Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE STATISTICHE

Ciclo 33

Settore Concorsuale: 13/D1 - STATISTICA

Settore Scientifico Disciplinare: SECS-S/01 - STATISTICA

FREQUENCY DOMAIN ANALYSIS OF STATIONARY TIME SERIES

Presentata da: Francesca Papagni

Coordinatore Dottorato

Monica Chiogna

Supervisore

Alessandra Luati

Esame finale anno 2021

Abstract

This thesis provides a necessary and sufficient condition for asymptotic efficiency of a nonparametric estimator of the generalised autocovariance function of a Gaussian stationary random process. The generalised autocovariance function is the inverse Fourier transform of a power transformation of the spectral density, and encompasses the traditional and inverse autocovariance functions. Its nonparametric estimator is based on the inverse discrete Fourier transform of the same power transformation of the pooled periodogram. The general result is then applied to the class of Gaussian stationary ARMA processes and its implications are discussed. We illustrate that for a class of contrast functionals and spectral densities, the minimum contrast estimator of the spectral density satisfies a Yule-Walker system of equations in the generalised autocovariance estimator. Selection of the pooling parameter, which characterizes the nonparametric estimator of the generalised autocovariance, controlling its resolution, is addressed by using a multiplicative periodogram bootstrap to estimate the finite-sample distribution of the estimator. A multivariate extension of recently introduced spectral models for univariate time series is considered, and an algorithm for the coefficients of a power transformation of matrix polynomials is derived, which allows to obtain the Wold coefficients from the matrix coefficients characterizing the generalised matrix cepstral models. This algorithm also allows the definition of the matrix variance profile, providing important quantities for vector time series analysis. A nonparametric estimator based on a transformation of the smoothed periodogram is proposed for estimation of the matrix variance profile.

Contents

Chapter 1

Introduction

Spectral analysis can be applied to any type of process which fluctuates in some form. Here the object of the analysis is the study of records from a random process.
The analysis of time series can be conducted by operating either in the time domain, or in the frequency domain. Spectral analysis, or frequency domain analysis, is equivalent to the time domain analysis, but it provides an alternative way of viewing the process, which can highlight some aspects of the series and be more illuminating for some applications (Brockwell, Davis, and Calder, 2002).
A time series $\{x_1, \ldots, x_N\}$ is a realization of a stochastic process $\{X_t\}, t \in T$, i.e. a sequence of random variables indexed by time, where T is a parametric space. A complete probabilistic time series model for the sequence of random variables $\{X_1, \ldots, X_n\}$ would specify all of the joint distributions of the random vectors $(X_1, X_2, \ldots, X_n)', \quad n = 1, 2, \ldots$. Since each time series is just one realization of the stochastic process, such a specification will contain far too many parameters to be estimated based on a small set of data, and is rarely used. Instead we specify only the first- and second-order moments of the joint distributions, i.e., the expected values $\mathbb{E}[X_t]$ and the expected products $\mathbb{E}[X_t X_{t+h}], t = 1, 2, \ldots, \quad h = 0, 1, 2, \ldots$, focusing on the properties of $\{X_t\}$ that depend only on these. This requires that the process is (weakly) stationary.
For this reason the autocovariance function plays an important role for characterizing a stationary process. The autocovariance function and its Fourier transform, the spectral density function, characterize the temporal dependence structure of a stationary stochas-

tic process, and are of fundamental importance in time series analysis and prediction. For Gaussian stationary processes they provide, along with the mean, a complete characterization of the probability distribution of the process as well as the basic ingredients for optimal (minimum mean square) prediction based on time series observations.
The autocovariance function is estimated nonparametrically by the sample autocovariance function. This estimator has a long tradition in time series analysis, and its properties are demonstrated and discussed in time series textbooks, such as, for instance, Brockwell and Davis (1991, ch. 7), where it is shown that under regularity conditions it has an asymptotically normal distribution and that the elements of the asymptotic covariance matrix are given by the celebrated Bartlett's formula.

The literature has further addressed the important question as to what classes of parametric linear processes admit the sample autocovariance as an asymptotically efficient estimator, i.e., an estimator whose variance achieves the Cramêr-Rao lower bound.

This issue has been investigated by Porat (1987) for Gaussian autoregressive (AR) moving average (MA) mixed processes, based on state-space representations and matrix Lyapunov equation theory. For Gaussian ARMA(r, q) processes with $r \geq q$ the sample autocovariances are asymptotically efficient only in a restricted number of cases, while if $q > r$ none of the sample autocovariances is asymptotically efficient. See also Walker (1995) for an alternative derivation of this result. The result implies that the variance and the first r autocovariances of a pure AR(r) process are efficiently estimated by the sample autocovariances, while for a pure MA process none of the sample autocovariances is asymptotically efficient.

Kakizawa and Taniguchi (1994) derived in the frequency domain a necessary and sufficient condition for asymptotic efficiency of the sample autocovariances that applies to the more general class of Gaussian stationary processes. Kakizawa (1999) extended the previous results to the case of vector processes. Boshnakov (2005) studied the efficiency of the sample autocovariances for processes obtained by a finite linear transformation of a pure autoregressive process.

The generalised autocovariance (GACV) function was defined in Proietti and Luati (2015) as the inverse Fourier transform of the pth power of the spectral density function. It encompasses the traditional autocovariance function ($p = 1$) and the inverse

autocovariance function (Cleveland, 1972), which is the sequence of the coefficients of the Fourier expansion of the inverse spectrum ($p = -1$).

In this thesis asymptotic efficiency of the nonparametric estimator of the GACV considered in Proietti and Luati (2015) is studied. Following Hannan and Nicholls (1977) and Luati et al. (2012), the estimator is based on powers of the pooled periodogram over m non-overlapping consecutive frequencies, where m is the *pooling parameter*. Proietti and Luati (2015) established consistency and asymptotic normality of the estimator.

We establish a necessary and sufficient condition for asymptotic efficiency in terms of the spectral density and its derivatives for general Gaussian stationary processes, which nests as a particular case the result of Kakizawa and Taniguchi (1994), which holds for $p = 1$.

After characterizing a class of processes for which the nonparametric estimator is fully efficient, we consider the case when the process is Gaussian ARMA(r, q). We show that asymptotic efficiency of the nonparametric estimator of the GACV for positive and negative integer powers p depends on the exisistence of a solution to a trigonometric polynomial equation. Investigation of conditions on the order of the trigonometric polynomials involved imply that if $p = 1$, the nonparametric estimator of the GACV is asymptotically efficient for $r \geq q$ and $0 \leq k \leq r - q$, which coincides with the result by Kakizawa and Taniguchi (1994). These results also show that the asymptotic variance of the nonparametric estimator achieves the Cramèr-Rao lower bound as $m \to \infty$ for $p = -1$ when $r = 0$ and $0 \leq k \leq q$, i.e. it estimates efficiently the first q inverse autocovariances when the true generating process is pure MA(q), thereby complementing the results by Bhansali (1980) and Battaglia (1988). The inverse autocovariance function is useful in interpolation problems and for the identification of ARMA models. The results obtained include as a special case the results for the sample autocovariance function by Porat (1987) and Kakizawa and Taniguchi (1994). Some numerical examples also illustrate the rate of convergence to the Cramèr-Rao bound.

We also illustrate that for a class of contrast functionals and spectral densities, the minimum contrast estimator of the spectral density satisfies a Yule-Walker system of equations in the generalised autocovariance estimator.

The pooling parameter m, which characterizes the nonparametric estimator of the GACV, plays a crucial role. It allows asymptotic efficiency of this estimator, since increasing m reduces the variance of the estimator. However, it inflates the finite-sample bias, determining a bias-variance trade-off. Its selection has been addressed in Proietti and Luati (2015) by the use of the Jackknife to estimate the MSE of the estimator. This thesis proposes the use of a multiplicative periodogram bootstrap (MPB) (Meyer, Paparoditis, and Kreiss, 2018) to estimate the MSE and to select the value of m that minimizes it. This procedure is motivated by viewing the estimator of the GACV as a periodogram-based estimator of a spectral mean. A real data application suggests that estimation of the GACV by the nonparametric estimator considered should proceed via a small value of $m > 1$.

Investigation of the properties of the generalised autocovariance function and its nonparametric estimator motivates further research to extend the generalised autocovariance function to the multivariate context, to study the properties of vector time series. Similar to the univariate case, the spectral density of a stationary vector process provides a complete characterization of the serial correlation structure of the process, and all information for prediction and interpolation. Frequency domain models for the spectrum of vector time series allow to avoid some difficulties arising with the time domain specification of models for stationary vector time series. Vector autoregressive moving-average models are often used to model vector stationary time series, and are specified in the time domain as difference equations. However, it requires to impose some restrictions on the coefficient matrices, which may increase computational cost. Frequency domain models for the matrix spectrum provide a recently extensively investigated alternative. In particular, Holan et al. (2017) defined the vector exponential model (VEXP), which extends to vector time series the exponential model by Bloomfield (1973). The VEXP model has some advantages over vector ARMA models, since it is stationary and invertible, with unconstrained parameters, allowed to take any real value. The VEXP model assumes that the logarithmic transformation of the spectral density matrix can be represented by a finite Fourier polynomial. The Fourier coefficients of the log-spectrum are called cepstral matrices, and their collection is the matrix cepstrum, in analogy with

the cepstral coefficients and the cepstrum for scalar time series. In the univariate case, the cepstral coefficients are related to the coefficients of the Wold representation of the process by a recursive formula by Pourahmadi (1983). A multivariate extension of this formula for the exponential and log transform of a matrix polynomial is non-intuitive, and is provided by Holan et al. (2017) to relate the cepstral matrices to the Wold coefficient matrices.

The VEXP models can be extended by defining the generalised matrix cepstral models Cavicchioli et al. (2020), which provide a multivariate version of the generalised cepstral models defined by Proietti and Luati (2019). The generalised matrix cepstral models specify a linear model for the Box-Cox transform of the spectrum, represented by a finite Fourier polynomial, characterized by the generalised cepstral matrices. These models include as special cases the VEXP model by Holan et al. (2017), and also AR and MA models, by varying the transformation parameter. In analogy with the generalised cepstral coefficients for scalar time series, the generalised cepstral matrices are connected to the multivariate extension of the generalised autocovariance function, which provides several important quantities for the analysis of vector time series. The generalised cepstral matrices are related to the Wold coefficient matrices that characterize the process. In the univariate case this relation is described by a recursive formula provided by (Gould, 1974), which is not easily generalised to the matrix coefficients in the multivariate case. This thesis provides an algorithm to obtain the coefficients of a power transformation of a matrix polynomial, which can be used to obtain the coefficients of the Wold representation of the process from the generalised cepstral matrices. The Wold coefficients are useful for forecasting and assessing goodness-of-fit of the model.

As will be illustrated in the following discussion, the generalised autocovariance function can be interpreted in terms of the traditional autocovariance of a power-transformed process, whose coefficients can be determined from the Wold coefficients characterizing the original process by a recursive formula by Gould (1974). Such a definition of an auxiliary process is useful to define and study the GACV and other quantities related to it, like the variance profile (Luati et al., 2012). Hence, to determine a multivariate extension of these quantities for vector time series, it is of interest to define a vector power-transformed process by its Wold coefficients. This aim is accomplished by determining a relation

between the coefficients of a given matrix polynomial and the coefficients of its power transformation, for any real power transform. Karampetakis and Tzekis (2005) derived a recursive relation for the coefficients of a power transformation of a matrix polynomial for positive integer powers, while Holan et al. (2017) derived these formulas for the exponential and logarithmic transform. Specification of the matrix variance profile in terms of an auxiliary process allows to give a useful interpretation to this measure, and analytical results in terms of the coefficients that characterize the process dynamics. The matrix variance profile nests as special cases the unconditional variance-covariance matrix of the process, the variance-covariance matrix of the one-step-ahead prediction error, and the variance-covariance matrix of the interpolation error. The power-transformed process is related with the definition of both the matrix variance profile and the generalised autocovariance matrix, and, if considering the inverse transform, it coincides with the inverse process mentioned in Heyse and Wei (1985) for the definition of the inverse autocovariance matrix function, which provides a useful tool for identification of ARMA models.

In Chapter 2 provides the basic concepts of spectral analysis, including the definition of the spectral density function, and the periodogram as its natural estimator. Asymptotic properties of the periodogram are presented, which suggest a smoothed version of the periodogram to overcome some drawbacks of the raw estimator of the spectrum. Some recent advances, including a description of the generalised cepstral models, and the definition of the generalised autocovariance function, and its nonparametric estimator are also presented in this chapter. A mathematical appendix shows analytical tools to express the asymptotic variance of the estimator of the GACV in terms of the spectral density function. Chapter 3 contains the main results of the project, concerning the asymptotic efficiency of the estimator of the GACV. A necessary and sufficient condition for asymptotic efficiency of the estimator is derived, and it is checked for the class of Gaussian ARMA processes. Selection of the pooling parameter is also addressed in this chapter by the use of a frequency domain bootstrap. The appendix at the end of the chapter shows the derivation of the Fisher information matrix for a Gaussian stationary process, which is used in the chapter to define the Cramèr-Rao bound. Also, the appendix derives the limit as $m \to \infty$ of a multiplicative factor that determines asymptotic efficiency of the

estimator. Chapter 4 concerns the multivariate extension of some concepts of spectral analysis introduced in Chapter 2 for scalar time series, including the generalised matrix cepstral models and the generalised autocovariance matrix. The relation between the coefficients of a given matrix polynomial and those of its power transformation, for any real power is investigates. This also leads to a multivariate extension of the variance profile, based on the definition of a power-transformed vector process. A nonparametric estimator of the matrix variance profile is proposed, based on a transformation of the smoothed periodogram matrix. Chapter 5 contains some concluding remarks, and possible future developments of the project.

Chapter 2

Spectral analysis

A time series can be analysed using two different approaches: one operating in the time domain, by means of the autocovariance function, and one operating in the frequency domain, in terms of the Fourier transform of the autocovariance function: the spectral density function.
Every stationary process admits both a time domain representation and a frequency domain representation, and its characteristics can be described equivalently by either the time domain or frequency domain approach.
Spectral analysis of time series refers to the analysis of stationary time series by means of their *spectral representation*, which decomposes a time series into a sum of sinusoidal components with uncorrelated random coefficients, that represent the latent components of the series. This allows to determine the relative contribution of each frequency component to the total variation of the process. The spectral representation of a time series is associated to the spectral decomposition of the autocovariance function.
We first introduce some important quantities for the spectral analysis of time series, which are at the basis of the definition of the generalised autocovariance function and its estimator.

2.1 The spectral density function

Let $\{X_t\}$ be a zero-mean stationary stochastic process with autocovariance function $\gamma_k = \mathbb{E}[X_t X_{t-k}]$, with $\gamma_k = \gamma_{-k}$, satisfying $\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty$. The spectral density function of $\{X_t\}$ is the function $f(\cdot)$ defined as the *Fourier transform* of the autocovariance function:

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-i\omega k} \gamma_k, \quad -\pi \leq \omega \leq \pi. \tag{2.1}$$

In turn, the autocovariance function is the *inverse Fourier transform* of the spectral density function,

$$\gamma_k = \int_{-\pi}^{\pi} f(\omega) e^{i\omega k} \, d\omega = \int_{-\pi}^{\pi} f(\omega) \cos \omega k \, d\omega, \tag{2.2}$$

that is, the autocovariances of a stationary time series with absolutely summable ACVF are the Fourier coefficients of the nonnegative even function $f(\cdot)$ defined in (2.1), (Brockwell et al., 2002).

For $k = 0$, one recovers $\gamma_0 = \int_{-\pi}^{\pi} f(\omega) \, d\omega$, the variance of the process. Thus, the spectral density function can be seen as a measure of the contribution to the total variance of the process, given by each periodic component. A large peak in the spectral density function, at a specific frequency ω indicates the presence in the time series of a strong sinusoidal component at (or near) ω.

The summability condition on the autocovariances implies that the spectral density function exists and is a real-valued continuous function of ω. By the Euler's formula:

$$e^{\pm i\omega} = \cos \omega \pm i \sin \omega, \tag{2.3}$$

and noting that for a stationary process $\gamma_k = \gamma_{-k}$, the spectral density function $f(\cdot)$ can also be written as:

$$f(\omega) = \frac{1}{2\pi} \{\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos \omega k\}, \quad -\pi \leq \omega \leq \pi. \tag{2.4}$$

Since the cosine function has period 2π, so also does $f(\omega)$, and it is also a nonnegative even function, i.e. $f(\omega) = f(-\omega)$, and $f(\omega) \geq 0$ for all $\omega \in (-\pi, \pi]$, and it suffices to confine attention to the values of f on the interval $[-\pi, \pi]$.

The autocovariance generating function $G_\gamma(z)$ represents a mathematical tool that allows us to derive an analytic formula for the spectral density function of linear processes, which is easy to compute. $G_\gamma(z)$ is defined as:

$$G_\gamma(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k = \gamma_0 + \sum_{k=1}^{\infty} \gamma_k (z^k + z^{-k}) \tag{2.5}$$

Hence, $\frac{1}{k!}\frac{\partial^k}{\partial L} G_\gamma(0) = \gamma_k$.
The autocovariance generating function evaluated at $e^{-i\omega}$ gives

$$G_\gamma(e^{-i\omega}) = \sum_{k=-\infty}^{\infty} \gamma_k (e^{-i\omega})^k = \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\omega k} = 2\pi f(\omega), \tag{2.6}$$

which is the spectral density function of the precess characterized by the autocovariance function γ_k.
Let us consider a zero-mean stationary process $\{X_t\}$, with Wold representation

$$X_t = \Psi(B)\epsilon_t = (1 + \psi_1 L + \psi_2 L^2 + \dots)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad \epsilon_t \sim WN(0, \sigma^2)$$

The autocovariance function of x_t is $\gamma_k = \sum_{j=0}^{\infty} \psi_j \psi_{j+k}$.
Then,

$$G_\gamma(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k = \sigma^2 \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} \psi_j \psi_{j+k} = \sigma^2 \Psi(L)\Psi(L^{-1}),$$

and thus, for the linear process $\{X_t\}$, we have

$$2\pi f(\omega) = G_\gamma(e^{-i\omega}) = \sigma^2 \Psi(e^{i\omega})\Psi(e^{-i\omega}) = \sigma^2 |\Psi(e^{-i\omega})|^2 \tag{2.7}$$

This formula provides a useful tool for the computation of the spectral density function of a specific model, by expressing it as a function of the parameters that govern the process.

One of the most popular estimators of the spectral density function is the periodogram, which is described next, and will be central in nonparametric estimation of the generalised autocovariance function.

2.2 The periodogram

If $\{X_t\}$ is a stationary time series with ACVF γ_k and spectral density $f(\cdot)$. Denote $\tilde{\gamma}_k$ the sample autocovariance function of the observations $\{x_1, x_2, \ldots, x_N\}$:

$$\tilde{\gamma}_k = \begin{cases} \frac{1}{N}\sum_{t=k+1}^{N}(x_t - \bar{x})(x_{t-k} - \bar{x}), & k = 0, \ldots N-1, \\ \tilde{\gamma}_{-k}, & k = -1, -2, \ldots, -N+1. \end{cases}$$

where $\bar{x}$ is the sample mean.
The periodogram of $\{x_1, x_2, \ldots, x_N\}$ is the function

$$I_N(\omega) = \frac{1}{2\pi N}\left|\sum_{t=1}^{N}(x_t - \bar{x})\exp(-i\omega t)\right|^2 \tag{2.8}$$

Then, just as the sample ACVF $\tilde{\gamma}_k$ of the observations $\{x_1, x_2, \ldots, x_N\}$, can be regarded as the sample analogue of γ_k, so also can the periodogram $I_N(\omega)$ of the observations be regarded as the sample analogue of $f(\omega)$ (Brockwell et al., 2002), as stated in the next proposition.

Proposition 1. *(Brockwell et al., 2002)*
If $\{x_1, x_2, \ldots, x_N\}$ are any real numbers and ω_k is any of the nonzero Fourier frequencies $\omega_k = \frac{2\pi k}{N}$ in $(-\pi, \pi]$, then

$$I_N(\omega_k) = \frac{1}{2\pi}\sum_{|h|<N}\tilde{\gamma}_h e^{-ih\omega_k}, \tag{2.9}$$

where $\tilde{\gamma}_h$ is the sample ACVF of $x_1, x_2, \ldots, x_N$. Hence $I_N(\omega)$ is a natural estimate of the spectral density $f(\omega)$.

The relationship between the periodogram and the sample autocovariance mimics that between the spectral density and the autocovariance function. $I_N(\omega)$ can equivalently be written as:

$$I_N(\omega) = \frac{1}{2\pi}\Big[\tilde{\gamma}_0 + 2\sum_{j=1}^{N-1}\tilde{\gamma}_j \cos\omega j\Big],$$

and the sample autocovariances are the Fourier coefficients of $I_N(\omega)$. Then,

$$\tilde{\gamma}_k = \int_{-\pi}^{\pi} I_N(\omega)\cos\omega k\, d\omega,$$

and the area under the periodogram gives the sample variance $\tilde{\gamma}_0$:

$$\tilde{\gamma}_0 = \int_{-\pi}^{\pi} I_N(\omega)\, d\omega$$

Similarly to the spectral density function, the periodogram is a nonnegative, even function, i.e. $I_N(-\omega) = I_N(\omega), \omega \in (0, \pi)$, with period 2π.
Some asymptotic distributional results for the periodogram are presented in the next section.

2.2.1 Large sample properties of the periodogram

Based on the previous discussion, one possible approach to estimate the spectral density function is provided by the periodogram of the observations. However, using the raw periodogram ordinates to estimate the spectral density function presents some limitations. Asymptotic distributional results for $I_N(\omega)$ are available. Fuller (1976) showed that for N sufficiently large, and $\omega \neq 0, \pi$:

$$\frac{2I_N(\omega)}{f(\omega)} \approx \chi^2(2) \tag{2.10}$$

where $\chi^2(2)$ is a Chi-square random variable with $df = 2$ degrees of freedom. Equivalently $\frac{I_N(\omega)}{f(\omega)}$ is approximately an exponential random variable with unit mean:

$$\frac{I_N(\omega)}{f(\omega)} \approx Exp(1)$$

While, at $\omega = 0, \pi$ the random variable $\frac{I_N(\omega)}{f(\omega)} \approx \chi^2(1)$ is approximately distributed as a Chi-square with 1 degree of freedom.
The random variables $\frac{2I_N(\omega)}{f(\omega)}$ and $\frac{2I_N(\lambda)}{f(\lambda)}$ are approximately independent.
By the properties of the Chi-square random variable with 2 degrees of freedom, or, by the properties of the standard exponential random variable, one has for $\omega \neq 0, \pi$:

$$\mathbb{E}[I_N(\omega)] \approx f(\omega)$$

Hence, the periodogram $I_N(\omega)$ is an asymptotically unbiased estimator of the spectral density function $f(\omega)$. However, it is an approximately unbiased estimator of the

spectral density function only for some processes and some frequencies. Moreover, the periodogram is not a consistent estimator of the spectral density function, since its variance does not reduce with increasing sample size N. Instead, as $N \to \infty$ its variance converges to a nonzero value as we are trying to estimate as many parameters as the observations available. To overcome this inconsistency problem in practice, smoothing techniques are often applied to the periodogram.

The periodogram is a central element in nonparametric estimation of the generalised autocovariance function, as will be illustrated in the following sections.
Based on the fundamental concepts of spectral analysis presented, let us introduce some recent advances in the context of frequency domain analysis of time series.

2.3 Recent advances

As illustrated in the previous chapter, the spectrum of a stationary process provides a complete characterization of the serial correlation structure of the process, and hence, all information needed for linear prediction and interpolation. One of the most popular models for the spectrum of a time series is the exponential model by Bloomfield (1973), which specifies a linear model for the log transform of the spectrum, represented by a finite Fourier polynomial. The Fourier coefficients of the expansion of the log-spectrum are called cepstral coefficients, and their collection forms the cepstrum, where cepstral and cepstrum are anagrams of spectral and spectrum (Bogert, 1963).
This idea is further developed by the introduction of the generalised linear cepstral models, defined by Proietti and Luati (2019).

2.3.1 Generalised linear cepstral models

The generalised linear cepstral models specify a linear model for the Box-Cox transformation of the spectral density $f(\cdot)$.
Let $\{X_t\}$ be a stationary stochastic process with autocovariance function $\gamma_k = \int_{-\pi}^{\pi} e^{i\omega k} f(\omega)\, d\omega$, $k = 0, \pm 1, \pm 2, \ldots$, where $f(\cdot)$ is the spectral density function, satisfying

$\int_{-\pi}^{\pi} \ln f(\omega)\, d\omega > -\infty$, and $\int_{-\pi}^{\pi} f(\omega)^{\lambda}\, d\omega < \infty$ for all $\lambda \in \mathbb{R}$. The Box-Cox transform of $2\pi f(\omega)$ is:

$$g(\omega) = \begin{cases} \frac{[2\pi f(\omega)]^{\lambda}-1}{\lambda}, & \lambda \neq 0, \\ \ln\left[2\pi f(\omega)\right], & \lambda = 0, \end{cases}$$

where $\lambda \in \mathbb{R}$. Assume that $g(\cdot)$ can be represented by a finite Fourier polynomial:

$$g(\omega) = c_{\lambda 0} + 2\sum_{k=1}^{K} c_{\lambda k} \cos \omega k \tag{2.11}$$

The generalised cepstral coefficients are defined as the inverse Fourier transform of $g(\omega)$:

$$c_{\lambda k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\omega) \cos\left(\omega k\right) d\omega, \quad k = 0, 1, \ldots, K. \tag{2.12}$$

For $\lambda = 0$ we obtain the exponential model for the spectrum (Bloomfield, 1973). c_{0k} are the cepstral coefficients, and their collection is the cepstrum (Bogert, 1963).

The generalised linear cepstral models nest several spectral models: the exponential model (obtained for the logarithmic link), the moving-average model (identity link), the autoregressive model (inverse link). The generalised cepstral coefficients are also used to evaluate the mutual information between the past and future of the process, and they are related to important characteristics of the process: $1/(1 - c_{-1,0})$ is the interpolation error variance, while $\exp\left(c_{0,0}\right)$ provides the one-step-ahead prediction error variance, by the Kolmogorov-Szego formula, and $c_{1,0} + 1 = \gamma_0$ is the unconditional variance of the process.

The generalised linear cepstral models can be estimated by maximization of the Whittle likelihood. Based on the distributional results for the periodogram $I(\omega)$ discussed in the previous section, the log-likelihood of the periodogram $I_N(\omega_j), j = 1, \ldots, [(N-1)/2]$ with $\omega_j = \frac{2\pi j}{N}$, is:

$$l(\lambda, \theta_\lambda) = -\sum_{j=1}^{N} \left[ln f(\omega_j) + \frac{I_N(\omega_j)}{f(\omega_j)} \right],$$

where $\theta_\lambda = (c_{\lambda,0}, c_{\lambda,1}, \ldots, c_{\lambda,K})'$ is the vector of parameters.

The maximum likelihood estimate of λ is obtained by maximization of the profile (partially maximized with respect to θ_λ) Whittle likelihood. The truncation parameter K is

selected by minimizing an information criterion, like AIC or BIC.

The generalised cepstral coefficients are linked to the generalised autocovariance function, defined as the inverse Fourier transform of the pth power of the spectral density function (Proietti and Luati, 2015), which is described in the next section.

2.3.2 The generalised autocovariance function

The generalised autocovariance function extends and includes as special cases the traditional autocovariance function and the inverse autocovariance function. Some of its main applications are in model identification, in cluster and discriminant analyses of time series, as it leads to a more general measure of distance. It also allows the definition of a class of White Noise tests with improved size and power properties, which can serve as goodness-of-fit tests. The GACV is related to the generalised cepstral coefficients (Proietti and Luati, 2019), which can be estimated by maximization of the Whittle likelihood. An alternative estimation method is represented by a nonparametric estimator, proposed in Proietti and Luati (2015), which we will consider in the next sections. First, let us provide the definition of the GACV.

Let $\{X_t\}_{t\in T}$, with $T \in \mathbb{N}$ a discrete time set, denote a zero mean stationary Gaussian process with autocovariance function $\gamma_k = \mathbb{E}(X_t X_{t-k}), k \in \mathbb{Z}$, and spectral density $f_\theta(\omega) = \sum_{k=-\infty}^{\infty} \gamma_k e^{-\imath\omega k}$, $\omega \in [-\pi, \pi]$, both depending on an $s \times 1$ vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)' \in \mathbb{R}^s$. Assume that the spectral density is positive, that $\int_{-\pi}^{\pi} \log f(\omega)\, d\omega > -\infty$, and that $\int_{-\pi}^{\pi} [f(\omega)]^p\, d\omega < \infty$.

For $p \in \mathbb{R}$ the generalised autocovariances, denoted γ_{pk}, are defined as the Fourier coefficients of $[2\pi f_\theta(\omega)]^p$ (Proietti and Luati, 2015), i.e.,

$$[2\pi f_\theta(\omega)]^p = \sum_{k=-\infty}^{\infty} \gamma_{pk} e^{-\imath\omega k} = \gamma_{p0} + 2\sum_{k=1}^{\infty} \gamma_{pk} \cos \omega k,$$

or, equivalently,

$$\gamma_{pk} = \frac{1}{2\pi}\int_{-\pi}^{\pi}[2\pi f_\theta(\omega)]^p \cos(k\omega)\, d\omega. \tag{2.13}$$

Obviously, $\gamma_{1k} = \gamma_k$, while $\gamma_{-1,k}$ is the inverse autocovariance function, see Cleveland (1972) and Battaglia (1983).

The generalised autocovariance function can be given a further interpretation in terms of the traditional autocovariance function, by defining a new process starting from the original one.

Let us consider a purely non-deterministic process $\{X_t\}$, with Wold representation $X_t = \Psi(L)\epsilon_t = (1 + \Psi_1 L + \Psi_2 L^2 + \dots)\epsilon_t$, with coefficients satisfying $\sum_{j=0}^{\infty}\Psi_j^2 < \infty$ and $\epsilon_t \sim WN(0, \sigma^2)$, where L is the lag operator ($L^j X_t = X_{t-j}$), and such that all the roots of the characteristic equation $\Psi(L) = 0$ are in modulus greater than 1.

The traditional autocovariance function of the process $\{X_t\}$ is

$$\gamma_k = \sigma^2 \sum_{j=0}^{\infty}\Psi_j\Psi_{j+k}, k = 0, 1, \dots, \qquad \gamma_k = \gamma_{-k}$$

The *power-transformed* process u_{pt} is defined as:

$$u_{pt} = \begin{cases} [\Psi(L)]^p\epsilon_t = [\Psi(L)]^p[\Psi(L)]^{-1}x_t, & \text{for } p \geq 0, \\ [\Psi(L^{-1})]^p\epsilon_t = [\Psi(L^{-1})]^p[\Psi(L)]^{-1}x_t, & \text{for } p < 0. \end{cases}$$

where L^{-1} is the forward operator. For any real p, the power transformation of the original polynomial is still a power series, with new coefficients determined recursively from the original ones (Gould, 1974). For $p > 0, [\Psi(L)]^p = \sum_{j=0}^{\infty}\varphi_j L^j$ and for $p < 0, [\Psi(L^{-1})]^p = \sum_{j=0}^{\infty}\varphi_j L^{-j}$, with:

$$\varphi_j = \frac{1}{j}\sum_{h=1}^{j}[h(p+1) - j]\Psi_h\varphi_{j-h},\ j > 0, \quad \varphi_0 = 1 \tag{2.14}$$

The spectral density function of the process u_{pt} is $f_u(\omega) = (2\pi)^{-1}\sigma^2|\Psi(e^{-i\omega})|^{2p} = (2\pi)^{-1}\sigma^2|\varphi(e^{-i\omega})|^2$. Then:

$$[2\pi f(\omega)]^p = \sigma^{2p}|\Psi(e^{-i\omega})|^{2p} = 2\pi f_u(\omega)(\sigma^2)^{p-1} \tag{2.15}$$

$$\gamma_{pk} = (\sigma^2)^{p-1}\gamma_u = \sigma^{2p}\sum_{j=0}^{\infty}\varphi_j\varphi_{j+k}, \tag{2.16}$$

where γ_u is the autocovariance function of the process u_{pt}.
This relation with the autocovariance function of the power-transformed process allows to give the GACV an analytic form, and to derive it for many stationary processes. We will use this to interpret the results about the asymptotic efficiency of the nonparametric estimator of the GACV.

The GACV is related to the *variance profile*, defined as the power mean of the spectrum of X_t (Luati et al., 2012):

$$v_p = \left\{ \frac{1}{2\pi} [2\pi f(\omega)]^p \right\}^{\frac{1}{p}}, \tag{2.17}$$

and for $p \neq 0$ we have $v_p = \gamma_{p0}^{\frac{1}{p}}$. Analogously we can observe that

$$v_p = \left\{ Var(u_{pt}) \frac{1}{\sigma^2} \right\}^{\frac{1}{p}} \sigma^2.$$

The variance profile provides, depending on the value of p the unconditional variance of the process, corresponding to the arithmetic mean of the spectral density ($p = 1$), the interpolation error variance, given by the harmonic mean of the spectrum ($p = -1$), or the one-step ahead prediction error variance, the geometric mean of the spectrum, obtained as a limiting case when $p \to 0$.

The generalised autocovariances are also linked to the generalised cepstral coefficients, defined as the Fourier coefficients of the Box-Cox transformation of the spectral density of the process (Proietti and Luati, 2019), introduced in the previous section. By simple algebra, one has:

$$c_{\lambda 0} = \frac{1}{\lambda}(\gamma_{\lambda 0} - 1), \quad c_{\lambda k} = \frac{1}{\lambda}\gamma_{\lambda k}, k \neq 0.$$

In Proietti and Luati (2019) estimation of the generalised cepstral coefficients by maximization of the Whittle likelihood is described. This represents an estimation method also for the GACVs, by their relation with the generalised cepstral coefficients, shown above.

An alternative option for estimation of the GACV is given by the nonparametric estimator proposed by Proietti and Luati (2015), which we describe next.

2.3.3 Nonparametric estimation of the GACV

The nonparametric estimator of the generalised autocovariance function proposed by Proietti and Luati (2015) is based on the same principles as the estimator of the variance profile, which generalises Hannan and Nicholls (1977) estimator of the prediction error variance. Hannan and Nicholls (1977) proposed to estimate the prediction error variance σ^2 by:

$$\hat{\sigma}^2(m) = m \exp\left\{ \frac{1}{M} \sum_{j=0}^{M-1} \log\left[\frac{1}{m} \sum_{k=1}^{m} 2\pi I(\omega_{jm+k}) \right] - \psi(m) \right\}, \tag{2.18}$$

where $M = \lfloor \frac{N-1}{2} \rfloor$, $\lfloor \cdot \rfloor$ denotes the integer part of the argument, and $\psi(\cdot)$ is the digamma function. By the distributional properties of the periodogram, $\psi(m)$ represents a bias correction term. The estimator $\hat{\sigma}^2(m)$ generalises the estimator proposed by Davis and Jones (1968) by replacing the raw periodogram ordinates ($m = 1$) with the non-overlapping averages of m consecutive ordinates, resulting in a smaller mean square estimation error. The estimator of the variance profile is based on a bias corrected power mean of the pooled periodogram over m non-overlapping consecutive frequencies.

$$\hat{v}_p(m) = m \left\{ \frac{1}{M} \sum_{j=0}^{M-1} \left[\frac{1}{m} \sum_{k=1}^{m} 2\pi I(\omega_{jm+k}) \right]^p \frac{\Gamma(m)}{\Gamma(m+p)} \right\}^{\frac{1}{p}}, \tag{2.19}$$

where $M = \lfloor \frac{N-1}{2m} \rfloor$. The introduction of $m > 1$ in (2.19) is needed for estimation of the interpolation error variance ($p = -1$) and, more generally, for estimation for negative p. Let us now introduce the nonparametric estimator of the GACV.

Given a time series of N consecutive observations, $\{x_t, t = 1, 2, \ldots, N\}$, we denote $I(\omega_j)$ the periodogram defined in (2.8) where ω_j is the Fourier frequency $\omega_j = \frac{2\pi j}{N} \in (0, \pi)$, $1 \leq j \leq \lfloor \frac{N-1}{2} \rfloor$.

Following Hannan and Nicholls (1977) and Luati et al. (2012), Proietti and Luati (2015) proposed the following nonparametric estimator of the generalised autocovariances

based on the inverse discrete Fourier transform of the pth power of the corrected pooled periodogram,

$$\hat{\gamma}_{pk} = \frac{1}{M}\sum_{j=0}^{M-1} Y_j^{(p)} \cos(\bar{\omega}_j k), \tag{2.20}$$

where $M = \lfloor \frac{N-1}{2m} \rfloor$ and, defining the pooled periodogram over $m \geq 1$ non-overlapping consecutive frequencies as

$$\bar{I}_j = \sum_{l=1}^{m} I(\omega_{jm+l}),$$

$$Y_j^{(p)} = (2\pi \bar{I}_j)^p \frac{\Gamma(m)}{\Gamma(m+p)}.$$

The frequencies $\bar{\omega}_j = \omega_{jm+(m+1)/2}$ are the mid range frequencies and m is the pooling parameter. The multiplicative factor $\frac{\Gamma(m)}{\Gamma(m+p)}$ is a bias-correction term, and its need is shown by the following distributional properties. It is known (Koopmans, 1974) that

$$(2\pi \bar{I}_j)^p = \Big(\sum_{k=1}^{m} 2\pi I(\omega_{jm+k})\Big)^p = [2\pi f_\theta(\omega_{jm+\frac{m+1}{2}})]^p X_j^p,$$

where $X_j \sim Ga(m,1)$, with $\mathbb{E}[X_j^p] = \frac{\Gamma(m+p)}{\Gamma(m)}$. Hence,

$$\mathbb{E}[Y_j^{(p)}] = \mathbb{E}[(2\pi \bar{I}_j)^p] \frac{\Gamma(m)}{\Gamma(m+p)} = [2\pi f_\theta(\omega_{jm+\frac{m+1}{2}})]^p$$

To ensure the existence of the second moment of the pth power of the gamma random variable with parameter $(m,1)$ some restrictions on m and p are needed In the Gaussian case we need $p > -m/2$, while in the more general IID case it has to be $p > 1 - m/4$.

Let $\boldsymbol{\gamma}_p = [\gamma_{p0}, \gamma_{p1}, \ldots, \gamma_{pK}]'$ be the vector of the generalised autocovariance functions up to lag K and $\hat{\boldsymbol{\gamma}}_p = [\hat{\gamma}_{p0}, \hat{\gamma}_{p1}, \ldots, \hat{\gamma}_{pK}]'$ the corresponding estimator. Under the stated assumptions and additional assumptions on m and on the coefficients of the Wold representation of the process Proietti and Luati (2015) showed that:

$$\sqrt{N}(\hat{\boldsymbol{\gamma}}_p - \boldsymbol{\gamma}_p) \underset{d}{\to} N(\mathbf{0}, \boldsymbol{V}) \tag{2.21}$$

where $\boldsymbol{V} = \{v_{kl}, k, l = 1, \ldots, K\}$, with

$$v_{kl} = m\,(C(m;p,p) - 1) \sum_{j=-\infty}^{\infty} (\gamma_{p,j+k}\gamma_{p,j+l} + \gamma_{p,j+k}\gamma_{p,j-l}) + k_4 \gamma_k \gamma_l, \tag{2.22}$$

or, equivalently,

$$v_{kl} = m\,(C(m;p,p) - 1)\,\frac{1}{\pi}\int_{-\pi}^{\pi}[2\pi f_\theta(\omega)]^{2p}\cos(\omega k)\cos(\omega l)\,d\omega, \tag{2.23}$$

where

$$C(m;p,p) = \frac{\Gamma(m+2p)\Gamma(m)}{\Gamma^2(m+p)},$$

and $\Gamma(\cdot)$ is the Gamma function. Equivalence between (2.22) and (2.23) is proved in the appendix 2.4. For $m = 1$ and $p = 1$ (2.22) and (2.23) coincide with the asymptotic covariance between the sample ACV at lag k and the sample ACV at lag l.
These asymptotic results are used to study asymptotic efficiency of the nonparametric estimator $\hat{\gamma}_{pk}$ of the GACV.

2.4 Appendix

Let us prove equivalence between equations (2.22) and (2.23) for the covariance between $\hat{\gamma}_{pk}$ and $\hat{\gamma}_{pk}$. Starting from (2.22) we derive Eq. (2.23), which is expressed in terms of the spectral density function, and is used in the derivation of results about asymptotic efficiency of $\hat{\gamma}_{pk}$ in Theorem 1.

The fourth cumulant k_4 is zero for a Gaussian process. Thus we have:

$$\begin{aligned}
&\lim_{N\to\infty}(N-m)cov\{\hat{\gamma}_{p,m},\hat{\gamma}_{p,n}\} \\
&= \sum_{l=-\infty}^{\infty}[\gamma_{p,l+m}\gamma_{p,l+n}+\gamma_{p,l+m}\gamma_{p,l-n}] \\
&= \sum_{l=-\infty}^{\infty}[\gamma_{p,l}\gamma_{p,l+n-m}+\gamma_{p,l}\gamma_{p,l-(n+m)}] \\
&= \sum_{l=-\infty}^{\infty}\gamma_{p,l}\gamma_{p,l+n-m}+\sum_{l=-\infty}^{\infty}[\gamma_{p,l}\gamma_{p,l-(n+m)} \qquad (2.24)
\end{aligned}$$

Now we know that the generalised autocovariance function is the inverse Fourier transform of the *pth* power of the spectral density function: $\gamma_{p,k} = \mathcal{F}^{-1}([2\pi f_\theta(\omega)]^p)$. Applying the *Convolution theorem* we can write:

$$\sum_{l=-\infty}^{\infty}\gamma_{p,l}\gamma_{p,l+n-m} = \mathcal{F}^{-1}([2\pi f_\theta(\omega)]^{2p}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}[2\pi f_\theta(\omega)]^{2p}\exp i\omega(n-m)\,d\omega.$$

By the Euler's formula, and some trigonometric identities, we obtain:

$$\sum_{l=-\infty}^{\infty} \gamma_{p,l}\gamma_{p,l+n-m} + \sum_{l=-\infty}^{\infty} \gamma_{p,l}\gamma_{p,l-(n+m)} =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} [2\pi f_\theta(\omega)]^{2p} \exp i\omega(n-m)\, d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi} [2\pi f_\theta(\omega)]^{2p} \exp -i\omega(n+m)\, d\omega =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} [2\pi f_\theta(\omega)]^{2p} [\exp i\omega(n-m) + \exp -i\omega(m+n)]\, d\omega =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} [2\pi f_\theta(\omega)]^{2p} [\cos\omega(n-m) + i\sin\omega(n-m) + \cos\omega(n+m) - i\sin\omega(n+m)]\, d\omega =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} [2\pi f_\theta(\omega)]^{2p} [\cos\omega(n-m) + \cos\omega(n+m)]\, d\omega =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} [2\pi f_\theta(\omega)]^{2p} [\cos n\omega\cos m\omega + \sin n\omega\sin m\omega + \cos n\omega\cos m\omega - \sin n\omega\sin m\omega]\, d\omega =$$

$$= (2\pi)^{2p-1}\int_{-\pi}^{\pi} [f_\theta(\omega)]^{2p}\, 2\cos(\omega n)\cos(\omega m)\, d\omega$$

which coincides with (2.23).

This result implies that the asymptotic variance of $\hat{\gamma}_{p,k}$ is:

$$\lim_{N\to\infty} N\, V\{\hat{\gamma}_{p,k}\} = (2\pi)^{2p-1}\int_{-\pi}^{\pi} [f_\theta(\omega)]^{2p}\, 2\cos^2(\omega k)\, d\omega \tag{2.25}$$

Chapter 3

Efficient nonparametric estimation of generalised autocovariances

Consistency and asymptotic normality of the estimator $\hat{\gamma}_{pk}$ of the GACV are established in Proietti and Luati (2015). Here we study asymptotic efficiency of this estimator. We establish a necessary and sufficient condition for asymptotic efficiency in terms of the spectral density and its derivatives for general Gaussian stationary processes, which nests as a particular case the result of Kakizawa and Taniguchi (1994), which holds for $p = 1$. The results also show that the nonparametric estimator $\hat{\gamma}_{pk}$ achieves the Cramèr-Rao lower bound as $m \to \infty$ for $p = -1$, i.e. it estimates efficiently the first q inverse autocovariances when the true generating process is pure MA(q). The inverse autocovariance function is useful in interpolation problems and for the identification of ARMA models.

We check this the condition for asymptotic efficiency for the class of Gaussian ARMA(r, q) processes. Analytical results for integer powers show that the nonparametric estimators of the GACV of order k, for $p = 1$ are asymptotically efficient for $r \geq q$ and $0 \leq k \leq r - q$, while for $p = -1$ asymptotic efficiency holds for $r = 0$ and $0 \leq k \leq q$. Hence, efficient estimation of the inverse autocovariance function by the nonparametric estimator of the GACV with $p = -1$ holds for MA processes. Some numerical results for noninteger powers are available, which also illustrate the rate of convergence to the Cramèr-Rao bound. The results obtained include as a special case the results for the

sample autocovariance function by Porat (1987) and Kakizawa and Taniguchi (1994). This chapter also illustrates that for a class of contrast functionals and spectral densities, the minimum contrast estimator of the spectral density satisfies a Yule-Walker system of equations in the generalised autocovariance estimator.
Selection of the pooling parameter m, which controls the resolution of the estimate, is discussed for large samples first. Considering asymptotic distributional results, we provide a preliminary overview of the behaviour of the asymptotic variance as m varies. Since the estimator is asymptotically unbiased, selection of m in large samples is based only on minimization of the asymptotic variance. Preliminary investigations and analytical and theoretical results suggest that varying m has a different impact on the variance of the estimator, depending on p. More specifically, increasing m implies a stronger variance reduction for estimation of the GACV for negative powers p.
However, a very large sample is required for asymptotic unbiasedness to hold. Since in finite samples its bias is nonzero, we estimate the finite-sample distribution of the estimator (bias and variance) and, hence, its MSE for selecting the optimal value of m. This is allowed by the multiplicative periodogram bootstrap, whose reasoning is illustrated in this chapter, together with a real data application.

3.1 Asymptotic efficiency of the estimator of the GACV

Let us first recall the definition of asymptotically efficient estimator of an unknown parameter.

Definition 1. A sequence of estimators T_n is **asymptotically efficient** for a parameter $\tau(\theta)$ if

$$\sqrt{n}[T_n - \tau(\theta)] \underset{d}{\to} N(0, v(\theta)) \tag{3.1}$$

and

$$v(\theta) = \frac{[\frac{d}{d\theta}\tau(\theta)]^2}{(\mathbb{E}_\theta(\frac{d}{d\theta}\log f(X|\theta))^2)} \tag{3.2}$$

that is, if the asymptotic variance of T_n achieves the *Cramér-Rao lower bound*.

Let $\{X_t\}_{t\in T}$, with $T \in \mathbb{N}$ a discrete time set, denote a zero mean stationary Gaussian process with autocovariance function $\gamma_k = \mathbb{E}(X_t X_{t-k}), k \in \mathbb{Z}$, and spectral density $f_\theta(\omega) = \sum_{k=-\infty}^{\infty} \gamma_k e^{-\imath\omega k}$, $\omega \in [-\pi, \pi]$, both depending on an $s \times 1$ vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)' \in \mathbb{R}^s$.
Impose the following assumptions:

Assumption 1. There exist two positive constants $\underline{c}$ and $\overline{c}$ such that
$0 < \underline{c} \leq f_\theta(\omega) \leq \overline{c} < \infty$, for $\omega \in [-\pi, \pi]$.

Assumption 2. The generalised autocovariances and their partial derivatives, $\partial\gamma_{pk}/\partial\theta_j$, satisfy the summability conditions $\sum_{k=1}^{\infty} k|\gamma_{pk}| < \infty$, $\sum_{k=1}^{\infty} k|\partial\gamma_{pk}/\partial\theta_j| < \infty$.

Assumption 3. The $s \times s$ matrix

$$\frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}'} \frac{d\omega}{f_\theta^2(\omega)}$$

is positive definite.

The first assumption restricts attention to short memory processes, ruling out long memory and non-invertible ones, see Hassler (2018). Assumption 2 implies that $\int_{-\pi}^{\pi} [f_\theta(\omega)]^p \, d\omega < \infty$ and $f_\theta(\omega)$ is differentiable with respect to θ_j, and $\partial f_\theta(\omega)/\partial\theta_j$ is continuous and differentiable with respect to ω, with continuous derivative.

Hence, the GACV estimator $\hat{\gamma}_{kp}$ in (2.20) is asymptotically efficient if its asymptotic variance, v_{kk} converges to the *Cramér-Rao lower bound*

$$CRB\{\hat{\gamma}_{pk}\} = \frac{\partial\gamma_{pk}}{\partial\boldsymbol{\theta}'} \mathfrak{I}_N^{-1}(\boldsymbol{\theta}) \frac{\partial\gamma_{pk}}{\partial\boldsymbol{\theta}}, \tag{3.3}$$

with

$$\frac{\partial\gamma_{pk}}{\partial\boldsymbol{\theta}} = (2\pi)^{p-1} \int_{-\pi}^{\pi} \frac{\partial[f_\theta(\omega)]^p}{\partial\boldsymbol{\theta}} \cos(k\omega)\, d\omega,$$

and $\mathfrak{I}_N(\boldsymbol{\theta})$ is the Fisher information matrix associated with $X_1, \ldots, X_N$,

$$\mathfrak{I}_N(\boldsymbol{\theta}) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}'} \frac{1}{f_\theta^2(\omega)}\, d\omega,$$

whose generic element $i_N(j_1, j_2)$ is:

$$\mathfrak{I}_N(j_1, j_2) = \mathbb{E}\left\{ \frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_{j_1}} \frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_{j_2}} \right\} \quad j_1, j_2 = 1, \ldots, s.$$

The generic element of the Fisher information matrix in terms of the spectral density function is derived in the appendix 3.6, and is given by:

$$i_N(j_1, j_2) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial f_\theta(\omega)}{\partial \theta_{j_1}} \frac{\partial f_\theta(\omega)}{\partial \theta_{j_2}} \frac{1}{f_\theta^2(\omega)} \, d\omega \tag{3.4}$$

Under assumptions $1-3$, and by (2.23), $v_{kk} \geq CRB\{\hat{\gamma}_{pk}\}$ gives the following inequality:

$$\begin{aligned} \lim_{N\to\infty} N\, V\{\hat{\gamma}_{pk}\} = m(C(m;p,p)-1)(2\pi)^{2p-1} \int_{-\pi}^{\pi} [f_\theta(\omega)]^{2p}\, 2\cos^2(\omega k)\, d\omega \geq \\ \lim_{N\to\infty} NCRB\{\hat{\gamma}_{pk}\} = \left\{ (2\pi)^{p-1} \int_{-\pi}^{\pi} \frac{\partial [f_\theta(\omega)]^p}{\partial \boldsymbol{\theta'}} \cos(k\omega)\, d\omega \right\} \\ \left\{ \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta'}} \frac{1}{f_\theta^2(\omega)} \, d\omega \right\}^{-1} \left\{ (2\pi)^{p-1} \int_{-\pi}^{\pi} \frac{\partial [f_\theta(\omega)]^p}{\partial \boldsymbol{\theta'}} \cos(k\omega)\, d\omega \right\}' \end{aligned} \tag{3.5}$$

We first recall the following Lemma by Kakizawa and Taniguchi (1994), which allows us to compare the asymptotic variance of $\hat{\gamma}_{pk}$ and its Cramèr-Rao lower bound:

Lemma 1. Let $A(\omega)$ and $B(\omega)$ be $r \times s, t \times s$ matrices, respectively, and let $g(\omega)$ be a function such that $g(\omega) > 0$ almost everywhere (a.e.) on $[-\pi, \pi]$. If the matrix

$$\left\{ \int_{-\pi}^{\pi} \frac{B(\omega)B(\omega)'}{g(\omega)} \, d\omega \right\}^{-1}$$

exists, then

$$\int_{-\pi}^{\pi} A(\omega)A(\omega)' g(\omega)\, d\omega \geq \left\{ \int_{-\pi}^{\pi} A(\omega)B(\omega)'\, d\omega \right\} \left\{ \int_{-\pi}^{\pi} \frac{B(\omega)B(\omega)'}{g(\omega)} \, d\omega \right\}^{-1} \left\{ \int_{-\pi}^{\pi} A(\omega)B(\omega)'\, d\omega \right\}'$$

where $\geq$ means the left-hand side minus the right-hand side is positive semi-definite. Here the equality holds if there exists an $r \times t$ matrix C which is independent of ω such that:

$$g(\omega)A(\omega) + CB(\omega) = 0.$$

This can be seen by substituting $A(\omega) = -\frac{CB(\omega)}{g(\omega)}$ in the inequality:

$$\int_{-\pi}^{\pi} \frac{CB(\omega)}{g(\omega)} \frac{B(\omega)'C'}{g(\omega)} g(\omega)\, d\omega \geq \left\{ \int_{-\pi}^{\pi} \frac{(-C)B(\omega)}{g(\omega)} B(\omega)'\, d\omega \right\} \times \times \left\{ \int_{-\pi}^{\pi} \frac{B(\omega)B(\omega)'}{g(\omega)}\, d\omega \right\}^{-1} \left\{ \int_{-\pi}^{\pi} B(\omega) \frac{(-1)B(\omega)'C'}{g(\omega)}\, d\omega \right\} \tag{3.6}$$

Since C does not depend on ω we can take it outside the integral, and write:

$$C\left(\int_{-\pi}^{\pi} \frac{B(\omega)}{g(\omega)} B(\omega)'\, d\omega \right) C' \geq C \left\{ \int_{-\pi}^{\pi} \frac{B(\omega)}{g(\omega)} B(\omega)'\, d\omega \right\} \times \times \cancel{\left\{ \int_{-\pi}^{\pi} \frac{B(\omega)B(\omega)'}{g(\omega)}\, d\omega \right\}^{-1}} \cancel{\left\{ \int_{-\pi}^{\pi} \frac{B(\omega)B(\omega)'}{g(\omega)}\, d\omega \right\}} C' = = C\left(\int_{-\pi}^{\pi} \frac{B(\omega)}{g(\omega)} B(\omega)'\, d\omega \right) C'$$

By noticing that:

$$\frac{\partial [f_\theta(\omega)]^p}{\partial \boldsymbol{\theta'}} = p[f_\theta(\omega)]^{p-1} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta'}},$$

and simplifying some constants, the inequality (3.5) becomes:

$$\frac{m(C(m;p,p)-1)}{p^2} \int_{-\pi}^{\pi} [f_\theta(\omega)]^{2p} \cos^2(\omega k)\, d\omega \geq \left\{ \int_{-\pi}^{\pi} [f_\theta(\omega)]^p \frac{\partial \ln f_\theta(\omega)}{\partial \boldsymbol{\theta'}} \cos(k\omega)\, d\omega \right\} \left\{ \int_{-\pi}^{\pi} \frac{\partial \ln f_\theta(\omega)}{\partial \boldsymbol{\theta}} \frac{\partial \ln f_\theta(\omega)}{\partial \boldsymbol{\theta'}}\, d\omega \right\}^{-1} \left\{ \int_{-\pi}^{\pi} [f_\theta(\omega)]^p \frac{\partial \ln f_\theta(\omega)}{\partial \boldsymbol{\theta'}} \cos(k\omega)\, d\omega \right\}'. \tag{3.7}$$

or, equivalently:

$$\frac{m(C(m;p,p)-1)}{p^2}\int_{-\pi}^{\pi} f_\theta^2(\omega)[f_\theta(\omega)]^{2(p-1)}\cos^2(\omega k)\,d\omega \geq \left\{\int_{-\pi}^{\pi}[f_\theta(\omega)]^{p-1}\frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta'}}\cos(k\omega)\,d\omega\right\}$$
$$\left\{\int_{-\pi}^{\pi}\frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}}\frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta'}}\frac{1}{f_\theta^2(\omega)}\,d\omega\right\}^{-1}\left\{\int_{-\pi}^{\pi}[f_\theta(\omega)]^{p-1}\frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta'}}\cos(k\omega)\,d\omega\right\}'. \tag{3.8}$$

Asymptotic efficiency of $\hat{\gamma}_{pk}$ occurs when v_{kk} achieves the CRB on the right hand side of (3.5), or, equivalently if in (3.8) equality holds. This requires a condition on the spectral density of the process that will be stated in Theorem 1.

To asses asymptotic efficiency of the estimator, note that inequality (3.8) takes the following form:

$$K\int_{-\pi}^{\pi} A(\omega)A(\omega)'g(\omega)\,d\omega \geq \left\{\int_{-\pi}^{\pi}A(\omega)B(\omega)'\,d\omega\right\}\left\{\int_{-\pi}^{\pi}\frac{B(\omega)B(\omega)'}{g(\omega)}\,d\omega\right\}^{-1}\left\{\int_{-\pi}^{\pi}A(\omega)B(\omega)'\,d\omega\right\}'. \tag{3.9}$$

with $A(\omega)=\cos(k\omega)[f_\theta(\omega)]^{p-1}$, $B(\omega)=\frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}}$, $g(\omega)=f_\theta^2(\omega)$, and $K=\frac{m(C(m;p,p)-1)}{p^2}$.

The attainment of the CRB thus depends also on the term $\frac{m(C(m;p,p)-1)}{p^2}$, involving both the power p and the pooling parameter m, and, as it is evident from (3.8), $\frac{m(C(m;p,p)-1)}{p^2}\to 1$ is required.

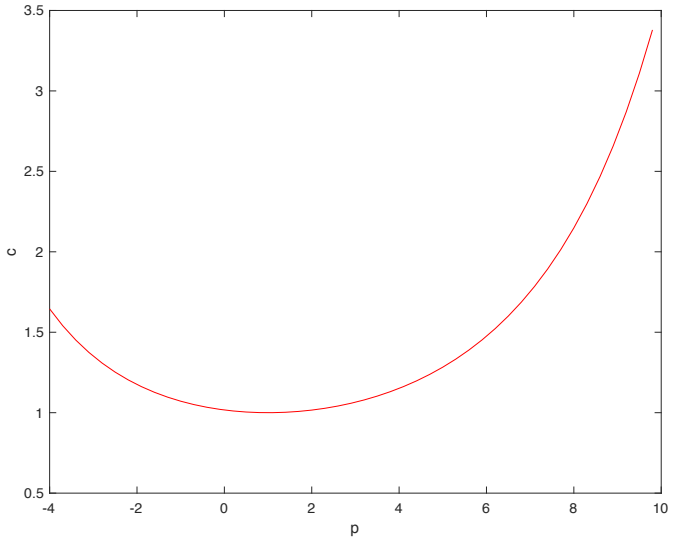

Figure 3.1: Plot of $c = \frac{m(C(m;p,p)-1)}{p^2}$ vs p with $m = 30$.

As shown in figure 3.1, this factor has a minimum point at $p = 1$, where it attains unity. It is possible to show that $K \to 1$ as $m \to \infty$, and the proof is provided in the appendix 3.6.

Hence, in the cases $p = 1$ or $p \neq 1$ and $m \to \infty$ $\frac{m(C(m;p,p)-1)}{p^2} = 1$. If $p \neq 1$ and $m \to \infty$, the Cramér-Rao inequality (3.8) becomes:

$$\int_{-\pi}^{\pi} f_\theta^2(\omega)[f_\theta(\omega)]^{2(p-1)} \cos^2(\omega k)\, d\omega \geq$$
$$\left\{ \int_{-\pi}^{\pi} [f_\theta(\omega)]^{p-1} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta'}} \cos(k\omega)\, d\omega \right\}$$
$$\left\{ \int_{-\pi}^{\pi} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta'}} \frac{1}{f_\theta^2(\omega)}\, d\omega \right\}^{-1} \left\{ \int_{-\pi}^{\pi} [f_\theta(\omega)]^{p-1} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta'}} \cos(k\omega)\, d\omega \right\}'. \tag{3.10}$$

Applying Lemma 1 by setting:

$$A(\omega) = \cos(k\omega)[f_\theta(\omega)]^{p-1}, \quad B(\omega) = \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}}, \quad g(\omega) = f_\theta^2(\omega),$$

implies the following result.

If $p \neq 1$ and $m \to \infty$, the following theorem provides a necessary and sufficient condition for asymptotic efficiency of $\hat{\gamma}_{pk}$ for Gaussian stationary processes:

Theorem 1. *Suppose that assumptions 1-3 are satisfied and that m and M are large enough for asymptotics and $\frac{m}{M}$ is small enough for f to be constant over frequency intervals of length $\frac{2\pi m}{M}$ and $m(C(m,p,p)-1) \to p^2$. Then, $\hat{\gamma}_{pk}$ is asymptotically efficient if and only if there exists an s-dimensional vector $\mathbf{c}$, independent of ω, such that:*

$$[f_\theta(\omega)]^{p+1} \cos(k\omega) + \mathbf{c}' \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}} = 0, \tag{3.11}$$

Theorem 1 provides a necessary and sufficient condition for asymptotic efficiency of $\hat{\gamma}_{pk}$ which is valid for general Gaussian stationary processes. It is expressed in terms of the spectral density function, which makes it easy to check for various models. This result embodies in a single equation the condition for asymptotic efficiency of the sample autocovariance function ($p = 1$), of the estimator $\hat{\gamma}_{-1,k}$ of the inverse autocovariance function ($p = -1$), which at lag $k = 0$ provides the inverse of the interpolation error variance, and of the estimator $\hat{\gamma}_{pk}$ for general real powers p.

We can use the relation with the autocovariance function of the power-transformed process, described in section 2.3.2 to interpret the results about the asymptotic efficiency of the nonparametric estimator of the GACV.
In view of the relation between x_t and u_{pt}, observe that:

$$AV\{\hat{\gamma}_{pk}\} = m(C(m;p,p)-1)(\sigma)^{2(p-1)} AV\{\tilde{\gamma}_{u,k}\} \tag{3.12}$$

$$I_N = \frac{1}{p^2} I_N(u_{pt})$$

$$CRB\{\hat{\gamma}_{pk}\} = (\sigma)^{2(p-1)} p^2 CRB\{\tilde{\gamma}_{u,k}\} \tag{3.13}$$

where $AV\{\tilde{\gamma}_{u,k}\}$ and $CRB\{\tilde{\gamma}_{u,k}\}$ are the asymptotic variance and CRB of the sample autocovariance of the process u_{pt}, denoted by $\tilde{\gamma}_{u,k}$. Then, one has:

$$\frac{AV\{\hat{\gamma}_{pk}\}}{CRB\{\hat{\gamma}_{pk}\}} = \frac{m(C(m;p,p)-1)}{p^2} \frac{AV\{\tilde{\gamma}_u\}}{CRB\{\tilde{\gamma}_u\}}$$

The relation between asymptotic efficiency of the nonparametric estimator of the GACV for a given process x_t and asymptotic efficiency of the sample estimator of the

ACVF of the power process u_{pt} is governed by the positive factor $\frac{m(C(m;p,p)-1)}{p^2}$. We know that its value decreases as m increases. This function has a minimum point at $p = 1$ and it increases as the distance of p from unity increases.
Some implications of Theorem 1 are given below.

Corollary 1. *Consider the process with spectral density function* $f_\theta(\omega) = \frac{1}{2\pi}\left[\frac{1}{\theta(\omega)}\right]^{\frac{1}{p}}$, *with* $\theta(\omega)$ *the trigonometric polynomial* $\theta(\omega) = \theta_0 + 2\sum_{j=1}^{K}\theta_j \cos(\omega j)$, *so that* $\frac{\partial\theta(\omega)}{\partial\boldsymbol{\theta}} = \boldsymbol{q}(\omega) = [1, 2\cos(\omega), 2\cos(2\omega), \ldots, 2\cos(\omega K)]'$. *Then,*

$$\frac{\partial f_\theta(\omega)}{\partial\boldsymbol{\theta}} = -(2\pi)^p \frac{1}{p}[f_\theta(\omega)]^{p+1}\boldsymbol{q}(\omega).$$

Condition (3.11) *in Theorem 1 becomes*

$$[f_\theta(\omega)]^{p+1}\left\{\cos(k\omega) - \frac{(2\pi)^p}{p}\mathbf{c}'\boldsymbol{q}(\omega)\right\} = 0,$$

which is satisfied if $\boldsymbol{c} = \left[0, 0, \ldots, \frac{p}{2(2\pi)^p}, 0, \ldots, 0\right]'$. *This implies that for* $p = -1$ *the process is moving-average of order* K *and the first* K *inverse autocovariances* $\boldsymbol{\gamma}_{-1,K} = [\gamma_{-1,1}, \ldots, \gamma_{-1,K}]'$ *and* $\gamma_{-1,0}$ *can be efficiently estimated as* $N \to \infty$ *by the estimator of the GACV* $\hat{\boldsymbol{\gamma}}_{-1,K}$ *with large* m.

Remark 1. The estimator (2.20) can be viewed in the wider context of estimation of functionals of the spectral density, which are related to many important quantities in stationary time series. Setting $m = 1$, for $p > 0$, Y_j^p is the *inverse Laplace transform* of $[2\pi f(\omega_j)]^{-(p+1)}$ evaluated at $2\pi I(\omega_j)$, proposed by Taniguchi (1980) for estimating $[2\pi f(\omega_j)]^p$. Asymptotic efficiency of this estimator is studied in Taniguchi (1981), who establishes that this estimator is asymptotically efficient if $p = 1$ and the spectral density is constant over $[-\pi, \pi]$. The nonparametric estimator $\hat{\gamma}_{pk}$ further generalises these results to any real power transform, including negative p. Furthermore, the introduction of the pooling parameter m allows asymptotically efficient estimates also for $p \neq 1$.

Remark 2. By setting the power p and the pooling parameter m to 1, inequality (3.8) reduces to the asymptotic Cramér-Rao inequality for the sample estimator of the autocovariance function analysed by Kakizawa and Taniguchi (1994). Note also that for

$p = 1$, by the properties of the Gamma function, the constant $m(C(m;p,p)-1)$ does not depend on the pooling parameter m. By the properties of the Gamma function we have:

$$m(C(m;1,1)-1) = m\Big(\frac{\Gamma(m+2p)\Gamma(m)}{\Gamma^2(m+p)} - 1\Big) = m\Big(\frac{\Gamma(m+2)\Gamma(m)}{\Gamma^2(m+1)} - 1\Big) =$$
$$= m\Big(\frac{(m+1)\Gamma(m+1)\Gamma(m)}{m^2\Gamma^2(m)} - 1\Big) =$$
$$= m\Big(\frac{(m+1)m\Gamma^2(m)}{m^2\Gamma^2(m)} - 1\Big) = \frac{m(m+1)}{m} - m = m + 1 - m = 1$$

Hence, $m(C(m;1,1)-1) = 1$. This implies that if we consider estimation of the traditional autocovariance function, the asymptotic variance of the nonparametric estimator $\hat{\gamma}_{1k}$ does not depend on the pooling parameter m. Indeed, $\hat{\gamma}_{1k}$ is the Riemannian sum approximation over the Fourier frequencies of the sample autocovariance at lag k, denoted by $\tilde{\gamma}_k$:

$$\lim_{N\to\infty} \frac{1}{\lfloor (N-1)/2 \rfloor} \sum_{j=1}^{\lfloor (N-1)/2 \rfloor} 2\pi I(\omega_j)\cos(\omega_j k) = \int_{-\pi}^{\pi} I(\omega)\cos(\omega k)\, d\omega = \tilde{\gamma}_k,$$

with $I(\omega) = \frac{1}{2\pi}\sum_{|h|<N} \tilde{\gamma}_h \cos(\omega h)$. Hence $\lim_{N\to\infty} \hat{\gamma}_{1k} = \tilde{\gamma}_k$, and their asymptotic variances, as $N \to \infty$, are equivalent. As a matter of fact, by setting $p = 1$, Theorem 1 provides the condition for asymptotic efficiency of the sample autocovariances by Kakizawa and Taniguchi (1994).

For $p = 1$ inequality (3.5) specialises to:

$$(4\pi)\int_{-\pi}^{\pi} [f_\theta(\omega)]^2 \cos^2(\omega k)\, d\omega \geq$$
$$\left\{\int_{-\pi}^{\pi} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta'}} \cos(k\omega)\, d\omega\right\}$$
$$\left\{\frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta'}} \frac{1}{f_\theta^2(\omega)}\, d\omega\right\}^{-1} \left\{\int_{-\pi}^{\pi} \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta'}} \cos(k\omega)\, d\omega\right\}' \tag{3.14}$$

In this case ($p = m = 1$), by setting in Lemma 1:

$$A(\omega) = \cos k\omega, \quad g(\omega) = f_\theta^2(\omega), \quad B(\omega) = \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}},$$

the condition for asymptotic efficiency of $\hat{\gamma}_{1k}$, and hence $\tilde{\gamma}_k$ is stated in the following theorem by Kakizawa and Taniguchi (1994):

Theorem 2. *(Kakizawa and Taniguchi, 1994)*

Suppose that assumptions 1-3 are satisfied. Then, the sample autocovariance $\tilde{\gamma}_k$ is asymptotically efficient if and only if there exists a s-dimensional vector $\mathbf{c}$ *which is independent of* ω *such that:*

$$f_\theta^2(\omega)\cos(k\omega) + \mathbf{c}'\frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}} = 0 \tag{3.15}$$

The validity of Theorem 1 can be analysed for stationary Gaussian ARMA(r,q) processes to establish the constraints on r, q, k allowing for asymptotic efficiency of $\hat{\gamma}_{pk}$.

3.2 Asymptotic efficiency of $\hat{\gamma}_{pk}$ for ARMA processes

Let $\{X_t\}_{t\in T}$ be a zero-mean stationary and invertible Gaussian ARMA(r,q) process, with spectral density function

$$f(\omega) = \frac{\sigma}{2\pi}\frac{\beta(e^{i\omega})\beta(e^{-i\omega})}{\phi(e^{i\omega})\phi(e^{-i\omega})},$$

where $\beta(e^{-i\omega}) = 1 + \sum_{j=1}^{q}\beta_j e^{-i\omega j}$, $1 + \sum_{j=1}^{q}\beta_j z^j \neq 0 \iff |z| \leq 1$, and $\phi(e^{-i\omega}) = 1 - \sum_{j=1}^{r}\phi_j e^{-i\omega j}$, $1 - \sum_{j=1}^{r}\phi_j z^j \neq 0 \iff |z| \leq 1$.

For an ARMA process we ask if $\hat{\gamma}_{pk}$ is an asymptotically efficient estimator as p varies.

By Theorem 1, asymptotic efficiency of $\hat{\gamma}_{pk}$ with sufficiently large m requires the existence of a s-dimensional vector $\mathbf{c}$ which is independent of ω such that Eq. (3.11) is satisfied. In this case $s = r+q+1$ is the number of parameters in $\boldsymbol{\theta} = (\sigma, \phi_1, \ldots, \phi_r, \beta_1, \ldots, \beta_q)'$.

First, we illustrate analytical results for positive and negative integer values of p. Then, in the next section, numerical results show how asymptotic efficiency of $\hat{\gamma}_{pk}$ varies with real powers p.

Let us first restrict our attention to positive integer values of p. Note that we can multiply both sides of Eq. (3.11) by the nonzero factor $[\phi(e^{i\omega})\phi(e^{-i\omega})]^{p+1}$, so that we only need to check whether there exists a vector $\boldsymbol{c}$ such that:

$$[(\frac{\sigma}{2\pi})^{p+1}\beta(e^{i\omega})\beta(e^{-i\omega})]^{p+1}\cos(k\omega) = -\mathbf{c}'[\phi(e^{i\omega})\phi(e^{-i\omega})]^{p+1}\frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}} \tag{3.16}$$

By Euler's formula and trigonometric identities we can express both the left-hand side and the right-hand side of (3.16) as Fourier polynomials. The existence of $\boldsymbol{c}$ depends on the degree of these polynomials. Defining the space of 2π-periodic functions admitting a finite Fourier polynomial representation as

$$M_l = \{F | F(\omega) = c_0 + c_1\cos\omega + \cdots + c_l\cos(l\omega), c_l \neq 0\},$$

for the specified ARMA(r, q) process we have:

$$[\beta(e^{i\omega})\beta(e^{-i\omega})]^{p+1}\cos(k\omega) \in M_{(p+1)q+k} \tag{3.17}$$

$$[\phi(e^{i\omega})\phi(e^{-i\omega})]^{p+1}\frac{\partial f_\theta(\omega)}{\partial\sigma} \in M_{pr+q} \tag{3.18}$$

$$[\phi(e^{i\omega})\phi(e^{-i\omega})]^{p+1}\frac{\partial f_\theta(\omega)}{\partial\phi_j} \in M_{max((p-1)r+q+j,pr+q-j)}, \quad j = 1,\ldots,r \tag{3.19}$$

$$[\phi(e^{i\omega})\phi(e^{-i\omega})]^{p+1}\frac{\partial f_\theta(\omega)}{\partial\beta_j} \in M_{max(pr+q-j,pr+j)}, \quad j = 1,\ldots,q. \tag{3.20}$$

These imply that (3.16) can be written as:

$$\boldsymbol{v}'(1,\cos\omega,\ldots,\cos\omega[(p+1)q+k])' = -\boldsymbol{c}'\mathbf{A}(1,\cos\omega,\ldots,\cos\omega(pr+q))' \tag{3.21}$$

where $\boldsymbol{v}$ has length $(p+1)q+k+1$ and contains the Fourier coefficients of $[\beta(e^{i\omega})\beta(e^{-i\omega})]^{p+1}\cos(k\omega)$ and the j-th row of the $s\times(pr+q+1)$ matrix $\mathbf{A}$ contains those of $[\phi(e^{i\omega})\phi(e^{-i\omega})]^{p+1}\partial f_\theta(\omega)/\partial\theta_j$. The left-hand side of (3.16) is:

$$[\beta(e^{i\omega})\beta(e^{-i\omega})]^{p+1}\cos(k\omega) = v_0 + v_1\cos\omega + \cdots + v_{(p+1)q+k}\cos[(p+1)q+k]\omega$$

To see this, first remind the following identities:

$$e^{\pm i\omega n} = \cos(\omega n) \pm i\sin(\omega n), \tag{3.22}$$

$$(e^{i\omega n} + e^{-i\omega n}) = 2\cos(\omega n), \tag{3.23}$$

$$\cos(\alpha)\cos(\beta) = \frac{1}{2}(\cos(\alpha+\beta) + \cos(\alpha-\beta)), \tag{3.24}$$

$$\cos(-\omega) = \cos(\omega). \tag{3.25}$$

Consider a polynomial $\beta(z) = \beta_0 + \beta_1 z + \beta_2 z^2 + \cdots + \beta_q z^q$. It is clear that $[\beta(z)]^n = b_0 + b_1 z + \cdots + b_{nq} z^{nq}$ for n a positive integer. Analogously, if p is a positive integer, $[\beta(z)]^{p+1} = e_0 + e_1 z + \cdots + e_{(p+1)q} z^{(p+1)q} = e(z)$. By simple algebra, it can be shown that $[\beta(z)\beta(z^{-1})] = a_0 + a_1(z + z^{-1}) + \cdots + a_q(z^q + z^{-q})$, and, similarly, $[\beta(z)]^{p+1}[\beta(z^{-1})]^{p+1} = g_0 + g_1(z + z^{-1}) + \cdots + g_{(p+1)q}(z^{(p+1)q} + z^{-(p+1)q})$. By (3.23):

$$[\beta(z)]^{p+1}[\beta(z^{-1})]^{p+1} = g_0 + g_1 2\cos\omega + \cdots + g_{(p+1)q} 2\cos\left([(p+1)q]\omega\right)$$

Then, if we consider $k \leq (p+1)q$, by (3.24):

$$\begin{aligned}
&[\beta(e^{i\omega})\beta(e^{-i\omega})]^{p+1}\cos(k\omega) = \\
&g_0\cos(\omega k) + g_1\big(\cos(k+1)\omega + \cos(k-1)\omega\big) + g_2\big(\cos(k+2)\omega + \cos(k-2)\omega\big) + \\
&+ \cdots + g_k\big(\cos(2k)\omega + \cos(k-k)\omega\big) + \cdots + g_{(p+1)q}\big(\cos((p+1)q+k)\omega + \cos((p+1)q-k)\omega\big) = \\
&= v_0 + v_1\cos\omega + v_2\cos 2\omega + \cdots + v_{(p+1)q+k}\cos((p+1)q+k)\omega, \qquad (3.26)
\end{aligned}$$

otherwise, if $k > (p+1)q$ the leading term in (3.26) is null. Eq. (3.26) coincides with (3.17). (3.18)- (3.20) can be derived analogously.

Relations (3.17) - (3.20) imply:

$$\boldsymbol{v}'(1, \cos\omega, \ldots, \cos\omega[(p+1)q+k])' = -\boldsymbol{c}'\mathbf{A}(1, \cos\omega, \ldots, \cos\omega(pr+q))', \qquad (3.27)$$

Consider $(p+1)q + k \leq pr + q$. Denote $\boldsymbol{d} = (1, \cos(\omega), \ldots, \cos\omega[(p+1)q+k])'$ and $\boldsymbol{d}_1 = (\cos\omega[(p+1)q+k+1], \ldots, \cos\omega(pr+q))'$. Then, Eq. (3.27) becomes:

$$\boldsymbol{v}'\boldsymbol{d} = -\boldsymbol{c}'\mathbf{A}(\boldsymbol{d}', \boldsymbol{d}_1')'. \qquad (3.28)$$

Note that if $p = 1$, the square matrix $\mathbf{A} \in \mathbb{R}^{s\times s}$ is nonsingular by assumption. In this case the condition for asymptotic efficiency is satisfied by setting $\boldsymbol{c} = -(A')^{-1}(\boldsymbol{v}', 0, \ldots, 0)'$. Hence, for $p = 1$, $\hat{\gamma}_{pk}$ is asymptotically efficient when $r \geq q$ and $0 \leq k \leq r - q$, which coincides with the results in Kakizawa and Taniguchi (1994).
On the other hand, if $p \neq 1, \mathbf{A} \in \mathbb{R}^{s\times(pr+q+1)}$ is not square in general. In particular, by assumption 3 $\mathbf{A} \in \mathbb{R}^{s\times(pr+q+1)}$ has full row-rank, $s = r + q + 1$. Investigation of the existence of a solution for to the matrix equation (3.27) allows to establish asymptotic

efficiency of the estimator for positive integer p.

Let us consider the case when the power $p+1$ is a negative integer ($p < -1$). By multiplying both sides of Eq. (3.11) by $[\beta(e^{i\omega})\beta(e^{-i\omega})]^{-(p+1)}$ the condition to be checked, that guarantees asymptotic efficiency of the estimator, becomes:

$$\left(\frac{\sigma}{2\pi}\right)^{p+1}[\phi(e^{i\omega})\phi(e^{-i\omega})]^{-(p+1)}\cos(k\omega) = -\mathbf{c}'[\beta(e^{i\omega})\beta(e^{-i\omega})]^{-(p+1)}\frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}} \tag{3.29}$$

By the same arguments used in the proof for positive integer p, the following relations are valid:

$$\begin{aligned}
&[\phi(e^{i\omega})\phi(e^{-i\omega})]^{-(p+1)}\cos(k\omega) \in M_{-(p+1)r+k}\\
&[\beta(e^{i\omega})\beta(e^{-i\omega})]^{-(p+1)}\frac{\partial f_\theta(\omega)}{\partial \sigma} \in M_{-pq-r}\\
&[\beta(e^{i\omega})\beta(e^{-i\omega})]^{-(p+1)}\frac{\partial f_\theta(\omega)}{\partial \phi_j} \in M_{max(-pq-r-j,-pq-2r+j)}, \quad j=1,\ldots,r\\
&[\beta(e^{i\omega})\beta(e^{-i\omega})]^{-(p+1)}\frac{\partial f_\theta(\omega)}{\partial \theta_j} \in M_{max(-pq-r-j,-pq-q-r+j)}, \quad j=1,\ldots,q.
\end{aligned}$$

Note that $-(p+1)r$ is a positive integer, and so is $-p$. Hence, the previous relations imply:

$$\boldsymbol{v}_1'(1,\cos\omega,\ldots,\cos\omega[-(p+1)r+k])' = -\boldsymbol{c}'\mathbf{B}(1,\cos\omega,\ldots,\cos\omega(-pq-r))', \tag{3.30}$$

where $\boldsymbol{v}_1$ has length $(-(p+1)r+k+1)$ and contains the Fourier coefficients of $[\phi(e^{i\omega})\phi(e^{-i\omega})]^{-(p+1)}\cos(k\omega$ and the j-th row of the matrix $\mathbf{B} \in \mathbb{R}^{s\times(-pq-r+1)}$ contains those of $[\beta(e^{i\omega})\beta(e^{-i\omega})]^{-(p+1)}\partial f_\theta(\omega)/\partial\theta_j$.

Analogous considerations as those for positive integer p imply that for p a negative integer asymptotic efficiency of $\hat{\gamma}_{pk}$ is determined by establishing the conditions of existence of a solution for the equation (3.30).
Note that for $p=-1$ Eq. (3.29) becomes:

$$\cos(k\omega) = \frac{\partial f_\theta(\omega)}{\partial \boldsymbol{\theta}}.$$

or, equivalently:

$$\boldsymbol{v}_1'(1, \cos\omega, \ldots, \cos\omega k)' = -\boldsymbol{c}'\mathbf{B}(1, \cos\omega, \ldots, \cos\omega(q-r))', \tag{3.31}$$

In Eq. (3.30),when $p = -1$, $\boldsymbol{v}_1' = (0, 0, \ldots, 1)$, and $\mathbf{B} \in \mathbb{R}^{s\times(q-r+1)}$ is a nonsingular square matrix if $r = 0$. In this case, for $0 \leq k \leq q$, we can specify $\boldsymbol{c} = -(\mathbf{B}')^{-1}(\boldsymbol{v}', 0, \ldots, 0)'$. This result is contained in Corollary 1, and implies that for a moving-average process of order q the nonparametric estimator $\hat{\gamma}_{pk}$ of the GACV with $p = -1$ and sufficiently large m is asymptotically efficient for $0 \leq k \leq q$ and, hence, it estimates efficiently the first q inverse autocovariances.

3.2.1 AR(1) example

Let us consider an example to show what the elements of $\mathbf{c}, \mathbf{v}$ and $\mathbf{A}$ are, and their role in assessing asymptotic efficiency of the estimator $\hat{\gamma}_{pk}$. For simplicity, consider an AR(1) process $\{X_t\}$, and asymptotic efficiency of $\hat{\gamma}_{1k}$. The spectral density function is:

$$f_\theta(\omega) = \frac{\sigma}{2\pi}\frac{1}{1 - 2\phi\cos\omega + \phi^2} \tag{3.32}$$

For a stationary AR(1) process the estimator $\hat{\gamma}_{1k}$ of the traditional autocovariance function is asymptotically efficient for $k = 0, 1$.
We find the elements of the vector $\boldsymbol{c}$ that satisfies (3.16).
Let us set $k = 0$. Consider the left-hand side of (3.16):

$$\begin{aligned}
\phi^2(e^{i\omega})\phi^2(e^{-i\omega})f_\theta^2(\omega)\cos(k\omega) = \\
= \phi^2(e^{i\omega})\phi^2(e^{-i\omega})\frac{\sigma^2}{4\pi^2}\frac{1}{\phi^2(e^{i\omega})\phi^2(e^{-i\omega})}\cos(k\omega) = \\
= \frac{\sigma^2}{4\pi^2}\cos(k\omega) = \frac{\sigma^2}{4\pi^2} = v
\end{aligned} \tag{3.33}$$

On the right-hand side of (3.16), the elements of the matrix $\mathbf{A}$ are:

$$\phi^2(e^{i\omega})\phi^2(e^{-i\omega})\frac{\partial f_\theta(\omega)}{\partial \sigma} = \frac{1}{2\pi}\phi(e^{-i\omega})\phi(e^{i\omega}) =$$
$$= \frac{1 - 2\phi\cos\omega + \phi^2}{2\pi} = a_{11} + a_{12}\cos\omega$$
$$a_{11} = \frac{1+\phi^2}{2\pi}, \quad a_{12} = \frac{-\phi}{\pi}$$
$$\phi^2(e^{i\omega})\phi^2(e^{-i\omega})\frac{\partial f_\theta(\omega)}{\partial \phi} = \frac{\sigma(2\cos\omega - 2\phi)}{2\pi} =$$
$$= \frac{\sigma}{\pi}\cos\omega - \frac{\sigma\phi}{\pi} = a_{21} + a_{22}\cos\omega$$
$$a_{21} = -\frac{\sigma\phi}{\pi}, \quad a_{22} = \frac{\sigma}{\pi}$$

Hence, $\mathbf{A}$ is:

$$\mathbf{A} = \begin{bmatrix} \frac{1+\phi^2}{2\pi} & \frac{-\phi}{\pi} \\ -\frac{\sigma\phi}{\pi} & \frac{\sigma}{\pi} \end{bmatrix}$$

And

$$(A')^{-1} = \begin{bmatrix} \frac{2\pi}{1-\phi^2} & \frac{2\pi\phi}{1-\phi^2} \\ \frac{2\pi\phi}{\sigma(1-\phi^2)} & \frac{\pi(1+\phi^2)}{\sigma(1-\phi^2)} \end{bmatrix}$$

Then, the vector $\mathbf{c}$ is given by:

$$\boldsymbol{c} = -(\mathbf{A}')^{-1}(v, 0)' =$$
$$= \begin{bmatrix} \frac{2\pi}{1-\phi^2} & \frac{2\pi\phi}{1-\phi^2} \\ \frac{2\pi\phi}{\sigma(1-\phi^2)} & \frac{\pi(1+\phi^2)}{\sigma(1-\phi^2)} \end{bmatrix} \begin{pmatrix} \frac{\sigma^2}{4\pi^2} \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{\sigma^2}{2\pi(1-\phi^2)} \\ -\frac{\sigma\phi}{2\pi(1-\phi^2)} \end{pmatrix}$$

Then, Eq. (3.11) for an AR(1) holds.

3.3 Numerical illustrations

Some specific cases of ARMA processes are considered and the performance of the nonparametric estimator of interest is compared to that of the sample estimator of the sample autocovariance function, asymptotically approximated by $\hat{\gamma}_{1k}$.

Let us start by analysing an AR(1) process $\{X_t\}_{t\in T}$. Its spectral density function is:

$$f_\theta(\omega) = \frac{\sigma}{2\pi}\frac{1}{1-2\phi\cos\omega+\phi^2}.$$

The parameter vector is $\boldsymbol{\theta} = (\phi, \sigma^2)'$. We denote the asymptotic variance of $\hat{\gamma}_{pk}$ by

$$AV\{\hat{\gamma}_{pk}\} = m(C(m;p,p)-1)\frac{\sigma^{2p}}{\pi}\int_{-\pi}^{\pi}\left(\frac{1}{1-2\phi\cos\omega+\phi^2}\right)^{2p}\cos^2(k\omega)\,d\omega.$$

Tables 3.1 and 3.2 refers to the AR(1) process with parameters parameters $\phi = 0.8$ and $\sigma^2 = 1$. It displays the values of $AV\{\hat{\gamma}_{pk}\}$, $CRB\{\hat{\gamma}_{pk}\}$ and their ratio for different values of m and p. Recall that for Gaussian processes it must be $m > -2p$. Values greater than one measure the inefficiency of the estimator (2.20). As we know, the sample autocovariance function $\hat{\gamma}_{1k}$ is asymptotically efficient for $k = 0, 1$. Except for this case, exact equality between the asymptotic variance and Cramér-Rao bound of $\hat{\gamma}_{pk}$ never holds, but it is approximated as m increases.

Table 3.1: Asymptotic efficiency of $\hat{\gamma}_{pk}$ for an AR(1) model with $\phi = 0.8$ and $\sigma^2 = 1$.

k	p	m	AV	CRB	AV/CRB	k	AV/CRB
1	2	1	$1.35 \cdot 10^5$	$9.94 \cdot 10^4$	1.36	2	1.38
		2	$1.26 \cdot 10^5$	$9.94 \cdot 10^4$	1.26		1.29
		30	$1.09 \cdot 10^5$	$9.94 \cdot 10^4$	1.10		1.12
	3/2	1	$3.09 \cdot 10^3$	$2.81 \cdot 10^3$	1.09		1.11
		2	$3.03 \cdot 10^2$	$2.81 \cdot 10^3$	1.07		1.09
		30	$2.91 \cdot 10^3$	$2.81 \cdot 10^3$	1.03		1.05
	1	1	$6.75 \cdot 10$	$6.75 \cdot 10$	1.00		1.01
	1/2	1	1.24	1.02	1.22		1.25
		2	1.20	1.02	1.17		1.20
		30	1.14	1.02	1.12		1.15
	-1/2	2	0.89	0.11	7.98		$1.20 \cdot 10^2$
		30	0.42	0.11	3.79		$5.70 \cdot 10$
	-1	3	$1.38 \cdot 10$	1.64	8.43		∞
		30	4.93	1.64	3.01		∞

Table 3.2: Asymptotic efficiency of $\hat{\gamma}_{pk}$ for an AR(1) model with $\phi = 0.8$ and $\sigma^2 = 1$.

k	p	m	AV/CRB	k	AV/CRB	k	AV/CRB
4	2	30	1.21	5	1.28	7	1.53
		50	1.20		1.27		1.52
	3/2	30	1.15		1.25		1.59
		50	1.15		1.25		1.59
	1	1	1.16		1.31		1.88
	1/2	30	1.56		1.99		3.58
		50	1.56		1.99		3.58
	-1/2	30	$5.02 \cdot 10^2$		$1.06 \cdot 10^3$		$3.9 \cdot 10^3$
		50	$4.94 \cdot 10^2$		$1.04 \cdot 10^3$		$3.85 \cdot 10^3$
	-1	30	1.61		8.48		3.46
		50	1.56		8.25		3.37

For positive values of p no constraint on the pooling parameter is needed, so we let m vary from 1 to 30 for $k = 1, 2$, to show the effect of pooling. As it is shown in Table 3.1, for $p = 2$ and $k = 1$ the asymptotic variance (and also the relative CRB) of the estimator of γ_{2k} has a strong increase with respect to the case for $p = 1$. However, pooling has a positive effect on the asymptotic efficiency of the estimator, as a value of $m = 30$ reduces its asymptotic variance by 18.7%. For negative powers, the pooling parameter m has a strong effect on the variance of $\hat{\gamma}_{pk}$ (keeping p fixed). If we consider estimation of the inverse autocovariance function $\gamma_{-1,k}$, it must be $m > 2$. For $m = 3$ the asymptotic variance of $\hat{\gamma}_{-1,1}$ is $AV\{\hat{\gamma}_{-1,1}\} = 13.82$, while, considering $m = 30$ gives $AV\{\hat{\gamma}_{-1,1}\} = 4.93$, with a reduction of 64.3%.

In general, from tables 3.1, 3.2 we can see that as the distance of p from unity increases the estimator of γ_{pk} gets far from the condition for asymptotic efficiency, as expected

from the previous considerations. For $p \neq 1$, pooling contributes in reducing the variance of the estimator, and especially when p is negative pooling has a sizeable impact on the variance, as shown by the important reduction achieved for $p = -1$.

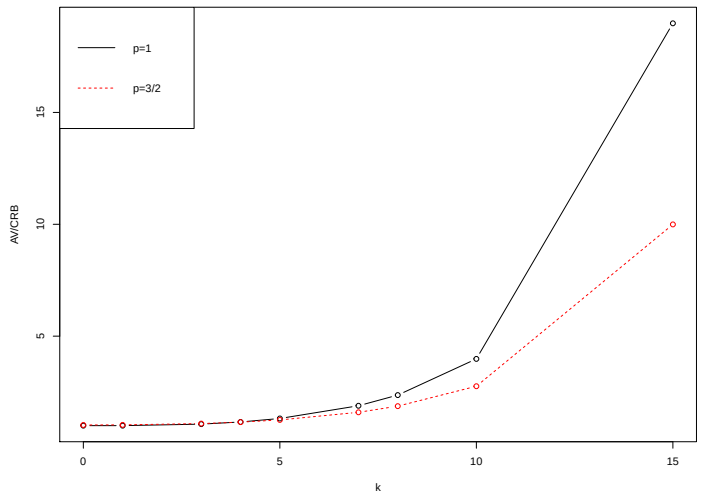

Figure 3.2: Plot of AV/CRB vs k for $p = 1.5$ and $p = 1$ for an AR(1) process with $\sigma = 1, \phi = 0.8$.

If we consider values of k larger than 1, neither the sample estimator $\tilde{\gamma}_k$ nor $\hat{\gamma}_{pk}$ is asymptotically efficient, and the results get worse as the lag k increases, *ceteris paribus*. For instance, for $k = 2$ the ratio between the asymptotic variance and CRB for $\hat{\gamma}_{pk}$ with $p = -1/2$ dramatically increases, and approaches infinity for $p = -1$. However, for $k > 1$ AV/CRB for some of the GACVs estimators tend to unity as m increases. Note also that for $k \geq 4$ the estimator of the GACV $\hat{\gamma}_{pk}$ for several values of p has better results than the sample estimator of the ACVF ($p = 1$), since the ratio AV/CRB relative to the first estimator is closer to unity, as shown in Table 2. For instance, with $k = 7$ the ratio AV/CRB for $\hat{\gamma}_{1,7}$ (or equivalently $\tilde{\gamma}_7$) is $AV/CRB = 1.88$, while the same ratio for $\hat{\gamma}_{2,7}$ and $\hat{\gamma}_{3/2,7}$ is 1.59 and 1.53 respectively, with $m = 30$. The ratios AV/CRB relative to $\hat{\gamma}_{pk}$ and $\hat{\gamma}_{1k}$ both increase as k increases, and the difference between them also becomes larger in favour of the estimator of the GACV for several values of p. This is clear from the plot in Figure 3.2 of the ratio AV/CRB against k, where the red dashed line refers

to the estimator of the GACV for $p = 3/2$ and the black solid line refers to the SACVF.

Let $\{X_t\}$ be a zero-mean Gaussian AR(2) stationary process. The spectral density function is:

$$f_\theta(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{(1 + \phi_1^2 + \phi_2^2 + (2\phi_1\phi_2 - 2\phi_1)\cos(\omega) - 2\phi_2\cos(2\omega))}$$

The vector of the parameters is $\boldsymbol{\theta} = (\phi_1, \phi_2, \sigma^2)'$. Let us investigate asymptotic efficiency of the estimator $\hat{\gamma}_{pk}$, as the power p and the pooling parameter m vary, and compare the results with those relative to an AR(1) process.

Table 3.3: Asymptotic efficiency of $\hat{\gamma}_{pk}$ for an AR(2) model with $\phi_1 = 0.7, \phi_2 = -0.1, \sigma = 1$.

k	p	m	AV/CRB	k	AV/CRB	k	AV/CRB
1	2	1	1.26	2	1.30	3	1.45
		2	1.18		1.21		1.36
		30	1.02		1.05		1.18
	3/2	1	1.06		1.08		1.22
		2	1.04		1.06		1.20
		30	1.007		1.02		1.15
	1	1	1.00		1.00		1.17
	1/2	1	1.09		1.17		1.56
		2	1.05		1.13		1.50
		30	1.004		1.08		1.43
	-1/2	2	2.34		6.35		54.70
		30	1.11		3.01		26.00
	-1	3	3.39		10.29		∞
		30	1.21		3.67		∞

From Table 3.3 we can see that the results are coherent with those relative to an AR(1) process: as p gets far from unity, the ratio between the asymptotic variance and the CRB of $\hat{\gamma}_{pk}$ increases. As before, for $p \neq 1$ the pooling effect is stronger for negative values of the power p, and as the lag k increases the results get worse. We can note that the estimators of the functions $\gamma_{3/2k}$ and $\gamma_{1/2k}$ at lag $k = 1$ perform very well, as the relative ratios between the asymptotic variance and CRB are respectively 1.007 and 1.004, very close to unity. In this case we also have better results for $\hat{\gamma}_{pk}$ with $p < 0$ and $k = 1, 2$.

It should be noted from Table 3.3 that for $k = 3$ the sample autocovariance, approximated by $\hat{\gamma}_{1,3}$, is not asymptotically efficient, as expected, and the ratio of its asymptotic variance and CRB is 1.17. In this case the estimator $\hat{\gamma}_{2,3}$ with $m = 30$ and the sample ACVF perform equally well. The ratio AV/CRB in this case decreases towards unity as m increases. The ratio AV/CRB relative to the estimator $\hat{\gamma}_{p3}$ with $p = 3/2$ is equal to 1.15. So in this case the nonparametric estimator $\hat{\gamma}_{pk}$ performs better than the sample estimator of the autocovariance function. Note that considering this AR(2) process, some other combination of the power p and the lag k produce better results in terms of efficiency, with respect to the AR(1) case: the estimators of $\gamma_{2k}, \gamma_{3/2k}, \gamma_{1/2k}$ at the lags $k = 1, 2$.

The following table refers to a moving-average process of order 1.

Table 3.4: Asymptotic efficiency of $\hat{\gamma}_{pk}$ for an MA(1) model with $\theta = 0.7$.

k	p	m	AV/CRB	k	AV/CRB
1	2	1	2.32	2	30.26
		30	1.89		24.60
		50	1.88		24.44
	3/2	1	1.97		78.73
		30	1.86		74.25
	1	1	1.85		2.02
	-1/2	2	2.30		4.36
		30	1.09		2.07
		50	1.07		2.04
	-1	3	3.00		4.15
		30	1.07		1.48
		50	1.04		1.44
	-2	5	6.62		7.78
		50	1.16		1.36

Table 3.4 shows a specific example of results about moving-average processes. In these cases results about asymptotic efficiency of $\hat{\gamma}_{pk}$ for positive and negative powers p are reversed with respect to the autoregressive case: the numerical results in the table above show good estimates for negative powers p, improving as m increases. As an example, the ratio AV/CRB for the estimator of the inverse autocovariance function ($p = -1$) is 1.07 for $m = 30$ and 1.04 with $m = 50$, very close to unity.

3.4 Minimum contrast estimation

Let us consider the process with spectral density function

$$[2\pi f_\theta(\omega)]^p = [\theta(\omega)]^{-1}, \tag{3.34}$$

where $\theta(\omega) > 0$ is the trigonometric polynomial $\theta_0 + 2\sum_{k=1}^{K} \theta_k \cos(\omega k)$. Writing $\theta(\omega) = \theta_0 |\phi(e^{-\imath\omega})|^2$, $\phi(e^{-\imath\omega}) = 1 - \sum_{j=1}^{K} \phi_j e^{-\imath\omega j}$, such that $\theta_k = \theta_0 \sum_{j=1}^{K-k} \phi_j \phi_{j+k}$, and setting $\sigma^2 = \theta_0^{-1}$, it can be seen by integrating both sides of (3.34) over $\omega \in [-\pi, \pi]$, that γ_{pk} is the autocovariance function of the AR(K) process $U_t = \sum_{j=1}^{s} \phi_j U_{t-j} + \sigma\epsilon_t, \epsilon_t \sim$ i.i.d. N$(0,1)$.

Following Taniguchi (1987), let us consider minimum contrast (MC) estimation of the spectral density $f_\theta(\omega)$ using the contrast functional

$$K(z;p) = \ln(z^p) + \frac{1}{z^p},$$

applied to $f_\theta(\omega)/g_N(\omega)$, where $g_N(\omega)$ is the corrected pooled periodogram, $g_N(\omega) = \bar{I}(\omega) \sqrt[p]{\Gamma(m)/\Gamma(m+p)}$ such that $\mathbb{E}\{[g_N(\omega)]^p\} = [f_\theta(\omega)]^p$.

Define

$$Y(\omega) = \frac{1}{2\pi} \sum_{-M+1}^{M-1} \hat{\gamma}_{pk} e^{-\imath\omega k}, \omega \in [-\pi, \pi],$$

so that $\hat{\gamma}_{pk} = \int_{-\pi}^{\pi} Y(\omega) e^{\imath\omega k} d\omega$, and $g_N(\omega) = [Y(\omega)]^{1/p}$.

The MC estimator of $(\phi_1, \ldots, \phi_K, \sigma^2)'$ is the minimizer of

$$\int_{-\pi}^{\pi} K\left(\frac{f_\theta(\omega)}{g_N(\omega)}, p\right) d\omega = \int_{-\pi}^{\pi} \left\{ \ln \sigma^2 - \ln |\phi(e^{-\imath\omega})|^2 - \ln Y(\omega) + \frac{1}{\sigma^2} Y(\omega) |\phi(e^{-\imath\omega})|^2 \right\} d\omega.$$

The MC estimator of σ^2 is $\hat{\sigma}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} Y(\omega) |\hat{\phi}(e^{-\imath\omega})|^2 d\omega$. Replacing in the contrast function (and noticing $\int_{-\pi}^{\pi} |\hat{\phi}(e^{-\imath\omega})|^2 d\omega = 0$), the MC estimator of $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_s)'$ is the minimizer of the criterion function

$$Q(\boldsymbol{\phi}) = \int_{-\pi}^{\pi} Y(\omega) |\phi(e^{-\imath\omega})|^2 d\omega.$$

Writing

$$|\phi(e^{-\imath\omega})|^2 = 1 - 2\boldsymbol{\phi}' \mathbf{b}(\omega) + \boldsymbol{\phi}' \mathbf{B}(\omega) \boldsymbol{\phi}$$

where $\mathbf{b}(\omega) = [\cos\omega, \cos(2\omega), \ldots, \cos(\omega K)]'$ and $\mathbf{B}(\omega) = \{\cos(\omega(h-k)), h, k = 1, 2, \ldots, s\}$, differentiating with respect to $\boldsymbol{\phi}$ and setting the derivatives equal to zero yields

$$\frac{\partial Q}{\partial \boldsymbol{\phi}} = \int_{-\pi}^{\pi} Y(\omega)(\mathbf{b}(\omega) - \mathbf{B}(\omega)\boldsymbol{\phi}) d\omega \equiv 0,$$

which is the generalised Yule-Walker system of equations:

$$\hat{\gamma}_{pk} = \sum_{j=1}^{K} \hat{\phi}_j \hat{\gamma}_{p,k-j}, k = 1, 2, \ldots, K.$$

Hence, an asymptotically efficient estimator of $(\boldsymbol{\phi}, \sigma^2)$ and thus of $\boldsymbol{\theta}$ can be obtained by solving a generalised Yule-Walker system base on the GACV estimator (2.20).

3.5 Choice of the pooling parameter m

As we have seen, an important issue to be addressed for estimation of the GACV is the choice of the pooling parameter m, since it allows $(m \to \infty)$ the asymptotic variance of the estimator $\hat{\gamma}_{pk}$ to achieve its lower bound. However this parameter also has an influence on other properties of the estimator, that have to be considered to have a good estimate of the GACV. These properties, and hence selection of m, will strongly depend on the power p. First of all, asymptotic theory requires M and m to be sufficiently large and $\frac{m}{M}$ sufficiently small. For negative powers p we need to impose a constraint on m to ensure the existence of the variance of the r.v. $\left(2\pi \sum_{l=1}^{m} I(\omega_{jm+l})\right)^p$. Moreover, the asymptotic properties of the estimator of the GACV depend on the power p, and for instance, its asymptotic variance is higher for negative p, in which case increasing m has a strong effect.

Note that the estimator of the GACV uses a Daniell-type estimator for the spectral density, which is plugged in the whole expression. Here m plays the role of a smoothing parameter, controlling the resolution of the estimator. In our specific case the pooled periodogram ordinates over m consecutive frequencies are used. Hence, we need $\frac{m}{M}$ to be sufficiently small for $f(\omega)$ to be effectively constant over frequency intervals of length $\frac{2\pi m}{M}$, and M and m to be sufficiently large for the asymptotics to be valid. For a fixed sample size N, increasing m reduces the variance of the estimator, producing a smoother estimate, but increases the finite sample bias, as the pooled periodogram ordinates over m frequencies are used instead of the raw ordinates at each frequency.

3.5.1 Choice of m in large samples

Let us consider a time series of length N large enough for asymptotics to be valid. The estimator $\hat{\gamma}_{pk}$ is asymptotically unbiased, hence the optimal selection of m will be based on minimizing the asymptotic variance, given by (3) with $k = l$, for a Gaussian stationary process. Minimization of this quantity consists in minimizing the multiplicative factor $m(C(m;p,p)-1)$, which is a nonincreasing function of the pooling parameter m for each given value of p. This can be easily seen graphically, by looking at the plot in Figure 3.3.

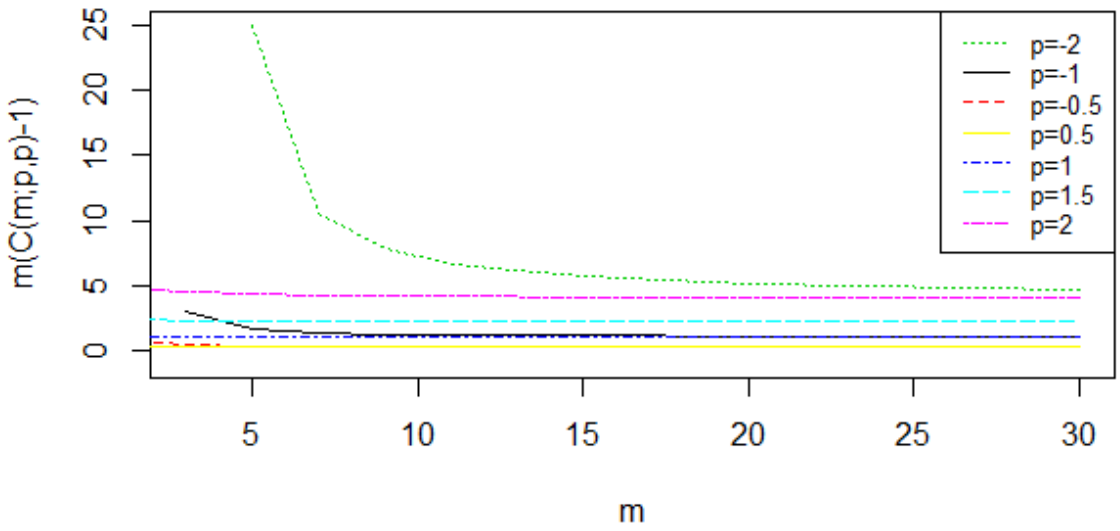

Figure 3.3: Plot of $m(C(m;p,p)-1)$ vs m at different values of the power p

The Gamma function $\Gamma(z)$ is an analytic function whose only finite singularities are $z = 0, -1, -2.\ldots$. In our case we consider $m > -2p$. The plot confirms that the factor $m(C(m;p,p)-1)$ (and hence the asymptotic variance of $\hat{\gamma}_{pk}$) decreases as the pooling parameter m increases. From the plot we can also observe that as the distance of the power p from unity increases the value of $m(C(m;p,p)-1)$ also increases, and the latter is greater for negative p than for positive p. This behaviour reflects on the variance of the estimator, as we saw in the previous ARMA examples.

The derivative of $m(C(m;p,p)-1)$ with respect to m, with $m > -2p$, is given by:

$$\frac{\Gamma(m+2p)\Gamma(m+1)}{[\Gamma(m+p)]^2}\left\{\Psi(m+2p)+\Psi(m+1)-2\Psi(m+p)\right\}-1$$

where $\Psi(z) = \frac{\partial ln\Gamma(z)}{\partial z} = \frac{\Gamma'(z)}{\Gamma(z)}$ is the Digamma function, which has simple poles at $z = 0, -1, -2, \ldots$. From Figure 3.3, $m(C(m;p,p)-1)$, for a given value of p, is a nonincreasing function of m, with a horizontal right asymptote. As shown in Appendix A, the horizontal asymptote $h(p)$ of $m(C(m;p,p)-1)$ as $m \to \infty$ is:

$$h(p) = p^2$$

The pooling parameter m can be selected as follows. We implement a code with R consisting of a function that increments the value of m by 1 unit, and evaluates the reduction in $m(C(m;p,p)-1)$. The optimal m selected is the one at which this reduction is small enough. Let us define

$$\Delta = |m_{i+1}(C(m_{i+1};p,p)-1) - m_i(C(m_i;p,p)-1)|,$$

We initialize Δ and set a tolerance level, say ϵ, small enough. The function takes as arguments an initial value for the parameter $m, (m > -2p)$, the value for the power p, and the maximum number of iterations that we want.

- Initialize m and the iteration number it;
- while $\Delta > \epsilon$ AND $it < it.max$
  - update the iteration number it;
  - update m;
  - evaluate $m(C(m;p,p)-1)$ at m;
  - evaluate Δ;
- return the value of m, $m(C(m;p,p)-1)$ and the iteration number.

Table 3.5 reports the selected values of the pooling parameter m for some values of the power p. The maximum number of iterations is 100 and the tolerance level ϵ is chosen as 1% of the range of the function for each value of p considered, as these functions have different slopes. For instance for $p = -2$ the curve is steeper than the others and we need $\epsilon = 0.21$. Note that the selected m strongly depends on the tolerance level chosen.

Table 3.5: Selected values of the pooling parameter m for $p = 2, 3/2, 1/2, -1/2, -1, -2$.

p	*min*	*max*	*range*	*epsilon*	m
2	4	5	1	0.01	14
3/2	2.25	2.39	0.14	0.014	5
1/2	0.25	0.27	0.02	0.002	5
−1/2	0.25	0.54	0.29	0.0029	12
−1	1	3	2	0.02	13
−2	4	25	21	0.21	15

The table shows that as p gets far from unity a higher value of m is selected, and the greatest value selected is 15 for estimating γ_{pk} with $p = -2$, while the lowest selected m is 5 for both $p = 3/2$ and $p = 1/2$.

3.5.2 Choice of m in finite samples

As we have seen, selection of m based on the asymptotic properties of the estimator of the GACV gives different results depending on p. However, these results are based on the assumption that the asymptotic distribution of $\hat{\gamma}_{pk}$ is valid to a good approximation, and we know that this requires a very large sample. As an example, with a size $N = 200$ and $m = 5$ M is only 19.

One method to select the optimal value of m based on a given sample is the use of

the Jackknife (Quenouille, 1949), as indicated by Luati et al. (2012). Here we propose the use of a generalization of the frequency domain bootstrap MPB (multiplicative periodogram bootstrap), as described in Meyer et al. (2018), which is valid for Gaussian stationary processes, and is based on the assumption of asymptotic independence of the periodogram ordinates. The MPB approach is often used to estimate the distribution of periodogram-based estimators $M(\varphi, I_N) = \int_{-\pi}^{\pi} \varphi(\omega) I(\omega)\, d\omega$ of a spectral mean $M(\varphi, f)$. The integral in the estimator $M(\varphi, I_N)$ is commonly approximated by a Riemann sum over the Fourier frequencies ω_j, $j \in \mathcal{G}(N)$, with $\mathcal{G}(N) := \big\{j \in \mathbb{Z} : 1 \leq |j| \leq [N/2]\big\}$.
We can view the estimator $\hat{\gamma}_{pk}$ of the GACV as an estimator $\hat{\gamma}_{pk} = \frac{1}{2\pi} M\left(\cos(\omega k), [2\pi \bar{I}_j]^p \frac{\Gamma(m)}{\Gamma(m+p)}\right)$ of the spectral mean $\gamma_{pk} = \frac{1}{2\pi} M(\cos(\omega k), [2\pi f(\omega)]^p)$. We know that for M and m large enough and $\frac{M}{m}$ is small enough, $2\pi \bar{I}_j = 2\pi f(\bar{\omega}_j) X_j$, with X_j i.i.d. Gamma random variables, $X_j \sim Ga(m, 1)$ and $\bar{\omega}_j$ are the mid range frequencies defined in section 2.1. By following the same reasoning as the standard MPB procedure, we can generate the pseudo-innovations U_j^* as i.i.d. $Ga(m, 1)$ r.v.s. Then define $U_j^{*p} = [U_j^*]^p \frac{\Gamma(m)}{\Gamma(m+p)}$, $j \in \mathcal{G}_1(N)$, where $\mathcal{G}_1(N) := \big\{j \in \mathbb{Z} : 0 \leq j \leq M - 1\big\}$.
Let $T_j^* = U_j^{*p} [2\pi \hat{f}_N(\bar{\omega}_j)]^p$, where $\hat{f}_N$ is a consistent (e.g. kernel-type) estimator of the spectral density f based on a sample of size N. The random variables $T^*(\omega_j)$ are then supposed to mimic the behaviour of $[2\pi \bar{I}_j]^p \frac{\Gamma(m)}{\Gamma(m+p)}$. The distribution of $L_{p,N} = \frac{1}{2\pi} M_{\mathcal{G}_1}(\varphi, [2\pi \bar{I}_j]^p \frac{\Gamma(m)}{\Gamma(m+p)}) - \frac{1}{2\pi} M(\varphi, [2\pi f]^p)$ is approximated by the distribution of

$$V_{p,N}^* = \frac{1}{2\pi} M_{\mathcal{G}_1}(\varphi, T^*) - \frac{1}{2\pi} M_{\mathcal{G}_2}(\varphi, [2\pi \hat{f}_N]^p),$$

with $\mathcal{G}_2(N) := \left\{ j \in \mathbb{Z} : 1 \leq j \leq \left[\frac{N-1}{2}\right] \right\}$.

3.5.3 Real data illustration

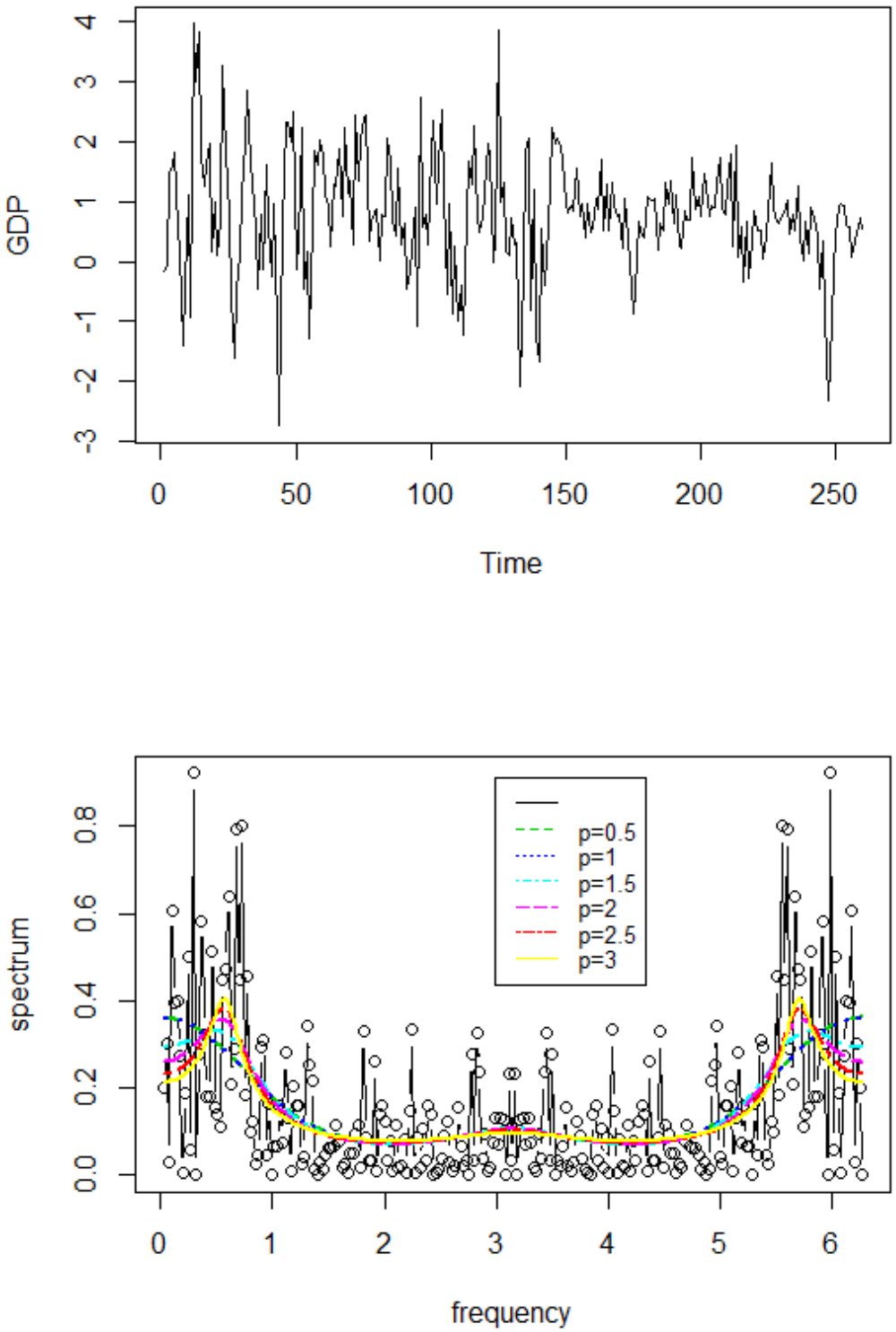

Figure 3.4: Quarterly growth rates of US Gross Domestic Product (1947.2-2012.1) and spectrum estimates.

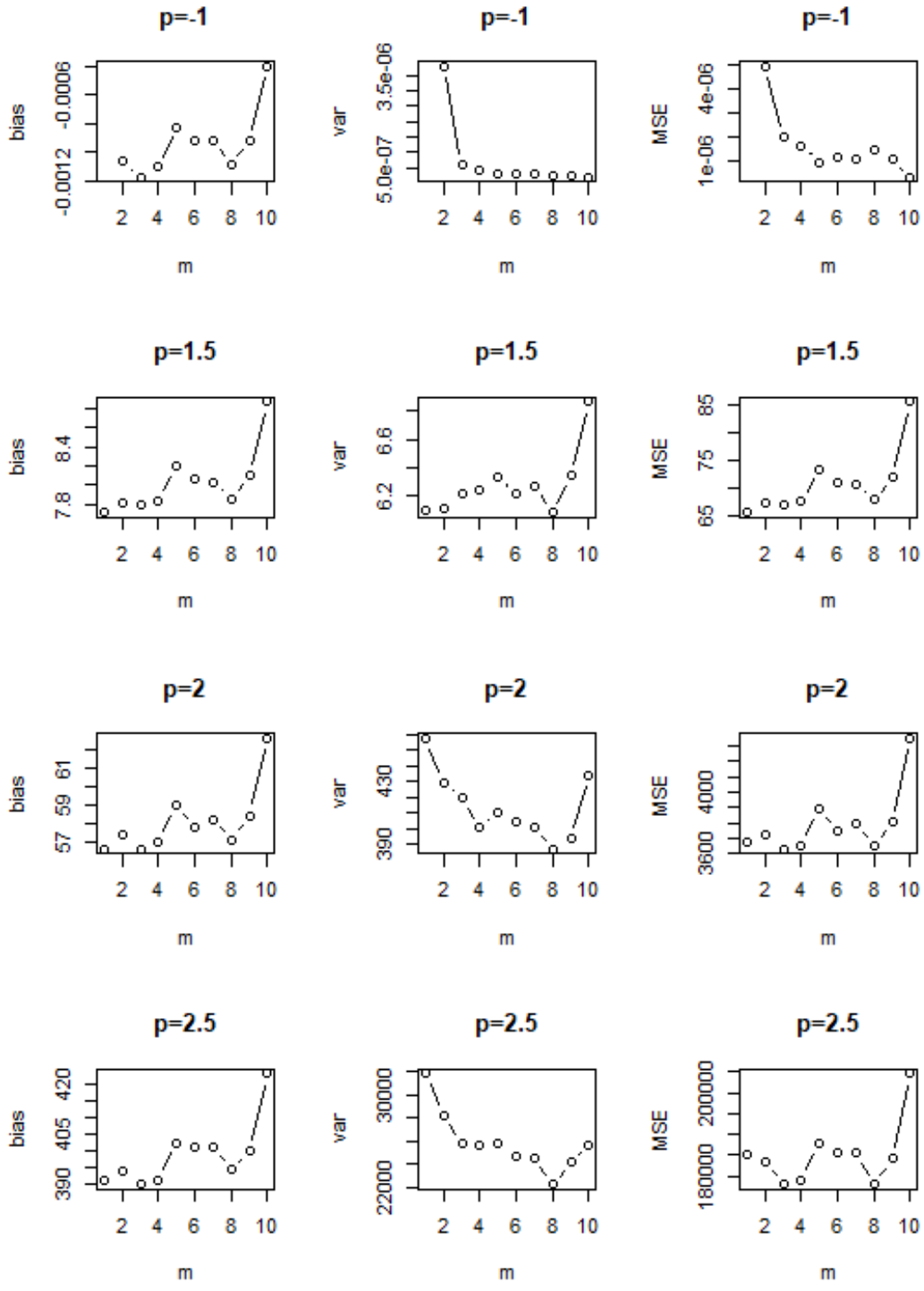

Figure 3.5: Plot of estimates of bias, variance and MSE of $\hat{\gamma_{pk}}$ for $k = 3$.

An empirical illustration of the theoretical results obtained is given by investigation of the quarterly growth rate series of the US Gross Domestic Product (GDP), whose cyclical nature has been largely inspected. Values of the power p different from 1 can help in identifying the cyclical features of the series. This can be achieved by solving a generalisation of the Yule-Walker system of equations. As p varies, the following expression describes a wider class of spectral models, from which the AR, MA, or fractional models are obtained as special cases:

$$2\pi f(\omega) = \Big[\frac{\sigma_p^2}{\phi_p(e^{i\omega})\phi_p(e^{-i\omega})}\Big]^{\frac{1}{p}}, \tag{3.35}$$

where $\phi_p(e^{i\omega}) = 1 - \phi_{p1}e^{i\omega} - \phi_{p2}e^{2i\omega} - \dots \phi_{pK}e^{Ki\omega}$. For positive values of p, $\phi_p(B)$ characterizes the autoregressive approximation of the process u_{pt}, from which the AR or MA approximation of the original process can be obtained recursively.
Given a time series realization, one can select the optimal value of p by minimizing a measure of deviance based on the Whittle's likelihood, $Dev(p) = \sum_{j=1}^{N}\big[\frac{I(\omega_j)}{\hat{f}_p(\omega_j)} + ln\hat{f}_p(\omega_j)\big]$. Estimates of the GACV for several powers p are needed to compute the spectral estimates $\hat{f}_p$ by the YW method. However, an important issue to be addressed before estimation of the GACV is the choice of the pooling parameter. Here we use the frequency domain bootstrap procedure described above for selecting the optimal level of pooling m for several values of p considered. The length of the series is 260, and we used $B = 10000$ bootstrap replications to estimate bias, variance and MSE of the estimators of the GACVs. A smoothed periodogram estimator with a Gaussian kernel is used as a consistent estimator $\hat{f}$. Figure 3.5 shows some of the results obtained, for $p = -1, 1.5, 2, 2.5$ and $k = 3$. From the plot, a suitable value of m would be 3 for $k = 3$, $p = 1.5$ and $p = 2$, or 4 for $p = 2.5$. The GACV estimates, obtained with optimal pooling parameters according to p, are used to solve the YW equation system to get the estimates of the parameters that characterize the AR approximation of the process u_{pt}. The corresponding estimates of the spectrum for some positive p are shown in Figure 3.4 together with the plot of the series. From the plot it is evident that powers p greater than 1 emphasize spectral peaks, revealing the cyclical nature of the series. By optimization of the above measure of deviance we select $p = 2.5$. The corresponding spectrum estimate characterizes a new model. It is given by the red line in Figure 3.4, with peaks showing

a cyclical pattern. The coefficients of the AR approximation of the power-transformed process for $p = 2.5$, estimated by the YW method, suggest an AR(3) model for u_{pt}.

3.6 Appendix

Let us derive the expression for the Fisher information matrix $\mathfrak{I}_N$ associated with $X_1, \ldots, X_N$.

Let us consider a Gaussian stationary process. The likelihood function is:

$$L(\boldsymbol{\theta}) = (2\pi)^{-N/2}|\Sigma_N|^{-1/2}\exp\{-\frac{1}{2}X_N'\Sigma_N^{-1}X_N\}$$

with

$$\Sigma_N = (\sigma)_{j_1,j_2}$$

where:

$$(\sigma)_{j_1,j_2} = \int_{-\pi}^{\pi} f_\theta(\omega)\exp\left[i(j_1-j_2)\omega\right]d\omega \qquad j_1, j_2 = 1, \ldots, N$$

The derivative of the log-likelihood function with respect to θ_j is:

$$\frac{\partial \log\{L(\theta)\}}{\partial\theta_j} = \frac{1}{2}X_N'\Sigma_N^{-1}\Sigma_N^{(j)}\Sigma_N^{-1}X_N - \frac{1}{2}tr\Sigma_N^{-1}\Sigma_N^{(j)}$$

where, for $j = 1, \ldots, s$,

$$\Sigma_N^{(j)} = (\sigma_{j_1,j_2}^{(j)})$$

$$\sigma_{j_1,j_2}^{(j)} = \frac{\partial\sigma_{j_1,j_2}}{\partial\theta_j} = \int_{-\pi}^{\pi}\exp\left[i(j_1-j_2)\omega\right]\frac{\partial f_\theta(\omega)}{\partial\theta_j}d\omega \qquad j_1, j_2 = 1, \ldots, N$$

The following lemmata are useful for the derivation of the matrix $\mathfrak{I}_N$

Lemma 2 ((J. R. Magnus and Neudecker, 1979)). Let A and B be symmetric non-random matrices of order N. Then,

$$\mathbb{E}\left\{(X_N'AX_N)(X_N'BX_N)\right\} = (trA\Sigma_N)(trB\Sigma_N) + 2tr(A\Sigma_NB\Sigma_N). \tag{3.36}$$

Theorem 3 ((J. Magnus and Neudecker, 1999)). *Let X be a r.v. with $\mathbb{E}[X] = \boldsymbol{\mu}$ and $V[X] = \Sigma$, and let A be a symmetric non-random matrix. Then*

$$\mathbb{E}\left[X'AX\right] = tr(A\Sigma) + \boldsymbol{\mu}'A\boldsymbol{\mu} \tag{3.37}$$

Lemma 3 ((Taniguchi, 1986)). Let D and D^* be spaces of functions on $[-\pi, \pi]$ defined by:

$$D = \left\{ f | f(\omega) = \sum_{u=-\infty}^{\infty} a(u) \exp -i\omega, a(u) = a(-u), \sum_{u=-\infty}^{\infty} |u||a(u)| < \infty \right\}$$

$$D^* = \left\{ f | f(\omega) \in D, 0 < F_1 \leq f(\omega) \leq F_2 < \infty for \quad \omega \in [-\pi, \pi]. \right\}$$

For $g_1, g_2 \in D^*$ we define $\Gamma_1, \Gamma_2, \Lambda_1, \Lambda_2$, the $N \times N$ Toeplitz matrices, by:

$$\Gamma_s = \left(\int_{-\pi}^{\pi} \exp\{i(j_1 - j_2)\omega\} f_s(\omega)\, d\omega \right)$$

$$\Lambda_s = \left(\int_{-\pi}^{\pi} \exp\{i(j_1 - j_2)\omega\} g_s(\omega)\, d\omega \right)$$

$j_1, j_2 = 1, \ldots, N, s = 1, 2$. Then

$$\frac{1}{N} tr \Gamma_1 {\Lambda_1}^{-1} \Gamma_2 {\Lambda_2}^{-1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\omega) f_2(\omega)}{g_1(\omega) g_2(\omega)}\, d\omega + O(N^{-1}). \tag{3.38}$$

Under the assumptions 1-3 we can derive the expression for the Fisher information matrix applying the definition given and by using the cited theorems and lemmata.

$$i_N(j_1,j_2) = \mathbb{E}\left\{ \frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_{j_1}} \frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_{j_2}} \right\} =$$

$$= \mathbb{E}\left[\left(\frac{1}{2}X'\Sigma_N^{-1}\Sigma_N^{(j_1)}\Sigma_N^{-1}X - \frac{1}{2}tr\Sigma_N^{-1}\Sigma_N^{(j_1)}\right)\left(\frac{1}{2}X'\Sigma_N^{-1}\Sigma_N^{(j_2)}\Sigma_N^{-1}X - \frac{1}{2}tr\Sigma_N^{-1}\Sigma_N^{(j_2)}\right)\right] =$$

$$= \frac{1}{4}\mathbb{E}\left[X'\Sigma_N^{-1}\Sigma_N^{(j_1)}\Sigma_N^{-1}XX'\Sigma_N^{-1}\Sigma_N^{(j_2)}\Sigma_N^{-1}X\right] +$$

$$-\frac{1}{4}\mathbb{E}\left[X'\Sigma_N^{-1}\Sigma_N^{(j_1)}\Sigma_N^{-1}Xtr\Sigma_N^{-1}\Sigma_N^{(j_2)}\right] +$$

$$-\frac{1}{4}\mathbb{E}\left[tr(\Sigma_N^{-1}\Sigma_N^{(j_1)})X'\Sigma_N^{-1}\Sigma_N^{(j_2)}\Sigma_N^{-1}X\right] +$$

$$+\frac{1}{4}tr(\Sigma_N^{-1}\Sigma_N^{(j_1)})tr(\Sigma_N^{-1}\Sigma_N^{(j_2)}) =$$

$$= \frac{1}{4}tr(\Sigma_N^{-1}\Sigma_N^{(j_1)}\Sigma_N^{-1}\Sigma_N)tr(\Sigma_N^{-1}\Sigma_N^{(j_2)}\Sigma_N^{-1}\Sigma_N) +$$

$$+\frac{1}{2}tr(\Sigma_N^{-1}\Sigma_N^{(j_1)}\Sigma_N^{-1}\Sigma_N\Sigma_N^{-1}\Sigma_N^{(j_2)}\Sigma_N^{-1}\Sigma_N) +$$

$$-\frac{1}{4}tr(\Sigma_N^{-1}\Sigma_N^{(j_2)})\mathbb{E}\left[X'\Sigma_N^{-1}\Sigma_N^{(j_1)}\Sigma_N^{-1}X\right] +$$

$$-\frac{1}{4}tr(\Sigma_N^{-1}\Sigma_N^{(j_1)})\mathbb{E}\left[X'\Sigma_N^{-1}\Sigma_N^{(j_2)}\Sigma_N^{-1}X\right] +$$

$$+\frac{1}{4}tr(\Sigma_N^{-1}\Sigma_N^{(j_1)})tr(\Sigma_N^{-1}\Sigma_N^{(j_2)}) =$$

Now we can apply Lemma 3 considering

$$\Gamma_1 = \Sigma_N^{(j_1)} = \int_{-\pi}^{\pi} \exp\{i(j_1-j_2)\omega\}\frac{\partial f_\theta(\omega)}{\partial \theta_{(j_1)}}\, d\omega, \Gamma_2 = \Sigma_N^{(j_2)},$$

$$\Lambda_1 = \Lambda_2 = \Sigma_N = \int_{-\pi}^{\pi} \exp\{i(j_1-j_2)\omega\} f_\theta(\omega)\, d\omega$$

So that $f_1(\omega) = \frac{\partial f_\theta(\omega)}{\partial \theta_{j_1}}$, $f_2(\omega) = \frac{\partial f_\theta(\omega)}{\partial \theta_{j_2}}$ and $g_1(\omega) = g_2(\omega) = f_\theta(\omega)$.
Applying the Lemma gives:

$$i_N(j_1,j_2) = \frac{1}{4}tr(\Sigma_N^{-1}\Sigma_N^{(j_1)})tr(\Sigma_N^{-1}\Sigma_N^{(j_2)}) + \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\partial f_\theta(\omega)}{\partial\theta_{j_1}}\frac{\partial f_\theta(\omega)}{\partial\theta_{j_2}}\frac{1}{f_\theta^2(\omega)}\,d\omega + \\ -\frac{1}{4}tr(\Sigma_N^{-1}\Sigma_N^{(j_2)})\mathbb{E}\left[X'\Sigma_N^{-1}\Sigma_N^{(j_1)}\Sigma_N^{-1}X\right] + \\ -\frac{1}{4}tr(\Sigma_N^{-1}\Sigma_N^{(j_2)})\mathbb{E}\left[X'\Sigma_N^{-1}\Sigma_N^{(j_2)}\Sigma_N^{-1}X\right] + \frac{1}{4}tr(\Sigma_N^{-1}\Sigma_N^{(j_1)})tr(\Sigma_N^{-1}\Sigma_N^{(j_2)}) = \tag{3.39}$$

Now we can use Theorem 3:

$$\mathbb{E}\left[X'AX\right] = tr(A\Sigma_N) + \boldsymbol{\mu}'A\boldsymbol{\mu}$$

where in our case we have: $\boldsymbol{\mu} = \mathbf{0}$. In the first expectation $A = \Sigma_N^{-1}\Sigma_N^{(j_1)}\Sigma_N^{-1}$, so $\mathbb{E}\left[X'\Sigma_N^{-1}\Sigma_N^{(j_1)}\Sigma_N^{-1}X\right] = tr(\Sigma_N^{-1}\Sigma_N^{(j_1)}\Sigma_N^{-1}\Sigma_N) + \mathbf{0}$. While in the second expectation $A = \Sigma_N^{-1}\Sigma_N^{(j_2)}\Sigma_N^{-1}$, giving $\mathbb{E}\left[X'\Sigma_N^{-1}\Sigma_N^{(j_2)}\Sigma_N^{-1}X\right] = tr(\Sigma_N^{-1}\Sigma_N^{(j_2)}\Sigma_N^{-1}\Sigma_N) + \mathbf{0}$. Using these results we obtain:

$$\begin{aligned} i_N(j_1,j_2) &= \frac{1}{2}tr(\Sigma_N^{-1}\Sigma_N^{(j_1)})tr(\Sigma_N^{-1}\Sigma_N^{(j_2)}) + \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\partial f_\theta(\omega)}{\partial\theta_{j_1}}\frac{\partial f_\theta(\omega)}{\partial\theta_{j_2}}\frac{1}{f_\theta^2(\omega)}\,d\omega + \\ &-\frac{1}{4}tr(\Sigma_N^{-1}\Sigma_N^{(j_2)})tr(\Sigma_N^{-1}\Sigma_N^{(j_1)}\Sigma_N^{-1}\Sigma_N) - \frac{1}{4}tr(\Sigma_N^{-1}\Sigma_N^{(j_1)})tr(\Sigma_N^{-1}\Sigma_N^{(j_2)}\Sigma_N^{-1}\Sigma_N) = \\ &= \frac{1}{2}tr(\Sigma_N^{-1}\Sigma_N^{(j_1)})tr(\Sigma_N^{-1}\Sigma_N^{(j_2)}) + \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\partial f_\theta(\omega)}{\partial\theta_{j_1}}\frac{\partial f_\theta(\omega)}{\partial\theta_{j_2}}\frac{1}{f_\theta^2(\omega)}\,d\omega + \\ &-\frac{1}{2}tr(\Sigma_N^{-1}\Sigma_N^{(j_1)})tr(\Sigma_N^{-1}\Sigma_N^{(j_2)}) = \\ &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\partial f_\theta(\omega)}{\partial\theta_{j_1}}\frac{\partial f_\theta(\omega)}{\partial\theta_{j_2}}\frac{1}{f_\theta^2(\omega)}\,d\omega \end{aligned}$$

which provides the generic element of the Fisher information matrix.

Asymptotic efficiency of the estimator of the GACV depends on the factor $\frac{m(C(m;p,p)-1)}{p^2}$. Here we prove its convergence to 1 as $m \to \infty$.

We start by using a result about the approximation of the ratio of two Gamma functions, obtained by the use of the Stirling's series (Erdélyi, Magnus, Oberhettinger, and Tricomi, 1954):

$$\frac{\Gamma(z+\alpha)}{\Gamma(z+\beta)} = z^{\alpha-\beta}\left[1 + \frac{(\alpha-\beta)(\alpha+\beta-1)}{2z} + O(|z|^{-2})\right] \tag{3.40}$$

as $z \to \infty$, where α and β are bounded. By using this approximation we can rewrite

$$\frac{\Gamma(m)\Gamma(m+2p)}{[\Gamma(m+p)][\Gamma(m+p)]} \approx m^{-p}\Big[1+\frac{(-p)(p-1)}{2m}\Big]m^{p}\Big[1+\frac{p(3p-1)}{2m}\Big] = \\ = \frac{4m^2+4mp^2-3p^4+4p^3-p^2}{4m^2}, \tag{3.41}$$

and substitute it in the whole expression:

$$m(C(m;p,p)-1) = m\Big(\frac{\Gamma(m)\Gamma(m+2p)}{[\Gamma(m+p)][\Gamma(m+p)]}-1\Big) \approx \\ \approx m\Big(\frac{4m^2+4mp^2-3p^4+4p^3-p^2}{4m^2}-1\Big) \tag{3.42}$$

By applying the change of variable $t = \frac{1}{m}$ we can use the De L'Hopital theorem to find the limit as $t \to 0^+$:

$$\lim_{t\to 0^+} \frac{\Big(\frac{4t^{-2}+4p^2t^{-1}-3p^4+4p^3-p^2}{4t^{-2}}-1\Big)}{t} = \\ = \lim_{t\to 0^+} \frac{\frac{\partial}{\partial t}\Big(\frac{4t^{-2}+4p^2t^{-1}-3p^4+4p^3-p^2}{4t^{-2}}-1\Big)}{\frac{\partial}{\partial t}t} = p^2 \tag{3.43}$$

Hence the horizontal asymptote $h(p)$ of $m(C(m;p,p)-1)$ as $m \to \infty$ is:

$$h(p) = p^2 \tag{3.44}$$

and, $\lim_{m\to\infty} \frac{m(C(m;p,p)-1)}{p^2} = 1$.

Chapter 4

Multivariate extensions

In studying a phenomenon, we often encounter many variables, $X_{i,t}, i = 1, \ldots, r$, and the observations are taken over time. For convenience we use a vector, $\boldsymbol{X}_t = (X_{1,t}, X_{2,t}, \ldots, X_{r,t})'$ to denote the set of these variables, where $X_{i,t}$ is the ith component of the random vector at time t. A realization from the vector process $\boldsymbol{X}_t$ is a vector (or multivariate) time series. We will consider discrete time series, with $t \in T$, where T is a discrete time set. The fundamental characteristic of a multivariate time series is that its observations depend both on component i and on time t (Wei, 2019).
We first introduce some basic concepts of multivariate time series analysis, in both the time domain and the frequency domain.

As shown in the previous chapters for scalar time series, the definition of an auxiliary process, based on a power transformation of the original process, allows the definition and investigation of many important quantities in time series analysis, nested in the GACV. Proietti and Luati (2015) defined the power-transformed process u_{pt}, whose Wold coefficients can be obtained recursively from the Wold coefficients of the original process (Gould, 1974). This allows to give the GACV (and the quantities related to it, including the variance profile) an analytical form in terms of the coefficients that govern the process.
Formulae for the coefficients of powers of Taylor series have been known since a long time. In the work by Gould (1974) several forms of formulae for the coefficients of any

real power of Taylor series are proved and put in a logical order. These formulae are useful in the definition of the generalised linear cepstral models (Proietti and Luati, 2019), characterized by the generalised cepstral coefficients, directly connected with the GACV, and hence the variance profile. The generalised linear cepstral models are linear models for the Box-Cox transform of the spectral density. The exponential model (Bloomfield, 1973), which arises as a special case, is a linear model for the log-spectrum. A recursive formula for the coefficients of the log-transform of a polynomial is provided by Pourahmadi (1983). A multivariate extension of the latter formula to matrix polynomials and of the exponential model is provided by Holan et al. (2017). We aim at deriving a method to get the coefficients of a power transformation of a matrix polynomial, valid for any real power. This allows analytical definition (in terms of the power-transformed process) of the matrix GACV, the matrix variance profile, and generalised linear cepstral models for vector time series.

In this chapter we will consider vector stationary time series. An $r$dimensional vector process $\boldsymbol{X}_t$ is a stationary process if each of its components is a univariate stationary process, and its first two moments are time-invariant.

4.1 Multivariate spectral analysis of time series

Let $\boldsymbol{X}_t = (X_{1,t}, X_{2,t}, \ldots, X_{r,t})'$ be a $r$dimensional real-valued vector process, with mean vector $\boldsymbol{\mu}$ and lag k covariance matrix function $\boldsymbol{\Gamma}_k = \mathbb{E}[(\boldsymbol{X}_t - \boldsymbol{\mu})(\boldsymbol{X}_{t+k} - \boldsymbol{\mu})'] = [\gamma_{i,j,k}]_{i,j=1}^{r}$, where $\gamma_{i,j,k}$ is the covariance between $X_{i,t}$ and $X_{j,t+k}$, while for $i = j$, $\gamma_{ii,k}$ is the autocovariance function for the ith component, $X_{i,t}$. It can be easily seen that $\boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}'_{-k}$, and that $\boldsymbol{\Gamma}_0$ is the variance-covariance matrix of the process$\boldsymbol{X}_t$.
The correlation matrix function of the process is defined by:

$$\boldsymbol{\rho}_k = \boldsymbol{D}^{-1/2}\boldsymbol{\Gamma}_k\boldsymbol{D}^{-1/2},$$

where $\boldsymbol{D}$ is a diagonal matrix in which the ith diagonal element is the variance of the i component $X_{i,t}$.

Let us consider the Wold representation of the vector process $\boldsymbol{X}_t : \boldsymbol{X}_t = \Psi(L)\boldsymbol{\epsilon}_t$, where $\{\boldsymbol{\epsilon}_t\}$ is a vector White Noise process with mean zero and covariance matrix Σ. Then, the covariance matrix function is given by:

$$\boldsymbol{\Gamma}_k = \sum_{j \geq 0} \Psi_{j+k} \Sigma \Psi_j^{'} \tag{4.1}$$

Given a realization of length N, $\boldsymbol{x}_t$, from the vector process $\boldsymbol{X}_t$, the covariance matrix function can be naturally estimated by the sample covariance matrix function, given by:

$$\tilde{\boldsymbol{\Gamma}}_k = \frac{1}{N} \sum_{t=1}^{N-k} (\boldsymbol{x}_t - \bar{\boldsymbol{x}})(\boldsymbol{x}_{t+k} - \bar{\boldsymbol{x}})', \quad 0 \leq k \leq N-1, \tag{4.2}$$

where $\bar{\boldsymbol{x}}$ is the sample mean.

Results for spectral analysis of univariate time series can be readily generalised to a r dimensional vector process.
The spectral representation of the covariance matrix function is given by:

$$\boldsymbol{\Gamma}_k = \int_{-\pi}^{\pi} e^{i\omega k} d\boldsymbol{F}(\omega), \tag{4.3}$$

where $\boldsymbol{F}(\omega)$ is the spectral distribution matrix of $\boldsymbol{X}_t$. The diagonal elements $F_{i,i}(\omega)$ are the spectral distribution functions of $X_{i,t}$, and the off-diagonal elements $F_{i,j}(\omega)$ are the cross-spectral distribution function between $X_{i,t}$, and $X_{j,t}$.
If the covariance matrix is absolutely summable, in the sense that each of the $r \times r$ sequence $\gamma_{i,j,k}$ is absolutely summable, then, the spectral density matrix exists and is given by:

$$\boldsymbol{f}(\omega)d\omega = d\boldsymbol{F}(\omega) = [f_{i,j}(\omega)d\omega]. \tag{4.4}$$

Thus,

$$\boldsymbol{\Gamma}_k = \int_{-\pi}^{\pi} e^{i\omega k} \boldsymbol{f}(\omega)\, d\omega, \tag{4.5}$$

and

$$\boldsymbol{f}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \boldsymbol{\Gamma}_k e^{-i\omega k} = [f_{i,j}(\omega)], \tag{4.6}$$

where

$$f_{i,j}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_{i,j,k} e^{-i\omega k}. \tag{4.7}$$

The diagonal elements $f_{i,i}(\omega)$ are the spectral density functions of $X_{i,t}$, and the off-diagonal elements $f_{i,j}(\omega)$ are the cross-spectrum of $X_{i,t}$, and $X_{j,t}$.

when $k = 0$ we have $\mathbf{\Gamma}_0 = \int_{-\pi}^{\pi} \boldsymbol{f}(\omega)\, d\omega$. hence, the area under the multivariate spectrum is the variance-covariance matrix of the process.It can be easily seen that the spectral density matrix function $\boldsymbol{f}(\omega)$ is positive semidefinite, that is $\boldsymbol{c'f}(\omega)\boldsymbol{c} \geq 0$ for any nonzero $r$dimensional complex vector $\boldsymbol{c'}$ (Wei, 2019). Also, $\boldsymbol{f}(\omega)$is Hermitian, i.e. $\boldsymbol{f}^*(\omega) = \boldsymbol{f}(\omega)$.

For a vector process$\{\boldsymbol{X}_t\}$ with Wold representation $\boldsymbol{X}_t = \Psi(L)\boldsymbol{\epsilon}_t$, where $\{\boldsymbol{\epsilon}_t\}$ is a vector White Noise process with mean zero and covariance matrix Σ, the spectral density matrix function is given by:

$$\boldsymbol{f}(\omega) = \Psi(e^{-i\omega})\Sigma\Psi'(e^{i\omega}). \tag{4.8}$$

Given a realization of length N $\boldsymbol{x}_t$ from the vector process $\boldsymbol{X}_t$, the periodogram matrix represents the sample analog of the spectral density matrix. Similar to the univariate case, the periodogram matrix, or sample spectrum matrix, is defined as:

$$\boldsymbol{I}_N(\omega_j) = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} \boldsymbol{X}_t e^{-i\omega_j t} \right|^2, \tag{4.9}$$

and it can be shown that:

$$\boldsymbol{I}_N(\omega_j) = \frac{1}{2\pi} \sum_{k=-(N-1)}^{N-1} \tilde{\mathbf{\Gamma}}_k e^{-i\omega_j k}, \tag{4.10}$$

The periodogram matrix, as the periodogram for scalar time series, is also a poor estimate of the spectral density matrix, and to overcome this issue, smoothing techniques for the periodogram matrix are often used.

Similar to univariate time series, models for the spectrum of a vector time series can be specified. Holan et al. (2017) extended to vector time series the exponential model

for scalar time series defined by Bloomfield (1973).

4.2 The vector exponential model

The vector exponential model (Holan et al., 2017) represents a flexible frequency domain model expressed in terms of the spectral density matrix, which extends to the multivariate case the exponential model for scalar time series of Bloomfield (1973).
The vector exponential model (VEXP) has some advantages over VAR and VARMA models. The VEXP model is always invertible, which implies that the spectral density matrix is non-singular at all frequencies, and it is also stationary. Although when estimation proceeds in an unconstrained fashion, VAR and VARMA processes need not be stationary and invertible, there are practical scenarios in which restrictions on VAR and VARMA processes are needed: non-invertibility of a vector process makes Whittle estimation intractable, and also, for non-invertible vector processes long-term forecasting filters are not well-defined (McCracken and McElroy, 2014). For a VEXP model stationarity and invertibility are automatic, and the parameters are unconstrained.

The definition of the VEXP model involves the notion of matrix exponential (Artin, 1991). Since the spectral density matrix is Hermitian, it can be written as $\boldsymbol{f}(\omega) = PAP^*$, where P is a unitary matrix,$P^* = P^{-1}$, and $A = A(\omega)$ is diagonal with positive diagonal elements. Then, $\exp \boldsymbol{f} = P \exp(A) P^*$, where $\exp(A)$ is a diagonal matrix with diagonal elements being the exponential of the diagonal elements of A. Hence, we can write:

$$\boldsymbol{f} = \exp[P \log(A) P^*].$$

Holan et al. (2017) noted that the matrix $P \log(A) P^*$ can be expanded as:

$$P \log(A) P^* = \sum_{k=-\infty}^{\infty} \Theta_k e^{-i\omega k}, \tag{4.11}$$

where Θ_k are the Fourier coefficients of $P \log(A) P^*$, and they are called cepstral matrices (cepstral coefficients when considering scalar time series).

Therefore, the spectral density matrix can be written as:

$$\boldsymbol{f} = \exp\Big(\sum_{k=-\infty}^{\infty} \Theta_k e^{-i\omega k}\Big) \tag{4.12}$$

To define the VEXP model, Holan et al. (2017) denoted $\Psi(z) = \exp\Omega(z)$, for some power series $\Omega(z) = \sum_{k\geq 1}\Omega_k z^k$. When the cepstral matrices are all commutative $\Theta(z) = \Omega(z)$. The VEXP model is obtained by truncating the power series $\Omega(z)$ to a polynomial. Then, the Wold representation of the order q VEXP model is defined to be:

$$\Psi(z) = \exp[\Omega]_1^q(z) \tag{4.13}$$

The White Noise process has covariance matrix $\Sigma = \exp\Omega_0$, and the spectral density function of the VEXP(q) process can be written as:

$$\boldsymbol{f}(\omega) = \exp\{[\Omega]_1^q(e^{-i\omega})\}\exp(\Omega_0)\exp\{[\Omega']_1^q(e^{i\omega})\} \tag{4.14}$$

For the VEXP(q) process a condition is needed on the Wold power series, which is that $det\Psi(z) \neq 0$ for all $z \in D = \{z \in \mathbb{C} : |z| \leq 1\}$. Conversely, whenever a cepstral power series is also well-defined for $z \in D$, then, $\exp\Omega(z)$ is well-defined. In particular, $[\Omega]_1^q(z)$ is always convergent on $\mathbb{C}$, so that the VEXP(q) is always invertible and stationary. As established in Holan et al. (2017), the VEXP(q) process with $q < \infty$ is stationary, invertible and identifiable.

In the univariate case, one may differentiate Eq. (4.13) with respect to z and match the coefficients to get the recurrence relations in (Pourahmadi, 1983). This approach is demonstrably false in the multivariate case, since differentiation of the matrix exponential must allow for the non-Abelian algebra, as pointed out in Holan et al. (2017). Instead, they provide formulas relating the Wold matrix coefficients to the cepstral matrices, by expanding the matrix exponential using a Taylor series. The Wold coefficients are obtained by:

$$\Psi_k = \sum_{l\geq 1}\frac{1}{l!}\Bigg(\sum_{\lambda\models k:|\lambda|=l}\prod_{j=1}^{l}\Omega_{i_j}\Bigg), \quad k \geq 1, \tag{4.15}$$

where $\lambda \models k$ denotes a partition of the integer k considering the order of the numbers occurring in the partition. This is due to the (general) non-commutativity of the cepstral

matrices, since $\Omega_1\Omega_2 \neq \Omega_2\Omega_1$ in general, and they must be accounted as distinct terms in the summation. Expanding the matrix logarithm, the relation of the cepstral matrices to the Wold coefficients is provided by:

$$\Omega_k = \sum_{l\geq 1} \frac{(-1)^l}{l} \left(\sum_{\lambda \models k : |\lambda| = l} \prod_{j=1}^{l} \Psi_{i_j} \right). \tag{4.16}$$

Since the number of partitions involved in the formulas above quickly grows, Holan et al. (2017) provide a simpler equivalent method to relate cepstral and Wold matrix coefficients.

Formula (4.15) allows to get the Wold coefficients characterizing the process, given the cepstral coefficients, by expanding the exponential transformation of the matrix polynomial $[\Omega]_1^q(z)$. The Wold coefficients are then used for computation of the spectral density matrix, for forecasting, or assessing goodness-of-fit of the model.
As previously cited, in the univariate case the relation between the Wold coefficients and the Fourier coefficients of the log-spectrum (the cepstral coefficients) is provided by a recursive formula involving previously computed Wold coefficients and a finite set of cepstral coefficients, described in Pourahmadi (1983).
Recursive formulae for the coefficients of transformations of Taylor series and polynomials have occupied attention of many scholars. In particular, recursive relations for the coefficients of powers of Taylor series have been known a long time.
Many authors have expresses the relation between the coefficients of a given Taylor series and its power transform in different implicit or explicit forms. Gould (1974) derived these formulae for power transformations of Taylor series, for any real power, and put them in a logical order. His paper provides recursive formulae also for powers of Dirichlet series.

In the next section we consider in the univariate case recursive formulae for transformations of polynomials, and a recursive formula is provided, which nests those by Pourahmadi (1983) and by Gould (1974).

4.3 Recursive formulae for powers of polynomials

Following Proietti et al. (2020), in the univariate case we derive an encompassing formula for the coefficients of the polynomial

$$\psi(z) = [1 + \lambda b(z)]^{\frac{1}{\lambda}}, \quad b(z) = \sum_{k=1}^{q} b_k z^k. \tag{4.17}$$

Denote $\theta(z) = 1 + \lambda b(z) = \sum_{k=0}^{q} \theta_k z^k, \theta_0 = 1, \theta_k = \lambda b_k, k \geq 1$.

Differentiating both sides of (4.17) with respect to z and denoting $\dot{\psi}(z) = \frac{d}{dz}\psi(z)$, $\dot{b}(z) = \frac{d}{dz}b(z)$, and $\dot{\theta}(z) = \frac{d}{dz}\theta(z)$,

$$\begin{aligned} \dot{\psi}(z) &= [1 + \lambda b(z)]^{\frac{1}{\lambda}-1}\dot{b}(z) \\ &= \psi(z)[1 + \lambda b(z)]^{-1}\dot{b}(z). \end{aligned}$$

Then, replacing $\dot{\theta}(z) = \lambda\dot{b}(z)$, gives the identity

$$\dot{\psi}(z)\theta(z) = \lambda^{-1}\psi(z)\dot{\theta}(z). \tag{4.18}$$

Replacing $\dot{\psi}(z) = \sum_{k=0}^{\infty}(k+1)\psi_{k+1}z^k$ and $\dot{\theta}(z) = \sum_{k=0}^{q-1}(k+1)\theta_{k+1}z^k$ and convoluting the polynomials on both sides of (4.18) yields

$$\sum_{k=0}^{\infty}\left[\sum_{j=0}^{k\wedge q}(k+1-j)\psi_{k+1-j}\theta_j\right]z^k = \sum_{k=0}^{\infty}\left[\sum_{j=1}^{k\wedge q}\frac{j}{\lambda}\psi_{k+1-j}\theta_j\right]z^k.$$

Equating the coefficients of z^k both sides

$$(K+1)\psi_{k+1} + \sum_{j=1}^{k\wedge q}(k+1-j)\theta_j\psi_{k+1-j} = \sum_{j=1}^{k\wedge q}\frac{j}{\lambda}\theta_j\psi_{k+1-j}, k \geq 1,$$

and rearranging yields

$$\psi_{k+1} = \frac{1}{k+1}\sum_{j=1}^{k\wedge q}\left[j(\lambda^{-1}+1) - (k+1)\right]\theta_j\psi_{k+1-j}, \quad \psi_0 = 1, \tag{4.19}$$

which is Gould's formula (Gould, 1974) $p = \lambda^{-1}$.

Also, replacing $\theta_j = \lambda b_j$, gives the encompassing formula

$$\psi_{k+1} = \frac{1}{k+1} \sum_{j=1}^{k \wedge q} [j(1+\lambda) - \lambda(k+1)]\, b_j \psi_{k+1-j}, \quad \psi_0 = 1, \tag{4.20}$$

which gives the formula by Pourahmadi (1983) as a limiting case, for $\lambda = 0$, providing the coefficients of $\psi(z) = \exp\{b(z)\}$ as $\psi_{k+1} = \sum_{j=1}^{k \wedge q} j b_j \psi_{k+1-j}$.

In the next section we are interested in finding the relation between the coefficients of a given matrix polynomial, and the coefficients of its power transformation, for any real power transform

4.4 Formulae for powers of matrix polynomials

Let $\Psi(z)$ be an $r \times r$ matrix polynomial of degree q, i.e., a polynomial in z with matrix coefficients:

$$\Psi(z) = \Psi_0 + \Psi_1 z + \Psi_2 z^2 + \cdots + \Psi_q z^q,$$

with

$$\Psi_0 = I_r.$$

This is equivalent to a square matrix whose elements are polynomials in z.
Let us consider:

$$[\Psi(z)]^p = \Phi_p(z) = \Phi_{p,0} + \Phi_{p,1} z + \Phi_{p,2} z^2 + \cdots + \Phi_{p,pq} z^{pq}, \tag{4.21}$$

where $\Phi_{p,0} = I_r$ and $p \in \Re$. The Taylor series expansion for the matrix polynomial $\Phi_p(z)$ is (Gohberg, Lancaster, and Rodman, 2005):

$$\begin{aligned} \Phi_p(z) = \Phi_p(z_0) + \Phi_p^{(1)}(z_0)(z - z_0) + \frac{1}{2!}\Phi_p^{(2)}(z_0)(z - z_0)^2 + \\ \cdots + \frac{1}{l!}\Phi_p^{(l)}(z_0)(z - z_0)^l + \cdots + \frac{1}{pq!}\Phi_p^{(pq)}(z_0)(z - z_0)^{pq}, \end{aligned} \tag{4.22}$$

where $\Phi_p^{(l)}(z)$ is the lth derivative of $\Phi_p(z)$.
Interest resides in finding a recursive relation between the coefficients of $\Phi_p(z)$ and the

coefficients of $\Psi(z)$. Note that $\Phi_1(z) = \Psi(z)$.

Let us start by first considering the case where $p = k$, a positive integer. The Taylor series expansion of $\Phi_k(z) = [\Psi(z)]^k$ about $z_0 = 0$ is:

$$\Phi_k(z) = [\Psi(z)]^k = \Phi_k(0) + \Phi_k^{(1)}(0)z + \frac{1}{2!}\Phi_k^{(2)}(0)z^2 + \cdots + \frac{1}{l!}\Phi_k^{(l)}(0)z^l + \cdots + \frac{1}{pq!}\Phi_k^{(kq)}(0)z^{kq} \tag{4.23}$$

where $\Phi_k^{(l)}(0) = \frac{D^l[\Psi(z)]^k}{Dz^l}|_{z=0}$ is the lth derivative of $[\Psi(z)]^k$ evaluated at $z = 0$. Hence, the lth coefficient of the power-transformed matrix polynomial $\Phi_k(z)$ is given by $\Phi_{k,l} = \frac{1}{l!}\Phi_k^{(l)}(0)$.
The first derivative of $[\Psi(z)]^k$ (for k a positive integer) is given by:

$$\frac{D[\Psi(z)]^k}{Dz} = \sum_{j=1}^{k} \Psi^{j-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{k-j}(z). \tag{4.24}$$

For $A(z)$ and $B(z)$ two $r \times r$ matrix polynomials, by the *product rule* for matrix functions:

$$\frac{D[A(z)B(z)]}{Dz} = \frac{DA(z)}{Dz}B(z) + A(z)\frac{DB(z)}{Dz}. \tag{4.25}$$

Iterated use of (4.24) and (4.25) allows to write the derivatives of $[\Psi(z)]^k$.

$$\frac{D^2[\Psi(z)]^k}{Dz^2} = \frac{D}{Dz}\Big[\frac{D\Psi^k(z)}{Dz}\Big] = \frac{D}{Dz}\Big[\sum_{j=1}^{k}\Psi^{j-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{k-j}(z)\Big].$$

By writing the derivative of a sum of functions as the sum of the derivatives of the functions, and using (4.24) and (4.25) we find that:

$$\frac{D^2[\Psi(z)]^k}{Dz^2} = \sum_{j=1}^{k}\Bigg\{\sum_{i=1}^{j-1}\Psi^{i-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{j-1-i}(z)\frac{D\Psi(z)}{Dz}\Psi^{k-j}(z) + \Psi^{j-1}(z)\frac{D^2\Psi(z)}{Dz^2}\Psi^{k-j}(z) + \\ + \Psi^{j-1}(z)\frac{D\Psi(z)}{Dz}\sum_{h=1}^{k-j}\Psi^{h-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{k-j-h}(z)\Bigg\}$$

Analogously the third derivative of $[\Psi(z)]^k$ is:

$$\frac{D^3[\Psi(z)]^k}{Dz^3} = \frac{D}{Dz}\Big[\frac{D^2\Psi^k(z)}{Dz^2}\Big] =$$
$$= \sum_{j=1}^{k}\Bigg\{\sum_{i=1}^{j-1}\Bigg[\sum_{u=1}^{i-1}\Psi^{u-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{i-1-u}(z)\frac{D\Psi(z)}{Dz}\Psi^{j-1-i}(z)\Bigg]\frac{\Psi(z)}{Dz}\Psi^{k-j}(z) +$$
$$+\sum_{i=1}^{j-1}\Bigg[\Psi^{i-1}(z)\frac{D^2\Psi(z)}{Dz^2}\Psi^{j-1-i}(z)\Bigg]\frac{D\Psi(z)}{Dz}\Psi^{k-j}(z) +$$
$$+\sum_{i=1}^{j-1}\Bigg[\Psi^{i-1}(z)\frac{D\Psi(z)}{Dz}\sum_{v=1}^{j-1-i}\Psi^{v-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{j-1-i-v}(z)\Bigg]\frac{D\Psi(z)}{Dz}\Psi^{k-j}(z) +$$
$$+\sum_{i=1}^{j-1}\Bigg[\Psi^{i-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{j-1-i}(z)\Bigg]\frac{D^2\Psi(z)}{Dz^2}\Psi^{k-j}(z) +$$
$$+\sum_{i=1}^{j-1}\Bigg[\Psi^{i-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{j-1-i}(z)\Bigg]\frac{D\Psi(z)}{Dz}\sum_{w=1}^{k-j}\Psi^{w-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{k-j-w}(z) +$$
$$+\sum_{p=1}^{j-1}\Bigg[\Psi^{p-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{j-1-p}(z)\Bigg]\frac{D^2\Psi(z)}{Dz^2}\Psi^{k-j}(z) + \Psi^{j-1}(z)\frac{D^3\Psi(z)}{Dz^3}\Psi^{k-j}(z) +$$
$$+\Psi^{j-1}(z)\frac{D^2\Psi(z)}{Dz^2}\sum_{m=1}^{k-j}\Psi^{m-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{k-j-m}(z) +$$
$$+\sum_{n=1}^{j-1}\Bigg[\Psi^{n-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{j-1-n}(z)\Bigg]\frac{D\Psi(z)}{Dz}\sum_{h=1}^{k-j}\Bigg[\Psi^{h-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{k-j-h}(z)\Bigg] +$$
$$+\Psi^{j-1}(z)\frac{D^2\Psi(z)}{Dz^2}\sum_{h=1}^{k-j}\Bigg[\Psi^{h-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{k-j-h}(z)\Bigg] +$$
$$+\Psi^{j-1}(z)\frac{D\Psi(z)}{Dz}\sum_{h=1}^{k-j}\Bigg[\sum_{o=1}^{h-1}\Psi^{o-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{h-1-o}(z)\frac{D\Psi(z)}{Dz}\Psi^{k-j-h}(z)\Bigg] +$$
$$+\Psi^{j-1}(z)\frac{D\Psi(z)}{Dz}\sum_{h=1}^{k-j}\Bigg[\Psi^{h-1}(z)\frac{D^2\Psi(z)}{Dz^2}\Psi^{k-j-h}(z)\Bigg] +$$
$$+\Psi^{j-1}(z)\frac{D\Psi(z)}{Dz}\sum_{h=1}^{k-j}\Bigg[\Psi^{h-1}(z)\frac{D\Psi(z)}{Dz}\sum_{g=1}^{k-j-h}\Psi^{g-1}(z)\frac{D\Psi(z)}{Dz}\Psi^{k-j-h-g}(z)\Bigg]\Bigg\}$$

The same proceeding can be applied to higher order derivatives.
To write the lth coefficient of $\Phi_k(z) = [\Psi(z)]^k$ note that:

$$\begin{aligned}\Psi(0) = I_r, [\Psi(0)]^k &= I_r \\ \Psi^{(1)}(0) &= \Psi_1 \\ \Psi^{(2)}(0) &= 2\Psi_2 \\ \Psi^{(3)}(0) &= 6\Psi_3 \\ &\dots\end{aligned}$$

Hence, by the previous results we get the coefficients of $\Phi_k(z)$ as $\Phi_{k,l} = \frac{1}{l!}\Phi_k^{(l)}(0)$:

$$\begin{aligned}\Phi_{k,0} &= I_r \\ \Phi_{k,1} &= k\Psi_1 \\ \Phi_{k,2} &= \frac{1}{2}k(k-1)(\Psi_1)^2 + k\Psi_2 \\ \Phi_{k,3} &= \frac{1}{6}k(k-1)(k-2)(\Psi_1)^3 + \frac{1}{3}\sum_{j=1}^{k}(2k-1-j)\Psi_2\Psi_1 + \frac{1}{3}\sum_{j=1}^{k}(k-2+j)\Psi_1\Psi_2 + k\Psi_3 \\ &\dots\end{aligned}$$

A recursive relation among the coefficients $\Phi_{k,l}$ is apparent, and it is given by

$$\Phi_{k,l} = \sum_{i=0}^{l}\Phi_{k-1,i}\Psi_{l-i}, \tag{4.26}$$

which coincides with the result by Karampetakis and Tzekis (2005).

Let us consider now $[\Psi(z)]^p = \Phi_p(z)$ with p a real number. In this case we can make use of the results in Holan et al. (2017) about the exponential transform and the log transform of a matrix polynomial.
For any complex-valued square matrix A, convergence of the matrix exponential $\exp\{A\}$ is guaranteed by Proposition 8.3 of Artin (1991). Let $\Psi(z) = \exp\{\Theta(z)\}$, with $\Theta(z)$ a finite degree matrix polynomial, and assume:

$$\| I_r - \Psi(z) \| < 1 \, \forall \, z \in D \tag{4.27}$$

for some matrix norm $\|\,\|$ and $D = \{z \in \mathbb{C} : |z| \leq 1\}$ (Holan et al., 2017). (4.27) guarantees invertibility of $\exp\{\Theta(z)\}$.

Let us get back to the power transformed matrix polynomial:

$$\Phi_p(z) = [\Psi(z)]^p = \Phi_{p,0} + \Phi_{p,1}z + \Phi_{p,2}z^2 + \Phi_{p,3}z^3 + \dots.$$

If we take the logarithm in the previous equation, by the properties of the log transform of matrix powers we have:

$$\ln\{\Phi_p(z)\} = \ln\{[\Psi(z)]^p\} = p\ln\{\Psi(z)\} = p\Theta(z). \tag{4.28}$$

At this stage we can derive the coefficients of $\Theta(z)$ based on the coefficients in $\Psi(z)$ using the recursion (4.16) in Holan et al. (2017):

$$\Theta_l = \sum_{k\geq 1} \frac{(-1)^k}{k} \left(\sum_{\lambda \models l : |\lambda| = k} \left(\prod_{j=1}^{k} \Psi_{i_j} \right) \right) \tag{4.29}$$

Let us set $p\Theta(z) = \Theta_p(z)$. Each coefficient of $\Theta_p(z)$ is obtained from the coefficients of $\Theta(z)$ by multiplying each element by p. Then:

$$\ln\{\Phi_p(z)\} = \Theta_p(z)$$

and

$$\Phi_p(z) = \exp\{\Theta_p(z)\}.$$

Hence, the coefficients of the matrix polynomial $\Phi_p(z)$ can be obtained by using the result (4.15) in Holan et al.(2017, pag. 29):

$$\Phi_{p,l} = \sum_{k\geq 1} \frac{1}{k!} \left(\sum_{\lambda \models l : |\lambda| = k} \left(\prod_{j=1}^{k} \Theta_{p,i_j} \right) \right) \tag{4.30}$$

Hence we are able to find the coefficients of the power transformed polynomial $\Phi_p(z)$ starting from the coefficients of $\Psi(z)$. The reverse proceeding is also allowed.

This two-step-procedure described above to obtain the coefficients of a power transformation of a matrix polynomial can be used for time series analysis purposes, in the same spirit as univariate formulae are used as discussed in the previous chapters and sections.

4.5 The multivariate generalised cepstral model

The univariate generalised cepstral models, defined by Proietti and Luati (2019), include as special cases various univariate models, as shown in chapter 2, as the univariate exponential model, or autoregressive and moving-average models. The generalised linear cepstral models for the spectrum of a scalar time series can be extended to the multivariate case to model the Box-Cox transform of the matrix spectrum of vector time series (Cavicchioli et al., 2020).

Let $\boldsymbol{X}_t = (X_{1,t}, X_{2,t}, \ldots, X_{r,t})'$ be a $r-$dimensional stationary vector process, with zero mean. As previously noted, the spectral density matrix is Hermitian, and it can be expressed as $\boldsymbol{f} = PAP^*$, where A is a diagonal matrix with positive diagonal elements, and $P^* = P^{-1}$. Then, the λth power, with $\lambda \in \mathbb{R}$, of the spectral density matrix is $\boldsymbol{f}^\lambda = PA^\lambda P^*$, and

$$\boldsymbol{f} = \left(PA^\lambda P^*\right)^{\frac{1}{\lambda}}.$$

The matrix function $\boldsymbol{f}^\lambda$ can be expanded as:

$$PA^\lambda P^* = \sum_{k=-\infty}^{\infty} \boldsymbol{\Gamma}_{\lambda k} e^{-i\omega k}, \tag{4.31}$$

where the matrix coefficients $\boldsymbol{\Gamma}_{\lambda k}$ can be obtained as the inverse Fourier transform of $PA^\lambda P^*$:

$$\boldsymbol{\Gamma}_{\lambda k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} PA^\lambda P^* e^{i\omega k}. \tag{4.32}$$

Equation above (4.32) provides a multivariate extension of the generalised autocovariance function for scalar time series, and are called generalised autocovariance matrices (Cavicchioli et al., 2020).

Let us introduce the Box-Cox transform $\boldsymbol{g}$ of the spectral density matrix $\boldsymbol{f}$:

$$\boldsymbol{g}(\omega) = \begin{cases} \frac{[\boldsymbol{f}(\omega)]^\lambda - I_r}{\lambda}, & \lambda \neq 0, \\ \ln\left[\boldsymbol{f}(\omega)\right], & \lambda = 0, \end{cases} \tag{4.33}$$

Cavicchioli et al. (2020) showed that:

$$\lim_{\lambda \to 0} \frac{[\boldsymbol{f}(\omega)]^{\lambda} - I_r}{\lambda} = \ln [\boldsymbol{f}(\omega)]$$

the matrix function $\boldsymbol{g}$ can be represented as:

$$\boldsymbol{g}(\omega) = \sum_{k=-\infty}^{\infty} \boldsymbol{C}_{\lambda k} e^{-i\omega k}. \tag{4.34}$$

In turn, the matrix coefficients $\boldsymbol{C}_{\lambda k}$ are the inverse Fourier transform of $\boldsymbol{g}(\omega)$, and are the generalised cepstral matrices:

$$\boldsymbol{C}_{\lambda k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \boldsymbol{g}(\omega) e^{i\omega k}. \tag{4.35}$$

The generalised matrix cepstral model of order K, with transformation parameter $\lambda \in \mathbb{R}$ arises by truncating the series in (4.34) and assuming that $\boldsymbol{g}$ can be represented by a finite Fourier polynomial:

$$\boldsymbol{g}(\omega) = \sum_{k=-K}^{K} \boldsymbol{C}_{\lambda k} e^{-i\omega k}. \tag{4.36}$$

Hence, the vector exponential model by Holan et al. (2017) is obtained as a special case when $\lambda = 0$. Also, for $\lambda = 1$ the order K generalised matrix cepstral model coincides with the vector moving-average model of order K, while for $\lambda = -1$ it gives the VAR(K) model.

Cavicchioli et al. (2020) noted that $\boldsymbol{C}_{\lambda,-k} = \boldsymbol{C}'_{\lambda,k}$, for every integer $k < 0$, and for $\lambda = k = 0$ the Kolmogorov-Szego formula for the prediction error variance, Σ, gives:

$$\exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln [\boldsymbol{f}(\omega)] d\omega\right) = \exp(\boldsymbol{C}_{00}) = \Sigma$$

For $\lambda \neq 0$ the generalised cepstral matrices are related to the generalised autocovariance matrix:

$$\boldsymbol{C}_{\lambda 0} = \frac{1}{\lambda}(\boldsymbol{\Gamma}_{\lambda 0} - I_r), \quad \boldsymbol{C}_{\lambda k} = \frac{1}{\lambda}\boldsymbol{\Gamma}_{\lambda k} \tag{4.37}$$

For $\lambda = -1$, $\boldsymbol{\Gamma}_{-1,k}$ provides the multivariate analog of the inverse autocovariance function, $\boldsymbol{\Gamma}_{0,k} = \boldsymbol{0}_{r \times r}$, and $\boldsymbol{\Gamma}_{0,0} = I_r$ up to a constant, while $I_r + \boldsymbol{C}_{1,0} = \boldsymbol{\Gamma}_{1,0}$ is the unconditional variance matrix of the process, and $(I_r - \boldsymbol{C}_{1,0})^{-1}$ is the interpolation error variance.

Let us consider the Wold representation of the vector process:

$$\boldsymbol{X}_t = \Psi(L)\boldsymbol{\epsilon}_t,$$

where $\{\boldsymbol{\epsilon}_t\}$ is a vector White Noise process with mean zero and covariance matrix Σ. The spectral density matrix is $\boldsymbol{f}(\omega) = \Psi(e^{-i\omega})\Sigma\Psi'(e^{i\omega})$. Then,

$$\boldsymbol{f}(\omega)^{\lambda} = \Phi_{\lambda}(e^{-i\omega})\Sigma\Phi_{\lambda}'(e^{i\omega}), \tag{4.38}$$

where

$$\Phi_{\lambda}(z) = \sum_{k=1}^{K} \Phi_{\lambda,k} z^k, \quad \Phi_{\lambda,0} = I_r.$$

Then, when $\lambda \neq 0$ the generalised cepstral matrices are:

$$\boldsymbol{C}_{\lambda 0} = \frac{1}{\lambda}\Big(\sum_{k=0}^{K} \Phi_{\lambda,k}\Sigma_{\lambda}\Phi_{\lambda,k}' - I_r\Big) \tag{4.39}$$

$$\boldsymbol{C}_{\lambda k} = \frac{1}{\lambda}\sum_{j=k}^{K} \Phi_{\lambda,j}\Sigma_{\lambda}\Phi_{\lambda,j-k}' \tag{4.40}$$

where we assume that $det\Phi_{\lambda}(z) \neq 0$ for $|z| \leq 1$. For $\lambda = 0$ the Wold coefficients can be obtained from the cepstral matrix coefficients $\boldsymbol{C}_{0k} = \Omega_k$, as shown by Holan et al. (2017), by formula (4.15) in section 4.2. As pointed out in section 4.2, Pourahmadi's recursive formula for the log-transform of scalar polynomials does not hold in the multivariate setting, since the algebra relating the Wold coefficients to the cepstral coefficients is no longer Abelian (Holan et al., 2017).

For $\lambda \neq 0$ we use the procedure described in the previous section, which relates the Wold coefficients to the coefficients of $\Phi_{\lambda}(z)$, instead of the formula provided at pag. 614 in (Cavicchioli et al., 2020).

First, note that $\Psi(z) = [\Phi_{\lambda}(z)]^{\frac{1}{\lambda}}, \lambda \in \mathbb{R}$, and $\ln[\Psi(z)] = \frac{1}{\lambda}\ln[\Phi_{\lambda}(z)] = \frac{1}{\lambda}\Theta(z)$, where we call $\ln[\Phi_{\lambda}(z)] = \Theta(z)$. Following the results by Holan et al. (2017), the coefficients of $\Theta(z)$ are obtained by:

$$\Theta_l = \sum_{k\geq 1} \frac{(-1)^k}{k}\Bigg(\sum_{\lambda \models l : |\lambda| = k}\Big(\prod_{j=1}^{k} \Phi_{\lambda,i_j}\Big)\Bigg). \tag{4.41}$$

Setting $\Theta_\lambda = \frac{1}{\lambda}\Theta(z)$, obtained by multiplying each element of $\Theta(z)$ by $\frac{1}{\lambda}$, the coefficients of the Wold polynomial $\Psi(z) = \exp\Theta_\lambda(z)$, are given by:

$$\Psi_l = \sum_{k\geq 1}\frac{1}{k!}\left(\sum_{\lambda\models l:|\lambda|=k}\left(\prod_{j=1}^{k}\Theta_{\lambda,i_j}\right)\right) \tag{4.42}$$

Hence, all the relevant information for prediction is available from the$K+1$ cepstral matrix coefficients.
Estimation of the matrix generalised cepstral models is addressed in Cavicchioli et al. (2020) by maximization of the Whittle likelihood.

4.6 The matrix variance profile

The notion of Variance Profile introduced by Luati et al. (2012) can be extended to the multivariate setting by the definition of an auxiliary vector process, derived from a power transformation of the original process.

Let $\boldsymbol{X}_t = \Psi(L)\boldsymbol{\epsilon}_t$ be a causal $r-$dimensional vector linear process, where $\boldsymbol{\epsilon}_t \sim i.i.d.N(\mathbf{0},\Sigma)$,
$\Psi(L) = I_r + \Psi_1 L + \Psi_2 L^2 + \ldots$, and L is the lag operator, $L^j\boldsymbol{X}_t = \boldsymbol{X}_{t-j}$. We assume that $|\Psi(z)| = 0 \iff |z| > 1$, that $\int_{-\pi}^{\pi}\log \boldsymbol{f}(\omega)\,d\omega > -\infty$, and that the powers $[\boldsymbol{f}(\omega)]^p$ exist. Define the power-transformed process:

$$\boldsymbol{u}_{pt} = \begin{cases}[\Psi(L)]^p\Sigma^{\frac{(p-1)}{2}}\boldsymbol{\epsilon}_t, & p\geq 0,\\ [\Psi(L^{-1})]^p\Sigma^{\frac{(p-1)}{2}}\boldsymbol{\epsilon}_t, & p<0,\end{cases}$$

which is equivalent to

$$\boldsymbol{u}_{pt} = \begin{cases}[\Psi(L)]^p\boldsymbol{\eta}_t, & p\geq 0,\\ [\Psi(L^{-1})]^p\boldsymbol{\eta}_t, & p<0,\end{cases}$$

with $\boldsymbol{\eta}_t \sim i.i.d.N(\mathbf{0},\Sigma^p)$. Note that the pth power of the matrix polynomial $\Psi(L)$ is still a matrix polynomial: $[\Psi(L)]^p = \Phi_p(L) = \sum_{j=0}^{\infty}\Phi_{p,j}L^j$, whose coefficients can be

obtained recursively as illustrated in section 4.4. The spectral density matrix of $\boldsymbol{u}_{pt}$ is

$$\begin{aligned}\boldsymbol{f}_u(\omega) = \Phi_p(e^{-i\omega})\,\Sigma^{\frac{(p-1)}{2}}\,\Sigma\,\Sigma^{\frac{(p-1)}{2}}\,\Phi_p'(e^{-i\omega}) = \\ = \Phi_p(e^{-i\omega})\,\Sigma^p\,\Phi_p'(e^{-i\omega}) = \\ = [\Psi(e^{-i\omega})]^p\,\Sigma^p\,[\Psi'(e^{-i\omega})]^p = [\boldsymbol{f}(\omega)]^p \end{aligned} \tag{4.43}$$

and its autocovariance matrix is

$$\boldsymbol{\Gamma}_{uk} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \boldsymbol{f}_u(\omega) e^{i\omega k}\, d\omega$$

Then, the matrix variance profile is defined as the power mean of the spectral density matrix of the process $\boldsymbol{X}$:

$$\begin{aligned}\boldsymbol{v}_p = \left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}[\boldsymbol{f}(\omega)]^p\, d\omega\right\}^{\frac{1}{p}} = \\ = \left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\boldsymbol{f}_u(\omega)\, d\omega\right\}^{\frac{1}{p}} = \left\{\boldsymbol{\Gamma}_{u0}\right\}^{\frac{1}{p}}\end{aligned} \tag{4.44}$$

The definition of the multivariate variance profile in terms of the auxiliary process turns out to be useful for interpretation and computational purposes.
The variance profile for specific values of p provides some important measures of variance for vector processes.

- $p = 1$:
  the auxiliary process is $\boldsymbol{u}_{pt} = \Psi(L)\boldsymbol{\epsilon}_t = \boldsymbol{X}_t$, with $\boldsymbol{\epsilon}_t \sim i.i.d.N(\boldsymbol{0}, \Sigma)$ and

  $$\boldsymbol{v}_1 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\boldsymbol{f}(\omega)\, d\omega = \boldsymbol{\Gamma}_0$$

  is the arithmetic mean of the spectral density matrix of the original process, and it gives the unconditional variance-covariance matrix $\boldsymbol{\Gamma}_0$ of the vector process $\boldsymbol{X}$.

- $p \to 0$:

  $$\lim_{p\to 0}\boldsymbol{v}_p = \lim_{p\to 0}\left\{\Gamma_{u0}\right\}^{\frac{1}{p}} = \lim_{p\to 0}\left\{\Sigma^p\right\}^{\frac{1}{p}} = \Sigma.$$

As $p \to 0, \boldsymbol{v}_p$ gives the variance-covariance matrix of the innovation process $\boldsymbol{\epsilon}_t$, and, indeed, the power process $\boldsymbol{u}_{pt} \to \boldsymbol{\eta}_t = \Sigma^{\frac{(p-1)}{2}} \boldsymbol{\epsilon}_t$ as $p \to 0$.
Hence,

$$\lim_{p\to 0} \boldsymbol{v}_p = \lim_{p\to 0} \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} [\boldsymbol{f}(\omega)]^p \, d\omega \right\}^{\frac{1}{p}} = \Sigma \tag{4.45}$$

Σ represents the variance-covariance matrix of the one-step-ahead prediction error, and its estimation provides useful measures of the the linear feedback and dependence between two multiple time series (Geweke, 1982) and is needed for computing the multivariate analog of the AIC and BIC (Lütkepohl, 2005).

- $p = -1$:
  in this case the power process $\boldsymbol{u}_{-1t}$ coincides with the inverse process considered in the definition of the inverse autocovariance matrix function in Heyse and Wei (1985), i.e. $\boldsymbol{u}_{-1t} = [\Psi(L^{-1})]^{-1}\boldsymbol{\eta}_t$, with $\boldsymbol{\eta}_t \sim i.i.d.N(\mathbf{0}, \Sigma^{-1})$. The spectral density matrix of $\boldsymbol{u}_{-1t}$ is $\boldsymbol{f}_u(\omega) = [\boldsymbol{f}(\omega)]^{-1}$, and its autocovariance matrix is:

$$\Gamma_{uk} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega k} [\boldsymbol{f}(\omega)]^{-1} \, d\omega = \Gamma_{ik},$$

  where Γ_{ik}, is the transpose of the inverse autocovariance matrix of $\boldsymbol{X}_t$, following the definition provided by Heyse and Wei (1985). Similar to the univariate case, the inverse autocorrelation matrix function provides a useful tool for model identification, as in specific cases it admits a dual relation with the autocorrelation matrix function. In particular, for vector autoregressive models the inverse autocorrelation matrix function shares the same cut-off property as its univariate analog: for autoregressive models of order r the inverse autocorrelation function for both univariate and multivariate time series are equal to zero for all lags greater than r. Heyse and Wei (1985) showed that for an ARMA(r, q) process $\boldsymbol{Z}_t, \Phi(B)\boldsymbol{Z}_t = \theta(B)\boldsymbol{e}_t$, where $\Phi(B)$ is the AR polynomial of order r in the backshift operator and $\theta(B)$ is the MA polynomial of order q, and $\boldsymbol{e}_t$ a White Noise process with zero mean and variance-covariance matrix Σ, the transpose of the inverse autocorrelation matrix function of $\boldsymbol{Z}_t$ coincides with the autocorrelation matrix function of the inverse process, defined as $\boldsymbol{U}_t = \Phi'(B)[\theta'(B)]^{-1}\boldsymbol{a}_t$, where $\boldsymbol{a}_t$ is a White Noise process with zero

mean and variance-covariance matrix Σ^{-1}. Hence, in the multivariate setting, duality between the inverse autocorrelation and the autocorrelation matrix functions holds when $\Phi'(B)[\theta'(B)]^{-1} = [\theta'(B)]^{-1}\Phi'(B)$. This means that, when the process is vector moving-average of order q the inverse autocorrelation matrix function will not cut-off at any lag. While, if the process is autoregressive of order r the inverse autocorrelation matrix function will equal zero for all lags greater than r.
For $p = -1$ the matrix variance profile is:

$$\boldsymbol{v}_{-1} = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} [\boldsymbol{f}(\omega)]^{-1} \, d\omega \right\}^{-1} = \left\{ \Gamma_{i0} \right\}^{-1},$$

which provides the interpolation error variance matrix. As in Pourahmadi (1993) (Theorem 3.1(b)), let $\tilde{\boldsymbol{X}}_0$ denote the linear interpolator of $\boldsymbol{X}_0$ based on $\{\boldsymbol{X}_t, t \neq 0\}$. Then, the variance-covariance matrix $\tilde{\Sigma}$ of the error of interpolation is given by:

$$\tilde{\Sigma} = COV(\boldsymbol{X}_0 - \tilde{\boldsymbol{X}}_0) = D_0^{-1}$$

where

$$D_0 = \frac{1}{2\pi} \int_0^{2\pi} [\boldsymbol{f}(\omega)]^{-1} d\omega.$$

This implies that $\boldsymbol{v}_{-1} = D_0^{-1} = \tilde{\Sigma}$ provides the variance matrix of the interpolation error.

Estimation of the matrix variance profile can be addressed by using smoothing techniques for the periodogram matrix, selecting a suitable kernel or spectral window.

4.7 Estimation of the Matrix Variance Profile

The matrix variance profile (4.44) can be estimated by a power transformation of the estimated spectral density matrix. A possible choice is to estimate the spectral density matrix nonparametrically by smoothing the periodogram matrix, to obtain a consistent estimate. A general form for the smoothed periodogram estimator of the spectrum matrix is:

$$\hat{\boldsymbol{f}}(\omega) = \int_{-\pi}^{\pi} \boldsymbol{I}(v) W_N(\omega - v) \, dv, \tag{4.46}$$

where $W()$ is a spectral window. A discrete approximation of the expression above is given by

$$\hat{\boldsymbol{f}}(\omega) = \frac{2\pi}{N} \sum_{j=-[N/2]}^{[N/2]} \boldsymbol{I}(\omega_j) W_N(\omega - \omega_j). \tag{4.47}$$

In particular, for a Daniell (i.e. rectangular) window with bandwidth $4\pi m/N$, (4.47) reduces to

$$\hat{\boldsymbol{f}}(\omega) = \frac{1}{(2m+1)} \sum_{j=-m}^{m} \boldsymbol{I}\left[\frac{2\pi\{v(\omega)+j\}}{N}\right], \tag{4.48}$$

where $v(\omega)$ is the largest integer such that $2\pi v(\omega)/N$ is closest to ω (Priestley, 1981). In (4.48) each diagonal element of the spectrum matrix is estimated by averaging $2m+1$ periodogram ordinates. If the spectrum is assumed to be roughly constant over the bandwidth of $W_N()$, then the distribution of $(2m+1)\hat{\boldsymbol{f}}(\omega)$ for $\omega \neq 0, \pi$ may be approximated by the complex Wishart distribution with parameters $2m+1, \boldsymbol{f}(\omega)$.
Further investigation suggests the following form for the estimator (4.48) (Brillinger, 2002):

$$\hat{\boldsymbol{f}}(\omega) = \frac{1}{m} \sum_{j=1}^{m} Re\boldsymbol{I}\left[\omega + \frac{2\pi j}{N}\right], \tag{4.49}$$

if $\omega = 0, \pm 2\pi, \ldots$, or if $\omega = \pm\pi, \pm 3\pi, \ldots$, and N is even, and the form:

$$\hat{\boldsymbol{f}}(\omega) = \frac{1}{m} \sum_{j=1}^{m} Re\boldsymbol{I}\left[\omega - \frac{\pi}{N} + \frac{2\pi j}{N}\right], \tag{4.50}$$

if $\omega = \pm\pi, \pm 3\pi, \ldots$, and N is odd.
Assuming that the autocovariance function of the process, $\boldsymbol{\Gamma}_k$, satisfies

$$\sum_{k=-\infty}^{\infty} |\boldsymbol{\Gamma}_k| < \infty,$$

and that the spectrum is nearly constant over the bandwidth of $W_N()$, then, each element of $\hat{\boldsymbol{f}}(\omega)$ is an asymptotically unbiased estimate of the corresponding element of $\boldsymbol{f}(\omega)$.

Estimation of the matrix variance profile is based on the Daniell-type estimator (4.48) of the spectrum matrix:

$$\hat{\boldsymbol{v}}_p = m\left\{ \frac{1}{M} \sum_{k=0}^{M-1} \left\{ \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{I}\left[\frac{2\pi\{km+j\}}{N} \right] \right\}^p BC \right\}^{\frac{1}{p}}, \tag{4.51}$$

where $M = \frac{N-1}{2m}$, andBC is a bias correction factor depending on the properties of powers of the complex Wishart distribution. When $p = 1$, the estimator (4.51) cioncides with (4.48), where the averages of $2m + 1$ periodogram ordinates are replaced by averages of periodogram ordinates over m consecutive non-overlapping frequencies. For the case $p = 1$ the corresponding estimator (4.51) is asymptotically unbiased, hence the bias correction factor equals unity, $BC = 1$. For $p \neq 1$, the diagonal elements of $\hat{\boldsymbol{v}}_p$ correspond, up to a multiplicative factor, to the estimates proposed by Luati et al. (2012) for the scalar variance profile. Hence, for $p \neq 1$, BC needs to be determined by computing the first order moment of powers of the complex Wishart distribution.

Chapter 5

Conclusions

This thesis has analysed asymptotic efficiency of a nonparametric estimator of the generalised autocovariance function, defined by Proietti and Luati (2015) as the inverse Fourier transform of the pth power of the spectral density function. The GACV includes as a special case the traditional autocovariance function, which is often estimated by the sample autocovariance, whose asymptotic properties have been demonstrated and discussed in the literature. Investigation of asymptotic efficiency of the sample autocovariance showed that it is asymptotically efficient only for some specific processes and in a limited number of cases. Porat (1987) studied asymptotic efficiency of the sample autocovariance for Gaussian autoregressive (AR) moving average (MA) mixed processes, based on state-space representations and matrix Lyapunov equation theory. These results showed that for Gaussian ARMA(r, q) processes with $r \geq q$ the lag k sample autocovariances are asymptotically efficient for $0 \leq k \leq r - q$, while if $q > r$ none of the sample autocovariances is asymptotically efficient. Kakizawa and Taniguchi (1994) derived in the frequency domain a necessary and sufficient condition for asymptotic efficiency of the sample autocovariances that applies to the more general class of Gaussian stationary processes.

This thesis has established a necessary and sufficient condition in the frequency domain for asymptotic efficiency of the nonparametric estimator of the GACV proposed by Proietti and Luati (2015). The result generalises the condition for asymptotic efficiency of the sample autocovariances provided by Kakizawa and Taniguchi (1994), and, hence,

embodies in a single equation the condition for asymptotic efficiency of the sample autocovariance ($p = 1$), of the estimator of the inverse autocovariances ($p = -1$), and of the estimators of the generalised autocovariances for general powers p. The results showed that asymptotic efficiency of the estimator can be achieved when $p = 1$, or when $p \neq 1$ and the pooling parameter m, which characterizes the estimator, is sufficiently large. The derived condition for asymptotic efficiency is expressed in terms of the spectral density function, which makes it easy to check for various models. Analytical results derived in this thesis showed that for Gaussian ARMA(r, q) processes, asymptotic efficiency of the nonparametric estimator of the GACV for positive and negative integer powers p depends on the existence of solutions to a trigonometric polynomial equation. Results show that the estimator of the GACV with $p = -1$ and large m is asymptotically efficient for $r = 0$ and $0 \leq k \leq q$. This result implies that the nonparametric estimator of the GACV estimates efficiently the first q inverse autocovariances when the true generating process is pure MA(q).

The results obtained showed that asymptotic efficiency of the nonparametric estimator of the GACV depends also on the pooling parameter m, which plays the role of a smoothing parameter. The estimator of the GACV is asymptotically unbiased, but in finite samples its bias arises as the pooled periodogram ordinates over m consecutive non-overlapping frequencies are used instead of the raw ordinates at each Fourier frequency. Increasing m reduces the variance but increases the finite-sample bias. Proietti and Luati (2015) proposed the use of the Jackknife to select the value of m. This thesis proposed the use the of multiplicative periodogram bootstrap (Meyer et al., 2018) to estimate the finite-sample distribution of the estimator of the GACV, and to select the optimal (minimum MMSE) value of the pooling parameter m. This procedure is based on asymptotic distributional results for the periodogram, and on the assumption of approximate independence of the periodogram ordinates. This procedure is motivated by the possibility of viewing the estimator of the GACV as a periodogram-based estimator of a spectral mean. Application to a real data examples showed that a small value of the pooling parameter m is preferable for estimation of the GACV.

This thesis has provided a description of fundamental concepts of spectral analysis of stationary scalar and vector time series, together with some of the most recent advances

in spectral modelling, including the vector exponential model (Holan et al., 2017), and the generalised matrix cepstral model (Cavicchioli et al., 2020). The VEXP model and the generalised matrix cepstral model are characterized by the cepstral and generalised cepstral matrices, respectively. The cepstral matrices are related to the Wold coefficients by a formula derived by Holan et al. (2017). This thesis provided an algorithm that relates the matrix coefficients of a given polynomial to the matrix coefficients of its power transform. This algorithm, based on the results by Holan et al. (2017), can be used to derive the matrix coefficients of the Wold representation of the process, from the generalised cepstral matrices. It is also used to define the matrix variance profile in terms of an auxiliary process, which allows to give it further interpretation. A nonparametric estimator of the matrix variance profile has been defined, based on the smoothed periodogram matrix using a Daniell window. Possible future development of this project could deal with the study of the distributional properties of this estimator. The study of power transformations of the complex Wishart distribution, and the derivation of the its moments allow to specify a bias correction factor for this estimator, as expected by distributional results concerning its univariate analog. Possible future work may include comparison of various smoothing techniques for the periodogram matrix, and comparison of nonparametric and parametric estimation for the matrix variance profile.

Bibliography

Artin, M. (1991). *Algebra.* Prentice-Hall: Englewood Cliffs, New Jersey.

Battaglia, F. (1983). Inverse autocovariances and a measure of linear determinism for a stationary process. *Journal of Time Series Analysis*, *4*(2), 79–87.

Battaglia, F. (1988). On the estimation of the inverse correlation function. *Journal of Time Series Analysis*, *9*(1), 1–10.

Bhansali, R. (1980). Autoregressive and window estimates of the inverse correlation function. *Biometrika*, *67*(3), 551–566.

Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika*, *60*(2), 217–226.

Bogert, B. P. (1963). The frequency alanalysis of time series for echoes; cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. *Time series analysis*, 209–243.

Boshnakov, G. N. (2005). On the asymptotic properties of multivariate sample autocovariances. *Journal of multivariate analysis*, *92*(1), 42–52.

Brillinger, D. R. (2002). *Time series: data analysis and theory.* SIAM.

Brockwell, P. J., and Davis, R. A. (1991). *Time series: theory and methods.* Springer Series in Statistics.

Brockwell, P. J., Davis, R. A., and Calder, M. V. (2002). *Introduction to time series and forecasting* (Vol. 2). Springer.

Cavicchioli, M., et al. (2020). Generalised cepstral models for the spectrum of vector time series. *Electronic Journal of Statistics*, *14*(1), 605–631.

Cleveland, W. S. (1972). The inverse autocorrelations of a time series and their applications. *Technometrics*, *14*(2), 277–293.

Davis, H. T., and Jones, R. H. (1968). Estimation of the innovation variance of a stationary time series. *Journal of the American Statistical Association*, *63*(321), 141–149.

Erdélyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F. (1954). Higher transcendental functions. *Bull. Amer. Math. Soc*, *60*, 405–408.

Fuller, W. A. (1976). Introduction to statistical time series, new york: Johnwiley. *FullerIntroduction to Statistical Time Series1976*.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American statistical association*, *77*(378), 304–313.

Gohberg, I., Lancaster, P., and Rodman, L. (2005). *Matrix polynomials*. Springer.

Gould, H. (1974). Coefficient identities for powers of taylor and dirichlet series. *American Mathematical Monthly*, 3–14.

Hannan, E., and Nicholls, D. (1977). The estimation of the prediction error variance. *Journal of the American Statistical Association*, *72*(360a), 834–840.

Hassler, U. (2018). *Time series analysis with long memory in view*. John Wiley & Sons.

Heyse, J., and Wei, W. (1985). Inverse and partial lag autocorrelation for vector time series. *American Statistical Association Proceedings of Business and Economic Statistics Section*, 233–237.

Holan, S. H., McElroy, T. S., and Wu, G. (2017). The cepstral model for multivariate time series: The vector exponential model. *Statistica Sinica*, *27*, 23–42.

Kakizawa, Y. (1999). Note on the asymptotic efficiency of sample covariances in gaussian vector stationary processes. *Journal of Time Series Analysis*, *20*(5), 551–558.

Kakizawa, Y., and Taniguchi, M. (1994). Asymptotic efficiency of the sample covariances in a gaussian stationary process. *Journal of Time Series Analysis*, *15*(3), 303–311.

Karampetakis, P., and Tzekis, P. (2005). On the computation of the minimal polynomial of a polynomial matrix. *International Journal of Applied Mathematics and Computer Science*, *15*, 339–349.

Koopmans, L. H. (1974). The spectral analysis of time series.

Luati, A., Proietti, T., and Reale, M. (2012). The variance profile. *Journal of the American Statistical Association*, *107*(498), 607–621.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science

& Business Media.

Magnus, J., and Neudecker, H. (1999). Matrix differential calculus with applications in statistics and econometrics (revised edition).

Magnus, J. R., and Neudecker, H. (1979). The commutation matrix: some properties and applications. *The Annals of Statistics*, 381–394.

McCracken, M. W., and McElroy, T. (2014). Multi-step ahead forecasting of vector time series. *Federal Reserve Bank of St. Louis Working Paper Series*(2012-060).

Meyer, M., Paparoditis, E., and Kreiss, J.-P. (2018). A frequency domain bootstrap for general stationary processes. *arXiv preprint arXiv:1806.06523*.

Porat, B. (1987). Some asymptotic properties of the sample covariances of gaussian autoregressive moving-average processes. *Journal of Time Series Analysis*, *8*(2), 205–220.

Pourahmadi, M. (1983). Exact factorization of the spectral density and its application to forecasting and time series analysis. *Communications in Statistics-Theory and Methods*, *12*(18), 2085–2094.

Pourahmadi, M. (1993). On relations between prediction error covariance of univariate and multivariate processes. *Statistics & probability letters*, *16*(5), 355–359.

Priestley, M. B. (1981). *Spectral analysis and time series: multivariate series, prediction and control* (Vol. 2). Academic Press.

Proietti, T., and Luati, A. (2015). The generalised autocovariance function. *Journal of Econometrics*, *186*(1), 245–257.

Proietti, T., and Luati, A. (2019). Generalised linear cepstral models for the spectrum of a time series. *Statistica Sinica*, *29*(3), 1561–1583.

Proietti, T., Luati, A., and Papagni, F. (2020). Recursive formulae for the coefficients of powers of matrix polynomials.

Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society. Series B (Methodological)*, *11*(1), 68–84.

Taniguchi, M. (1980). On estimation of the integrals of certain functions of spectral density. *Journal of Applied Probability*, 73–83.

Taniguchi, M. (1981). An estimation procedure of parameters of a certain spectral density model. *Journal of the Royal Statistical Society: Series B (Methodological)*,

43(1), 34–40.

Taniguchi, M. (1986). Third order asymptotic properties of maximum likelihood estimators for gaussian arma processes. *Journal of multivariate analysis*, *18*(1), 1–31.

Taniguchi, M. (1987). Minimum contrast estimation for spectral densities of stationary processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, *49*(3), 315–325.

Walker, A. (1995). On results of porat concerning asymptotic efficiency of sample covariances of gaussian arma processes. *Journal of Time Series Analysis*, *16*(2), 237–248.

Wei, W. W. (2019). *Multivariate time series analysis and applications*. John Wiley & Sons.

Acknowledgement

I would like to thank Professor Alessandra Luati, my supervisor, for her support and guidance during these years, and for her sympathy. This work would have not been completed without her unwavering support. I would also like to thank Professor Tommaso Proietti for his illuminating advices. All the Professors I met during this path deserve my thanks.

I would like to thank Raffaele for his support, for always believing in me. Lastly, I would not have completed this project without the love of my family. Heartfelt thanks to my parents and my sister for their love and for always supporting me, for the endless calls and talks to always support me and for believing in me. I could not have done it without you!