \omega' \in A_T(\nu)$ then*

$$\left|\omega - \omega' + h_T(\omega) - h_T(\omega')\right| < k\left|\omega - \omega'\right|.$$

Condition 6.2 $|h_T(\omega_0)| < (1-k)\,T^{-\nu}$.

In order to show that these conditions hold, first note that

$$\xi_t = \frac{1}{\sin\omega}\sum_{j=0}^{t}\sin\{(j+1)\,\omega\}\,X_{t-j}. \tag{6.26}$$

Let

$$d_T(\omega) = \sin\omega\sum_{t=0}^{T-1}\xi'_{t-1}\Omega X_t$$

and

$$e_T(\omega) = \sin^2\omega\sum_{t=0}^{T-1}\xi'_{t-1}\Omega\xi_{t-1}.$$

Then

$$h_T(\omega) = \sin\omega d_T(\omega)\, e_T^{-1}(\omega).$$

Following the approach of Theorem 16 of Quinn and Hannan (2001), which shows the above conditions hold in the univariate case, we will analyse $d_T(\omega)$ and $e_T(\omega)$ separately using an expansion around ω_0. We will make use of the fact that

$$
\begin{aligned}
&\sum_{j=0}^{t} \sin(j\omega)\cos\{(t-j)\omega_0\} \\
&= \frac{1}{2}\sum_{j=0}^{t} \sin\{j(\omega-\omega_0)+t\omega_0\} + \frac{1}{2}\sum_{j=0}^{t} \sin\{j(\omega+\omega_0)-t\omega_0\} \\
&= \frac{1}{2}\cos(t\omega_0)\sum_{j=0}^{t} \sin\{j(\omega-\omega_0)\} + \frac{1}{2}\sin(t\omega_0)\sum_{j=0}^{t} \cos\{j(\omega-\omega_0)\} + O(1) \\
&= \frac{1}{2}\cos(t\omega_0)\sum_{k=0}^{\infty}(-1)^k \frac{(\omega-\omega_0)^{2k+1}}{(2k+1)!}\sum_{j=0}^{t} j^{2k+1} \\
&\quad + \frac{1}{2}\sin(t\omega_0)\sum_{k=0}^{\infty}(-1)^k \frac{(\omega-\omega_0)^{2k}}{(2k)!}\sum_{j=0}^{t} j^{2k} + O(1) \\
&= \frac{1}{2}\cos(t\omega_0)\sum_{k=0}^{\infty}(-1)^k \frac{(\omega-\omega_0)^{2k+1}}{(2k+2)!}\left\{t^{2k+2}+O\left(t^{2k+1}\right)\right\} \\
&\quad + \frac{1}{2}\sin(t\omega_0)\sum_{k=0}^{\infty}(-1)^k \frac{(\omega-\omega_0)^{2k}}{(2k+1)!}\left\{t^{2k+1}+O\left(t^{2k}\right)\right\} + O(1)
\end{aligned}
$$

and, similarly,

$$
\begin{aligned}
&\sum_{j=0}^{t} \sin(j\omega)\sin\{(t-j)\omega_0\} \\
&= \frac{1}{2}\sin(t\omega_0)\sum_{k=0}^{\infty}(-1)^k \frac{(\omega-\omega_0)^{2k+1}}{(2k+2)!}\left\{t^{2k+2}+O\left(t^{2k+1}\right)\right\} \\
&\quad - \frac{1}{2}\cos(t\omega_0)\sum_{k=0}^{\infty}(-1)^k \frac{(\omega-\omega_0)^{2k}}{(2k+1)!}\left\{t^{2k+1}+O\left(t^{2k}\right)\right\} + O(1).
\end{aligned}
$$

Also,

$$
\begin{aligned}
&\sum_{j=0}^{t} j\cos(j\omega)\cos\{(t-j)\omega_0\} \\
&= \frac{1}{2}\cos(t\omega_0)\sum_{k=0}^{\infty}(-1)^k \frac{(\omega-\omega_0)^{2k}}{(2k+2)(2k)!}\left\{t^{2k+2}+O\left(t^{2k+1}\right)\right\} \\
&\quad - \frac{1}{2}\sin(t\omega_0)\sum_{k=0}^{\infty}(-1)^k \frac{(\omega-\omega_0)^{2k+1}}{(2k+3)(2k+1)!}\left\{t^{2k+3}+O\left(t^{2k+2}\right)\right\} + O(1)
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{j=0}^{t} j\cos(j\omega)\sin\{(t-j)\omega_0\} \\
&= \frac{1}{2}\cos(t\omega_0)\sum_{k=0}^{\infty}(-1)^k \frac{(\omega-\omega_0)^{2k+1}}{(2k+3)(2k+1)!}\left\{t^{2k+3}+O\left(t^{2k+2}\right)\right\} \\
&\quad + \frac{1}{2}\sin(t\omega_0)\sum_{k=0}^{\infty}(-1)^k \frac{(\omega-\omega_0)^{2k}}{(2k+2)(2k)!}\left\{t^{2k+2}+O\left(t^{2k+1}\right)\right\} + O(1).
\end{aligned}
$$

Furthermore,

$$\sum_{t=0}^{T-1} t^j e^{i\omega t} \varepsilon_t = O\left\{T^{j+1/2} (\log T)^{1/2}\right\}, \tag{6.27}$$

which follows from Theorem 2.4.

We begin with $e_T(\omega)$ which, from (6.26), is equal to

$$\begin{aligned}
&\sum_{t=0}^{T-1} \left\{\sum_{j=0}^{t} \sin(j\omega)\, \varepsilon'_{t-j}\right\} \Omega \left\{\sum_{k=0}^{t} \sin(k\omega)\, \varepsilon_{t-k}\right\} \\
&+ \sum_{t=0}^{T-1} \left[\alpha'_0 \sum_{j=0}^{t} \sin(j\omega) \cos\{(t-j)\,\omega_0\} + \beta'_0 \sum_{j=0}^{t} \sin(j\omega) \sin\{(t-j)\,\omega_0\}\right] \Omega \\
&\qquad \times \left[\alpha_0 \sum_{j=0}^{t} \sin(j\omega) \cos\{(t-j)\,\omega_0\} + \beta_0 \sum_{j=0}^{t} \sin(j\omega) \sin\{(t-j)\,\omega_0\}\right] \\
&+ 2\sum_{t=0}^{T-1} \left[\alpha'_0 \sum_{j=0}^{t} \sin(j\omega) \cos\{(t-j)\,\omega_0\} + \beta'_0 \sum_{j=0}^{t} \sin(j\omega) \sin\{(t-j)\,\omega_0\}\right] \Omega \\
&\qquad \times \left\{\sum_{j=0}^{t} \sin(j\omega)\, \varepsilon_{t-j}\right\}.
\end{aligned} \tag{6.28}$$

From the mean value theorem, the first term in (6.28) is

$$\begin{aligned}
&\sum_{t=0}^{T-1} \left\{\sum_{j=0}^{t} \sin(j\omega_0)\, \varepsilon'_{t-j}\right\} \Omega \left\{\sum_{k=0}^{t} \sin(k\omega_0)\, \varepsilon_{t-k}\right\} \\
&+ 2\,(\omega-\omega_0) \sum_{t=0}^{T-1} \left\{\sum_{j=0}^{t} j\cos(j\omega^*)\, \varepsilon'_{t-j}\right\} \Omega \left\{\sum_{k=0}^{t} \sin(k\omega^*)\, \varepsilon_{t-k}\right\},
\end{aligned}$$

where ω^* is a point on the line segment between ω and ω_0, which is

$$\sum_{t=0}^{T-1} \left\{\sum_{j=0}^{t} \sin(j\omega_0)\, \varepsilon'_{t-j}\right\} \Omega \left\{\sum_{k=0}^{t} \sin(k\omega_0)\, \varepsilon_{t-k}\right\} + (\omega-\omega_0)\, O\left(T^3 \log T\right) = O\left(T^2 \log T\right).$$

The third term in (6.28) is

$$
\begin{aligned}
2 \sum_{t=0}^{T-1} & \left[\alpha_0' \sum_{j=0}^{t} \sin (j\omega) \cos \{(t-j)\,\omega_0\} + \beta_0' \sum_{j=0}^{t} \sin (j\omega) \sin \{(t-j)\,\omega_0\} \right] \Omega \\
& \times \left\{ \sum_{j=0}^{t} \sin (j\omega_0)\, \varepsilon_{t-j} \right\} \\
+ 2 (\omega - \omega_0) & \sum_{t=0}^{T-1} \left[\alpha_0' \sum_{j=0}^{t} j \cos (j\omega^*) \cos \{(t-j)\,\omega_0\} + \beta_0' \sum_{j=0}^{t} j \cos (j\omega^*) \sin \{(t-j)\,\omega_0\} \right] \Omega \\
& \times \left\{ \sum_{j=0}^{t} \sin \{(t-j)\,\omega^*\}\, \varepsilon_j \right\} \\
+ 2 (\omega - \omega_0) & \sum_{t=0}^{T-1} \left[\alpha_0' \sum_{j=0}^{t} \sin (j\omega^*) \cos \{(t-j)\,\omega_0\} + \beta_0' \sum_{j=0}^{t} \sin (j\omega^*) \sin \{(t-j)\,\omega_0\} \right] \Omega \\
& \times \left\{ \sum_{j=0}^{t} (t-j) \cos \{(t-j)\,\omega^*\}\, \varepsilon_j \right\},
\end{aligned}
$$

which is

$$
\begin{aligned}
& 2 \sum_{t=0}^{T-1} \left[\alpha_0' \sum_{j=0}^{t} \sin (j\omega_0) \cos \{(t-j)\,\omega_0\} + \beta_0' \sum_{j=0}^{t} \sin (j\omega_0) \sin \{(t-j)\,\omega_0\} \right] \Omega \\
& \qquad \times \left\{ \sum_{j=0}^{t} \sin (j\omega_0)\, \varepsilon_{t-j} \right\} + (\omega - \omega_0)\, O \left\{ T^{7/2} (\log T)^{1/2} \right\} \\
& = O \left\{ T^{5/2} (\log T)^{1/2} \right\}.
\end{aligned}
$$

The second term in (6.28) is

$$
\begin{aligned}
\sum_{t=0}^{T-1} & \left[\alpha_0' \sum_{j=0}^{t} \sin (j\omega_0) \cos \{(t-j)\,\omega_0\} + \beta_0' \sum_{j=0}^{t} \sin (j\omega_0) \sin \{(t-j)\,\omega_0\} \right] \Omega \\
& \times \left[\alpha_0 \sum_{j=0}^{t} \sin (j\omega_0) \cos \{(t-j)\,\omega_0\} + \beta_0 \sum_{j=0}^{t} \sin (j\omega_0) \sin \{(t-j)\,\omega_0\} \right] \\
+ 2 (\omega - \omega_0) & \sum_{t=0}^{T-1} \left[\alpha_0' \sum_{j=0}^{t} j \cos (j\omega^*) \cos \{(t-j)\,\omega_0\} + \beta_0' \sum_{j=0}^{t} j \cos (j\omega^*) \sin \{(t-j)\,\omega_0\} \right] \Omega \\
& \times \left[\alpha_0 \sum_{j=0}^{t} \sin (j\omega^*) \cos \{(t-j)\,\omega_0\} + \beta_0 \sum_{j=0}^{t} \sin (j\omega^*) \sin \{(t-j)\,\omega_0\} \right],
\end{aligned}
$$

which is

$$\sum_{t=0}^{T-1}\left[\alpha_0'\sum_{j=0}^{t}\sin(j\omega_0)\cos\{(t-j)\omega_0\}+\beta_0'\sum_{j=0}^{t}\sin(j\omega_0)\sin\{(t-j)\omega_0\}\right]\Omega$$
$$\times\left[\alpha_0\sum_{j=0}^{t}\sin(j\omega_0)\cos\{(t-j)\omega_0\}+\beta_0\sum_{j=0}^{t}\sin(j\omega_0)\sin\{(t-j)\omega_0\}\right]$$
$$+(\omega-\omega_0)O\left(T^3\right)$$
$$=\frac{\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0}{24}T^3+O\left(T^2\right).$$

Thus

$$e_T(\omega)=e_T(\omega_0)+(\omega-\omega_0)O\left\{T^{7/2}(\log T)^{1/2}\right\} \tag{6.29}$$

and also

$$e_T(\omega)=\frac{\alpha_0'\Sigma\alpha_0+\beta_0'\Omega\beta_0}{24}T^3+O\left\{T^{5/2}(\log T)^{1/2}\right\}. \tag{6.30}$$

Next we analyse $d_T(\omega)$, which is

$$\sum_{t=0}^{T-1}\varepsilon_t'\Omega\sum_{j=0}^{t}\sin(j\omega)\varepsilon_{t-j}$$
$$+\sum_{t=0}^{T-1}\left\{\alpha_0'\cos(t\omega_0)+\beta_0'\sin(t\omega_0)\right\}\Omega\sum_{j=0}^{t}\sin(j\omega)\varepsilon_{t-j}$$
$$+\sum_{t=0}^{T-1}\varepsilon_t'\Omega\sum_{j=0}^{t}\left[\alpha_0\sin(j\omega)\cos\{(t-j)\omega_0\}+\beta_0\sin(j\omega)\sin\{(t-j)\omega_0\}\right]$$
$$+\sum_{t=0}^{T-1}\left\{\alpha_0'\cos(t\omega_0)+\beta_0'\sin(t\omega_0)\right\}\Omega$$
$$\times\sum_{j=0}^{t}\left[\alpha_0\sin(j\omega)\cos\{(t-j)\omega_0\}+\beta_0\sin(j\omega)\sin\{(t-j)\omega_0\}\right]. \tag{6.31}$$

Again using the mean value theorem, the first term in (6.31) is

$$\sum_{t=0}^{T-1}\varepsilon_t'\Omega\sum_{j=0}^{t}\sin(j\omega_0)\varepsilon_{t-j}+(\omega-\omega_0)\sum_{t=0}^{T-1}\varepsilon_t'\Omega\sum_{j=0}^{t}j\cos(j\omega^*)\varepsilon_{t-j},$$

where ω^* is a point on the line segment between ω and ω_0. By the Cauchy–Schwarz inequality,

$$\left\{\sum_{t=0}^{T-1}\varepsilon_t'\Omega\sum_{j=0}^{t}j\cos(j\omega^*)\varepsilon_{t-j}\right\}^2$$
$$\leqslant\left\{\sum_{t=0}^{T-1}\varepsilon_t'\Omega\varepsilon_t\right\}\sum_{t=0}^{T-1}\left\{\sum_{j=0}^{t}j\cos(j\omega^*)\varepsilon_{t-j}'\right\}\Omega\left\{\sum_{j=0}^{t}j\cos(j\omega^*)\varepsilon_{t-j}\right\}$$
$$=O\left(T^5\log T\right)$$

and so the first term in (6.31) is

$$\sum_{t=0}^{T-1} \varepsilon_t' \Omega \sum_{j=0}^{t} \sin(j\omega_0)\, \varepsilon_{t-j} + (\omega - \omega_0)\, O\left\{T^{5/2} (\log T)^{1/2}\right\} = O\left\{T^{3/2} (\log T)^{1/2}\right\}.$$

The second term in (6.31) is

$$\begin{aligned}
&\sum_{t=0}^{T-1} \varepsilon_t' \Omega \sum_{j=t}^{T-1} \sin\{(j-t)\,\omega_0\}\{\alpha_0 \cos(j\omega_0) + \beta_0 \sin(j\omega_0)\} \\
&+ (\omega - \omega_0) \sum_{t=0}^{T-1} \varepsilon_t' \Omega \sum_{j=t}^{T-1} (j-t) \cos\{(j-t)\,\omega^*\}\{\alpha_0 \cos(j\omega_0) + \beta_0 \sin(j\omega_0)\} \\
&= \sum_{t=0}^{T-1} \varepsilon_t' \Omega \sum_{j=t}^{T-1} \sin\{(j-t)\,\omega_0\}\{\alpha_0 \cos(j\omega_0) + \beta_0 \sin(j\omega_0)\} \\
&\quad + (\omega - \omega_0) \sum_{t=0}^{T-1} \varepsilon_t' \Omega \sum_{j=0}^{T-1-t} j \cos(j\omega^*)\,[\alpha_0 \cos\{(t+j)\,\omega_0\} + \beta_0 \sin\{(t+j)\,\omega_0\}],
\end{aligned}$$

which is

$$\begin{aligned}
&\sum_{t=0}^{T-1} \varepsilon_t' \Omega \sum_{j=t}^{T-1} \sin\{(j-t)\,\omega_0\}\{\alpha_0 \cos(j\omega_0) + \beta_0 \sin(j\omega_0)\} + (\omega - \omega_0)\, O\left\{T^{5/2} (\log T)^{1/2}\right\} \\
&= O\left\{T^{3/2} (\log T)^{1/2}\right\},
\end{aligned}$$

since

$$\begin{aligned}
&\sum_{j=0}^{T-1-t} j \cos(j\omega) \cos\{(t+j)\,\omega_0\} \\
&= \frac{1}{2} \cos(t\omega_0) \sum_{k=0}^{\infty} (-1)^k \frac{(\omega - \omega_0)^{2k}}{(2k+2)\,(2k)!} \left[(T-t)^{2k+2} + O\left\{(T-t)^{2k+1}\right\}\right] \\
&\quad + \frac{1}{2} \sin(t\omega_0) \sum_{k=0}^{\infty} (-1)^k \frac{(\omega - \omega_0)^{2k+1}}{(2k+3)\,(2k+1)!} \left[(T-t)^{2k+3} + O\left\{(T-t)^{2k+2}\right\}\right] + O(1)
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{j=0}^{T-1-t} j \cos(j\omega) \sin\{(t+j)\,\omega_0\} \\
&= \frac{1}{2} \sin(t\omega_0) \sum_{k=0}^{\infty} (-1)^k \frac{(\omega - \omega_0)^{2k}}{(2k+2)\,(2k)!} \left[(T-t)^{2k+2} + O\left\{(T-t)^{2k+1}\right\}\right] \\
&\quad - \frac{1}{2} \cos(t\omega_0) \sum_{k=0}^{\infty} (-1)^{2k+1} \frac{(\omega - \omega_0)^{2k+1}}{(2k+3)\,(2k+1)!} \left[(T-t)^{2k+3} + O\left\{(T-t)^{2k+2}\right\}\right] + O(1).
\end{aligned}$$

Similarly, the third term in (6.31) is

$$\begin{aligned}
&\sum_{t=0}^{T-1} \varepsilon_t' \Omega \sum_{j=0}^{t} [\alpha_0 \sin(j\omega_0) \cos\{(t-j)\,\omega_0\} + \beta_0 \sin(j\omega_0) \sin\{(t-j)\,\omega_0\}] \\
&+ (\omega - \omega_0)\, O\left\{T^{5/2} (\log T)^{1/2}\right\} \\
&= O\left\{T^{3/2} (\log T)^{1/2}\right\}.
\end{aligned}$$

The fourth term in (6.31) is

$$\begin{aligned}
&\sum_{t=0}^{T-1}\left\{\alpha_0'\cos(t\omega_0)+\beta_0'\sin(t\omega_0)\right\}\Omega\\
&\qquad\times\sum_{j=0}^{t}\left[\alpha_0\sin(j\omega_0)\cos\{(t-j)\omega_0\}+\beta_0\sin(j\omega_0)\sin\{(t-j)\omega_0\}\right]\\
&+(\omega-\omega_0)\sum_{t=0}^{T-1}\left\{\alpha_0'\cos(t\omega_0)+\beta_0'\sin(t\omega_0)\right\}\Omega\\
&\qquad\qquad\times\sum_{j=0}^{t}\left[\alpha_0 j\cos(j\omega^*)\cos\{(t-j)\omega_0\}+\beta_0 j\cos(j\omega^*)\sin\{(t-j)\omega_0\}\right]\\
&=O(T)+(\omega-\omega_0)\left\{\frac{\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0}{24}T^3+O\left(T^2\right)\right\}. \qquad (6.32)
\end{aligned}$$

Thus

$$d_T(\omega)=d_T(\omega_0)+(\omega-\omega_0)\left[\frac{\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0}{24}T^3+O\left\{T^{5/2}(\log T)^{1/2}\right\}\right]. \qquad (6.33)$$

Now,

$$h_T(\omega)-h_T(\omega')=d_T(\omega)e_T^{-1}(\omega)-d_T(\omega')e_T^{-1}(\omega'),$$

and so

$$e_T(\omega)e_T(\omega')\left\{h_T(\omega)-h_T(\omega')\right\}=\left\{d_T(\omega)-d_T(\omega')\right\}e_T(\omega)-\left\{e_T(\omega)-e_T(\omega')\right\}d_T(\omega')$$

which is, from (6.29), (6.30) and (6.33),

$$\begin{aligned}
&(\omega-\omega')\left[\frac{\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0}{24}T^3+O\left\{T^{5/2}(\log T)^{1/2}\right\}\right]\\
&\times\left[\frac{\alpha_0'\Sigma\alpha_0+\beta_0'\Omega\beta_0}{24}T^3+O\left\{T^{5/2}(\log T)^{1/2}\right\}\right]\\
&-(\omega-\omega')O\left\{T^{7/2}(\log T)^{1/2}\right\}O\left\{T^{3/2}(\log T)^{1/2}\right\}\\
&=(\omega-\omega')\left[\left(\frac{\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0}{24}\right)^2T^6+O\left\{T^{11/2}(\log T)^{1/2}\right\}\right].
\end{aligned}$$

Thus

$$\begin{aligned}
h_T(\omega)-h_T(\omega')&=(\omega-\omega')\frac{\left[\left(\frac{\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0}{24}\right)^2T^6+O\left\{T^{11/2}(\log T)^{1/2}\right\}\right]}{\left[\frac{\alpha_0'\Sigma\alpha_0+\beta_0'\Omega\beta_0}{24}T^3+O\left\{T^{5/2}(\log T)^{1/2}\right\}\right]^2}\\
&=(\omega-\omega')\left[1+O\left\{T^{-1/2}(\log T)^{1/2}\right\}\right] \qquad (6.34)
\end{aligned}$$

and so

$$\omega-\omega'-h_T(\omega)+h_T(\omega')=(\omega-\omega')O\left\{T^{-1/2}(\log T)^{1/2}\right\}.$$

Furthermore,

$$\begin{aligned}
h_T(\omega_0) &= d_T(\omega_0)\, e_T^{-1}(\omega_0) \\
&= \frac{O\left\{T^{3/2}(\log T)^{1/2}\right\}}{\frac{\alpha_0'\Sigma\alpha_0+\beta_0'\Omega\beta_0}{24}T^3 + O\left\{T^{5/2}(\log T)^{1/2}\right\}} \\
&= O\left\{T^{-3/2}(\log T)^{1/2}\right\}. \qquad (6.35)
\end{aligned}$$

From (6.34) and (6.35), Conditions 6.1 and 6.2 hold for $1 < \nu < 3/2$. Since $A_T(\nu)$ is expanding as v decreases, it follows that there is a unique point $\widehat{\omega}_T$ such that $h(\widehat{\omega}_T) = 0$ for which $T^{-\nu}(\widehat{\omega}_T - \omega_0) \to 0$ as $T \to \infty$ for any $\nu < 3/2$.

6.A.8 Proof of Theorem 6.7

We first consider the case where $1 < \nu < 3/2$. From (6.34), letting $\omega' = \widehat{\omega}_T$,

$$\omega - \widehat{\omega}_T - h_T(\omega) + h_T(\widehat{\omega}_T) = (\omega - \widehat{\omega}_T)\, O\left\{T^{-1/2}(\log T)^{1/2}\right\}. \qquad (6.36)$$

But $h_T(\widehat{\omega}_T) = 0$ and $\omega_j - h_T(\omega_j) = \omega_{j+1}$ so

$$\omega_{j+1} - \widehat{\omega}_T = (\omega_j - \widehat{\omega}_T)\, O\left\{T^{-1/2}(\log T)^{1/2}\right\}.$$

Putting $\omega = \omega_0$ in (6.36),

$$\begin{aligned}
\omega_0 - \widehat{\omega}_T &= h_T(\omega_0) + o\left(T^{-3/2}\right) \qquad (6.37) \\
&= O\left\{T^{-3/2}(\log T)^{1/2}\right\}.
\end{aligned}$$

We now analyse $e_T(\omega)$ and $d_T(\omega)$ when $1/2 < \nu \leqslant 1$. We begin with $e_T(\omega)$, which is given by (6.28). The first term is $O\left(T^2 \log T\right)$. The second term is

$$\begin{aligned}
&\alpha'\Sigma^{-1}\alpha \sum_{t=0}^{T-1}\left[\sum_{j=0}^{t}\sin(j\omega)\cos\{(t-j)\omega_0\}\right]^2 + \beta'\Sigma^{-1}\beta \sum_{t=0}^{T-1}\left[\sum_{j=0}^{t}\sin(j\omega)\sin\{(t-j)\omega_0\}\right]^2 \\
&+ 2\alpha'\Sigma^{-1}\beta \sum_{t=0}^{T-1}\left[\sum_{j=0}^{t}\sin(j\omega)\cos\{(t-j)\omega_0\}\right]\left[\sum_{j=0}^{t}\sin(j\omega)\sin\{(t-j)\omega_0\}\right] \\
&= \frac{\alpha'\Sigma^{-1}\alpha}{4}\sum_{t=0}^{T-1}\left[\sum_{j=0}^{t}\sin\{j(\omega-\omega_0)+t\omega_0\} + O(1)\right]^2 \\
&\quad + \frac{\beta'\Sigma^{-1}\beta}{4}\sum_{t=0}^{T-1}\left[\sum_{j=0}^{t}\cos\{j(\omega-\omega_0)+t\omega_0\} + O(1)\right]^2 \\
&\quad + \frac{\alpha'\Sigma^{-1}\beta}{2}\sum_{t=0}^{T-1}\left[\sum_{j=0}^{t}\sin\{j(\omega-\omega_0)+t\omega_0\} + O(1)\right]\left[\sum_{j=0}^{t}\cos\{j(\omega-\omega_0)+t\omega_0\} + O(1)\right].
\end{aligned}$$

Letting $\lambda = (\omega - \omega_0)/2$,

$$\sum_{j=0}^{t} \sin\{j(\omega - \omega_0) + t\omega_0\} = \sin(t\omega_0 + t\lambda)\frac{\sin\{(t+1)\lambda\}}{\sin\lambda} \tag{6.38}$$

and, similarly,

$$\sum_{j=0}^{t} \cos\{j(\omega - \omega_0) + t\omega_0\} = \cos(t\omega_0 + t\lambda)\frac{\sin\{(t+1)\lambda\}}{\sin\lambda}. \tag{6.39}$$

Thus

$$\begin{aligned}
\sum_{t=0}^{T-1}\left[\sum_{j=0}^{t}\sin\{j(\omega-\omega_0)+t\omega_0\}\right]^2 &= \sum_{t=0}^{T-1}\sin^2(t\omega_0+t\lambda)\frac{\sin^2\{(t+1)\lambda\}}{\sin^2\lambda} \\
&= \frac{1}{4\sin^2\lambda}\sum_{t=0}^{T-1}\{1-\cos(2t\omega_0+2t\lambda)\}\left[1-\cos\{2(t+1)\lambda\}\right] \\
&= \frac{1}{4\sin^2\lambda}\left[T - \mathrm{Re}\left(\sum_{t=0}^{T-1}e^{2i(t+1)\lambda}\right)+O(1)\right] \\
&= \frac{1}{4\sin^2\lambda}\left[T - \cos\{(T+1)\lambda\}\frac{\sin(T\lambda)}{\sin\lambda}+O(1)\right] \\
&= \frac{1}{4\sin^2\lambda}\left[T + \frac{1}{2} - \frac{\sin\{(2T+1)\lambda\}}{2\sin\lambda}+O(1)\right] \\
&= \frac{1}{8}g_{2T+1}(\lambda) + \frac{1}{\sin^2\lambda}O(1),
\end{aligned}$$

where

$$g_T(x) = \frac{T}{\sin^2 x}\left\{1 - \frac{\sin(Tx)}{T\sin x}\right\}.$$

Similarly,

$$\sum_{t=0}^{T-1}\left[\sum_{j=0}^{t}\cos\{j(\omega-\omega_0)+t\omega_0\}\right]^2 = \frac{1}{8}g_{2T+1}(\lambda) + \frac{1}{\sin^2\lambda}O(1)$$

and

$$\sum_{t=0}^{T-1}\left[\sum_{j=0}^{t}\sin\{j(\omega-\omega_0)+t\omega_0\}\right]\left[\sum_{j=0}^{t}\cos\{j(\omega-\omega_0)+t\omega_0\}\right] = \frac{1}{\sin^2\lambda}O(1).$$

Now,

$$g_{2T+1}(\lambda) = O\left(T^{1+2\nu}\right)$$

and

$$\frac{1}{\sin^2\lambda} = O\left(T^{2\nu}\right).$$

Thus the second term of $e_T(\omega)$ is

$$\frac{\alpha_0'\Omega\alpha_0 + \beta_0'\Omega\beta_0}{32}g_{2T+1}(\lambda) + O\left(T^{2\nu}\right).$$

By the Cauchy–Schwarz inequality, the third term of $e_T(\omega)$ is $O\left\{T^{3/2+\nu}(\log T)^{1/2}\right\}$ and so

$$e_T(\omega) = \frac{\alpha_0'\Omega\alpha_0 + \beta_0'\Omega\beta_0}{32} g_{2T+1}(\lambda) + O\left\{T^{3/2+\nu}(\log T)^{1/2}\right\}. \tag{6.40}$$

We next analyse $d_T(\omega)$, which is given by (6.31), for $1/2 < \nu \leqslant 1$. Let

$$C_\varepsilon(j) = T^{-1}\sum_{t=j}^{T-1}\varepsilon_{t-j}\varepsilon_t'.$$

Then, for any ω,

$$\begin{aligned}
T^{-1}\left|\sum_{j=0}^{T-1}\sum_{t=j}^{T-1}\sin(j\omega)\,\varepsilon_t'\Omega\varepsilon_{t-j}\right| &= T^{-1}\left|\sum_{j=0}^{T-1}\sum_{t=j}^{T-1}\sin(j\omega)\operatorname{tr}\left(\Omega\varepsilon_{t-j}\varepsilon_t'\right)\right| \\
&= \left|\sum_{j=0}^{T-1}\sin(j\omega)\operatorname{tr}\left(\Omega T^{-1}\sum_{t=j}^{T-1}\varepsilon_{t-j}\varepsilon_t'\right)\right| \\
&= \left|\sum_{j=0}^{T-1}\sin(j\omega)\operatorname{tr}\left[\Omega\left\{C_\varepsilon(j)-\Gamma_\varepsilon(j)\right\}\right] + \sum_{j=0}^{T-1}\sin(j\omega)\operatorname{tr}\left\{\Omega\Gamma_\varepsilon(j)\right\}\right| \\
&\leqslant \sup_{j=0,\ldots,T-1}\left|C_\varepsilon(j)-\Gamma_\varepsilon(j)\right| + O(1) \\
&= O(1),
\end{aligned} \tag{6.41}$$

from An et al. (1982). Thus the first term in (6.31) is $O(T)$. The third term in (6.31) is

$$\sum_{t=0}^{T-1}\varepsilon_t'\Omega\alpha_0\sum_{j=0}^{t}\sin(j\omega)\cos\{(t-j)\omega_0\} + \sum_{t=0}^{T-1}\varepsilon_t'\Omega\beta_0\sum_{j=0}^{t}\sin(j\omega)\sin\{(t-j)\omega_0\}$$

which, from (6.38) and (6.39), is equal to

$$\begin{aligned}
&\frac{1}{4\sin\lambda}\sum_{t=0}^{T-1}\varepsilon_t'\Omega\alpha_0\left[\cos(t\omega_0-\lambda)-\cos\{t\omega_0+(2t+1)\lambda\}\right] \\
&+\frac{1}{4\sin\lambda^*}\sum_{t=0}^{T-1}\varepsilon_t'\Omega\alpha_0\left[\cos(-t\omega_0-\lambda^*)-\cos\{-t\omega_0+(2t+1)\lambda^*\}\right] \\
&+\frac{1}{4\sin\lambda}\sum_{t=0}^{T-1}\varepsilon_t'\Omega\beta_0\left[\sin\{t\omega_0+(2t+1)\lambda\}-\sin(t\omega_0-\lambda)\right] \\
&+\frac{1}{4\sin\lambda^*}\sum_{t=0}^{T-1}\varepsilon_t'\Omega\beta_0\left[\sin\{-t\omega_0+(2t+1)\lambda^*\}-\sin(-t\omega_0-\lambda^*)\right],
\end{aligned} \tag{6.42}$$

where $\lambda^* = (\omega+\omega_0)/2$. The second and fourth terms in (6.42) are $O\left\{T^{1/2}(\log T)^{1/2}\right\}$, and the first and third terms in (6.42) are $\frac{1}{\sin\lambda}O\left\{T^{1/2}(\log T)^{1/2}\right\}$. Thus the third term in (6.31) is $\frac{1}{\sin\lambda}O\left\{T^{1/2}(\log T)^{1/2}\right\}$. Similarly, the second term in (6.31) is also $\frac{1}{\sin\lambda}O\left\{T^{1/2}(\log T)^{1/2}\right\}$.

The fourth term in (6.31) is

$$\begin{aligned}
&\alpha_0'\Omega\alpha_0\sum_{t=0}^{T-1}\cos(t\omega_0)\sum_{j=0}^{t}\sin(j\omega)\cos\{(t-j)\,\omega_0\}\\
&+\beta_0'\Omega\beta_0\sum_{t=0}^{T-1}\sin(t\omega_0)\sum_{j=0}^{t}\sin(j\omega)\sin\{(t-j)\,\omega_0\}\\
&+\alpha_0'\Omega\beta_0\sum_{t=0}^{T-1}\cos(t\omega_0)\sum_{j=0}^{t}\sin(j\omega)\sin\{(t-j)\,\omega_0\}\\
&+\beta_0'\Omega\alpha_0\sum_{t=0}^{T-1}\sin(t\omega_0)\sum_{j=0}^{t}\sin(j\omega)\cos\{(t-j)\,\omega_0\}\,. \qquad (6.43)
\end{aligned}$$

From (6.38) and (6.39),

$$\begin{aligned}
&\sum_{t=0}^{T-1}\cos(t\omega_0)\sum_{j=0}^{t}\sin(j\omega)\cos\{(t-j)\,\omega_0\}\\
&=\frac{1}{2}\sum_{t=0}^{T-1}\cos(t\omega_0)\sum_{j=0}^{t}\sin(2j\lambda+t\omega_0)+O(T)\\
&=\frac{1}{2\sin\lambda}\sum_{t=0}^{T-1}\cos(t\omega_0)\sin(t\omega_0+\lambda)\sin\{(t+1)\,\lambda\}+O(T)\\
&=\frac{1}{4\sin\lambda}\sum_{t=0}^{T-1}\sin(2t\omega_0+t\lambda)\sin\{(t+1)\,\lambda\}+\frac{1}{4\sin\lambda}\sum_{t=0}^{T-1}\sin(t\lambda)\sin\{(t+1)\,\lambda\}+O(T)\,.
\end{aligned} \qquad (6.44)$$

The first term in (6.44) is $O(T)$. The second term in (6.44) is

$$\begin{aligned}
\frac{1}{4\sin\lambda}\sum_{t=0}^{T-1}\sin(t\lambda)\sin\{(t+1)\,\lambda\}&=\frac{1}{8\sin\lambda}\left[T\cos\lambda-\sum_{t=0}^{T-1}\cos\{(2t+1)\,\lambda\}\right]\\
&=\frac{1}{8\sin\lambda}\left\{T\cos\lambda-\frac{\sin(2T\lambda)}{2\sin\lambda}\right\}\\
&=\frac{T}{8\sin\lambda}\left\{\cos\lambda-1+1-\frac{\sin(2T\lambda)}{2T\sin\lambda}\right\}\\
&=-\frac{T}{8}\tan(\lambda/2)+\frac{\sin\lambda}{16}g_{2T}(\lambda)\\
&=O\left(T^{1-\nu}\right)+\frac{\sin\lambda}{16}g_{2T}(\lambda)\,.
\end{aligned}$$

Treating the other terms in (6.43) in the same way, the fourth term in (6.31) is

$$\frac{\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0}{16}\sin\lambda g_{2T}(\lambda)+O\left(T^{1-\nu}\right).$$

Therefore

$$d_T(\omega)=\frac{\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0}{16}\sin\lambda g_{2T}(\lambda)+\frac{1}{\sin\lambda}O\left\{T^{1/2}(\log T)^{1/2}\right\}. \qquad (6.45)$$

It follows from (6.40) and (6.45) that

$$\frac{d_T(\omega)}{e_T(\omega)} = (\omega-\omega_0)\left[1+O\left\{T^{1/2-\nu}(\log T)^{1/2}\right\}\right]+O\left\{T^{-1/2-\nu}(\log T)^{1/2}\right\}.$$

Thus

$$h_T(\omega)+\omega-\widehat{\omega}_T = (\omega-\widehat{\omega}_T)\,O\left\{T^{1/2-\nu}(\log T)^{1/2}\right\}+O\left\{T^{1/2-2\nu}(\log T)^{1/2}\right\}$$

and so

$$\omega_{j+1}-\widehat{\omega}_T = (\omega_j-\widehat{\omega}_T)\,O\left\{T^{-1/2-\nu}(\log T)^{1/2}\right\}+O\left\{T^{-1/2-2\nu}(\log T)^{1/2}\right\}.$$

Also,

$$h_T(\omega_0)+\omega_0-\widehat{\omega}_T = O\left\{T^{-1/2-2\nu}(\log T)^{1/2}\right\},$$

and so

$$\begin{aligned}\omega_0-\widehat{\omega}_T &= h(\omega_0)+o\left(T^{-3/2}\right)\\ &= O\left\{T^{-3/2}(\log T)^{1/2}\right\},\end{aligned}$$

as before.

To prove the final part of the theorem, let $\omega_{j+1}-\widehat{\omega}_T = O\left(T^{-\phi_{j+1}}\right)$ and $\phi_0=\nu$. Then

$$\phi_{j+1} = 2\phi_j-\frac{1}{2}+\kappa,$$

where κ is arbitrarily small. Thus

$$\phi_j = 2^{j-1}\left(\nu-\frac{1}{2}\right)+\frac{1}{2}+\kappa$$

and so $\phi_j>1$ when

$$2^{j-1}(2\nu-1)>1,$$

that is when

$$j \geqslant \lfloor 2-\log(2\nu-1)/\log(2)\rfloor.$$

One more iteration is then required for to be within $o\left(T^{-3/2}\right)$ of $\widehat{\omega}_T$.

6.A.9 Proof of Theorem 6.8

From (6.37),

$$\widehat{\omega}_T-\omega_0 = -d_T(a_0)\,e_T^{-1}(a_0)\{1+o(1)\}.$$

From (6.30),

$$T^{-3}e_T(a) \to \frac{\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0}{24}.$$

It remains to find the asymptotic distribution of $T^{-3/2}d_T(a)$, which, from (6.31), (6.32) and (6.41), is

$$T^{-3/2}\sum_{t=0}^{T-1}\left\{\alpha_0'\cos(t\omega_0)+\beta_0'\sin(t\omega_0)\right\}\Omega\sum_{j=0}^{t}\sin(j\omega_0)\varepsilon_{t-j}$$
$$+T^{-3/2}\sum_{t=0}^{T-1}\varepsilon_t'\Omega\sum_{j=0}^{t}\left[\alpha_0\sin(j\omega)\cos\{(t-j)\omega_0\}+\beta_0\sin(j\omega_0)\sin\{(t-j)\omega_0\}\right]$$
$$+O\left(T^{-1/2}\right). \tag{6.46}$$

The first and second terms in (6.46) are

$$T^{-3/2}\sum_{t=0}^{T-1}\varepsilon_t'\Omega\alpha_0\left[\sum_{j=0}^{t}\sin(j\omega_0)\cos\{(t-j)\omega_0\}+\sum_{j=t}^{T-1}\cos(j\omega_0)\sin\{(j-t)\omega_0\}\right]$$
$$+T^{-3/2}\sum_{t=0}^{T-1}\varepsilon_t'\Omega\beta_0\left[\sum_{j=0}^{t}\sin(j\omega_0)\sin\{(t-j)\omega_0\}+\sum_{j=t}^{T-1}\sin(j\omega_0)\sin\{(j-t)\omega_0\}\right]$$

which is equal to

$$\frac{1}{2}T^{-3/2}\sum_{t=0}^{T-1}\varepsilon_t'\Omega\alpha_0\left\{(t+1)\sin(t\omega_0)-(T-t)\sin(t\omega_0)+O(1)\right\}$$
$$+\frac{1}{2}T^{-3/2}\sum_{t=0}^{T-1}\varepsilon_t'\Omega\beta_0\left\{-(t+1)\cos(t\omega_0)+(T-t)\cos(t\omega_0)+O(1)\right\}$$
$$=\frac{1}{2}T^{-3/2}\sum_{t=0}^{T-1}\varepsilon_t'\Omega\alpha_0(2t-T)\sin(t\omega_0)+\frac{1}{2}T^{-3/2}\sum_{t=0}^{T-1}\varepsilon_t'\Omega\beta_0(2t-T)\cos(t\omega_0)+O\left(T^{-1/2}\right)$$
$$=-\frac{1}{2\sqrt{2}}\sqrt{2}\alpha_0'\Omega T^{-1/2}\sum_{t=0}^{T-1}\varepsilon_t\left(1-2tT^{-1}\right)\sin(t\omega_0)$$
$$-\frac{1}{2\sqrt{2}}\beta_0'\Omega\sqrt{2}T^{-1/2}\sum_{t=0}^{T-1}\varepsilon_t\left(1-2tT^{-1}\right)\cos(t\omega_0)+O\left(T^{-1/2}\right).$$

It was shown in the proof of Theorem 6.1 that

$$\sqrt{2}T^{-1/2}\sum_{t=0}^{T-1}\varepsilon_t\left(1-2tT^{-1}\right)\sin(t\omega_0)$$

and

$$\sqrt{2}T^{-1/2}\sum_{t=0}^{T-1}\varepsilon_t\left(1-2tT^{-1}\right)\cos(t\omega_0)$$

are both asymptotically normal with mean zero and variance $2\pi f_\varepsilon(\omega_0)/3$. Thus (6.46) is asymptotically normal with mean zero and variance

$$\frac{2\pi}{24}\left\{\alpha_0'\Omega f_\varepsilon(\omega_0)\Omega\alpha_0+\beta_0'\Omega f_\varepsilon(\omega_0)\Omega\beta_0\right\}.$$

Hence $T^{3/2}(\widehat{\omega}-\omega_0)$ is asymptotically normal with mean zero and variance

$$48\pi\frac{\alpha_0'\Omega f_\varepsilon(\omega_0)\Omega\alpha_0+\beta_0'\Omega f_\varepsilon(\omega_0)\Omega\beta_0}{(\alpha_0'\Omega\alpha_0+\beta_0'\Omega\beta_0)^2}.$$

7

Discriminating Between Time Series With Periodic Components

7.1 Introduction

In this chapter we consider the problem of discriminating between two or more time series which are generated by stochastic processes which contain periodic components. As in Chapter 6, we model such processes as the sum of sinusoids at fixed frequencies and stationary noise. One null hypothesis of interest is that the fixed frequencies of each process are the same and the noise processes have the same spectral shape. Another is that the fixed frequencies of the processes are the same without any restriction on the spectral densities of the noise processes. This second test would be relevant if the signal of interest was defined by the periodic component, and the stationary part was considered to be produced entirely by background noise. An application for these tests arises, for example, in underwater sonar, where we may have two sonar recordings obtained in different areas and/or at different times and we wish to determine if they have been produced by the same man-made object, for example a submarine.

Both tests described above were considered by Quinn (2006) for the case of discriminating between two univariate time series. Algorithms were given for computing estimators of the common frequencies under the null hypotheses and test statistics were derived using the pseudo-likelihood ratio approach described in previous chapters of this thesis. The noise processes were modelled as autoregressions and the autoregressive orders were estimated using information criteria (see Section 3.3.4). In this chapter we propose modified algorithms which incorporate the work of previous chapters, in particular around modelling the noise processes using fixed order autoregressive approximation (see Section 3.6). We derive the asymptotic properties of the estimators and extend the procedures for the case of discriminating between more than two time series. We then propose extensions for the case where the time series are multivariate, drawing on the work of Chapter 6. The results of simulation studies are presented which demonstrate the behaviour of the test statistics under the null hypotheses as well as the power of the tests in detecting differences in the fixed frequencies of two or more time series.

7.2 Discriminating Between Two Time Series With Periodic Components

Let $\{X_t\}$ and $\{Y_t\}$ be univariate processes which are made up of a deterministic periodic component and a stationary stochastic component. We model the periodic components as the sum of f_X and f_Y sinusoids, respectively and the stationary components as autoregressions of order p_X and p_Y, respectively. We therefore consider the models

$$X_t = \sum_{j=1}^{f_X} \left\{ \alpha_{X,j} \cos\left(\omega_{X,j} t\right) + \beta_{X,j} \sin\left(\omega_{X,j} t\right) \right\} + E_t \tag{7.1}$$

and

$$Y_t = \sum_{j=1}^{f_Y} \left\{ \alpha_{Y,j} \cos\left(\omega_{Y,j} t\right) + \beta_{Y,j} \sin\left(\omega_{Y,j} t\right) \right\} + U_t, \tag{7.2}$$

where

$$E_t + \sum_{j=1}^{p_X} \delta_{X,j} E_{t-1} = \varepsilon_t$$

and

$$U_t + \sum_{j=1}^{p_Y} \delta_{Y,j} U_{t-1} = u_t.$$

Note that we are assuming that $\{X_t\}$ and $\{Y_t\}$ have zero means and in practice we will mean correct the data. We make the usual assumptions on $\{\varepsilon_t\}$ and $\{u_t\}$, that is, that they are independent sequences of martingale differences and that

$$E\left(\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right) = \sigma_\varepsilon^2 \qquad \text{and} \qquad E\left(u_t^2 \mid \mathcal{G}_{t-1}\right) = \sigma_u^2,$$

where $\mathcal{F}_t$ and $\mathcal{G}_t$ are the σ-fields generated by $\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$ and $\{u_t, u_{t-1}, \ldots\}$, respectively.

The first null hypothesis we consider is that $\{X_t\}$ and $\{Y_t\}$ have sinusoids at the same frequencies and that $\{E_t\}$ and $\{U_t\}$ have the same spectral shape. This is equivalent to

$$H_0^{(1)} : \omega_{X,j} = \omega_{Y,j} \ \forall j, \ \delta_{X,k} = \delta_{Y,k} \ \forall k.$$

The second null hypothesis we consider is that $\{X_t\}$ and $\{Y_t\}$ have sinusoids at the same frequencies with no restriction on the spectral densities of $\{E_t\}$ and $\{U_t\}$. This is equivalent to

$$H_0^{(2)} : \omega_{X,j} = \omega_{Y,j} \ \forall j.$$

In both cases, the alternative hypothesis, H_A, is the complement of the null. Procedures will first be developed in order to test $H_0^{(1)}$ and then these will be modified for testing $H_0^{(2)}$, which is the simpler of the two. Therefore throughout the chapter, when we refer to the null hypothesis, we will mean $H_0^{(1)}$, unless otherwise specified.

The test statistics will be derived using the same pseudo-likelihood ratio procedure as in previous chapters. As in Chapter 3, we let $p_X = p_Y = p$ and derive test statistics, parameter estimators and their asymptotic properties for fixed p. In practice we let $p = \lfloor (\log T_{\min})^c \rfloor$, where $T_{\min} = \min(T_1, T_2)$ and $c > 1$. In this way, we do not need to assume that $\{E_t\}$ and $\{U_t\}$ truly are autoregressive and instead use long-order autoregressive approximation (see Section 3.6).

Let $d_\delta(z) = 1 + \sum_{j=1}^{p} \delta_j z^j$. Then $E_t = d_{\delta_X}^{-1}(z)\, \varepsilon_t$ and

$$d_{\delta_X}(z)\, X_t = \sum_{j=1}^{f_X} d_{\delta_X}(z) \left\{ \alpha_{X,j} \cos(\omega_{X,j} t) + \beta_{X,j} \sin(\omega_{X,j} t) \right\} + \varepsilon_t.$$

Since

$$\begin{aligned} d_{\delta_X}(z)\, e^{i\omega_j t} &= e^{i\omega_j t} + \sum_{j=1}^{p} \delta_{X,j} e^{i\omega_j (t-j)} \\ &= e^{i\omega_j t} d_{\delta_X}\left(e^{-i\omega_j}\right), \end{aligned}$$

(7.1) can be rewritten as

$$X_t + \sum_{j=1}^{p} \delta_{X,j} X_{t-j} = \sum_{j=1}^{f_X} \left\{ \widetilde{\alpha}_{X,j} \cos(\omega_{X,j} t) + \widetilde{\beta}_{X,j} \sin(\omega_{X,j} t) \right\} + \varepsilon_t,$$

where $\widetilde{\alpha}_{X,j}$ and $\widetilde{\beta}_{X,j}$ are related to $\alpha_{X,j}$ and $\beta_{X,j}$ by the identity

$$\widetilde{\alpha}_{X,j} + i\widetilde{\beta}_{X,j} = d_{\delta_X}\left(e^{i\omega}\right) \left(\alpha_{X,j} + i\beta_{X,j}\right),$$

$j = 1, \ldots, f_X$. Similarly, (7.2) can be rewritten as

$$Y_t + \sum_{j=1}^{p} \delta_{Y,j} Y_{t-j} = \sum_{j=1}^{f_Y} \left\{ \widetilde{\alpha}_{Y,j} \cos\left(\omega_{Y,j} t\right) + \widetilde{\beta}_{Y,j} \sin\left(\omega_{Y,j} t\right) \right\} + u_t,$$

where

$$\widetilde{\alpha}_{Y,j} + i\widetilde{\beta}_{Y,j} = d_{\delta_Y}\left(e^{i\omega}\right)\left(\alpha_{Y,j} + i\beta_{Y,j}\right),$$

$j = 1, \ldots, f_Y$. Let

$$X = \left[\begin{array}{ccc} X_0 & \cdots & X_{T_1 - 1} \end{array}\right], \qquad Y = \left[\begin{array}{ccc} Y_0 & \cdots & Y_{T_2 - 1} \end{array}\right],$$

$$\delta_X = \left[\begin{array}{ccc} \delta_{X,1} & \cdots & \delta_{X,p} \end{array}\right]', \qquad \delta_Y = \left[\begin{array}{ccc} \delta_{Y,1} & \cdots & \delta_{Y,p} \end{array}\right]',$$

$$\omega_X = \left[\begin{array}{ccc} \omega_{X,1} & \cdots & \omega_{X,f_X} \end{array}\right]', \qquad \omega_Y = \left[\begin{array}{ccc} \omega_{Y,1} & \cdots & \omega_{Y,f_Y} \end{array}\right]',$$

$$\theta_X = \left[\begin{array}{cccccc} \widetilde{\alpha}_{X,1} & \cdots & \widetilde{\alpha}_{X,f_X} & \widetilde{\beta}_{X,1} & \cdots & \widetilde{\beta}_{X,f_X} \end{array}\right]',$$

$$\theta_Y = \left[\begin{array}{cccccc} \widetilde{\alpha}_{Y,1} & \cdots & \widetilde{\alpha}_{Y,f_X} & \widetilde{\beta}_{Y,1} & \cdots & \widetilde{\beta}_{Y,f_Y} \end{array}\right]',$$

and $M_{T,f}(\omega)$ be the $(T-p) \times 2f$ matrix with $(t-p+1)$th row

$$\left[\begin{array}{cccccc} \cos\left(\omega_1 t\right) & \cdots & \cos\left(\omega_f t\right) & \sin\left(\omega_1 t\right) & \cdots & \sin\left(\omega_f t\right) \end{array}\right],$$

$t = p, \ldots, T-1$. The Gaussian log-likelihoods are then

$$l_X\left(\omega_X, \theta_X, \delta_X, \sigma_\varepsilon^2\right) = -\frac{T_1}{2}\log\left(2\pi\sigma_\varepsilon^2\right) - \frac{1}{2\sigma_\varepsilon^2} s_X\left(\omega_X, \theta_X, \delta_X,\right) \tag{7.3}$$

and

$$l_Y\left(\omega_Y, \theta_Y, \delta_Y, \sigma_u^2\right) = -\frac{T_1}{2}\log\left(2\pi\sigma_u^2\right) - \frac{1}{2\sigma_u^2} s_Y\left(\omega_Y, \theta_Y, \delta_Y\right), \tag{7.4}$$

where

$$s_X\left(\omega_X, \theta_X, \delta_X\right) = \left\{d_{\delta_X}(z) X - M_{T_1,f_X}\left(\omega_X\right)\theta_X\right\}' \left\{d_{\delta_X}(z) X - M_{T_1,f_X}\left(\omega_X\right)\theta_X\right\}$$

and

$$s_Y\left(\omega_Y, \theta_Y, \delta_Y\right) = \left\{d_{\delta_Y}(z) Y - M_{T_2,f_Y}\left(\omega_Y\right)\theta_Y\right\}' \left\{d_{\delta_Y}(z) Y - M_{T_2,f_Y}\left(\omega_Y\right)\theta_Y\right\}.$$

Under H_A, (7.3) and (7.4) can be maximised separately. The estimators of ω_X and θ_X which maximise (7.3) can be computed using the methods discussed in Section 6.2. The estimated sinusoids can be removed from the time series by regression and then δ_X and σ_ε^2 can be estimated by fitting autoregressions of order p to the residuals. The parameter estimators which maximise (7.4) can be computed in the same way. The maximised Gaussian log-likelihood is the sum of the individual ones, which is

$$\widehat{l}_A = -\frac{T_1+T_2}{2}\{1+\log(2\pi)\} - \frac{T_1}{2}\log\left(\widehat{\sigma}^2_{\varepsilon;A}\right) - \frac{T_2}{2}\log\left(\widehat{\sigma}^2_{u;A}\right), \tag{7.5}$$

where $\widehat{\sigma}^2_{\varepsilon;A}$ and $\widehat{\sigma}^2_{u;A}$ are the estimators of σ^2_ε and σ^2_u, respectively.

Under $H_0^{(1)}$, the frequencies and autoregressive parameters of the two processes are the same. Let $f = f_X = f_Y$, $\omega = \omega_X = \omega_Y$ and $\delta = \delta_X = \delta_Y$ denote the common parameters. The Gaussian log-likelihood is then

$$-\frac{T_1}{2}\log\left(2\pi\sigma^2_\varepsilon\right) - \frac{T_2}{2}\log\left(2\pi\sigma^2_u\right) - \frac{1}{2\sigma^2_\varepsilon}s_X(\omega,\theta_X,\delta) - \frac{1}{2\sigma^2_u}s_Y(\omega,\theta_Y,\delta). \tag{7.6}$$

Let $\widehat{\omega}$, $\widehat{\theta}_X$, $\widehat{\theta}_Y$ and $\widehat{\delta}$ be the maximisers of (7.6) with respect to ω, θ_X, θ_Y and δ, respectively. The maximisers of (7.6) with respect to σ^2_ε and σ^2_u are then

$$\widehat{\sigma}^2_{\varepsilon;0} = T_1^{-1}s_X\left(\widehat{\omega}_X,\widehat{\theta}_X,\widehat{\delta}\right) \qquad \text{and} \qquad \widehat{\sigma}^2_{u,0} = T_2^{-1}s_Y\left(\widehat{\omega}_Y,\widehat{\theta}_Y,\widehat{\delta}\right),$$

respectively. Therefore the maximised Gaussian log-likelihood under the null hypothesis is

$$\widehat{l}_0 = -\frac{T_1+T_2}{2}\{1+\log(2\pi)\} - \frac{T_1}{2}\log\left(\widehat{\sigma}^2_{\varepsilon;0}\right) - \frac{T_2}{2}\log\left(\widehat{\sigma}^2_{u,0}\right). \tag{7.7}$$

Methods for computing $\widehat{\omega}$, $\widehat{\theta}_X$, $\widehat{\theta}_Y$ and $\widehat{\delta}$ are given in Section 7.3.

The test statistic is $2\left(\widehat{l}_A - \widehat{l}_0\right)$ which, from (7.5) and (7.7), is

$$\Lambda_f = T_1\log\left(\frac{\widehat{\sigma}^2_{\varepsilon;0}}{\widehat{\sigma}^2_{\varepsilon;A}}\right) + T_2\log\left(\frac{\widehat{\sigma}^2_{u;0}}{\widehat{\sigma}^2_{u;A}}\right).$$

It will be shown that the asymptotic distribution of Λ_f under $H_0^{(1)}$ is chi-squared with $p+f$ degrees of freedom. Therefore $H_0^{(1)}$ is rejected at significance level α when Λ is greater than the $100(1-\alpha)$th percentile of the χ^2 distribution with $p+f$ degrees of freedom.

7.3 Parameter Estimation Under the Null Hypothesis

7.3.1 Estimating a Single Common Frequency

We begin by considering the case where $f = 1$, that is where

$$X_t = \alpha_X\cos(\omega t) + \beta_X\sin(\omega t) + E_t \tag{7.8}$$

and

$$Y_t = \alpha_Y\cos(\omega t) + \beta_Y\sin(\omega t) + U_t. \tag{7.9}$$

As in the case of a single time series, we will motivate the estimation techniques by first considering the case where $\{E_t\}$ and $\{U_t\}$ are Gaussian and white. Putting $p = 0$, the conditional Gaussian log-likelihood is

$$-\frac{T_1}{2}\log\left(2\pi\sigma_\varepsilon^2\right) - \frac{T_2}{2}\log\left(2\pi\sigma_u^2\right) - \frac{1}{2\sigma_\varepsilon^2}\left\{X - M_{T_1,1}(\omega)\,\theta_X\right\}'\left\{X - M_{T_1,1}(\omega)\,\theta_X\right\}$$
$$-\frac{1}{2\sigma_u^2}\left\{Y - M_{T_2,1}(\omega)\,\theta_Y\right\}'\left\{Y - M_{T_2,1}(\omega)\,\theta_Y\right\}. \tag{7.10}$$

For fixed ω, the maximisers of (7.10) with respect to θ_X and θ_Y have the same asymptotic properties as

$$\widehat{\theta}_X(\omega) = \begin{bmatrix} 2T_1^{-1}\sum_{t=0}^{T_1-1} X_t\cos(\omega t) \\ 2T_1^{-1}\sum_{t=0}^{T_1-1} X_t\sin(\omega t) \end{bmatrix} \quad \text{and} \quad \widehat{\theta}_Y(\omega) = \begin{bmatrix} 2T_2^{-1}\sum_{t=0}^{T_2-1} Y_t\cos(\omega t) \\ 2T_2^{-1}\sum_{t=0}^{T_2-1} Y_t\sin(\omega t) \end{bmatrix},$$

respectively, from (6.6) and (6.7). It follows that the maximisers of (7.10) with respect to σ_ε^2 and σ_u^2 have the same asymptotic properties as

$$\widehat{\sigma}_\varepsilon^2(\omega) = T_1^{-1}\left\{\sum_{t=0}^{T_1-1} X_t^2 - I_{T_1,X}(\omega)\right\} \quad \text{and} \quad \widehat{\sigma}_u^2(\omega) = T_2^{-1}\left\{\sum_{t=0}^{T_2-1} Y_t^2 - I_{T_2,Y}(\omega)\right\},$$

respectively. Thus the maximiser of (7.10) with respect to ω has the same asymptotic properties as the maximiser of

$$\widetilde{l}_0(\omega) = -\frac{T_1}{2}\log\left\{\widehat{\sigma}_\varepsilon^2(\omega)\right\} - \frac{T_2}{2}\log\left\{\widehat{\sigma}_u^2(\omega)\right\}. \tag{7.11}$$

The asymptotic theory below shows that the maximiser of $\widetilde{l}_0(\omega)$ is a strongly consistent estimator of the common frequency in (7.8) and (7.9). That is, we can maximise the likelihood as though $\{E_t\}$ and $\{U_t\}$ are Gaussian and white and the resulting estimator will be strongly consistent even if that is not the case. We will also establish the central limit theorem for the estimator. Note that for the central limit theorem, we need to make the assumption that $T_2 = \kappa T_1$ for some constant κ.

Let

$$\widehat{\omega} = \arg\max_{\omega} \widetilde{l}_0(\omega).$$

Also let $f_E(\omega)$ and $f_U(\omega)$ be the spectral densities of $\{E_t\}$ and $\{U_t\}$, respectively, and

$$\rho_X^2 = \alpha_X^2 + \beta_X^2 \qquad \text{and} \qquad \rho_Y^2 = \alpha_Y^2 + \beta_Y^2.$$

In the theorems below and their proofs, a parameter written with a 0 in the subscript will denote the true value of that parameter. The proofs of the theorems are in the Appendix.

Theorem 7.1 $T_{\min}(\widehat{\omega} - \omega_0) \to 0$ *almost surely as* $T_1, T_2 \to \infty$.

Theorem 7.2 *Let $\kappa = T_2/T_1$. Then the distribution of $T_1^{3/2}(\widehat{\omega} - \omega_0)$ converges to the normal distribution with mean zero and variance*

$$48\pi \left\{ \frac{\rho_{X0}^2}{f_E(\omega_0)} + \frac{\kappa^3 \rho_{Y0}^2}{f_U(\omega_0)} \right\}^{-1}$$

as $T_1 \to \infty$.

In practice, we can maximise $\widetilde{l}_0(\omega)$ using the Gauss–Newton algorithm by modifying the procedure given in Section 6.2.3. Alternatively, we can estimate ω using a generalisation of the Quinn–Fernandes technique which was proposed by Quinn (2006). We describe these two methods below and also show that the estimator from the Quinn–Fernandes technique is strongly consistent and follows the same central limit theorem as $\widehat{\omega}$. For both methods we introduce the parameter $\lambda = \sigma_\varepsilon^2/\sigma_u^2$ and show how to maximise $\widetilde{l}_0(\omega)$ for a given λ. In the algorithms which follow we incorporate the estimation of λ.

Estimating the Common Frequency Using the Gauss–Newton Algorithm

Following the method of Section 6.2.3, we reparametrise (7.8) and (7.9) and let

$$X = M_{T_1}^*(\omega)\,\theta_X^* + \varepsilon$$

and

$$Y = M_{T_2}^*(\omega)\,\theta_Y^* + u,$$

where $\theta_X^* = \begin{bmatrix} \alpha_X^* & \beta_X^* \end{bmatrix}'$, $\theta_Y^* = \begin{bmatrix} \alpha_Y^* & \beta_Y^* \end{bmatrix}'$ and $M_T^*(\omega)$ is the $T \times 3$ matrix with $(t+1)$th row given by (6.15). Also let

$$e_X(\omega) = X - M_{T_1}^*(\omega)\,\widehat{\theta}_X^*(\omega) \qquad \text{and} \qquad e_Y(\omega) = Y - M_{T_2}^*(\omega)\,\widehat{\theta}_Y^*(\omega),$$

where

$$\widehat{\theta}_X^*(\omega) = \left\{ M_{T_1}^{*\prime}(\omega)\,M_{T_1}^*(\omega) \right\}^{-1} \left\{ M_{T_1}^{*\prime}(\omega)\,X \right\}$$

and

$$\widehat{\theta}_Y^*(\omega) = \left\{ M_{T_2}^{*\prime}(\omega)\,M_{T_2}^*(\omega) \right\}^{-1} \left\{ M_{T_2}^{*\prime}(\omega)\,Y \right\}.$$

The estimator of ω is the minimiser of

$$S(\omega) = R_X(\omega) + \lambda R_Y(\omega),$$

where

$$R_X(\omega) = e_X'(\omega)\,e_X(\omega) \qquad \text{and} \qquad R_Y(\omega) = e_Y'(\omega)\,e_Y(\omega).$$

The derivative of $S(\omega)$ is

$$\frac{d}{d\omega}S(\omega) = 2e_X'(\omega)\frac{d}{d\omega}e_X(\omega) + 2\lambda e_Y'(\omega)\frac{d}{d\omega}e_Y(\omega).$$

Thus, given λ and a current estimate of ω, denoted $\widetilde{\omega}$, the Gauss–Newton algorithm updates the estimate by

$$\widetilde{\omega} - \frac{e_X'(\widetilde{\omega})\frac{d}{d\omega}e_X(\widetilde{\omega}) + \lambda e_Y'(\widetilde{\omega})\frac{d}{d\omega}e_Y(\widetilde{\omega})}{\frac{d}{d\omega}e_X'(\widetilde{\omega})\frac{d}{d\omega}e_X(\widetilde{\omega}) + \lambda\frac{d}{d\omega}e_Y'(\widetilde{\omega})\frac{d}{d\omega}e_Y(\widetilde{\omega})}$$

and repeats until convergence. The functions $e_X(\omega)$, $e_Y(\omega)$ and their derivatives can be computed using the methods given in Section 6.2.3. For an initial estimate we use the maximiser of

$$I_{T^*,X}(\omega) + \lambda I_{T^*,Y}(\omega)$$

over the Fourier frequencies, where $T^* = 4\max(T_1, T_2)$. That is, the periodograms are computed by first zero-padding the time series to four times the length of the longest series.

Estimating the Common Frequency Using the Quinn–Fernandes Technique

Quinn (2006) proposed a generalisation to the Quinn–Fernandes technique to estimate the common frequency from two time series with equal variances. Algorithm 7.1 presents a modified version of this for the case where the variances of the time series are not necessarily the same but where their ratio is known. That is, the algorithm estimates ω for a given λ. For an initial estimate of ω, we use the maximiser of

$$I_{T,X}(\omega) + \lambda I_{T,Y}(\omega)$$

over the Fourier frequencies.

The asymptotic theory below shows that the estimator produced by Algorithm 7.1 is strongly consistent and follows the same central limit theorem as the maximiser of $\widetilde{l}_0(\omega)$. The procedure for establishing the asymptotic properties follows closely that of the corresponding theorems in Quinn and Hannan (2001) for the case of a single time series. Note that we make the assumption that $T_2 = \kappa T_1$.

Let

$$h_{T_1,T_2}(a) = \frac{\sum_{t=0}^{T_1-1} X_t\xi_{t-1}(a) + \lambda\sum_{t=0}^{T_2-1} Y_t\eta_{t-1}(a)}{\sum_{t=0}^{T_1-1}\xi_{t-1}^2(a) + \lambda\sum_{t=0}^{T_2-1}\eta_{t-1}^2(a)},$$

where

$$\xi_t(a) = a\xi_{t-1}(a) - \xi_{t-2}(a) + X_t \qquad \text{and} \qquad \eta_t(a) = a\eta_{t-1}(a) - \eta_{t-2}(a) + Y_t.$$

Also let $a_{j+1} = a_j + 2h_{T_1,T_2}(a_j)$, $\omega_j = \cos^{-1}(a_j/2)$ and $A_{T_1}(\nu) = \left\{a : |a - a_0| < cT_1^{-\nu}\right\}$, where $a_0 = 2\cos(\omega_0)$. Theorem 7.3 shows that the sequence $\{a_j\}$ converges to a unique

Algorithm 7.1 Estimating a common frequency from two time series using the Quinn–Fernandes technique for a given λ

1. Put $\widehat{a} = 2\cos(\widehat{\omega})$, where $\widehat{\omega}$ is an initial estimate of ω.
2. For $t = 0, \ldots, T_1 - 1$, put
$$\xi_t = \widehat{a}\xi_{t-1} - \xi_{t-2} + X_t,$$
where $\xi_{-1} = \xi_{-2} = 0$, and for $t = 0, \ldots, T_2 - 1$, put
$$\eta_t = \widehat{a}\eta_{t-1} - \eta_{t-2} + Y_t,$$
where $\eta_{-1} = \eta_{-2} = 0$.
3. Replace $\widehat{a}$ by $\widehat{a} + \nu$, where
$$\nu = 2\frac{\sum_{t=0}^{T_1-1} X_t\xi_{t-1} + \lambda\sum_{t=0}^{T_2-1} Y_t\eta_{t-1}}{\sum_{t=0}^{T_1-1} \xi_{t-1}^2 + \lambda\sum_{t=0}^{T_2-1} \eta_{t-1}^2}.$$
4. Repeat steps 2 and 3 until $|\nu|$ converges to 0.
5. Put $\widehat{\omega} = \cos^{-1}(\widehat{a}/2)$.

point $\widehat{a}_{T_1} \in A_{T_1}(\nu)$ which is such that $h_{T_1,T_2}(\widehat{a}_{T_1}) = 0$. It follows that there is a unique point, $\widehat{\omega}_{T_1}$, such that $T_1^{\nu}(\widehat{\omega}_{T_1} - \omega_0) \to 0$ almost surely for any $\nu < 3/2$. Theorem 7.4 shows how many iterates are needed to converge to the fixed point and that, provided the initial estimator is within $O\left(T_1^{-1}\right)$ of ω_0, only two iterates are needed. Theorem 7.5 establishes the central limit theorem for the estimator. The proofs of the theorems are in the Appendix.

Theorem 7.3 *Let $\kappa = T_2/T_1$ and let $A_{T_1}(\nu) = \left\{a : |a - a_0| < cT_1^{-\nu}\right\}$, for fixed constant $c > 0$. Then there exists a unique point $\widehat{a}_{T_1} \in A_{T_1}(\nu)$ such that $h_{T_1,T_2}(\widehat{a}_{T_1}) = 0$ for $1 < \nu < 3/2$. Thus there is a unique solution to $h_{T_1}(a_0) = 0$ for which $T_1^{\nu}(\widehat{\omega}_{T_1} - \omega_0) \to 0$ almost surely as $T_1 \to \infty$ for all $\nu < 3/2$, where $\widehat{\omega}_{T_1} = \cos^{-1}(\widehat{a}_{T_1}/2)$.*

Theorem 7.4 *Let $a_1 \in A_{T_1}(\nu)$. If $1 < \nu < 3/2$ then*
$$a_{j+1} - \widehat{a}_{T_1} = (a_j - \widehat{a}_{T_1})\, O\left\{T_1^{-1/2}(\log T_1)^{1/2}\right\},$$
while if $1/2 < \nu \leqslant 1$ then
$$a_{j+1} - \widehat{a}_{T_1} = (a_j - \widehat{a}_{T_1})\, O\left\{T_1^{1/2-\nu}(\log T_1)^{1/2}\right\} + O\left\{T_1^{1/2-2\nu}(\log T_1)^{1/2}\right\}.$$
Also
$$\widehat{a}_{T_1} - a_0 = O\left\{T_1^{-3/2}(\log T_1)^{1/2}\right\}$$

and

$$a_k - \widehat{a}_{T_1} = o\left(T_1^{-3/2}\right)$$

for

$$k \geqslant \lfloor 3 - \log(2\nu - 1) / \log 2 \rfloor .$$

Theorem 7.5 *Let $\kappa = T_2/T_1$. Then the distribution of $T_1^{3/2}\left(\widehat{\omega}_{T_1} - \omega_0\right)$ converges to the normal distribution with mean zero and variance*

$$48\pi \left\{ \frac{\rho_{X0}^2}{f_E(\omega_0)} + \frac{\kappa^3 \rho_{Y0}^2}{f_U(\omega_0)} \right\}^{-1}$$

as $T_1 \to \infty$.

7.3.2 Estimating More Than One Common Frequency

When $f > 1$ we have, under $H_0^{(1)}$,

$$X_t = \sum_{j=1}^{f} \left\{ \alpha_{X,j} \cos(\omega_j t) + \beta_{X,j} \sin(\omega_j t) \right\} + E_t$$

and

$$Y_t = \sum_{j=1}^{f} \left\{ \alpha_{Y,j} \cos(\omega_j t) + \beta_{Y,j} \sin(\omega_j t) \right\} + U_t .$$

Following the same calculations as above, the common frequencies, $\omega_1, \ldots, \omega_f$, may be estimated by the f greatest maximisers of $\widetilde{l}_0(\omega)$, ignoring sidelobes, where now

$$\widehat{\sigma}_\varepsilon^2(\omega) = T_1^{-1} \left\{ \sum_{t=0}^{T_1-1} X_t^2 - \sum_{j=1}^{f} I_{T_1,X}(\omega_j) \right\}$$

and

$$\widehat{\sigma}_u^2(\omega) = T_2^{-1} \left\{ \sum_{t=0}^{T_2-1} Y_t^2 - \sum_{j=1}^{f} I_{T_2,Y}(\omega_j) \right\} .$$

It follows that we can estimate the frequencies one at a time, just as in the case of a single time series. That is, after each frequency is estimated, the sinusoid at that frequency is removed from both time series by regression and the next frequency is estimated from the residuals.

In order to estimate the number of frequencies, we can use an information criterion along the lines of that given in Section 6.2.4. In this case, the estimation of f frequencies involves estimators of $2f$ parameters, from θ_X, which have asymptotic standard error of $O\left(T_1^{-1/2}\right)$, $2f$ parameters, from θ_Y, which have asymptotic standard error of $O\left(T_2^{-1/2}\right)$, p parameters, from δ, which have asymptotic standard error of $O\left\{(T_1 + T_2)^{-1/2}\right\}$ and f parameters, $\omega_1, \ldots, \omega_f$, which have asymptotic standard error of $O\left\{(T_1+T_2)^{-3/2}\right\}$. We therefore use the information criterion

$$\phi_0(f) = T_1 \log\left\{\widehat{\sigma}^2_{\varepsilon;0}(f)\right\} + T_2 \log\left\{\widehat{\sigma}^2_{u;0}(f)\right\} + (p+7f)\log(T_1+T_2),$$

where $\widehat{\sigma}^2_{\varepsilon;0}(f)$ and $\widehat{\sigma}^2_{u;0}(f)$ are the estimators of σ^2_ε and σ^2_u, respectively, obtained by fitting autoregressions with common autoregressive parameters to the time series after removing f common sinusoids by regression. The estimated number of frequencies is then the minimiser of $\phi_0(f)$ over $f = 0, \ldots, F$, where F is assumed to be greater than the true number of frequencies.

7.4 The Test Statistic

The asymptotic distribution of the test statistic under $H_0^{(1)}$, for a given f, is given in Theorem 7.6. The proof of the theorem is in the Appendix.

Theorem 7.6 *Under $H_0^{(1)}$, the distribution of Λ_f converges to the χ^2 distribution with $p+f$ degrees of freedom as $T_1, T_2 \to \infty$.*

The full procedure for computing the parameter estimators under both H_A and $H_0^{(1)}$ for a given f is given in Algorithm 7.2. Note that at each step in the algorithm where frequencies are estimated, under both the null and alternative hypotheses, the sinusoids at these frequencies are removed from the time series by regression, and an autoregression of order p is fitted to the residuals. The frequencies are then re-estimated on the time series filtered by the autoregressive parameter estimates. This strengthens the algorithm particularly if a root of the auxiliary equation is close to the unit circle or if the sample sizes are small. Note also that the frequencies can be estimated either by maximising the Gaussian white log-likelihood using the Gauss–Newton algorithm or by using the Quinn–Fernandes technique. As shown above, the estimators are asymptotically equivalent, and even in small samples simulations suggest the results will only differ negligibly. It is important, however, that the same method is chosen for estimating the parameters under both the null and alternative hypotheses. In practice, the Quinn–Fernandes technique is computationally faster and so is generally preferred.

The test statistic is computed by applying Algorithm 7.2 for each $f = 0, \ldots, F$ and calculating

$$\widehat{f}_0 = \arg\min_f \phi_0(f)$$

Algorithm 7.2 Computing the parameter estimators under $H_0^{(1)}$ and H_A for a given f

1. Use Algorithm 6.1 to fit f sinusoids to $\{X_t\}$. Remove the estimated sinusoids by regression and fit an autoregression of order p to the residuals, denoting the parameter estimates by $\widehat{\delta}_X$ and $\widehat{\sigma}^2_{\varepsilon;A}(f)$. Repeat using $\left\{d_{\widehat{\delta}_X}(z)\,X_t\right\}$ in place of $\{X_t\}$.
2. Use Algorithm 6.1 to fit f sinusoids to $\{Y_t\}$. Remove the estimated sinusoids by regression and fit an autoregression of order p to the residuals, denoting the parameter estimates by $\widehat{\delta}_Y$ and $\widehat{\sigma}^2_{u;A}(f)$. Repeat using $\left\{d_{\widehat{\delta}_Y}(z)\,Y_t\right\}$ in place of $\{Y_t\}$.
3. Put $\widehat{\lambda} = \widehat{\sigma}^2_{\varepsilon;A}(f)/\widehat{\sigma}^2_{u;A}(f)$, $\left\{\widehat{E}_t\right\} = \{X_t\}$ and $\left\{\widehat{U}_t\right\} = \{Y_t\}$.
4. Estimate a common frequency using the methods of Section 7.3.1 to $\left\{\widehat{E}_t\right\}$ and $\left\{\widehat{U}_t\right\}$ with $\lambda = \widehat{\lambda}$. Remove the corresponding sinusoid from both time series using regression and denote the residuals by $\left\{\widehat{E}_t\right\}$ and $\left\{\widehat{U}_t\right\}$.
5. Repeat step 4 until f common frequencies have been estimated.
6. Use the methods of Chapter 3 to fit autoregressions of order p with the same autoregressive parameters to $\{X_t\}$ and $\{Y_t\}$ with $\lambda = \widehat{\lambda}$, denoting the parameter estimates by $\widehat{\delta}_{\widehat{\lambda}}$ and $\widetilde{\sigma}^2_{\varepsilon;\widehat{\lambda}}$. Let $\left\{\widehat{E}_t\right\} = \left\{d_{\widehat{\delta}_{\widehat{\lambda}}}(z)\,X_t\right\}$ and $\left\{\widehat{U}_t\right\} = \left\{d_{\widehat{\delta}_{\widehat{\lambda}}}(z)\,Y_t\right\}$.
7. Repeat steps 4–6 once.
8. Update $\widehat{\lambda}$ by
$$\frac{T_2\widetilde{\sigma}^2_{\varepsilon;\widehat{\lambda}}}{\sum_{t=p}^{T_2-1}\left\{d_{\widehat{\delta}_{\widehat{\lambda}}}(z)\,Y_t\right\}^2}$$
and put $\left\{\widehat{E}_t\right\} = \{X_t\}$ and $\left\{\widehat{U}_t\right\} = \{Y_t\}$.
9. Repeat steps 4–8 until $\widehat{\lambda}$ converges.
10. Put $\widehat{\sigma}^2_{\varepsilon;0}(f) = \widetilde{\sigma}^2_{\varepsilon;\widehat{\lambda}}$ and $\widehat{\sigma}^2_{u;0}(f) = \widetilde{\sigma}^2_{\varepsilon;\widehat{\lambda}}/\widehat{\lambda}$.

as well as

$$\widehat{f}_X = \arg\min_f \phi_X(f) \qquad \text{and} \qquad \widehat{f}_Y = \arg\min_f \phi_Y(f),$$

where

$$\phi_X(f) = T_1 \log\left\{\widehat{\sigma}^2_{\varepsilon;A}(f)\right\} + (p+5f)\log T_1$$

and

$$\phi_Y(f) = T_2 \log\left\{\widehat{\sigma}^2_{u;A}(f)\right\} + (p+5f)\log T_2.$$

The test statistic is then

$$\Lambda_{\widehat{f}_0,\widehat{f}_X,\widehat{f}_Y} = T_1 \log\left\{\frac{\widehat{\sigma}^2_{\varepsilon;0}\left(\widehat{f}_0\right)}{\widehat{\sigma}^2_{\varepsilon;A}\left(\widehat{f}_X\right)}\right\} + T_2 \log\left\{\frac{\widehat{\sigma}^2_{u;0}\left(\widehat{f}_0\right)}{\widehat{\sigma}^2_{u;A}\left(\widehat{f}_Y\right)}\right\}$$

and $H_0^{(1)}$ is rejected at significance level α when $\Lambda_{\widehat{f}_0,\widehat{f}_X,\widehat{f}_Y}$ is greater than the $100(1-\alpha)$th percentile of the χ^2 distribution with $p+\widehat{f}_0$ degrees of freedom.

7.5 Discriminating Between the Fixed Frequencies Only

Under $H_0^{(2)}$, the autoregressive orders of $\{E_t\}$ and $\{U_t\}$ are not necessarily the same. Furthermore, the estimators of σ_ε^2 and σ_u^2 are calculated by fitting independent autoregressions to the residuals obtained by removing f common frequencies from $\{X_t\}$ and $\{Y_t\}$, respectively. That is, δ_X and δ_Y are not necessarily the same. The procedure for computing the parameter estimators under both H_A and $H_0^{(2)}$ for a given f is given in Algorithm 7.3. The algorithm assumes that the autoregressive orders of $\{X_t\}$ and $\{Y_t\}$ are fixed at p_X and p_Y, respectively. In practice we let $p_X = \lfloor (\log T_1)^{c_1} \rfloor$ and $p_Y = \lfloor (\log T_2)^{c_2} \rfloor$ where $c_1, c_2 > 1$.

Algorithm 7.3 Computing the parameter estimators under $H_0^{(2)}$ and H_A for a given f

1. Use Algorithm 6.1 to fit f sinusoids to $\{X_t\}$. Remove the estimated sinusoids by regression and fit an autoregression of order p_X to the residuals, denoting the parameter estimates by $\widehat{\delta}_X$ and $\widehat{\sigma}^2_{\varepsilon;A}(f)$. Repeat using $\left\{d_{\widehat{\delta}_X}(z)\,X_t\right\}$ in place of $\{X_t\}$.
2. Use Algorithm 6.1 to fit f sinusoids to $\{Y_t\}$. Remove the estimated sinusoids by regression and fit an autoregression of order p_Y to the residuals, denoting the parameter estimates by $\widehat{\delta}_Y$ and $\widehat{\sigma}^2_{u;A}(f)$. Repeat using $\left\{d_{\widehat{\delta}_Y}(z)\,Y_t\right\}$ in place of $\{Y_t\}$.
3. Put $\widehat{\lambda} = \widehat{\sigma}^2_{\varepsilon;A}(f)/\widehat{\sigma}^2_{u;A}(f)$, $\left\{\widehat{E}_t\right\} = \{X_t\}$ and $\left\{\widehat{U}_t\right\} = \{Y_t\}$.
4. Estimate a common frequency by applying the methods of Section 7.3.1 to $\left\{\widehat{E}_t\right\}$ and $\left\{\widehat{U}_t\right\}$ with $\lambda = \widehat{\lambda}$. Remove the corresponding sinusoid from both time series using regression and denote the residuals by $\left\{\widehat{E}_t\right\}$ and $\left\{\widehat{U}_t\right\}$.
5. Repeat step 4 until f common frequencies have been estimated.
6. Fit an autoregression of order p_X to $\left\{\widehat{E}_t\right\}$, denoting the parameter estimates by $\widehat{\delta}_X$ and $\widehat{\sigma}^2_{\varepsilon;0}(f)$, and fit an autoregression of order p_Y to $\left\{\widehat{U}_t\right\}$, denoting the parameter estimates by $\widehat{\delta}_Y$ and $\widehat{\sigma}^2_{u;0}(f)$. Let $\left\{\widehat{E}_t\right\} = \left\{d_{\widehat{\delta}_X}(z)\,X_t\right\}$ and $\left\{\widehat{U}_t\right\} = \left\{d_{\widehat{\delta}_Y}(z)\,Y_t\right\}$.
7. Repeat steps 4–6 once.
8. Update $\widehat{\lambda}$ by $\widehat{\sigma}^2_{\varepsilon;0}(f)/\widehat{\sigma}^2_{u;0}(f)$ and put $\left\{\widehat{E}_t\right\} = \{X_t\}$ and $\left\{\widehat{U}_t\right\} = \{Y_t\}$.
9. Repeat steps 4–8 until $\widehat{\sigma}^2_{\varepsilon;0}(f)$ and $\widehat{\sigma}^2_{u;0}(f)$ converge.

The test statistic is computed by applying Algorithm 7.3 for each $f = 0, \ldots, F$ and calculating

$$\widehat{f}_0 = \arg\min_f \phi_0^{(2)}(f)$$

as well as

$$\widehat{f}_X = \arg\min_f \phi_X(f) \qquad \text{and} \qquad \widehat{f}_Y = \arg\min_f \phi_Y(f),$$

where

$$\phi_0^{(2)}(f) = T_1 \log\left\{\widehat{\sigma}^2_{\varepsilon;A}(f)\right\} + T_2 \log\left\{\widehat{\sigma}^2_{u;A}(f)\right\} + p_X \log T_1 + p_Y \log T_2 + 7f \log(T_1 + T_2).$$

The test statistic is then

$$\Lambda_{\widehat{f}_0,\widehat{f}_X,\widehat{f}_Y} = T_1 \log\left\{\frac{\widehat{\sigma}^2_{\varepsilon;0}\left(\widehat{f}_0\right)}{\widehat{\sigma}^2_{\varepsilon;A}\left(\widehat{f}_X\right)}\right\} + T_2 \log\left\{\frac{\widehat{\sigma}^2_{u;0}\left(\widehat{f}_0\right)}{\widehat{\sigma}^2_{u;A}\left(\widehat{f}_Y\right)}\right\}$$

and $H_0^{(2)}$ is rejected at significance level α when $\Lambda_{\widehat{f}_0,\widehat{f}_X,\widehat{f}_Y}$ is greater than the $100(1-\alpha)$th percentile of the χ^2 distribution with $\widehat{f}_0$ degrees of freedom.

7.6 Comparing More Than Two Time Series

The methods above are easily extended to the case where we wish to compare more than two time series with periodic components. Suppose we have samples of size T_k from $\{X_{k,t}\}$, $k = 1, \ldots, n$. We fit the models

$$X_{k,t} = \sum_{j=1}^{f_k} \left\{\alpha_{k,j} \cos(\omega_{k,j} t) + \beta_{k,j} \sin(\omega_{k,j} t)\right\} + E_{k,t},$$

where

$$E_{k,t} + \sum_{j=1}^{p_k} \delta_{k,j} E_{k,t-1} = \varepsilon_{k,t}.$$

We make the usual assumptions on the noise processes, that is that they are independent sequences of martingale differences and that

$$E\left(\varepsilon^2_{k,t} \mid \mathcal{F}_{k,t-1}\right) = \sigma^2_k,$$

$k = 1, \ldots, n$, where $\mathcal{F}_{k,t}$ is the σ-field generated by $\{\varepsilon_{k,t}, \varepsilon_{k,t-1}, \ldots\}$. The first null hypothesis of interest is that the processes have sinusoids at the same frequencies and that the noise processes have the same spectral shape. This is equivalent to

$$H_0^{(1)} : \omega_{1,j} = \cdots = \omega_{n,j},\ \forall j,\ \delta_{1,k} = \cdots = \delta_{n,k},\ \forall k.$$

The second null hypothesis of interest is that the processes have sinusoids at the same frequencies with no restriction on the spectral densities of the noise processes. This is equivalent to

$$H_0^{(2)} : \omega_{1,j} = \cdots = \omega_{n,j},\ \forall j.$$

In each case the alternative hypothesis, H_A, is the complement of the null hypothesis. Under both H_A, the n time series are independent and their parameters can be estimated separately using the methods discussed in Section 6.2.

Under $H_0^{(1)}$, we derive the test procedure letting $p_1 = \cdots = p_n = p$, where p is fixed. In practice we let $p = \lfloor \log (T_{\min})^c \rfloor$, where $T_{\min} = \min (T_1, \ldots, T_n)$ and $c > 1$. We need to estimate common frequencies from n time series. Suppose $f = 1$ and denote the common frequency by ω. The Gaussian log-likelihood is

$$-\sum_{j=1}^{n} \left[\frac{T_j}{2} \log \left(2\pi\sigma_j^2 \right) + \frac{1}{2\sigma_j^2} \left\{ X_j - M_{T_j,1} (\omega) \theta_j \right\}' \left\{ X_j - M_{T_j,1} (\omega) \theta_j \right\} \right], \tag{7.12}$$

where

$$X_k = \left[\begin{array}{ccc} X_{k,0} & \cdots & X_{k,T_k - 1} \end{array} \right],$$

$$\theta_k = \left[\begin{array}{cc} \widetilde{\alpha}_k & \widetilde{\beta}_k \end{array} \right]'$$

and

$$\widetilde{\alpha}_k + i\widetilde{\beta}_k = d_{\delta_k} \left(e^{i\omega} \right) \left(\alpha_k + i\beta_k \right),$$

$k = 1, \ldots, n$. Letting $\lambda_1 = 1$ and $\lambda_k = \widehat{\sigma}_1^2 / \widehat{\sigma}_k^2$, $k = 2, \ldots, n$, we can rewrite (7.12) as

$$-\sum_{j=1}^{n} \left[\frac{T_j}{2} \log \left(2\pi\lambda_j\sigma_1^2 \right) + \frac{1}{2\lambda_j\sigma_1^2} \left\{ X_j - M_{T_j,1} (\omega) \theta_j \right\}' \left\{ X_j - M_{T_j,1} (\omega) \theta_j \right\} \right].$$

This can be maximised with respect to ω, for given $\lambda_1, \ldots, \lambda_n$, using the Gauss–Newton algorithm as follows. Let

$$e_k (\omega) = X_k - M_{T_k}^* (\omega) \widehat{\theta}_k^* (\omega),$$

where

$$\widehat{\theta}_k^* (\omega) = \left\{ M_{T_k}^{*\prime} (\omega) M_{T_k}^* (\omega) \right\}^{-1} \left\{ M_{T_k}^{*\prime} (\omega) X_k \right\}.$$

Given a current estimate of ω, denoted $\widetilde{\omega}$, the Gauss–Newton method updates the estimate by

$$\widetilde{\omega} - \frac{\sum_{j=1}^{n} \lambda_j e_j' (\widetilde{\omega}) \frac{d}{d\omega} e_j' (\widetilde{\omega})}{\sum_{j=1}^{n} \lambda_j \frac{d}{d\omega} e_j' (\widetilde{\omega}) \frac{d}{d\omega} e_j (\widetilde{\omega})},$$

where $e_k (\widetilde{\omega})$, $k = 1, \ldots, n$, and their derivatives can be computed using the methods given in Section 7.3.1. An initial estimate can be obtained by maximising

$$\sum_{j=1}^{n} \lambda_j I_{T^*, X_j} (\omega)$$

over the Fourier frequencies, where $T^* = 4 \max (T_1, \ldots, T_n)$.

An alternative method for estimating ω is to use an extension of the Quinn–Fernandes technique given in Section 7.3.1. Algorithm 7.4 estimates ω for given $\lambda_1, \ldots, \lambda_n$. The algorithm can be initialised by maximising

$$\sum_{j=1}^{n} \lambda_j I_{T,X_j}(\omega)$$

over the Fourier frequencies.

Algorithm 7.4 Estimating a common frequency from more than two time series for given $\lambda_1, \ldots, \lambda_n$ using the Quinn–Fernandes technique

1. Put $\widehat{a} = 2\cos(\widehat{\omega})$, where $\widehat{\omega}$ is an initial estimate of ω.
2. For $k = 1, \ldots, n$, $t = 0, \ldots, T_k - 1$, put
$$\xi_{k,t} = \widehat{a}\xi_{k,t-1} - \xi_{k,t-2} + X_{k,t},$$
where $\xi_{k,-1} = \xi_{k,-2} = 0$.
3. Replace $\widehat{a}$ by $\widehat{a} + \nu$, where
$$\nu = 2\frac{\sum_{j=1}^{n} \lambda_j \sum_{t=0}^{T_j-1} X_{j,t}\xi_{t-1}}{\sum_{k=1}^{n} \lambda_k \sum_{t=0}^{T_k-1} \xi_{t-1}^2}.$$
4. Repeat steps 2 and 3 until $|\nu|$ converges to 0.
5. Put $\widehat{\omega} = \cos^{-1}(\widehat{a}/2)$.

If $f > 1$, the common frequencies can be estimated one at a time using the methods given above. After each frequency is estimated, the corresponding sinusoid is removed from each time series using regression and the next frequency is estimated from the residuals. Once all the common frequencies have been removed, autoregressions with common autoregressive parameters can be fitted to the residuals using the methods of Section 3.8. As before, the frequencies can be re-estimated using the time series filtered by the autoregressive parameters. The full procedure for computing the parameter estimates for a given f under both H_A and $H_0^{(1)}$ is given in Algorithm 7.5. Note that the algorithm uses the Quinn–Fernandes technique to estimate frequencies. This is asymptotically equivalent to maximising the Gaussian likelihood and is computationally faster (see the discussion in Section 7.4).

The test statistic is computed by applying Algorithm 7.5 for each $f = 0, \ldots, F$ and calculating

$$\widehat{f}_0 = \arg\min_f \phi_0(f)$$

as well as

$$\widehat{f}_k = \arg\min_f \phi_k(f),$$

Algorithm 7.5 Computing the parameter estimators under $H_0^{(1)}$ and H_A with more than two time series for a given f

1. For $k = 1, \ldots, n$, use Algorithm 6.1 to fit f sinusoids to $\{X_{k,t}\}$. Remove the estimated sinusoids by regression and fit an autoregression of order p to the residuals, denoting the parameter estimates by $\widehat{\delta}_k$ and $\widehat{\sigma}^2_{k;A}(f)$. Repeat using $\left\{d_{\widehat{\delta}_k}(z)\, X_{k,t}\right\}$ in place of $\{X_{k,t}\}$.
2. For $k = 1, \ldots, n$, put $\widehat{\lambda}_k = \widehat{\sigma}^2_{1;A}(f)/\widehat{\sigma}^2_{k;A}(f)$ and $\left\{\widehat{E}_{k,t}\right\} = \{X_{k,t}\}$.
3. Estimate a common frequency by applying Algorithm 7.4 to $\left\{\widehat{E}_{1,t}\right\}, \ldots, \left\{\widehat{E}_{n,t}\right\}$ with $\lambda_k = \widehat{\lambda}_k$, $k = 1, \ldots, n$. Remove the corresponding sinusoid from each time series using regression and denote the residuals by $\left\{\widehat{E}_{1,t}\right\}, \ldots, \left\{\widehat{E}_{n,t}\right\}$.
4. Repeat step 3 until f frequencies have been estimated.
5. Use the methods of Chapter 3 to fit autoregressions of order p with the same autoregressive parameters to $\left\{\widehat{E}_{1,t}\right\}, \ldots, \left\{\widehat{E}_{n,t}\right\}$ with $\lambda_k = \widehat{\lambda}_k$, $k = 1, \ldots, n$, denoting the parameter estimates by $\widehat{\delta}_{\widehat{\lambda}}$ and $\widetilde{\sigma}^2_{\varepsilon;\widehat{\lambda}}$. Let $\left\{\widehat{E}_{k,t}\right\} = \left\{d_{\widehat{\delta}_{\widehat{\lambda}}}(z)\, X_{k,t}\right\}$, $k = 1, \ldots, n$.
6. Repeat steps 3–5 once.
7. For $k = 2, \ldots, n$, update $\widehat{\lambda}_k$ by
$$\frac{T_k \widetilde{\sigma}^2_{\varepsilon;\widehat{\lambda}}}{\sum_{t=p}^{T_k-1} \left\{d_{\widehat{\delta}_{\widehat{\lambda}}}(z)\, X_{k,t}\right\}^2}$$
and put $\left\{\widehat{E}_{k,t}\right\} = \{X_{k,t}\}$, $k = 1, \ldots, n$.
8. Repeat steps 3–7 until $\widehat{\lambda}_k$, $k = 1, \ldots, n$, converge.
9. Put $\widehat{\sigma}^2_{1;0}(f) = \widetilde{\sigma}^2_{\varepsilon;\widehat{\lambda}}$ and $\widehat{\sigma}^2_{k;0}(f) = \widetilde{\sigma}^2_{\varepsilon;\widehat{\lambda}}/\widehat{\lambda}_k$, $k = 2, \ldots, n$.

$k = 1, \ldots, n$, where

$$\phi_0(f) = \sum_{j=1}^{n} T_j \log\left\{\widehat{\sigma}^2_{j;0}(f)\right\} + \{p + (2n+3)f\} \log\left(\sum_{j=1}^{n} T_j\right)$$

and

$$\phi_k(f) = T_k \log\left\{\widehat{\sigma}^2_{k;A}(f)\right\} + (p + 5f)\log T_k.$$

The test statistic is then

$$\Lambda_{\widehat{f}_0, \widehat{f}_1, \ldots, \widehat{f}_n} = \sum_{j=1}^{n} \log\left\{\frac{\widehat{\sigma}^2_{j;0}\left(\widehat{f}_0\right)}{\widehat{\sigma}^2_{j;A}\left(\widehat{f}_j\right)}\right\}$$

and $H_0^{(1)}$ is rejected at significance level α when $\Lambda_{\widehat{f}, \widehat{f}_1, \ldots, \widehat{f}_n}$ is greater than the $100(1-\alpha)$th percentile of the χ^2 distribution with $(n-1)\left(p + \widehat{f}_0\right)$ degrees of freedom.

In order to test $H_0^{(2)}$, the procedure is almost the same except that autoregressions of different orders, denoted by p_k, are fitted independently to $\{\varepsilon_{k,t}\}$, $k = 1, \ldots, n$, under both the null and alternative hypotheses. The full procedure for computing the parameter estimates for a given f under both H_A and $H_0^{(2)}$ is given in Algorithm 7.6. The algorithm assumes that $p_1, \ldots, p_n$ are fixed. In practice we let $p_k = \lfloor (\log T_k)^{c_k} \rfloor$, where $c_k > 1$, $k = 1, \ldots, n$.

Algorithm 7.6 Computing the parameter estimators under $H_0^{(2)}$ and H_A with more than two time series for a given f

1. For $k = 1, \ldots, n$, use Algorithm 6.1 to fit f sinusoids to $\{X_{k,t}\}$. Remove the estimated sinusoids by regression and fit an autoregression of order p_k to the residuals, denoting the parameter estimates by $\widehat{\delta}_k$ and $\widehat{\sigma}^2_{k;A}(f)$. Repeat using $\left\{d_{\widehat{\delta}_k}(z) X_{k,t}\right\}$ in place of $\{X_{k,t}\}$.
2. For $k = 1, \ldots, n$, put $\widehat{\lambda}_k = \widehat{\sigma}^2_{1;A}(f) / \widehat{\sigma}^2_{k;A}(f)$ and $\left\{\widehat{E}_{k,t}\right\} = \left\{\widehat{X}_{k,t}\right\}$.
3. Estimate a common frequency by applying Algorithm 7.4 to $\left\{\widehat{E}_{1,t}\right\}, \ldots, \left\{\widehat{E}_{n,t}\right\}$ with $\lambda_k = \widehat{\lambda}_k$, $k = 1, \ldots, n$. Remove the corresponding sinusoid from each time series using regression and denote the residuals by $\left\{\widehat{E}_{1,t}\right\}, \ldots, \left\{\widehat{E}_{n,t}\right\}$.
4. Repeat step 3 until f frequencies have been estimated.
5. For $k = 1, \ldots, n$, fit an autoregression of order p_k to $\left\{\widehat{E}_{k,t}\right\}$, denoting the parameter estimates by $\widehat{\delta}_k$ and $\widehat{\sigma}^2_{k;0}(f)$. Let $\left\{\widehat{E}_{k,t}\right\} = \left\{d_{\widehat{\delta}_k}(z) X_{k,t}\right\}$.
6. Repeat steps 3–5 once.
7. For $k = 1, \ldots, n$, update $\widehat{\lambda}_k$ by $\widehat{\sigma}^2_{1;0}(f) / \widehat{\sigma}^2_{k;0}(f)$ and put $\left\{\widehat{E}_{k,t}\right\} = \{X_{k,t}\}$.
8. Repeat steps 3–7 until $\widehat{\sigma}^2_{k;0}(f)$, $k = 1, \ldots, n$, converge.

The test statistic is computed by applying Algorithm 7.6 for each $f = 0, \ldots, F$ and calculating

$$\widehat{f}_0 = \arg\min_f \phi_0^{(2)}(f),$$

where

$$\phi_0^{(2)}(f) = \sum_{j=1}^n T_j \log\left\{\widehat{\sigma}^2_{j;0}(f)\right\} + \sum_{j=1}^n p_j \log T_j + (2n+3) f \log\left(\sum_{j=1}^n T_j\right),$$

as well as

$$\widehat{f}_k = \arg\min_f \phi_k(f),$$

$k = 1, \ldots, n$. The test statistic is then

$$\Lambda_{\widehat{f}_0, \widehat{f}_1, \ldots, \widehat{f}_n} = \sum_{j=1}^n \log\left\{\frac{\widehat{\sigma}^2_{j;0}\left(\widehat{f}_0\right)}{\widehat{\sigma}^2_{j;A}\left(\widehat{f}_j\right)}\right\}$$

and the null hypothesis is rejected at significance level α when $\Lambda_{\widehat{f}_0,\widehat{f}_1,\ldots,\widehat{f}_n}$ is greater than the $100(1-\alpha)$th percentile of the χ^2 distribution with $(n-1)\,\widehat{f}_0$ degrees of freedom.

7.7 Comparing Multivariate Time Series With Periodic Components

Using the results of Chapter 6 we can extend the procedures given in this chapter to the case where $\{X_{k,t}\}$, $k=1,\ldots,n$, are vector processes with periodic components. We model $\{X_{k,t}\}$ using the multichannel sinusoidal model. That is, for $k=1,\ldots,n$,

$$X_{k,t}=\sum_{j=1}^{f_k}\left\{\alpha_{k,j}\cos\left(\omega_{k,j}t\right)+\beta_{k,j}\sin\left(\omega_{k,j}t\right)\right\}+E_{k,t},$$

where

$$E_{k,t}=\sum_{j=1}^{p_k}\delta_{k,j}E_{k,t-1}=\varepsilon_{k,t},$$

$\alpha_{k,j}$, $\beta_{k,j}$, $j=1,\ldots,f$, are $d\times 1$ and $\{X_{k,t}\}$, $\{E_{k,t}\}$ and $\{\varepsilon_{k,t}\}$ are d-dimensional. We make the usual assumptions on the noise processes, that is that they are independent sequences of martingale differences and that

$$E\left(\varepsilon_{k,t}\varepsilon_{k,t}'\mid\mathcal{F}_{k,t-1}\right)=\Sigma_k,$$

$k=1,\ldots,n$, where $\mathcal{F}_{k,t}$ is the σ-field generated by $\{\varepsilon_{k,t},\varepsilon_{k,t-1},\ldots\}$.

The first null hypothesis that we consider is that the processes have sinusoids at the same frequencies and the spectral densities of the noise processes differ only by a common scale. This is equivalent to

$$H_0^{(1)}:\omega_{1,j}=\cdots=\omega_{n,j},\ \forall j,\ \delta_{1,k}=\cdots=\delta_{n,k},\ \forall k,\ \Sigma_1=\lambda_2\Sigma_2=\cdots=\lambda_n\Sigma_n,$$

for positive constants $\lambda_2,\ldots,\lambda_n$. The condition placed on the noise spectral densities is the same as that for the second null hypothesis considered in Chapter 5. The second null hypothesis that we consider is that the processes have sinusoids at the same frequencies with no restrictions on the spectral densities of the noise processes. This is equivalent to

$$H_0^{(2)}:\omega_{1,j}=\cdots=\omega_{n,j},\ \forall j.$$

In each case the alternative hypothesis, H_A, is the complement of the null hypothesis. Under both H_A, the parameters can be estimated separately for each $\{X_{k,t}\}$, $k=1,\ldots,n$, using Algorithm 6.3.

Under $H_0^{(1)}$, we derive the test statistic letting $p_1 = \cdots = p_n = p$, where p is fixed. In practice we let $p = \lfloor \log (T_{\min})^c \rfloor$, where $T_{\min} = \min (T_1, \ldots, T_n)$ and $c > 1$. A common frequency can be estimated by maximising

$$\sum_{j=1}^{n} \widetilde{J}_{T_j,\Omega_j} (\omega), \tag{7.13}$$

where

$$\widetilde{J}_{T_k,\Omega_k} (\omega) = F_{T_k}^{*} (\omega) \Omega_k F_{T_k} (\omega),$$
$$F_{T_k} (\omega) = \sqrt{2} T_k^{-1/2} \sum_{t=0}^{T_k - 1} e^{-i\omega t} X_{k,t}$$

and $\Omega_1, \ldots, \Omega_n$ are suitable positive definite symmetric matrices. For given $\lambda_1, \ldots, \lambda_n$, the estimation is performed in two stages. In the first stage, Ω_k is set to $\lambda_k I_d$, $k = 1, \ldots, n$, and the f common frequencies are estimated one at a time. For each estimated frequency, the corresponding sinusoid is removed from each time series by regression and the next frequency is estimated from the residuals. Once f sinusoids have been removed, vector autoregressions with common autoregressive parameters and common order p are fitted to the residuals using the methods of Section 5.6. In the second stage, Ω_k is set to λ_k multiplied by the inverse of the estimated spectral density, $k = 1, \ldots, n$ and the f common frequencies are then re-estimated with the updated $\Omega_1, \ldots, \Omega_n$.

In order to maximise (7.13), a modified version of the multivariate Quinn–Fernandes technique can be used. Algorithm 7.7 computes the maximiser of (7.13) using the Quinn–Fernandes technique for given $\Omega_1, \ldots, \Omega_n$. The algorithm can be initialised by maximising (7.13) over the Fourier frequencies. The full procedure for computing the parameter estimates for a given f under both H_A and $H_0^{(1)}$ is given in Algorithm 7.8.

The test statistic is computed by applying Algorithm 7.8 for each $f = 0, \ldots, F$ and calculating

$$\widehat{f}_0 = \arg\min_f \phi_0 (f)$$

as well as

$$\widehat{f}_k = \arg\min_f \phi_k (f),$$

$k = 1, \ldots, n$, where

$$\phi_0 (f) = \sum_{j=1}^{n} T_j \log \left| \widehat{\Sigma}_{j;0} (f) \right| + \left\{ d^2 p + (2nd + 3) f \right\} \log \left(\sum_{j=1}^{n} T_j \right)$$

and

$$\phi_k (f) = T_k \log \left| \widehat{\Sigma}_{k;A} (f) \right| + \left\{ d^2 p + (2d + 3) f \right\} \log T_k.$$

Algorithm 7.7 Estimating a common frequency from n time series for given $\Omega_1, \ldots, \Omega_n$ using the Quinn–Fernandes technique

1. Put $\widehat{a} = 2\cos(\widehat{\omega})$, where $\widehat{\omega}$ is an initial estimate of ω.
2. For $k = 1, \ldots, n$, $t = 0, \ldots, T_k - 1$, put
$$\xi_{k,t} = \widehat{a}\xi_{k,t-1} - \xi_{k,t-2} + X_{k,t},$$
where $\xi_{k,-1} = \xi_{k,-2} = 0$.
3. Replace $\widehat{a}$ by $\widehat{a} + \nu$, where
$$\nu = 2\frac{\sum_{j=1}^{n}\sum_{t=0}^{T_j-1}\xi'_{j,t-1}\Omega_j X_{j,t}}{\sum_{k=1}^{n}\sum_{t=0}^{T_k-1}\xi'_{k,t-1}\Omega_j \xi_{k,t-1}}.$$
4. Repeats steps 2 and 3 until $|\nu|$ converges to 0.
5. Put $\widehat{\omega} = \cos^{-1}(\widehat{a}/2)$.

Test statistic is then
$$\Lambda_{\widehat{f}_0,\widehat{f}_1,\ldots,\widehat{f}_n} = \sum_{j=1}^{n} T_j \log\left\{\frac{\left|\widehat{\Sigma}_{j;0}\left(\widehat{f}_0\right)\right|}{\left|\widehat{\Sigma}_{j;A}\left(\widehat{f}_j\right)\right|}\right\}$$
and $H_0^{(1)}$ is rejected when $\Lambda_{\widehat{f}_0,\widehat{f}_1,\ldots,\widehat{f}_n}$ is greater than the $100(1-\alpha)$th percentile of the χ^2 distribution with
$$(n-1)\left\{d^2 p + \widehat{f}_0 + d(d+1)/2 - 1\right\}$$
degrees of freedom.

Under $H_0^{(2)}$, a common frequency can be estimated by maximising (7.13) in two stages. In the first stage, Ω_k is set to I_d for $k = 1, \ldots, n$, and f common frequencies are estimated one at a time. As usual, the sinusoid corresponding to each estimated frequency is removed from each time series by regression before estimating the next. Independent vector autoregressions of order p_k are then fitted to the residuals, and Ω_k is set to the inverse of the estimated spectral density of the kth autoregression. In the second stage, the f common frequencies are then re-estimated with the updated Ω_k. The full procedure for computing the parameter estimators under both H_A and $H_0^{(2)}$ for a given f is given in Algorithm 7.9. The algorithm assumes that $p_1, \ldots, p_n$ are fixed and in practice we let $p_k = \lfloor(\log T_k)^{c_k}\rfloor$, where $c_k > 1$, $k = 1, \ldots, n$.

The test statistic is computed by applying Algorithm 7.9 for each $f = 0, \ldots, F$ and calculating
$$\widehat{f}_0 = \arg\min_f \phi_0^{(2)}(f),$$

Algorithm 7.8 Computing the estimators under both $H_0^{(1)}$ and H_A with n multivariate time series for a given f

1. For $k = 1, \ldots, n$, use Algorithm 6.3 to fit f sinusoids to $\{X_{k,t}\}$, fitting vector autoregressions of order p. Denote the estimate of the residual covariance matrix by $\widehat{\Sigma}_{k;A}(f)$.
2. For $k = 1, \ldots, n$, put $\Omega_k = \widehat{\lambda}_k I_d$, where
$$\widehat{\lambda}_k = d / \operatorname{tr}\left\{\widehat{\Sigma}_{1;A}(f)\, \widehat{\Sigma}_{k;A}^{-1}(f)\right\},$$
and put $\left\{\widehat{E}_{k,t}\right\} = \{X_{k,t}\}$.
3. Estimate a common frequency by applying Algorithm 7.7 to $\left\{\widehat{E}_{1,t}\right\}, \ldots, \left\{\widehat{E}_{n,t}\right\}$. Remove the corresponding sinusoid from each time series using regression and denote the residuals by $\left\{\widehat{E}_{1,t}\right\}, \ldots, \left\{\widehat{E}_{n,t}\right\}$.
4. Repeat step 3 until f frequencies have been estimated.
5. Use the methods of Section 5.6 to fit vector autoregressions of order p with the same autoregressive parameters to $\left\{\widehat{E}_{1,t}\right\}, \ldots, \left\{\widehat{E}_{n,t}\right\}$ subject to $\Sigma_1 = \widehat{\lambda}_k \Sigma_k$, $k = 2, \ldots, n$. Denote the parameter estimates by $\widehat{\delta}_{\widehat{\lambda}}$ and $\widetilde{\Sigma}_{\widehat{\lambda}}$.
6. For $k = 1, \ldots, n$, let
$$\Omega_k = 2\pi\widehat{\lambda}_k \left(I_d + \sum_{j=1}^{p} \widehat{\delta}_{\widehat{\lambda},j} e^{-ij\widetilde{\omega}}\right)^{*} \widetilde{\Sigma}_{\widehat{\lambda}}^{-1} \left(I_d + \sum_{j=1}^{p} \widehat{\delta}_{\widehat{\lambda},j} e^{-ij\widetilde{\omega}}\right),$$
where
$$\widehat{\delta}_{\widehat{\lambda}} = \left[\begin{array}{ccc} \widehat{\delta}_{\widehat{\lambda},1} & \cdots & \widehat{\delta}_{\widehat{\lambda},p} \end{array}\right].$$
7. Repeat steps 3-5 once.
8. For $k = 2, \ldots, n$, update $\widehat{\lambda}_k$ by
$$\widehat{\lambda}_k = \frac{dT_k}{\operatorname{tr}\left[\widetilde{\Sigma}_{\widehat{\lambda}}^{-1} \sum_{t=p}^{T_k-1} \left\{\left(X_{k,t} + \sum_{j=1}^{p} \widehat{\delta}_{\widehat{\lambda},j} X_{k,t-j}\right)\left(X_{k,t} + \sum_{j=1}^{p} \widehat{\delta}_{\widehat{\lambda},j} X_{k,t-j}\right)'\right\}\right]}$$
and put $\left\{\widehat{E}_{k,t}\right\} = \{X_{k,t}\}$, $k = 1, \ldots, n$.
9. Repeat steps 3-8 until $\widehat{\lambda}_k$, $k = 1, \ldots, n$, converge.
10. Put $\widehat{\Sigma}_{1;0}(f) = \widetilde{\Sigma}_{\widehat{\lambda}}$ and $\widehat{\Sigma}_{k;0}(f) = \widetilde{\Sigma}_{\widehat{\lambda}} / \widehat{\lambda}_k$, $k = 2, \ldots, n$.

Algorithm 7.9 Computing the estimators under both $H_0^{(2)}$ and H_A with n multivariate time series for a given f

1. For $k = 1, \ldots, n$, use Algorithm 6.3 to fit f sinusoids to $\{X_{k,t}\}$, fitting autoregressions of order p_k. Denote the estimate of the residual covariance matrix by $\widehat{\Sigma}_{k;A}(f)$.
2. For $k = 1, \ldots, n$, put $\Omega_k = I_d$ and $\left\{\widehat{E}_{k,t}\right\} = \{X_{k,t}\}$.
3. Estimate a common frequency by applying Algorithm 7.7 to $\left\{\widehat{E}_{1,t}\right\}, \ldots, \left\{\widehat{E}_{n,t}\right\}$. Remove the corresponding sinusoid from each time series using regression and denote the residuals by $\left\{\widehat{E}_{1,t}\right\}, \ldots, \left\{\widehat{E}_{n,t}\right\}$.
4. Repeat step 3 until f frequencies have been estimated.
5. For $k = 1, \ldots, n$, fit a vector autoregression of order p_k to $\left\{\widehat{E}_{k,t}\right\}$, denoting the parameter estimates by $\widehat{\delta}_k$ and $\widehat{\Sigma}_{k;0}(f)$.
6. For $k = 1, \ldots, n$ let
$$\Omega_k = 2\pi \left(I_d + \sum_{j=1}^{p_k} \widehat{\delta}_{k,j} e^{-ij\widetilde{\omega}} \right)^* \widehat{\Sigma}_{k;0}^{-1}(f) \left(I_d + \sum_{j=1}^{p_k} \widehat{\delta}_{k,j} e^{-ij\widetilde{\omega}} \right),$$
where
$$\widehat{\delta}_k = \left[\begin{array}{ccc} \widehat{\delta}_{k,1} & \cdots & \widehat{\delta}_{k,p_k} \end{array} \right].$$
7. Repeat steps 3-5 once.

where
$$\phi_0^{(2)}(f) = \sum_{j=1}^{n} T_j \log \left| \widehat{\Sigma}_{j;0}(f) \right| + d^2 \sum_{j=1}^{n} p_j \log(T_j) + (2nd + 3) f \log \left(\sum_{j=1}^{n} T_j \right),$$
as well as
$$\widehat{f}_k = \arg\min_f \phi_k(f),$$
$k = 1, \ldots, n$. Test statistic is then
$$\Lambda_{\widehat{f}_0, \widehat{f}_1, \ldots, \widehat{f}_n} = \sum_{j=1}^{n} T_j \log \left\{ \frac{\left| \widehat{\Sigma}_{j;0}\left(\widehat{f}_0\right) \right|}{\left| \widehat{\Sigma}_{j;A}\left(\widehat{f}_j\right) \right|} \right\}$$
and $H_0^{(2)}$ is rejected when $\Lambda_{\widehat{f}_0, \widehat{f}_1, \ldots, \widehat{f}_n}$ is greater than the $100(1-\alpha)$th percentile of the χ^2 distribution with $(n-1)\widehat{f}_0$ degrees of freedom.

7.8 Simulations

The first simulation study presented in this section examines the behaviour of the test statistic under the null hypotheses. The number of time series compared, n, was 2 or 3, the dimension,

d, was 1 or 2 and the true number of frequencies, f, was 1 or 2. When $n = 2$ time series were simulated from the models

$$X_t = \alpha_1 \cos(\omega_1 t) + \beta_1 \sin(\omega_1 t) + \alpha_2 \cos(\omega_2 t) + \beta_2 \sin(\omega_2 t) + E_t,$$

and

$$Y_t = \alpha_1 \cos(\omega_1 t) + \beta_1 \sin(\omega_1 t) + \alpha_2 \cos(\omega_2 t) + \beta_2 \sin(\omega_2 t) + U_t,$$

with sample sizes $1,000$ and $1,250$, respectively. When $n = 3$ a third time series was simulated from the model

$$Z_t = \alpha_1 \cos(\omega_1 t) + \beta_1 \sin(\omega_1 t) + \alpha_2 \cos(\omega_1 t) + \beta_2 \sin(\omega_1 t) + W_t,$$

with sample size $1,500$. The noise processes were either generated from white noise (WN), that is,

$$E_t = \varepsilon_t, \qquad U_t = u_t \qquad \text{and} \qquad W_t = w_t,$$

the autoregressive models

$$E_t + \delta_1 E_{t-1} = \varepsilon_t, \qquad U_t + \delta_1 U_{t-1} = u_t \qquad \text{and} \qquad W_t + \delta_1 W_{t-1} = w_t,$$

or the moving average models

$$E_t = \varepsilon_t + \delta_2 \varepsilon_{t-1}, \qquad U_t = u_t + \delta_2 u_{t-1} \qquad \text{and} \qquad W_t = w_t + \delta_2 w_{t-1}.$$

The frequencies were $\omega_1 = \pi/3$ and $\omega_2 = 0$ or $2\pi/3$. For the $d = 1$ case the parameter values were

$$\delta_1 = 0.7, \qquad \delta_2 = 0.8, \qquad \alpha_1 = \alpha_2 = \frac{1}{\sqrt{2}}, \qquad \beta_1 = -\frac{1}{\sqrt{2}} \qquad \text{and} \qquad \beta_2 = 0,$$

and $\{\varepsilon_t\}$, $\{u_t\}$ and $\{w_t\}$ were simulated from the normal distribution with means zero and variances 1, $\frac{1}{2}$ and $\frac{3}{4}$, respectively. For the $d = 2$ case the parameter values were

$$\delta_1 = \begin{bmatrix} 0.7 & 0.1 \\ -0.1 & 0.7 \end{bmatrix}, \qquad \delta_2 = \begin{bmatrix} 0.8 & 0.1 \\ -0.1 & 0.8 \end{bmatrix},$$

$$\alpha_1 = \alpha_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \qquad \beta_1 = -\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \qquad \text{and} \qquad \beta_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

and $\{\varepsilon_t\}$, $\{u_t\}$ and $\{w_t\}$ were simulated from the multivariate normal distribution with means zero and covariance matrices I_2, $\frac{1}{2}I_2$ and $\frac{3}{4}I_2$, respectively. Figure 7.1 shows the component spectral densities, and their coherency, of $\{E_t\}$ for the autoregressive case when $d = 2$. The

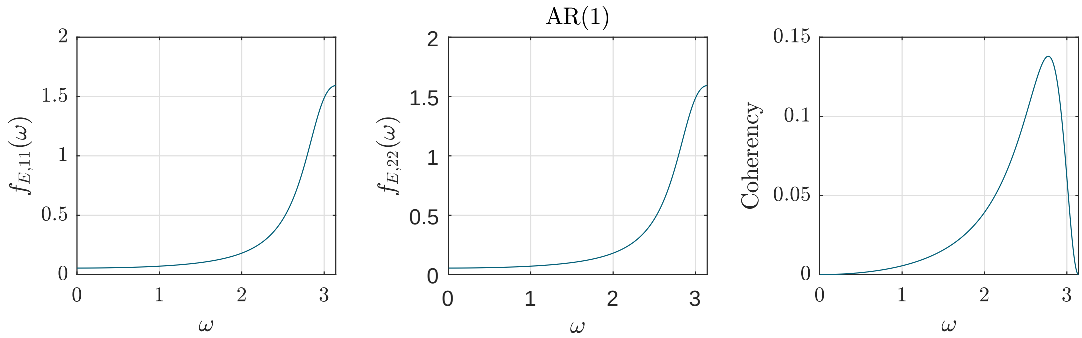

Figure 7.1: The spectral densities of each component, as well as their coherency, of $\{E_t\}$, when $\{E_t\}$ is autoregressive and $d = 2$.

MA(1) parts of Figure 5.1 show the component spectral densities, and their coherency, for the moving average case when $d = 2$.

The tests for both $H_0^{(1)}$ and $H_0^{(2)}$ were applied to 10,000 replications of these simulations. The maximum number of frequencies was set to one more than the true value. The simulations were also repeated with all sample sizes doubled. Tables 7.1–7.8 summarise the means, variances and Type I error rates of the resulting test statistics.

For the univariate cases, when testing $H_0^{(1)}$, the means and variances were very close to their theoretical values and the Type I error rates were all close to 0.05. When testing $H_0^{(2)}$, the means were very close to their theoretical values but the variances were a little higher, and some of the Type I error rates were closer to 0.06 for the smaller sample sizes. When the sample sizes were doubled, the variances generally closer to their theoretical values and the Type I error rates were closer to 0.05 (see table 7.4).

For the multivariate cases, when testing both $H_0^{(1)}$ and $H_0^{(2)}$, the means were very close to their theoretical values but the variances and Type I error rates were a little higher for the smaller sample sizes. When the sample sizes were doubled, the variances were closer to their theoretical values and the Type I error rates were closer to 0.05.

Note that when testing $H_0^{(2)}$, some of the test statistics were negative, but only slightly. This could only occur if the likelihood maximised under the null is more than that under the alternative. This may be due to numerical problems of convergence under the alternative or underestimation of order estimates using BIC. Further simulations, not reported here, show that the number of negative values is reduced as the sample sizes increase, and eventually goes to zero.

Table 7.1: Summary of simulations under $H_0^{(1)}$ when $d = 1$.

		$f = 1$				$f = 2$			
		df	mean	var	Type I	df	mean	var	Type I
$n = 2$	WN	8	8.045	16.707	0.052	9	9.010	17.927	0.048
	AR(1)	8	8.007	16.177	0.048	9	8.962	18.450	0.053
	MA(1)	8	8.024	16.900	0.053	9	9.074	19.670	0.052
$n = 3$	WN	16	15.965	32.397	0.050	18	18.151	36.766	0.055
	AR(1)	16	16.072	33.238	0.053	18	18.034	36.523	0.048
	MA(1)	16	15.968	33.761	0.053	18	17.973	37.566	0.049

Table 7.2: Summary of simulations under $H_0^{(2)}$ when $d = 1$.

		$f = 1$				$f = 2$			
		df	mean	var	Type I	df	mean	var	Type I
$n = 2$	WN	1	1.014	2.366	0.054	2	2.029	4.214	0.050
	AR(1)	1	1.033	2.259	0.053	2	2.025	4.518	0.054
	MA(1)	1	1.108	2.973	0.064	2	2.087	5.064	0.060
$n = 3$	WN	2	2.037	4.377	0.054	4	4.107	8.442	0.054
	AR(1)	2	2.032	4.083	0.050	4	4.118	8.658	0.058
	MA(1)	2	2.240	5.653	0.068	4	4.222	9.841	0.064

Table 7.3: Summary of simulations under $H_0^{(1)}$ when $d = 1$ and the sample sizes were doubled.

		$f = 1$				$f = 2$			
		df	mean	var	Type I	df	mean	var	Type I
$n = 2$	WN	8	7.988	16.057	0.050	9	9.034	18.257	0.052
	AR(1)	8	7.986	15.984	0.051	9	9.012	18.235	0.050
	MA(1)	8	7.910	15.821	0.046	9	9.048	19.261	0.055
$n = 3$	WN	16	16.011	31.881	0.049	18	18.129	36.406	0.050
	AR(1)	16	16.077	32.157	0.051	18	18.026	36.303	0.050
	MA(1)	16	16.037	33.149	0.049	18	18.002	37.719	0.050

Table 7.4: Summary of simulations under $H_0^{(2)}$ when $d = 1$ and the sample sizes were doubled.

		$f = 1$				$f = 2$			
		df	mean	var	Type I	df	mean	var	Type I
$n = 2$	WN	1	1.015	2.230	0.052	2	2.065	4.428	0.054
	AR(1)	1	1.005	2.044	0.052	2	2.018	4.215	0.051
	MA(1)	1	1.074	2.466	0.060	2	2.051	4.275	0.054
$n = 3$	WN	2	2.012	3.965	0.049	4	4.134	8.937	0.055
	AR(1)	2	2.031	4.123	0.053	4	4.065	8.344	0.052
	MA(1)	2	2.081	4.918	0.059	4	4.111	9.092	0.058

Table 7.5: Summary of simulations under $H_0^{(1)}$ when $d = 2$.

		$f = 1$				$f = 2$			
		df	mean	var	Type I	df	mean	var	Type I
$n = 2$	WN	31	31.503	65.042	0.060	32	32.490	65.590	0.057
	AR(1)	31	31.531	64.708	0.060	32	32.551	67.182	0.057
	MA(1)	31	31.636	65.188	0.059	32	32.590	67.032	0.058
$n = 3$	WN	62	62.567	126.923	0.054	64	64.632	128.962	0.055
	AR(1)	62	62.312	124.759	0.051	64	64.684	134.290	0.061
	MA(1)	62	62.756	128.658	0.059	64	64.737	132.614	0.061

Table 7.6: Summary of simulations under $H_0^{(2)}$ when $d = 2$.

		$f = 1$				$f = 2$			
		df	mean	var	Type I	df	mean	var	Type I
$n = 2$	WN	1	1.010	2.187	0.054	2	2.058	4.398	0.056
	AR(1)	1	1.003	2.130	0.053	2	1.941	4.694	0.056
	MA(1)	1	1.035	2.479	0.061	2	1.978	4.570	0.057
$n = 3$	WN	2	1.998	4.240	0.056	4	4.102	8.719	0.055
	AR(1)	2	2.016	4.379	0.055	4	3.800	8.612	0.052
	MA(1)	2	2.123	5.488	0.067	4	4.007	8.847	0.056

Table 7.7: Summary of simulations under $H_0^{(1)}$ when $d = 2$ and the sample sizes were doubled.

		$f = 1$				$f = 2$			
		df	mean	var	Type I	df	mean	var	Type I
$n = 2$	WN	31	31.267	63.135	0.054	32	32.385	65.639	0.055
	AR(1)	31	31.163	63.732	0.054	32	32.136	64.099	0.051
	MA(1)	31	31.267	62.760	0.052	32	32.131	65.266	0.051
$n = 3$	WN	62	62.161	121.031	0.049	64	63.953	129.494	0.052
	AR(1)	62	62.011	123.721	0.050	64	64.044	130.945	0.052
	MA(1)	62	61.992	126.453	0.050	64	64.172	131.361	0.052

Table 7.8: Summary of simulations under $H_0^{(2)}$ when $d = 2$ and the sample sizes were doubled.

		$f = 1$				$f = 2$			
		df	mean	var	Type I	df	mean	var	Type I
$n = 2$	WN	1	1.032	2.195	0.052	2	1.995	4.068	0.052
	AR(1)	1	0.989	2.203	0.053	2	1.936	4.258	0.052
	MA(1)	1	1.021	2.236	0.054	2	2.037	4.448	0.056
$n = 3$	WN	2	2.016	4.079	0.052	4	4.057	8.458	0.056
	AR(1)	2	2.024	4.342	0.054	4	3.859	8.434	0.050
	MA(1)	2	2.071	4.586	0.058	4	4.043	8.477	0.055

The second simulation study in this section demonstrates the power of the test in detecting differences in a fixed frequency in two time series. Time series of equal length were simulated from the models

$$X_t = \alpha_1 \cos(\omega_X t) + \beta_1 \sin(\omega_X t) + E_t$$

and

$$Y_t = \alpha_1 \cos(\omega_Y t) + \beta_1 \sin(\omega_Y t) + U_t,$$

where

$$\alpha_1 = \frac{1}{\sqrt{2}}, \qquad \beta_1 = -\frac{1}{\sqrt{2}}$$

and $\{E_t\}$ and $\{U_t\}$ were generated by the same white noise, autoregressive and moving average processes as above. The case where $d = 2$ and $\{E_t\}$ and $\{U_t\}$ were autoregressive was also considered. In this case

$$\alpha_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \qquad \text{and} \qquad \beta_1 = -\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

The frequency ω_X was fixed at $\pi/3$ and the frequency ω_Y varied from

$$\frac{\pi}{3} + \frac{\pi}{T}(-1, -0.98, \ldots, 1),$$

where T is the common sample size. The test for $H_0^{(1)}$ was applied to 10,000 replications of these simulations. Figure 7.2 plots the empirical powers when the sample sizes were 250 and Figure 7.3 plots the empirical powers when the sample sizes were 500.

For the smaller sample size, the test performed well when the noise processes were from white noise and autoregressions, although the Type I error rates for the autoregressive cases were over 0.05 (0.106 and 0.076 for the $d = 1$ and $d = 2$ cases, respectively). The test did not perform well when the noise processes were from moving average processes, with a Type I error rate of 0.339. When the sample sizes were increased to 500, the test performed much better for the moving average case with a Type I error rate of 0.057. The Type I error rates were also closer to 0.05 for the autoregressive cases.

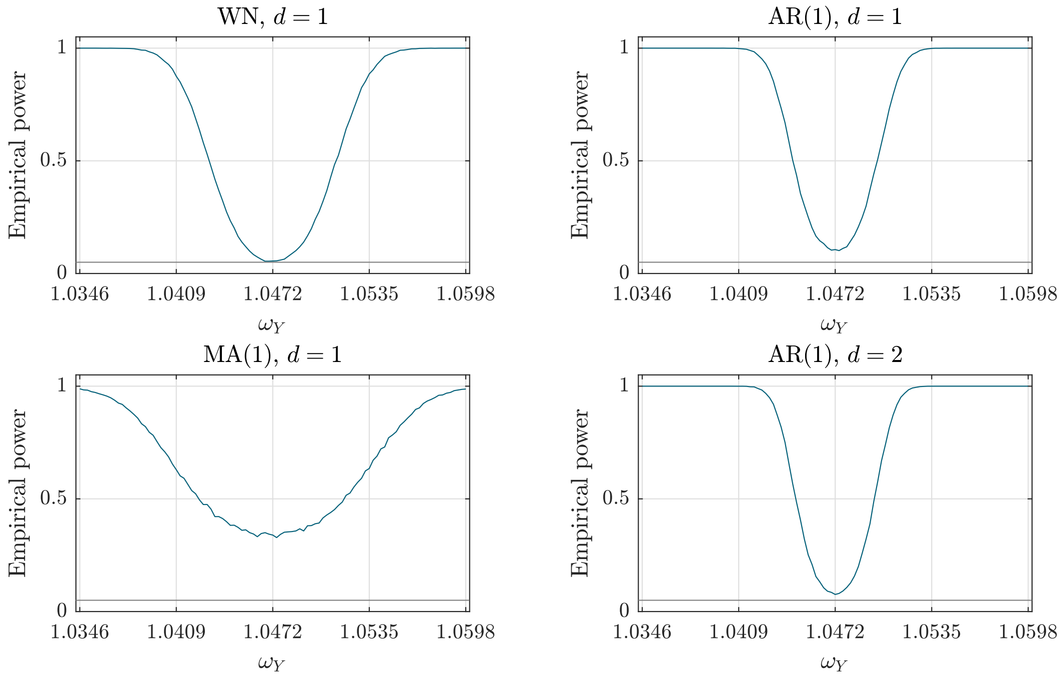

Figure 7.2: Empirical powers when the sample sizes were 250. The lowest horizontal line in each plot indicates the significance level of 0.05.

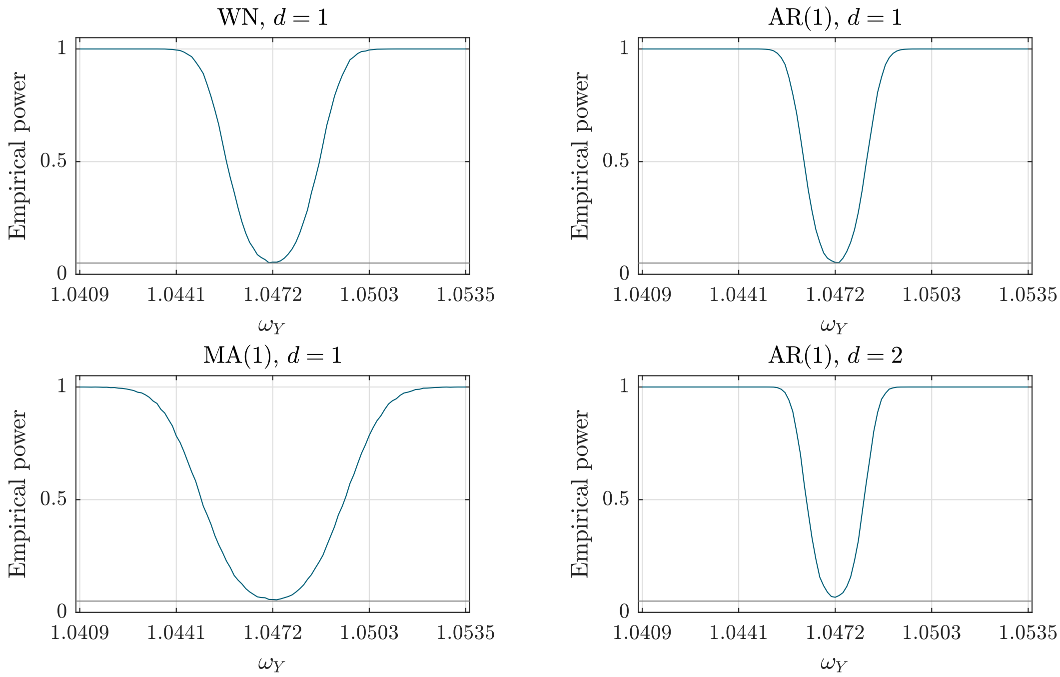

Figure 7.3: Empirical powers when the sample sizes were 500. The lowest horizontal line in each plot indicates the significance level of 0.05.

7.A Appendix

In what follows, where convergence is indicated, it will mean convergence in the almost sure sense, unless otherwise stated. Where order notation is used, it will also mean orders in the almost sure sense.

7.A.1 Proof of Theorem 7.1

From (7.11)

$$\begin{aligned}\widetilde{l}_0(\omega_0) - \widetilde{l}_0(\omega) &= \frac{T_1}{2}\log\left\{\frac{S_X - I_{T_1,X}(\omega)}{S_X - I_{T_1,X}(\omega_0)}\right\} + \frac{T_2}{2}\log\left\{\frac{S_Y - I_{T_1,Y}(\omega)}{S_Y - I_{T_1,Y}(\omega_0)}\right\} \\ &= \frac{T_1}{2}\log\left\{1 + \frac{I_{T_1,X}(\omega_0) - I_{T_1,X}(\omega)}{S_X - I_{T_1,X}(\omega_0)}\right\} + \frac{T_2}{2}\log\left\{1 + \frac{I_{T_2,Y}(\omega_0) - I_{T_2,Y}(\omega)}{S_Y - I_{T_2,Y}(\omega_0)}\right\},\end{aligned}$$

where

$$S_X = \sum_{t=0}^{T_1-1} X_t^2 \qquad \text{and} \qquad S_Y = \sum_{t=0}^{T_2-1} Y_t^2.$$

Now,

$$\begin{aligned}\sum_{t=0}^{T_1-1} X_t e^{-i\omega t} &= \sum_{t=0}^{T_1-1} E_t e^{-i\omega t} + \alpha_{X0}\sum_{t=0}^{T_1-1}\cos(\omega_0 t)\, e^{-i\omega t} + \beta_{X0}\sum_{t=0}^{T_1-1}\sin(\omega_0 t)\, e^{-i\omega t} \\ &= \sum_{t=0}^{T_1-1} E_t e^{-i\omega t} + \frac{\alpha_{X0}}{2}\sum_{t=0}^{T_1-1}\left(e^{i\omega_0 t} + e^{-i\omega_0 t}\right)e^{-i\omega t} + \frac{\beta_{X0}}{2i}\sum_{t=0}^{T_1-1}\left(e^{i\omega_0 t} - e^{-i\omega_0 t}\right)e^{-i\omega t} \\ &= U_{T_1}(\omega) + \left(\frac{\alpha_{X0}}{2} + \frac{\beta_{X0}}{2i}\right)\frac{e^{i(\omega_0-\omega)T_1} - 1}{e^{i(\omega_0-\omega)} - 1} + O(1),\end{aligned}$$

where

$$U_{T_1}(\omega) = \sum_{t=0}^{T_1-1} E_t e^{-i\omega t}.$$

Thus

$$T_1^{-1} I_{T_1,X}(\omega) = 2T_1^{-2}\left|U_{T_1}(\omega) + \left(\frac{\alpha_{X0}}{2} + \frac{\beta_{X0}}{2i}\right)\frac{e^{i(\omega_0-\omega)T_1} - 1}{e^{i(\omega_0-\omega)} - 1} + O(1)\right|^2. \tag{7.14}$$

But, from Theorem 2.3,

$$\sup_\omega |U_{T_1}(\omega)|^2 = O(T_1 \log T_1) \tag{7.15}$$

and, also,

$$T_1^{-2}\left|\frac{e^{i(\omega_0-\omega)T_1} - 1}{e^{i(\omega_0-\omega)} - 1}\right|^2 = \frac{\sin^2\{T_1(\omega_0-\omega)/2\}}{T_1^2\sin^2\{(\omega_0-\omega)/2\}}.$$

Therefore, from (7.14) and (7.15),

$$T_1^{-1} I_{T_1,X}(\omega) = \frac{\rho_{X0}^2}{2}\frac{\sin^2\{T_1(\omega_0-\omega)/2\}}{T_1^2\sin^2\{(\omega_0-\omega)/2\}} + o(1).$$

Also,

$$\sum_{t=0}^{T_1-1} X_t e^{-i\omega_0 t} = U_{T_1}(\omega_0) + \left(\frac{\alpha_{X0}}{2} + \frac{\beta_{X0}}{2i}\right) T + O(1)$$

and so

$$T_1^{-1} I_{T_1,X}(\omega_0) = \frac{\rho_{X0}^2}{2} + o(1).$$

Therefore

$$T_1^{-1}\left\{I_{T_1,X}(\omega_0) - I_{T_1,X}(\omega)\right\} = \frac{\rho_{X0}^2}{2}\left[1 - \frac{\sin^2\{T_1(\omega_0-\omega)/2\}}{T_1^2 \sin^2\{(\omega_0-\omega)/2\}}\right] + o(1).$$

Furthermore,

$$\begin{aligned} T_1^{-1} S_X &= T_1^{-1} \sum_{t=0}^{T_1-1} \left\{\alpha_{X0}\cos(\omega_0 t) + \beta_{X0}\sin(\omega_0 t) + E_t\right\}^2 \\ &= \gamma_E(0) + \frac{\rho_{X0}^2}{2} + o(1), \end{aligned}$$

where $\gamma_E(0) = E\left(E_t^2\right)$, and so

$$T_1^{-1}\left\{S_X - I_{T_1,X}(\omega_0)\right\} = \gamma_E(0) + o(1). \tag{7.16}$$

Let $\omega = \omega_0 + \kappa/T_1$ for some $\kappa > 0$. Then

$$T_1^{-1}\left\{I_{T_1,X}(\omega_0) - I_{T_1,X}(\omega_0 + \kappa/T_1)\right\} = \frac{\rho_{X0}^2}{2}\left[1 - \frac{\sin^2(\kappa/2)}{T_1^2 \sin^2\{\kappa/(2T_1)\}}\right] + o(1).$$

Consider the function

$$f_T(x) = \frac{\sin^2 x}{T^2 \sin^2(x/T)}.$$

For any $x \neq 0$,

$$f_T(x) \to \frac{\sin^2 x}{x^2} < 1$$

as $T \to \infty$. Therefore

$$\begin{aligned} \lim_{T_1\to\infty} \frac{I_{T_1,X}(\omega_0) - I_{T_1,X}(\omega_0+\kappa/T_1)}{S_X - I_{T_1,X}(\omega_0)} &> \frac{\rho_{X0}^2}{2\gamma_E(0)} \\ &> 0. \end{aligned}$$

Similarly,

$$T_2^{-1}\left\{S_Y - I_{T_2,Y}(\omega_0)\right\} = \gamma_U(0) + o(1) \tag{7.17}$$

and

$$\begin{aligned} \lim_{T_2\to\infty} \frac{I_{T_2,Y}(\omega_0) - I_{T_2,Y}(\omega_0+\kappa/T_2)}{S_Y - I_{T_2,Y}(\omega_0)} &> \frac{\rho_{Y0}^2}{2\gamma_U(0)} \\ &> 0, \end{aligned}$$

where $\gamma_U(0) = E\left(U_t^2\right)$. Thus

$$\liminf_{T_1,T_2\to\infty} \inf_{|\omega-\omega_0|>\kappa/T_{\min}} \left\{\widetilde{l}_0(\omega_0) - \widetilde{l}_0(\omega)\right\} > 0$$

and it follows from Theorem 2.5 that $T_{\min}(\widehat{\omega} - \omega_0) \to 0$ as $T_1, T_2 \to \infty$.

7.A.2 Proof of Theorem 7.2

From the mean value theorem,

$$0 = \frac{d\widetilde{l}_0(\omega)}{d\omega} = \frac{d\widetilde{l}_0(\omega_0)}{d\omega} + \frac{d^2\widetilde{l}_0(\widetilde{\omega})}{d\omega^2}(\widehat{\omega} - \omega_0),$$

where $\widetilde{\omega}$ is a point on the line segment between ω_0 and $\widehat{\omega}$. Since $T_{\min}(\widehat{\omega} - \omega_0) \to 0$ as $T_1, T_2 \to \infty$, $\widehat{\omega} - \omega_0$ has the same asymptotic distribution as

$$-\left\{\frac{d^2\widetilde{l}_0(\omega_0)}{d\omega^2}\right\}^{-1} \frac{d\widetilde{l}_0(\omega_0)}{d\omega}.$$

The first and second derivatives of $\widetilde{l}_0(\omega)$ are

$$\frac{d\widetilde{l}_0(\omega)}{d\omega} = -\frac{T_1}{2}\frac{\frac{d}{d\omega}I_{T_1,X}(\omega)}{S_X - I_{T_1,X}(\omega)} - \frac{T_2}{2}\frac{\frac{d}{d\omega}I_{T_2,Y}(\omega)}{S_Y - I_{T_2,Y}(\omega)}$$

and

$$\begin{aligned}\frac{d^2\widetilde{l}_0(\omega)}{d\omega^2} = &\frac{T_1}{2}\left[\frac{\frac{d^2}{d\omega^2}I_{T_1,X}(\omega)}{S_X - I_{T_1,X}(\omega)} + \left\{\frac{\frac{d}{d\omega}I_{T_1,X}(\omega)}{S_X - I_{T_1,X}(\omega)}\right\}^2\right] \\ &+ \frac{T_2}{2}\left[\frac{\frac{d^2}{d\omega^2}I_{T_2,Y}(\omega)}{S_Y - I_{T_2,Y}(\omega)} + \left\{\frac{\frac{d}{d\omega}I_{T_2,Y}(\omega)}{S_Y - I_{T_2,Y}(\omega)}\right\}^2\right].\end{aligned}$$

From Hannan (1973b), $T_1^{-3/2} dI_{T_1,X}(\omega_0)/d\omega$ is asymptotically normal with mean zero and variance

$$\frac{2\pi f_E(\omega_0)\rho_{X0}^2}{6}$$

and also $T_1^{-3} d^2 I_{T_1,X}(\omega)/d\omega^2 \to -\rho_{X0}^2/12$. Similarly, $T_2^{-3/2} dI_{T_2,Y}(\omega_0)/d\omega$ is asymptotically normal with mean zero and variance

$$\frac{2\pi f_U(\omega_0)\rho_{Y0}^2}{6}$$

and also $T_2^{-3} d^2 I_{T_2,Y}(\omega)/d\omega^2 \to -\rho_{Y0}^2/12$. Thus, from (7.16) and (7.17), $T_1^{-3/2} d\widetilde{l}_0(\omega_0)/d\omega$ is asymptotically normal with mean zero and variance

$$\frac{2\pi f_E(\omega_0)\rho_{X0}^2}{24\gamma_E^2(0)} + \frac{2\pi f_U(\omega_0)\kappa^3\rho_{Y0}^2}{24\gamma_U^2(0)}.$$

Also,

$$T_1^{-3}\frac{d^2\widetilde{l}_0(\omega_0)}{d\omega^2} \to -\frac{\rho_{X0}^2}{24\gamma_E(0)} - \frac{\kappa^3\rho_{Y0}^2}{24\gamma_U(0)}$$

since

$$\left\{\frac{\frac{d}{d\omega}I_{T_1,X}(\omega_0)}{S_X - I_{T_1,X}(\omega_0)}\right\}^2 = O(T_1) \quad \text{and} \quad \left\{\frac{\frac{d}{d\omega}I_{T_2,Y}(\omega_0)}{S_Y - I_{T_2,Y}(\omega_0)}\right\}^2 = O(T_2).$$

Thus $T_1^{3/2}(\widehat{\omega}-\omega_0)$ is asymptotically normal with mean zero and variance

$$\begin{aligned}&48\pi\left\{\frac{f_E(\omega_0)\rho_{X0}^2}{\gamma_E^2(0)}+\frac{f_U(\omega_0)\kappa^3\rho_{Y0}^2}{\gamma_U^2(0)}\right\}\left\{\frac{\rho_{X0}^2}{\gamma_E(0)}+\frac{\kappa^3\rho_{Y0}^2}{\gamma_U(0)}\right\}^{-2}\\&=48\pi\left\{\frac{\rho_{X0}^2}{f_E(0)}+\frac{\kappa^3\rho_{Y0}^2}{f_U(0)}\right\}^{-1}\end{aligned}$$

since, under $H_0^{(1)}$,

$$\frac{f_E(\omega_0)}{\gamma_E(0)}=\frac{f_U(\omega_0)}{\gamma_U(0)}.$$

7.A.3 Proof of Theorem 7.3

Following the proof of Theorem 16 of Quinn and Hannan (2001) we use Dieudonné's fixed point theorem (Dieudonné, 1960, Section 10.1). We must show that the following two conditions are met as $T_1\to\infty$.

Condition 7.1 *There exists k, $0\leqslant k<1$, such that if $a,a'\in A_{T_1}(\nu)$ then*

$$\left|a-a'+2h_{T_1}(a)-2h_{T_1}(\omega')\right|<k\left|a-a'\right|.$$

Condition 7.2 $|2h_T(a_0)|<(1-k)T^{-\nu}$.

Let

$$\begin{aligned}d_{T_1,X}(a)&=\sin\omega\sum_{t=0}^{T_1-1}X_t\xi_{t-1}(a),\ d_{T_2,Y}(a)=\sin\omega\sum_{t=0}^{T_2-1}Y_t\eta_{t-1}(a),\\e_{T_1,X}(a)&=\sin^2(\omega)\sum_{t=0}^{T_1-1}\xi_{t-1}^2(a),\ e_{T_2,Y}(a)=\sin^2(\omega)\sum_{t=0}^{T_2-1}\eta_{t-1}^2(a).\end{aligned}$$

Then

$$h_{T_1,T_2}(a)=\sin\omega\left\{d_{T_1,X}(a)+\lambda d_{T_2,Y}(a)\right\}\left\{e_{T_1,X}(a)+\lambda e_{T_2,Y}(a)\right\}^{-1}.$$

From the proof of Theorem 16 of Quinn and Hannan (2001),

$$e_{T_1,X}(a)=\frac{\rho_X^2}{24}T_1^3+O\left\{T_1^{5/2}(\log T_1)^{1/2}\right\},\tag{7.18}$$

$$e_{T_2,Y}(a)=\frac{\rho_Y^2}{24}T_2^3+O\left\{T_2^{5/2}(\log T_2)^{1/2}\right\},\tag{7.19}$$

and

$$\begin{aligned}e_{T_1,X}(a)&=e_{T_1,X}(a_0)+(\omega-\omega_0)O\left\{T_1^{7/2}(\log T_1)^{1/2}\right\},\\e_{T_2,Y}(a)&=e_{T_2,Y}(a_0)+(\omega-\omega_0)O\left\{T_2^{7/2}(\log T_2)^{1/2}\right\}.\end{aligned}$$

Also

$$d_{T_1,X}(a)=d_{T_1,X}(a_0)+(\omega-\omega_0)\left\{\frac{\rho_X^2}{24\gamma_\varepsilon(0)}T_1^3+O\left(T_1^2\right)\right\}$$

and

$$d_{T_2,Y}(a) = d_{T_2,Y}(a_0) + (\omega - \omega_0)\left\{\frac{\rho_Y^2}{24\gamma_u(0)}T_2^3 + O\left(T_2^2\right)\right\},$$

which are both $O\left\{T_1^{3/2}(\log T_1)^{1/2}\right\}$. Let $a = 2\cos\omega$ and $a' = 2\cos\omega'$. Then $h_{T_1,T_2}(a) - h_{T_1,T_2}(a')$ is equal to

$$\begin{aligned}&\sin\omega\left\{d_{T_1,X}(a) + \lambda d_{T_2,Y}(a)\right\}\left\{e_{T_1,X}(a) + \lambda e_{T_2,Y}(a)\right\}^{-1}\\&- \sin\omega'\left\{d_{T_1,X}\left(a'\right) + \lambda d_{T_2,Y}\left(a'\right)\right\}\left\{e_{T_1,X}\left(a'\right) + \lambda e_{T_2,Y}\left(a'\right)\right\}^{-1}.\end{aligned}$$

Thus

$$\begin{aligned}&\left\{e_{T_1,X}(a) + \lambda e_{T_2,Y}(a)\right\}\left\{e_{T_1,X}\left(a'\right) + \lambda e_{T_2,Y}\left(a'\right)\right\}\left\{h_{T_1,T_2}(a) - h_{T_1,T_2}\left(a'\right)\right\}\\&= \left(\sin\omega - \sin\omega'\right)\left\{d_{T_1,X}(a) + \lambda d_{T_2,Y}(a)\right\}\left\{e_{T_1,X}\left(a'\right) + \lambda e_{T_2,Y}\left(a'\right)\right\}\\&+ \sin\omega'\left\{d_{T_1,X}(a) + \lambda d_{T_2,Y}(a) - d_{T_1,X}\left(a'\right) - \lambda d_{T_2,Y}\left(a'\right)\right\}\left\{e_{T_1,X}(a) + \lambda e_{T_2,Y}(a)\right\}\\&- \sin\omega'\left\{d_{T_1,X}(a) + \lambda d_{T_2,Y}(a)\right\}\left\{e_{T_1,X}(a) + \lambda e_{T_2,Y}(a) - e_{T_1,X}\left(a'\right) - \lambda e_{T_2,Y}\left(a'\right)\right\}\end{aligned}$$

which is equal to

$$\begin{aligned}&\left(\omega - \omega'\right)O\left\{T_1^{9/2}(\log T_1)^{1/2}\right\} + \left(\omega - \omega'\right)O\left(T_1^5\log T_1\right)\\&+ \sin\omega'\left(\omega - \omega'\right)\left(\frac{\rho_X^2 + \lambda\kappa^3\rho_Y^2}{24}\right)^2 T_1^6\left[1 + O\left\{T_1^{-1/2}(\log T_1)^{1/2}\right\}\right].\end{aligned}$$

Thus

$$\begin{aligned}&\left\{e_{T_1,X}(a) + \lambda e_{T_2,Y}(a)\right\}\left\{e_{T_1,X}\left(a'\right) + \lambda e_{T_2,Y}\left(a'\right)\right\}\left\{h_{T_1,T_2}(a) - h_{T_1,T_2}\left(a'\right)\right\}\\&= \sin\omega'\left(\omega - \omega'\right)\left(\frac{\rho_X^2 + \lambda\kappa^3\rho_Y^2}{24}\right)^2 T_1^6\left[1 + O\left\{T_1^{-1/2}(\log T_1)^{1/2}\right\}\right],\end{aligned}$$

and so

$$h_{T_1,T_2}(a) - h_{T_1,T_2}\left(a'\right) = \sin\omega'\left(\omega - \omega'\right)\left[1 + O\left\{T_1^{-1/2}(\log T_1)^{1/2}\right\}\right].$$

Now,

$$a - a' = 2\cos(\omega) - 2\cos\left(\omega'\right) = -2\sin\left(\omega'\right)\left(\omega - \omega'\right) + o\left(T_1^{-2}\right),$$

and so

$$a - a' + 2h_{T_1,T_2}(a) - 2h_{T_1,T_2}\left(a'\right) = \left(a - a'\right)O\left\{T_1^{-1/2}(\log T_1)^{1/2}\right\} \tag{7.20}$$

Also,

$$\begin{aligned}2h_{T_1,T_2}(a_0) &= 2\sin\omega_0 O\left\{T_1^{3/2}(\log T_1)^{1/2}\right\}\left[\left(\frac{\rho_X^2 + \lambda\kappa^3\rho_Y^2}{24}\right)^2 T_1^3 + O\left\{T_1^{5/2}(\log T_1)^{1/2}\right\}\right]^{-1}\\&= 2\sin\omega_0 O\left\{T_1^{-3/2}(\log T_1)^{1/2}\right\}.\end{aligned} \tag{7.21}$$

From (7.20) and (7.21), Conditions 7.1 and 7.2 are met and therefore the first part of the theorem is proved. It follows that $T_1^{\nu}\left(\widehat{a}_{T_1} - a_0\right) \to 0$ for all $\nu < 3/2$, and that, since $\widehat{a}_{T_1} = 2\cos\widehat{\omega}_{T_1}$, $T_1^{\nu}\left(\widehat{\omega} - \omega_0\right) \to 0$.

7.A.4 Proof of Theorem 7.4

Denote the kth iterate of the algorithm by a_k. That is, $a_{k+1} = a_k + 2h_{T_1,T_2}(\nu)$. Let $a_1 \in A_{T_1}(\nu)$ where $\nu > 1$. From (7.20), putting $a' = \widehat{a}_{T_1}$,

$$\begin{aligned} h_{T_1,T_2}(a) &= -\frac{1}{2}(a - \widehat{a}_{T_1}) + \frac{1}{2}(a - \widehat{a}_{T_1}) O\left\{T_1^{-1/2}(\log T_1)^{1/2}\right\} \\ &= -\frac{1}{2}(a - \widehat{a}_{T_1})\left[1 + O\left\{T_1^{-1/2}(\log T_1)^{1/2}\right\}\right]. \end{aligned} \tag{7.22}$$

Thus

$$\begin{aligned} 2h_{T_1,T_2}(a) + a - \widehat{a}_{T_1} &= (a - \widehat{a}_{T_1}) O\left\{T_1^{-1/2}(\log T_1)^{1/2}\right\} \\ &= O\left\{T_1^{-1/2-\nu}(\log T_1)^{1/2}\right\}. \end{aligned}$$

But $2h_{T_1,T_2}(a_j) + a_j = a_{j+1}$ and so

$$a_{j+1} - \widehat{a}_{T_1} = (a_j - \widehat{a}_{T_1}) O\left\{T_1^{-1/2}(\log T_1)^{1/2}\right\}.$$

Thus the theorem is true for $\nu > 1$ since

$$\lfloor 3 - \log(2\nu - 1)/\log 2 \rfloor = 2.$$

If $1/2 < \nu \leqslant 1$ then, from the proof of Theorem 17 of Quinn and Hannan (2001),

$$e_{T_1,X}(a) = \frac{\rho_X^2 T_1}{4(\omega - \omega_0)^2}\left[1 - \frac{\sin\{T_1(\omega - \omega_0)\}}{T_1(\omega - \omega_0)}\right]\left[1 + O\left\{T_1^{1/2-\nu}(\log T_1)^{1/2}\right\}\right],$$

$$e_{T_2,Y}(a) = \frac{\rho_X^2 \kappa T_1}{4(\omega - \omega_0)^2}\left[1 - \frac{\sin\{(\omega - \omega_0)T_1\}}{2(\omega - \omega_0)T_1}\right]\left[1 + O\left\{T_1^{1/2-\nu}(\log T_1)^{1/2}\right\}\right],$$

and

$$d_{T_1,X}(a) = \frac{\rho_X^2 T_1}{4(\omega - \omega_0)}\left[1 - \frac{\sin\{T_1(\omega - \omega_0)\}}{T_1(\omega - \omega_0)}\right]\left\{1 + O\left(T_1^{-\nu}\right)\right\} + O\left\{T_1^{3/2}(\log T_1)^{1/2}\right\},$$

$$d_{T_2,Y}(a) = \frac{\rho_Y^2 \kappa T_1}{4(\omega - \omega_0)}\left[1 - \frac{\sin\{T_1(\omega - \omega_0)\}}{T_1(\omega - \omega_0)}\right]\left\{1 + O\left(T_1^{-\nu}\right)\right\} + O\left\{T_1^{3/2}(\log T_1)^{1/2}\right\}.$$

Thus

$$\begin{aligned} &e_{T_1,X}(a) + \lambda e_{T_2,Y}(a) \\ &= \frac{T_1}{4(\omega - \omega_0)^2}\left(\rho_X^2 + \lambda\kappa\rho_Y^2\right)\left[1 - \frac{\sin\{T_1(\omega - \omega_0)\}}{T_1(\omega - \omega_0)}\right]\left[1 + O\left\{T_1^{1/2-\nu}(\log T_1)^{1/2}\right\}\right] \end{aligned}$$

and

$$\begin{aligned} &d_{T_1,X}(a) + \lambda d_{T_2,Y}(a) \\ &= \frac{T_1}{4(\omega - \omega_0)}\left(\rho_X^2 + \lambda\kappa\rho_Y^2\right)\left[1 - \frac{\sin\{T_1(\omega - \omega_0)\}}{T_1(\omega - \omega_0)}\right]\left[1 + O\left\{T_1^{1/2-\nu}(\log T_1)^{1/2}\right\}\right]. \end{aligned}$$

Also, $\left\{e^{-1}_{T_1,X}(a)+\lambda^{-1}e^{-1}_{T_2,Y}(a)\right\}^{-1}$ is $O\left(T_1^{-1-2\nu}\right)$ and it follows that

$$h_{T_1,T_2}(a)=-\frac{1}{2}(a-\widehat{a}_{T_1})\left[1+O\left\{T_1^{1/2-\nu}(\log T_1)^{1/2}\right\}\right]+O\left\{T_1^{1/2-2\nu}(\log T_1)^{1/2}\right\}.$$

Thus

$$2h_{T_1,T_2}(a)+a-\widehat{a}_{T_1}=(a-\widehat{a}_{T_1})\,O\left\{T_1^{1/2-\nu}(\log T_1)^{1/2}\right\}+O\left\{T_1^{1/2-2\nu}(\log T_1)^{1/2}\right\}$$

and so

$$a_{j+1}-\widehat{a}_{T_1}=(a_j-\widehat{a}_{T_1})\,O\left\{T_1^{1/2-\nu}(\log T_1)^{1/2}\right\}+O\left\{T_1^{1/2-2\nu}(\log T_1)^{1/2}\right\}.$$

Let $n=\nu$, then $n_{j+1}=2n_j-1/2+\delta$, where δ is arbitrarily small. Then

$$n_k=2^{k-1}\left(\nu-\frac{1}{2}\right)+\frac{1}{2}+\delta$$

and $n_k>1$ when

$$k>1-\frac{\log(2\nu-1)}{\log 2}.$$

The next iterate is then $o\left(T_1^{-3/2}\right)$. If $\nu=1$, then $a_3-\widehat{a}_{T_1}=o\left(T_1^{-3/2}\right)$.

7.A.5 Proof of Theorem 7.5

From (7.22),

$$h_{T_1,T_2}(a_0)=\frac{1}{2}(\widehat{a}_{T_1}-a_0)\left[1+O\left\{T_1^{-1/2}(\log T_1)^{1/2}\right\}\right]$$

and so

$$\widehat{a}_{T_1}-a_0=2h_{T_1}(a_0)+o\left(T^{-3/2}\right).$$

Since

$$\widehat{a}_{T_1}-a_0=2\cos\widehat{\omega}-2\cos\omega_0=-2\sin\omega_0(\widehat{\omega}-\omega_0)+o\left(T_1^{-2}\right),$$

it follows that $\widehat{\omega}-\omega_0$ has the same asymptotic distribution as

$$-\frac{d_{T_1,X}(a_0)+\lambda d_{T_2,Y}(a_0)}{e_{T_1,X}(a_0)+\lambda e_{T_2,Y}(a_0)}.$$

Now, from (7.18) and (7.19),

$$T_1^{-3}\left\{e_{T_1,X}(a_0)+e_{T_2,Y}(a_0)\right\}\to\frac{\rho_X^2+\lambda\kappa^3\rho_Y^2}{24}.$$

Also,

$$T_1^{-3/2}\left\{d_{T_1,X}(a_0)+\lambda d_{T_2,Y}(a_0)\right\}$$

has the same asymptotic distribution as

$$T_1^{-3/2}\frac{\alpha_{X0}}{2}\sum_{t=0}^{T_1-1}(2t-T)\sin(t\omega_0)E_t - T_1^{-3/2}\frac{\beta_{X0}}{2}\sum_{t=0}^{T_1-1}(2t-T)\cos(t\omega_0)E$$
$$+T_1^{-3/2}\frac{\lambda\kappa^{3/2}\alpha_{Y0}}{2}\sum_{t=0}^{T_1-1}(2t-T)\sin(t\omega_0)U_t - T_1^{-3/2}\frac{\lambda\kappa^{3/2}\beta_{Y0}}{2}\sum_{t=0}^{T_1-1}(2t-T)\cos(t\omega_0)E_t.$$

Using Theorem 2.6 and the same argument as in the proof of Theorem 6.2, this is asymptotically normal with mean zero and covariance matrix

$$\frac{2\pi\rho_X^2 f_\varepsilon(\omega_0)}{24}+\frac{2\pi\lambda^2\kappa^3\rho_Y^2 f_u(\omega_0)}{24}.$$

Thus $T_1^{3/2}(\widehat{\omega}-\omega_0)$ converges to the normal distribution with mean zero and covariance matrix

$$48\pi\left\{\frac{\rho_X^2}{f_E(\omega_0)}+\frac{\kappa^3\rho_Y^2}{f_U(\omega_0)}\right\}^{-1}$$

as $T_1\to\infty$ since, under $H_0^{(1)}$,

$$\lambda=\frac{f_E(\omega_0)}{f_U(\omega_0)}.$$

7.A.6 Proof of Theorem 7.6

Let

$$\theta=\left[\begin{array}{cccccc}\theta_1' & \theta_2' & \sigma_\varepsilon^2 & \sigma_u^2 & \theta_3' & \theta_4'\end{array}\right],$$

where $\theta_1=\delta_X$, $\theta_2=\omega_X$, $\theta_3=\delta_Y-\delta_X$ and $\theta_4=\omega_Y-\omega_X$. The hypothesis test is then

$$H_0:\left[\begin{array}{cc}\theta_3' & \theta_4'\end{array}\right]'=0$$
$$H_A:\left[\begin{array}{cc}\theta_3' & \theta_4'\end{array}\right]'\neq 0.$$

The test statistic is

$$\Lambda_f=2\left\{\sup_\theta l(\theta)-\sup_{\theta_3=\theta_4=0} l(\theta)\right\},$$

where

$$l(\theta)=l_X\left(\omega_X,\theta_X,\delta_X,\sigma_\varepsilon^2\right)+l_Y\left(\omega_Y,\theta_Y,\delta_Y,\sigma_u^2\right),$$

which is given by (7.3) and (7.4). Let $\theta_*=\left[\begin{array}{cccc}\theta_1' & \theta_2' & \sigma_\varepsilon^2 & \sigma_u^2\end{array}\right]$ and denote the true value of θ under H_0 by $\theta_0=\left[\begin{array}{cc}\theta_{*0}' & 0\end{array}\right]'$. The estimators under H_0 and H_A, denoted $\widehat{\theta}_0$ and $\widehat{\theta}_A$, respectively, satisfy

$$0=\frac{\partial l\left(\widehat{\theta}_A\right)}{\partial\theta}\quad\text{and}\quad 0=\frac{\partial l\left(\widehat{\theta}_0\right)}{\partial\theta},$$

where $\widehat{\theta}_0 = \left[\begin{array}{cc} \widehat{\theta}'_{*0} & 0 \end{array} \right]'$. From the mean value theorem, letting θ_j, θ_{*j} and θ_{Aj} denote the jth components of θ, θ_* and θ_A, respectively,

$$0 = \frac{\partial l\left(\widehat{\theta}_A\right)}{\partial \theta_j} = \frac{\partial l\left(\theta_0\right)}{\partial \theta_j} + \frac{\partial^2 l\left(\widetilde{\theta}_{Aj}\right)}{\partial \theta_j \partial \theta'}\left(\widehat{\theta}_A - \theta_0\right) \tag{7.23}$$

and

$$0 = \frac{\partial l\left(\widehat{\theta}_0\right)}{\partial \theta_{*j}} = \frac{\partial l\left(\theta_0\right)}{\partial \theta_{*j}} + \frac{\partial^2 l\left(\widetilde{\theta}_{0j}\right)}{\partial \theta_{*j} \partial \theta'_*}\left(\widehat{\theta}_{*0} - \theta_{*0}\right), \tag{7.24}$$

where the $\widetilde{\theta}_{Aj}$ are points on the line segment between θ_0 and $\widehat{\theta}_A$ and the $\widetilde{\theta}_{0j}$ are points on the line segment between θ_0 and $\widehat{\theta}_0$. Now, $\delta_X = \theta_1$, $\delta_Y = \theta_1 + \theta_3$, $\omega_X = \theta_2$ and $\omega_Y = \theta_2 + \theta_4$. Thus the first derivatives of $l\left(\theta\right)$ with respect to $\theta_1, \ldots, \theta_4$ at θ_0 are

$$\begin{aligned}
\frac{\partial l\left(\theta_0\right)}{\partial \theta_1} &= \frac{\partial l\left(\theta_0\right)}{\partial \delta_X} + \frac{\partial l\left(\theta_0\right)}{\partial \delta_Y} = -T_1^{1/2} z_1 - T_2^{1/2} z_2, \\
\frac{\partial l\left(\theta_0\right)}{\partial \theta_2} &= \frac{\partial l\left(\theta_0\right)}{\partial \omega_X} + \frac{\partial l\left(\theta_0\right)}{\partial \omega_Y} = -T_1^{3/2} w_1 - T_2^{3/2} w_2, \\
\frac{\partial l\left(\theta_0\right)}{\partial \theta_3} &= \frac{\partial l\left(\theta_0\right)}{\partial \delta_Y} = -T_2^{1/2} z_2, \\
\frac{\partial l\left(\theta_0\right)}{\partial \theta_4} &= \frac{\partial l\left(\theta_0\right)}{\partial \omega_Y} = -T_2^{3/2} w_2,
\end{aligned}$$

where

$$z'_1 = \frac{T_1^{-1/2}}{\sigma_\varepsilon^2} \sum_{t=0}^{T_1-1} \varepsilon_t \left[\begin{array}{c} X_{t-1} \\ \vdots \\ X_{t-p} \end{array} \right], \qquad z'_2 = \frac{T_2^{-1/2}}{\sigma_u^2} \sum_{t=0}^{T_2-1} u_t \left[\begin{array}{c} Y_{t-1} \\ \vdots \\ Y_{t-p} \end{array} \right],$$

and w_1 and w_2 are the $f \times 1$ vectors with jth element

$$-\frac{T_1^{-3/2}}{\sigma_\varepsilon^2} \sum_{t=0}^{T_1-1} \varepsilon_t \left\{ \alpha_{X,j} t \sin\left(\omega_{X0,j} t\right) - \beta_{X,j} t \cos\left(\omega_{X0,j} t\right) \right\}$$

and

$$-\frac{T_2^{-3/2}}{\sigma_u^2} \sum_{t=0}^{T_2-1} u_t \left\{ \alpha_{Y,j} t \sin\left(\omega_{Y0,j} t\right) - \beta_{Y,j} t \cos\left(\omega_{Y0,j} t\right) \right\},$$

respectively. That is,

$$\frac{\partial l\left(\theta_0\right)}{\partial \theta} = -N_{T_1,T_2} Z,$$

where

$$N_{T_1,T_2} = \left[\begin{array}{ccccc}
T_1^{1/2} I_p & 0 & 0 & T_2^{1/2} I_p & 0 \\
0 & T_1^{3/2} I_f & 0 & 0 & T_2^{3/2} I_f \\
0 & 0 & I_2 & 0 & 0 \\
0 & 0 & 0 & T_2^{1/2} I_p & 0 \\
0 & 0 & 0 & 0 & T_2^{3/2} I_f
\end{array} \right],$$

$$Z = \left[\begin{array}{ccccc} z_1' & w_1' & a' & z_2' & w_2' \end{array} \right]'$$

and

$$a = \left[\begin{array}{cc} \frac{\partial l(\theta_0)}{\partial \sigma_\varepsilon^2} & \frac{\partial l(\theta_0)}{\partial \sigma_u^2} \end{array} \right]'.$$

Now, $\partial^2 l\left(\theta_0\right)/\partial \omega_{X,j} \partial \omega_{X,k}$ is equal to

$$\begin{aligned}
&-\frac{1}{\sigma_\varepsilon^2} \sum_{t=0}^{T_1-1} \left\{\alpha_{X,j} t \sin\left(\omega_{X0,j} t\right) - \beta_{X,j} t \cos\left(\omega_{X0,j} t\right)\right\} \left\{\alpha_{X,k} t \sin\left(\omega_{X0,k} t\right) - \beta_{X,k} t \cos\left(\omega_{X0,k} t\right)\right\} \\
&= O\left(T^2\right),
\end{aligned}$$

if $j \neq k$, and equal to

$$\begin{aligned}
&-\frac{1}{\sigma_\varepsilon^2} \sum_{t=0}^{T_1-1} \left\{\alpha_{X,j} t \sin\left(\omega_{X0,j} t\right) - \beta_{X,j} t \cos\left(\omega_{X0,j} t\right)\right\}^2 \\
&-\frac{1}{\sigma_\varepsilon^2} \sum_{t=0}^{T_1-1} \varepsilon_t \left\{\alpha_{X,j} t^2 \cos\left(\omega_{X0,j} t\right) + \beta_{X,j} t^2 \sin\left(\omega_{X0,j} t\right)\right\},
\end{aligned}$$

if $j = k$. Let d_X and d_Y be the $f \times f$ matrices with jth diagonal elements

$$\begin{aligned}
T_1^{-3} \sum_{t=0}^{T_1-1} \Big[&\left\{\alpha_{X,j} t \sin\left(\omega_{X0,j} t\right) - \beta_{X,j} t \cos\left(\omega_{X0,j} t\right)\right\}^2 \\
&+\varepsilon_t \left\{\alpha_{X,j} t^2 \cos\left(\omega_{X0,j} t\right) + \beta_{X,j} t^2 \sin\left(\omega_{X0,j} t\right)\right\}\Big]
\end{aligned}$$

and

$$\begin{aligned}
T_2^{-3} \sum_{t=0}^{T_2-1} \Big[&\left\{\alpha_{Y,j} t \sin\left(\omega_{Y0,j} t\right) - \beta_{Y,j} t \cos\left(\omega_{Y0,j} t\right)\right\}^2 \\
&+u_t \left\{\alpha_{Y,j} t^2 \cos\left(\omega_{Y0,j} t\right) + \beta_{Y,j} t^2 \sin\left(\omega_{Y0,j} t\right)\right\}\Big],
\end{aligned}$$

respectively, and all other elements equal to $o(1)$. Let C_X and C_Y be the $p \times p$ matrices with (i,j)th elements

$$T_1^{-1} \sum_{t=p}^{T_1-1} X_{t-i} X_{t-j} \qquad \text{and} \qquad T_2^{-1} \sum_{t=p}^{T_2-1} Y_{t-i} Y_{t-j},$$

respectively. The second derivatives of $l(\theta)$ with respect to $\theta_1, \ldots, \theta_4$ at θ_0 are therefore

$$\begin{aligned}
\frac{\partial^2 l\left(\theta_0\right)}{\partial \theta_1 \partial \theta_1'} &= \frac{\partial^2 l\left(\theta_0\right)}{\partial \delta_X \partial \delta_X'} + \frac{\partial^2 l\left(\theta_0\right)}{\partial \delta_Y \partial \delta_Y'} = -\frac{T_1}{\sigma_\varepsilon^2} C_X - \frac{T_2}{\sigma_u^2} C_Y, \\
\frac{\partial^2 l\left(\theta_0\right)}{\partial \theta_2 \partial \theta_2'} &= \frac{\partial^2 l\left(\theta_0\right)}{\partial \omega_X \partial \omega_X'} + \frac{\partial^2 l\left(\theta_0\right)}{\partial \omega_Y \partial \omega_Y'} = -\frac{T_1^3}{\sigma_\varepsilon^2} d_X - \frac{T_2^3}{\sigma_u^2} d_Y, \\
\frac{\partial^2 l\left(\theta_0\right)}{\partial \theta_3 \partial \theta_3'} &= \frac{\partial^2 l\left(\theta_0\right)}{\partial \theta_1 \partial \theta_3'} = \frac{\partial^2 l\left(\theta_0\right)}{\partial \theta_3 \partial \theta_1'} = -\frac{T_2}{\sigma_u^2} C_Y,
\end{aligned}$$

$$\frac{\partial^2 l(\theta_0)}{\partial\theta_4\partial\theta_4'} = \frac{\partial^2 l(\theta_0)}{\partial\theta_2\partial\theta_4'} = \frac{\partial^2 l(\theta_0)}{\partial\theta_4\partial\theta_2'} = -\frac{T_2^3}{\sigma_u^2} d_X,$$

with all other second derivatives equal to zero. Using a second order Taylor expansion of $l\left(\widehat{\theta}_A\right)$ around θ_0, and since $\widehat{\theta}_A \to \theta_0$, $l\left(\widehat{\theta}_A\right) - l(\theta_0)$ has the same asymptotic properties as

$$\frac{\partial l(\theta_0)}{\partial\theta'}\left(\widehat{\theta}_A - \theta_0\right) + \frac{1}{2}\left(\widehat{\theta}_A - \theta_0\right)' \frac{\partial^2 l(\theta_0)}{\partial\theta\partial\theta'}\left(\widehat{\theta}_A - \theta_0\right),$$

which, because of (7.23), is asymptotically equivalent to

$$\begin{aligned}
&-\frac{1}{2}\frac{\partial l(\theta_0)}{\partial\theta'}\left\{\frac{\partial^2 l(\theta_0)}{\partial\theta\partial\theta'}\right\}^{-1}\frac{\partial l(\theta_0)}{\partial\theta} \\
&= -\frac{1}{2}Z'\left\{N_{T_1,T_2}^{-1}\frac{\partial^2 l(\theta_0)}{\partial\theta\partial\theta'}\left(N_{T_1,T_2}^{-1}\right)'\right\}^{-1} Z \\
&= \frac{1}{2}Z'\begin{bmatrix}
\frac{1}{\sigma_\varepsilon^2}C_X & 0 & 0 & 0 & 0 \\
0 & \frac{1}{\sigma_\varepsilon^2}d_X & 0 & 0 & 0 \\
0 & 0 & A & 0 & 0 \\
0 & 0 & 0 & \frac{1}{\sigma_u^2}C_Y & 0 \\
0 & 0 & 0 & 0 & \frac{1}{\sigma_u^2}d_Y
\end{bmatrix}^{-1} Z,
\end{aligned}$$

where

$$A = \begin{bmatrix} \frac{\partial^2 l(\theta_0)}{\partial(\sigma_\varepsilon^2)^2} & 0 \\ 0 & \frac{\partial^2 l(\theta_0)}{\partial(\sigma_u^2)^2} \end{bmatrix}.$$

Similarly, from (7.24), $l\left(\widehat{\theta}_0\right) - l(\theta_0)$ has the same asymptotic properties as

$$\frac{1}{2}Z'\begin{bmatrix}
\frac{1}{\sigma_\varepsilon^2}C_X & 0 & 0 & 0 & 0 \\
0 & \frac{1}{\sigma_\varepsilon^2}d_X & 0 & 0 & 0 \\
0 & 0 & A & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix} Z.$$

Thus Λ_f has the same asymptotic distribution as

$$\begin{bmatrix} z_2' & w_2' \end{bmatrix}\begin{bmatrix} \left(\frac{1}{\sigma_u^2}C_Y\right)^{-1} & 0 \\ 0 & \left(\frac{1}{\sigma_u^2}d_Y\right)^{-1} \end{bmatrix}\begin{bmatrix} z_2 \\ w_2 \end{bmatrix}.$$

But z_2 is asymptotically normal with mean zero and covariance matrix Ω and

$$\frac{1}{\sigma_u^2}C_Y \to \Omega,$$

where Ω is the $p \times p$ matrix with (i,j)th element $\gamma_Y(|i-j|)/\sigma_u^2$. Also, from Theorem 2.6, w_2 is asymptotically normal with mean zero and covariance matrix V, where V is the $f \times f$

diagonal matrix with jth diagonal element

$$\frac{\alpha_{Y,j}^2 + \beta_{Y,j}^2}{6\sigma_u^2}.$$

Now, from (6.27),

$$\sum_{t=0}^{T_1-1} t^2 u_t e^{i\omega_{Y0,j}t} = O\left\{T_2^{5/2} (\log T_2)^{1/2}\right\}$$

and so the jth diagonal element of d_Y is equal to

$$\begin{aligned} &T_2^{-3} \sum_{t=0}^{T_2-1} \{\alpha_{Y,j} t \sin(\omega_{Y0,j}t) - \beta_{Y,j} t \cos(\omega_{Y0,j}t)\}^2 + o(1) \\ &= \frac{1}{6}\left(\alpha_{Y,j}^2 + \beta_{Y,j}^2\right) + o(1). \end{aligned}$$

Thus

$$\frac{1}{\sigma_u^2} d_Y \to V,$$

and Λ_f has asymptotically the χ^2 distribution with $p+f$ degrees of freedom, since $\dim \Omega = p$ and $\dim V = f$.

8
Conclusion

In this thesis we have developed new methods for discriminating between time series on the basis of the spectral densities of their underlying processes. The general approach we have taken is to fit parametric models to the time series and then derive test statistics using a pseudo-likelihood ratio procedure. The approach differs from existing methods which are nonparametric. We first considered the case of discriminating between time series from univariate stationary processes, which is the case considered in most of the existing literature. We then showed how the approach can be extended to discriminating between time series from a wider range of processes.

In Chapter 3 we developed a test for discriminating between univariate stationary processes based on fitting autoregressions to the time series and comparing model parameters. Parameters were estimated by maximising Gaussian log-likelihood functions, and it was shown that the resulting estimators are strongly consistent and follow a central limit theorem even when the processes are not Gaussian. Since we do not wish to assume that the processes truly are autoregressive, we proposed fitting fixed order autoregressions to the time series where the autoregressive orders are a function of the sample sizes. It was shown that this fixed order autoregressive approximation is effective even when the time series are not from autoregressions. In a simulation study, the parametric test had higher empirical power than nonparametric methods.

In Chapter 4 we extended the pseudo-likelihood ratio test of Chapter 3 to fitting autoregressive-moving average (ARMA) models. It was shown using simulations that if the autoregressive and moving average orders are known, then fitting ARMA models improves the power of the test. However, if the orders are not known and are estimated, the test that fits fixed order autoregressions performs better. A new procedure was developed which fits ARMA models to two time series with the same model parameters, which was based on an extension of the Hannan–Rissanen procedure. This procedure may have applications in, for example, fitting ARMA models to repeatedly observed time series. A topic of future research therefore will be to generalise this procedure for the case of more than two time series and to study its asymptotic properties.

In Chapter 5 we considered methods for comparing two or more time series from multivariate stationary processes. Three null hypotheses were considered. The first was that the time series are from vector autoregressions with the same autoregressive parameters. Unlike the univariate case, this does not necessarily mean that the processes have the same spectral shape. The second was that the time series are from stationary vector processes with spectral densities which differ only in scale. The third was that the time series are from stationary vector processes where each of the corresponding vector components has the same spectral shape. Techniques were given for estimating the parameters under each of the null hypotheses which were based on maximising Gaussian likelihood functions. It was shown that the estimators are strongly consistent even when the processes are not Gaussian. The tests performed well in simulations, although in some cases fairly large sample sizes were required for the Type I error rate to be at the significance level. This was especially the case when the time series were of high dimension.

In Chapter 6 we considered the estimation of frequency in time series from processes which contain periodic components. In particular, we developed new procedures for estimating frequency in the multichannel sinusoidal model. The estimation procedures we developed were motivated by maximising log-likelihood functions assuming that the stationary components are Gaussian and white. We then showed that the resulting estimators are strongly consistent and follow central limit theorems even when the stationary components are not Gaussian, and coloured. We presented a multivariate version of the Quinn–Fernandes technique, which can be incorporated into the frequency estimation procedures and is computationally faster than, for example, maximising the Gaussian white log-likelihood using the Gauss–Newton algorithm. The estimator produced by the Quinn–Fernandes technique was shown to be strongly consistent and follow the same central limit theorem as the maximiser of the Gaussian white log-likelihood. The estimation procedures performed well in simulations, particularly with relatively large samples and high signal-to-noise ratios. The simulations demonstrated where the threshold effect occurs as the signal-to-noise ratio decreases, for different sample sizes and different noise processes.

The procedures developed in Chapter 6 are for estimating a number of independent frequencies in the multichannel sinusoidal model. The procedures may not work if two frequencies are close together. New estimation techniques will be required for this case, along the lines of Hannan and Quinn (1989). Also not considered here is where the frequencies are all harmonics of a fundamental frequency, which was considered by Quinn and Thomson (1991) for the univariate case. These will be topics for future research.

In Chapter 7 we developed procedures for discriminating between two or more time series from processes which contain periodic components. Both the univariate and multivariate cases were considered, and the procedures incorporated the work of each of the previous chapters. The tests that were produced were generally shown to have good power properties for detecting differences in the fixed frequencies of two time series with periodic components in the presence of noise. Some strengthening of the algorithms may be possible, since a small number of negative test statistics were produced.

The thesis has established a general framework for a parametric approach to time series discrimination. Future research could look at applying this framework to further classes of models. For example, ARMA models with exogenous variables or linear regression models with ARMA errors. Other cases which could be of practical importance are where the time series are sampled at different time intervals or where the time series have missing values.

References

Abramowitz, M. and Stegun, I. A., editors (1965). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover Publications, New York.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, 21:243–247.

Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, 30(1):9–14.

An, H.-Z., Chen, Z.-G., and Hannan, E. J. (1982). Autocorrelation, autoregression and autoregressive approximation. *The Annals of Statistics*, 10(3):926–936.

An, H.-Z., Chen, Z.-G., and Hannan, E. J. (1983). The maximum of the periodogram. *Journal of Multivariate Analysis*, 13(3):383–400.

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society*, 8:27–41.

Bartlett, M. S. (1954). Problèmes de l'analyse spectrale des séries temporelles stationnaires. *Publ. Inst. Statist. Univ. Paris*, 3:119–134.

Bassily, H., Lund, R., and Wagner, J. (2009). Fault detection in multivariate signals with applications to gas turbines. *IEEE Transactions on Signal Processing*, 57:835–842.

Billingsley, P. (1961). The Lindeberg-Lévy theorem for martingales. *Proceedings of the American Mathematical Society*, 12(5):788–792.

Bowden, R. S. and Clarke, B. R. (2012). A single series representation of multiple independent ARMA processes. *Journal of Time Series Analysis*, 33(2):304–311.

Bowden, R. S. and Clarke, B. R. (2017). Using multivariate time series methods to estimate location and climate change effects on temperature readings employed in electricity demand simulation. *Australian & New Zealand Journal of Statistics*, 59(4):413–431.

Coates, D. S. and Diggle, P. J. (1986). Tests for comparing two estimated spectral densities. *Journal of Time Series Analysis*, 7(1):7–20.

De Souza, P. and Thomson, P. (1982). LPC distance measures and statistical tests with particular reference to the likelihood ratio. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 30(2):304–315.

Decowski, J. and Li, L. (2015). Wavelet-based tests for comparing two time series with unequal lengths. *Journal of Time Series Analysis*, 36(2):189–208.

Dette, H., Kinsvater, T., and Vetter, M. (2011). Testing non-parametric hypotheses for stationary processes by estimating minimal distances. *Journal of Time Series Analysis*, 32(5):447–461.

Dieudonné, J. A. (1960). *Foundations of Modern Analysis*. Academic Press, New York.

Diggle, P. J. and Fisher, N. I. (1991). Nonparametric comparison of cumulative periodograms. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 40(3):423–434.

Durbin, J. (1959). Efficient estimation of parameters in moving-average models. *Biometrika*, 46:306–316.

Durbin, J. (1960). The fitting of time-series models. *Revue de l'Institut International de Statistique*, 28(3):233–244.

Fokianos, K. and Savvides, A. (2008). On comparing several spectral densities. *Technometrics*, 50(3):317–331.

Gentle, J. E. (2017). *Matrix Algebra : Theory, Computations and Applications in Statistics*. Springer, 2nd edition.

Grant, A. J. and Quinn, B. G. (2017). Parametric spectral discrimination. *Journal of Time Series Analysis*, 38(6):838–864.

Hannan, E. J. (1970). *Multiple Time Series*. Wiley, New York.

Hannan, E. J. (1973a). Central limit theorems for time series regression. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 26(2):157–170.

Hannan, E. J. (1973b). The estimation of frequency. *Journal of Applied Probability*, 10(3):510–519.

Hannan, E. J. (1979). The central limit theorem for time series regression. *Stochastic Processes and their Applications*, 9(3):281–289.

Hannan, E. J. and Kavalieris, L. (1984a). A method for autoregressive-moving average estimation. *Biometrika*, 71(2):273–280.

Hannan, E. J. and Kavalieris, L. (1984b). Multivariate linear time series models. *Advances in Applied Probability*, 16(3):492–561.

Hannan, E. J. and Mackisack, M. (1986). A law of the iterated logarithm for an estimate of frequency. *Stochastic Processes and their Applications*, 22(1):103–109.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2):190–195.

Hannan, E. J. and Quinn, B. G. (1989). The resolution of closely adjacent spectral lines. *Journal of Time Series Analysis*, 10(1):13–31.

Hannan, E. J. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69(1):81–94.

Jin, L. (2011). A data-driven test to compare two or multiple time series. *Computational Statistics & Data Analysis*, 55(6):2183–2196.

Jin, L. (2015). Comparing autocorrelation structures of multiple time series via the maximum distance between two groups of time series. *Journal of Statistical Computation and Simulation*, 85(17):3535–3548.

Jin, L. and Wang, S. (2016). A new test for checking the equality of the correlation structures of two time series. *Journal of Time Series Analysis*, 37(3):355–368.

Kakizawa, Y., Shumway, R. H., and Taniguchi, M. (1998). Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association*, 93(441):328–340.

Kavalieris, L. and Hannan, E. J. (1994). Determining the number of terms in a trigonometric regression. *Journal of Time Series Analysis*, 15(6):613–625.

Levinson, N. (1947). The Wiener RMS (root mean square) error criterion in filter design and prediction. *Journal of Mathematics and Physics*, 25:261–278.

Lu, K. and Li, L. (2013). On Fan's adaptive Neyman tests for comparing two spectral densities. *Journal of Statistical Computation and Simulation*, 83(9):1585–1601.

Lukacs, E. (1970). *Characteristic functions.* Griffin, London, 2nd edition.

Lund, R., Bassily, H., and Vidakovic, B. (2009). Testing equality of stationary autocovariances. *Journal of Time Series Analysis*, 30(3):332–348.

Maharaj, E. A. (1999). Comparison and classification of stationary multivariate time series. *Pattern Recognition*, 32(7):1129–1138.

Neudecker, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *Journal of the American Statistical Association*, 64(327):953–963.

Preuß, P. and Hildebrandt, T. (2013). Comparing spectral densities of stationary time series with unequal sample sizes. *Statistics & Probability Letters*, 83(4):1174–1183.

Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Academic Press, London.

Quinn, B. G. (1980). Order determination for a multivariate autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2):182–185.

Quinn, B. G. (1988). A note on AIC order determination for multivariate autoregressions. *Journal of Time Series Analysis*, 9(3):241–245.

Quinn, B. G. (1989). Estimating the number of terms in a sinusoidal regression. *Journal of Time Series Analysis*, 10(1):71–75.

Quinn, B. G. (2006). Statistical methods of spectrum change detection. *Digital Signal Processing*, 16(5):588–596.

Quinn, B. G. and Fernandes, J. M. (1991). A fast efficient technique for the estimation of frequency. *Biometrika*, 78(3):489–497.

Quinn, B. G. and Hannan, E. J. (2001). *The Estimation and Tracking of Frequency.* Cambridge University Press, Cambridge.

Quinn, B. G. and Kootsookos, P. J. (1994). Threshold behavior of the maximum likelihood estimator of frequency. *IEEE Transactions on Signal Processing*, 42(11):3291–3294.

Quinn, B. G., McKilliam, R. G., and Clarkson, I. V. L. (2008). Maximizing the periodogram. In *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, pages 1–5.

Quinn, B. G. and Thomson, P. J. (1991). Estimating the frequency of a periodic function. *Biometrika*, 78(1):65–74.

Ravishanker, N., Hosking, J. R. M., and Mukhopadhyay, J. (2010). Spectrum-based comparison of stationary multivariate time series. *Methodology and Computing in Applied Probability*, 12(4):749–762.

Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York.

Rice, J. A. and Rosenblatt, M. (1988). On frequency estimation. *Biometrika*, 75(3):477–484.

Sakai, H. (1993). The determination of the number of terms in a multichannel sinusoidal regression. *Journal of Time Series Analysis*, 14(6):621–628.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, 63(1):117–126.

Stout, W. F. (1970). The Hartman-Wintner law of the iterated logarithm for martingales. *The Annals of Mathematical Statistics*, 41(6):2158–2160.

Tugnait, J. K. (2013). Wireless user authentication via comparison of power spectral densities. *IEEE Journal on Selected Areas in Communications*, 31(9):1791–1802.

Tugnait, J. K. (2016). Comparing multivariate complex random signals: Algorithm, performance analysis and application. *IEEE Transactions on Signal Processing*, 64(4):934–947.

Walker, A. M. (1971). On the estimation of a harmonic component in a time series with stationary independent residuals. *Biometrika*, 58(1):21–36.

Wang, X. (1993). An AIC type estimator for the number of cosinusoids. *Journal of Time Series Analysis*, 14(4):433–440.

Whittle, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*, 50(1–2):129–134.

Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics*, 9(3):501–513.