STUDIES IN MULTIDIMENSIONAL STOCHASTIC PROCESSES: MULTIVARIATE LONG-RANGE DEPENDENCE AND SYNTHESIS OF GAUSSIAN RANDOM FIELDS

Stefanos Kechagias

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research (Statistics).

Chapel Hill
2015

Approved by:

Vladas Pipiras

Shankar Bhamidi

Amarjit Budhiraja

Chuanshu Ji

M. Ross Leadbetter

ABSTRACT

STEFANOS KECHAGIAS: Studies in multidimensional stochastic processes: multivariate long-range dependence and synthesis of Gaussian random fields
(Under the direction of Vladas Pipiras)

This thesis is concerned with the study of multidimensional stochastic processes with special dependence structures. It is comprised of 3 parts. The first two parts concern multivariate long-range dependent time series. These are stationary multivariate time series exhibiting long-range dependence in the sense that the impact of past values of the series to the future ones dies out slowly with the increasing lag. In contrast to the univariate case, where long-range dependent time series are well understood and applied across a number of research areas such as Economics, Finance, Computer Networks, Physics, Climate Sciences and many others, the study of multivariate long-range dependent time series has not matured yet. This thesis sets proper theoretical foundations of such series and examines their statistical inference under novel models. The third part of the thesis is concerned with two-dimensional stationary Gaussian random fields. In particular, a fast algorithm is proposed for exact synthesis of such fields based on convex optimization and is shown to outperform existing approaches.

To my parents, Anatoli and Thanasis...

ACKNOWLEDGEMENTS

It is with a great sense of gratitude and appreciation that I would like to thank my advisor Professor Vladas Pipiras. His unending support and guidance have been instrumental to my academic growth and to the successful completion of this thesis. Constantly pursuing excellence, Professor Pipiras instilled in me his uncompromising scientific standards. I will be forever grateful to him for believing in me, for sharing with me his ideas and passion for Mathematics during our countless meetings, and most importantly for inspiring me to follow my dreams to the fullest.

I would also like to thank the other members of my dissertation committee, Professors Shankar Bhamidi, Amarjit Budhiraja, Chuanshu Ji and M. Ross Leadbetter for their immeasurable contribution in teaching me the notions of measure and probability. Their encouragement and support pushed me to become a better researcher, while their helpful comments and suggestions improved the quality of this thesis. Moreover, I am deeply thankful to Professor Edward Carlstein and to all the faculty members at the Department of Statistics and Operations Research who mentored me in my first steps as a Statistics teacher, and were always willing to offer me their valuable advice.

Next, I wish to offer my sincere thanks to Dr. Mark Little for giving me the opportunity to work as a Research Fellow at the Advanced Analytics and Research Division at SAS. There, I have been fortunate to work alongside Dr. Xilong Chen, an exceptional economist and statistician. In particular, my discussions with Dr. Chen have helped me gain a broader perspective on the role of Statistics in real world problems and have also contributed to the material of Chapter 3.

On a more personal level, I want to thank Professor Nikolaos Artavanis for introducing me to the wonderful world of Mathematics and for motivating me to pursue a graduate degree at UNC. Additionally, I will always be grateful to Apostolis and Matina Karkalemi for their invaluable teachings and their parental love. I also want to express my special

thanks and deep appreciation to Lily Bowers for her love and faith in me. Her continuing patience and support were paramount to the completion of this thesis. Moreover, I am particularly thankful to her parents Tom and Abby Bowers who supported me in every step of this journey.

Finally, my words fall short to express my gratitude to my parents, Anatoli Choupanidou and Thanasis Kechagias. With their uncompromising moral values and their unparalleled work ethic they have always set an example for me to follow. From the bottom of my heart, I thank them for their self-sacrifice, for their profound kindness and for their endless love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
AR	Autoregressive
BIC	Bayesian Information Criterion
CME	Circulant Matrix Embedding
CLDL	Conditional Likelihood Durbin Levinson
CG	Conjugate Gradient
CCE	Cutoff Matrix Embedding
DFT	Discrete Fourier Transform
DL	Durbin Levinson
FARIMA	Fractionally Autoregressive Integrated Moving Average
FFT	Fast Fourier Transform
FIVARMA	Fractionally Integrated Vector Autoregressive Moving Average
IN	Innovations
LRD	Long-Range Dependence/Dependent
MA	Moving Average
OCE	Optimal Circulant Embedding
PCG	Preconditioned Conjugate Gradient
PLB	Primal Log-Barrier
SRD	Short-Range Dependence/Dependent
SWCE	Smoothing Windows Circulant Embedding
VAR	Vector Autoregressive
VARFIMA	Vector Autoregressive Fractionally Integrated Moving Average

CHAPTER 1
Introduction

Since the introduction of Brownian motion in the late 19th century, stochastic processes have been used extensively to model the evolution of random phenomena over time and space. Specifically in recent years, technological advancements in monitoring systems and data collection methods have motivated the use of multidimensional stochastic processes for simultaneous modeling of data originating from multiple sources. This thesis focuses on two classes of multidimensional stochastic processes that have special dependence structures, namely, multivariate (vector-valued) time series exhibiting *long-range dependence* (LRD) (Chapters 2 and 3) and multidimensional (scalar-valued) stationary Gaussian random fields (Chapter 4).

Long-range dependent (LRD) time series models appear in the analysis of time series data, where the dependence between observations that are increasingly far apart in time dies out slowly. This behavior is in contrast to the one displayed by the well-studied family of vector autoregressive (VAR) series which are short-range dependent, in the sense that their memory dies out exponentially fast for observations that are increasingly far apart in time. In the univariate case, LRD time series models have been studied extensively in theory and have been popular in a wide range of application areas such as Statistical Physics, Computer Networks, Finance, Hydrology and Geosciences. They are commonly defined as discrete-time stationary time series models characterized by slowly decaying autocorrelations in the time domain and in the frequency domain by a spectral density function which explodes at the zero frequency. Bivariate and, more generally, multivariate LRD time series models have also been considered by a number of researchers. But proper theoretical foundations for a general class of such models were laid only recently in Kechagias and Pipiras (2015), a work that we present in Chapter 2.

In particular, the focus of Chapter 2 is on the role of the so-called *phase*, an important notion that emerges in the study of multivariate LRD time series but which erringly has received little attention until now. More specifically, the phase parameter is the argument ϕ of the complex constant $ce^{i\phi}$ that appears in the cross spectrum at the zero frequency and controls the asymmetry (time non-reversibility) in the series at large time lags. A time-reversible bivariate time series, for example, has a (matrix-valued) autocovariance function $\gamma(n)$ that satisfies the symmetry condition $\gamma(n) = \gamma(-n)$, $n \in \mathbb{Z}$. Consequently, such a series will have real-valued spectrum and thus the phase parameter $\phi = 0$. In many empirical applications, however, the observed LRD series are not time-reversible calling for multivariate LRD models that allow for general ϕ. In Chapter 2, we construct such models by taking noncausal (two-sided) linear representations with matrix-valued coefficients whose entries decay as a power law. Moreover, we propose a parametric model which is a noncausal multivariate extension of the celebrated FARIMA series that admits such a representation and give an exact form for its autocovariance function. Finally, we develop a causal multivariate LRD model that allows for general phase by introducing a new type of coefficients which we will refer to as *trigonometric power-law* coefficients.

In Chapter 3, we turn our attention to statistical inference. As the practical usefulness of the causal linear representations that we constructed using the trigonometric power-law coefficients is yet to be determined, we focus here on the noncausal fractionally integrated model discussed in Chapter 2. We begin by proposing suitable reparametrizations under which the model is identifiable in the sense that the (Gaussian) likelihood function does not take the same value for different parameter choices. In particular, these parameterizations seem to apply only in the bivariate case, and so the focus throughout Chapter 3 is on bivariate series. Our second contribution in Chapter 3 is the enrichment of the parametrized FARIMA class with models that allow for autoregressive (AR, for short) and moving average (MA, for short) components while still being identifiable and of course leading to a general phase parameter. This extension is far from simple compared to the univariate case, since the matrix-valued LRD and SRD filters do not commute in general. For example, taking nondiagonal AR and MA components yields fractionally cointegrated series which are a lot more delicate to deal with.

Parametric estimation is performed using a conditional maximum likelihood approach that is based on the multivariate Durbin-Levinson algorithm (a recursive time series algorithm commonly used to solve the so-called Yule-Walker equations). This method is appealing in our case as it only requires the knowledge of the autocovariance function of models with no AR component whose exact form we provide. Moreover, we show that switching from the Durbin-Levinson to a computationally slower algorithm called the Innovations algorithm, allows for multistep ahead forecasting based on a finite information set. Despite the efficiency loss, this approach is particularly useful as the usual infinite past forecasting approaches break down in the noncausal case. The performance of the estimation method is tested in a simulation study, where we fit a number of different models on synthetic data. Our findings show that the method works well in most cases with a few exceptions that raise interesting questions for future work. We conclude Chapter 3 with an application to the annualized U.S. inflation rates on goods and services, which show evidence of both long-range dependence and asymmetry. The proposed noncausal model outperforms other models previously used to analyze this data.

In Chapter 4, we switch gears from multivariate long-range dependent time series to multidimensional stationary Gaussian random fields. Such fields are fundamental modeling tools in many application areas including Environmental Sciences, Image Analysis, Geostatistics and Machine Learning. The computational cost of synthesizing such data using conventional methods (e.g. Cholesky decomposition), however, is often prohibitive, restricting empirical studies to small sample sizes. Moreover, in practice, spatial datasets are often large with many missing observations which may need to be simulated in Monte Carlo EM or MCMC algorithms, calling for fast (Gaussian field) data-generation algorithms. In particular, in Chapter 4, we propose a new method that can be used for the exact synthesis of (large) stationary Gaussian two-dimensional random fields with a prescribed covariance structure.

One of the most popular methods for fast and exact generation of stationary Gaussian random fields is the so-called standard circulant matrix embedding (CME). The idea of the method is to use a suitable periodization to embed a covariance matrix of interest into a larger circulant matrix whose eigenvalues can be computed efficiently using the fast Fourier

transform (FFT). If all the eigenvalues are nonnegative, which in theory is true for large enough sample size, it is straightforward to construct a stationary Gaussian random field with the desired covariance structure. Nevertheless, some of the eigenvalues often remain negative for computationally feasible large sample sizes and many covariance structures of practical interest.

A possible reason that some eigenvalues are negative is that the covariance embedding is not smooth at the boundary of periodization. To deal with this issue, two approaches have been suggested in the literature, namely, the cutoff circulant embedding (CCE) and the smoothing windows circulant embedding (SWCE). In the CCE method, the initial covariance is extended suitably in a parametric fashion, based on the model at hand, to a larger domain, leading to a covariance embedding with nonnegative eigenvalues. While such extensions have been found for several classes of models of stationary fields, their construction is often nontrivial. The SWCE method, on the other hand, introduces a transition region around the boundary of periodization, over which the covariance function is then extended using a smoothing kernel which smooths out the discontinuities. The SWCE method is fast, independent of the model at hand and outperforms the CCE method in the sense that it requires smaller transition regions.

Our contribution to this body of work is a new circulant embedding method called optimal optimal circulant embedding (OCE). Given a covariance function r on a finite square lattice, the OCE method finds the embedding covariance $\widetilde{r}$ that is closest to r and has nonnegative eigenvalues by solving a quadratic program with linear inequality constraints. The objective function of the problem is the quadratic distance between r and $\widetilde{r}$ taken at all points outside the transition region, while the constraints consist of all the eigenvalues of $\widetilde{r}$ being nonnegative.

The OCE method is appealing in several ways. First, the method proposes a novel approach based on quadratic constrained optimization. This is quite fitting given the growing integration of optimization tools in Statistics. Second, several key components of the optimization procedure can be implemented more efficiently using FFT. In particular, we show that direct matrix-vector product calculations of complexity $O(N^4)$, where N is the side length of the field to be constructed, can be performed using FFT which reduces the

complexity to $O(N^2 \log N)$. Moreover, these calculations need to be repeated over multiple iterations and thus our approach reduces the computational cost significantly. Third, the OCE outperforms the SWCE and CCE methods for several practical covariance functions as it needs smaller transition regions to lead to embeddings with nonnegative eigenvalues.

CHAPTER 2

Definitions and representations of multivariate long-range dependent time series

2.1 Introduction

Long-range dependent (LRD, in short) time series models have been studied extensively in theory and have been popular in a wide range of applications (Beran (2013), Doukhan, Oppenheim and Taqqu (2003), Giraitis, Koul and Surgailis (2012), Park and Willinger (2000), Robinson (2003)). They are defined as (second-order) stationary time series models satisfying one of the following non-equivalent conditions. In the time domain, the autocovariance function $\gamma(h) = \mathrm{Cov}(X_0, X_h)$ of a LRD time series $\{X_n\}_{n\in\mathbb{Z}}$ is such that

$$\gamma(k) = L_1(k)k^{2d-1}, \quad \text{as } k \to \infty, \tag{2.1}$$

where $d \in (0, 1/2)$ is the *long-range dependence (LRD) parameter* and L_1 is a slowly varying function at infinity. In the spectral domain, the spectral density function $f(\lambda)$ of $\{X_n\}_{n\in\mathbb{Z}}$ is such that

$$f(\lambda) = L_2(\lambda)\lambda^{-2d}, \quad \text{as } \lambda \to 0^+, \tag{2.2}$$

where L_2 is a slowly varying function at 0. Another common way to define a LRD time series is through a causal (one-sided) linear representation

$$X_n = \mu + \sum_{k=0}^{\infty} \psi_k \epsilon_{n-k}, \tag{2.3}$$

where $\{\epsilon_n\}_{n\in\mathbb{Z}}$ is a white noise series, μ is a constant mean and the sequence $\{\psi_k\}_{k\geq 0}$ satisfies

$$\psi_k = L_3(k)k^{d-1}, \quad \text{as } k \to \infty, \tag{2.4}$$

where L_3 is a slowly varying function at infinity.

In this chapter, we are interested in the notion of LRD for multivariate time series, that is, $\mathbb{R}^p$-valued time series $X_n = (X_n^1, \ldots, X_n^p)'$, where prime indicates transpose. Second-order stationary multivariate series are now characterized by matrix-valued autocovariance function $\gamma(h) = \mathbb{E}X_0X_h' - \mathbb{E}X_0\mathbb{E}X_h'$ and matrix-valued spectral density $f(\lambda)$. In the multivariate case, $\gamma(h)$ does *not* necessarily satisfy the symmetry (time-reversibility) condition $\gamma(h) = \gamma(-h)$ and, moreover, $f(\lambda)$ can have complex-valued entries in general (Hannan (1970), Reinsel (1997), Lütkepohl (2005)). Several forms of multivariate LRD, not surprisingly, have already been considered in the literature. The goal of this chapter is to clarify this notion. As will be seen below, there are a number of new interesting issues that arise in the multivariate but not in the univariate case, and which have not been studied in greater detail yet.

The most general form of the *bivariate* LRD appears in Robinson (2008) who supposed that the spectral density matrix satisfies

$$f(\lambda) \sim \begin{pmatrix} \omega_{11}|\lambda|^{-2d_1} & \omega_{12}|\lambda|^{-(d_1+d_2)}e^{-i\mathrm{sign}(\lambda)\phi} \\ \omega_{21}|\lambda|^{-(d_1+d_2)}e^{i\mathrm{sign}(\lambda)\phi} & \omega_{22}|\lambda|^{-2d_2} \end{pmatrix}, \quad \text{as } \lambda \to 0, \tag{2.5}$$

where $\omega_{11}, \omega_{12}, \omega_{21}, \omega_{22} \in \mathbb{R}$, $d_1, d_2 \in (0, 1/2)$ and $\phi \in (-\pi, \pi]$. The parameter ϕ is called a *phase* parameter and is unique to the bivariate LRD case (see also Section 2.2 below). It controls asymmetry (time non-reversibility) in the series at large time lags. Many results of this chapter will be related directly to this parameter and its role. The definition of LRD with

$$\phi = 0 \tag{2.6}$$

was considered in Lobato (1999), Velasco (2003), Marinucci and Robinson (2003), Christensen and Nielsen (2006), Nielsen (2004; 2007). The value $\phi = 0$ is associated with LRD time series which are symmetric (time-reversible). The case of

$$\phi = \frac{\pi}{2}(d_1 - d_2) \tag{2.7}$$

is another special case and was considered in Lobato (1997), Robinson (2002; 2008), Shimotsu (2007), Nielsen (2011), Nielsen and Frederiksen (2011). For example, a natural extension of the FARIMA$(0, d, 0)$ series to the bivariate case as

$$\begin{pmatrix} (I-B)^{d_1} & 0 \\ 0 & (I-B)^{d_2} \end{pmatrix} X_n = \epsilon_n, \tag{2.8}$$

where $\{\epsilon_n\}$ is a bivariate white noise and B is the backshift operator, corresponds to ϕ given by (2.7). The case of general ϕ is considered in Robinson (2008) as indicated above, and without referring to ϕ explicitely in Robinson (1995). Multivariate LRD models also appear in Robinson (1994), Chan and Terrin (1995), Marinucci and Robinson (2001; 2003), Robinson and Yajima (2002), Chen and Hurvich (2003; 2006) in the context of fractional cointegration, and in Achard, Bassett, Meyer-Lindenberg and Bullmore (2008), Wendt, Scherrer, Abry and Achard (2009) in the context of fractal connectivity.

We contribute to the understanding of the notion of multivariate LRD in the following three ways. First, we extend the definition (2.5) to the multivariate case and consider its analogue in the time domain (as in (2.1)) and relationships between them. This contribution is somewhat standard, but also necessary to set a proper foundation for dealing with multivariate LRD. Again, much of the discussion will focus on the role played by the phase parameters (ϕ in (2.5) in the bivariate case). In the bivariate case, similar results can be found in Robinson (2008).

Our second contribution is more original. It concerns linear representations of multivariate LRD series of the form

$$X_n = \sum_{k=-\infty}^{\infty} \Psi_k \epsilon_{n-k}, \tag{2.9}$$

where Ψ_k are $p \times p$ matrices and $\{\epsilon_n\}_{n\in\mathbb{Z}}$ is a p-variate white noise. Even more specifically, we are interested in *causal (one-sided)* representations, that is, (2.9) with $\Psi_k = 0$ when $k < 0$. It is not too difficult to construct *non-causal (two-sided)* representations (2.9) having general phase parameters in the definition of LRD by having Ψ_k decay as suitable power-law functions as $k \to \infty$ and $k \to -\infty$. But it is not obvious how to construct such causal representations. For example, taking Ψ_k to behave as a power-law function in a

causal representation leads necessarily to the phase parameters $\frac{\pi}{2}(d_{j_1} - d_{j_2})$, where d_{j_1}, d_{j_2} are the LRD parameters of component series (Section 2.3 below).

We show that *causal* multivariate LRD series of general phase can be constructed taking the elements of Ψ_k as linear combinations of, what we will call, *trigonometric power-law coefficients*

$$\begin{aligned} c_k^{a,b} &= k^{-b}\cos(2\pi k^a), \\ s_k^{a,b} &= k^{-b}\sin(2\pi k^a), \qquad k \geq 0, \end{aligned} \tag{2.10}$$

where $0 < a < 1$ and $\frac{1}{2} < b \leq 1 - \frac{1}{2}a$. (By convention, $0^p = 0$ for $p \in \mathbb{R}$, so that $c_0^{a,b} = s_0^{a,b} = 0$.) The use of such coefficients can be traced back at least to Wainger (1965). What makes them special and relevant for LRD is that their discrete Fourier transform satisfies, for example,

$$\sum_{k=0}^{\infty} c_k^{a,b} e^{-ik\lambda} \sim c_1 |\lambda|^{-\frac{1-b-a/2}{1-a}} e^{i(c_2|\lambda|^{-\frac{a}{1-a}} - \frac{\pi}{4})}, \quad \text{as} \quad \lambda \to 0, \tag{2.11}$$

where c_1, c_2 are two non-zero constants (Wainger (1965) and Appendix A.2 below). Thus, even in the univariate case, the time series

$$X_n = \sum_{k=0}^{\infty} c_k^{a,b} \epsilon_{n-k} \tag{2.12}$$

is LRD (cf. (2.2)) with the LRD parameter

$$d = \frac{1 - b - a/2}{1 - a}. \tag{2.13}$$

Though the trigonometric power-law coefficients can be used to construct new univariate and multivariate causal LRD series, their statistical inference and practical usefulness remain to be explored in the future.

Lastly and third, we provide a natural multivariate extension of FARIMA$(0, d, 0)$ having a general phase. As indicated above, the extension (2.8) of FARIMA$(0, d, 0)$ series to the bivariate case has necessarily the phase parameter $\phi = (d_1 - d_2)\pi/2$. For example, a

bivariate extension with a general phase can be obtained with

$$X_n = \begin{pmatrix} (I-B)^{-d_1} & 0 \\ 0 & (I-B)^{-d_2} \end{pmatrix} Q_+ \epsilon_n + \begin{pmatrix} (I-B^{-1})^{-d_1} & 0 \\ 0 & (I-B^{-1})^{-d_2} \end{pmatrix} Q_- \epsilon_n, \tag{2.14}$$

where $Q_+,\ Q_-$ are 2×2 matrices with real-valued entries. We provide explicit formulas for the autocovariance functions of this extension, including the multivariate case, in Section 2.5.

The structure of Chapter 2 is as follows. The definitions of multivariate LRD in the time and spectral domains are given in Section 2.2. Non-causal representations of multivariate LRD series are studied in Section 2.3. Section 2.4 concerns causal linear representations. Multivariate FARIMA series are considered in Section 2.5. Conclusions and future directions can be found in Section 2.6. Technical proofs are moved to Appendix A.1, and the behavior of the Fourier series of the trigonometric power-law coefficients (2.10) is presented in Appendix A.2.

2.2 Definitions in the time and spectral domains

We begin with the definitions of multivariate LRD in the time and spectral domains, extending conditions (2.1) and (2.2). We shall suppose for simplicity that all slowly varying functions behave asymptotically as constants. This is the relevant case for statistical inference. Moreover, these slowly varying functions would appear below in a matrix form which, along with a matrix regular variation, is only now receiving a closer look (e.g. Meerschaert and Scheffler (2013)).

The definitions below use the following notation. For $a > 0$ and a diagonal matrix $M = \mathrm{diag}(m_1, \ldots, m_p)$, we write $a^M = \mathrm{diag}(a^{m_1}, \ldots, a^{m_p})$. The autocovariance matrix function of a second-order stationary series $X = \{X_n\}_{n \in \mathbb{Z}}$ is defined as $\gamma(h) = (\gamma_{jk}(h))_{j,k=1,\ldots,p} = \mathbb{E}X_0 X_h' - \mathbb{E}X_0 \mathbb{E}X_h'$, $h \in \mathbb{Z}$, and the corresponding spectral density matrix function, if it exists, is denoted $f(\lambda) = (f_{jk}(\lambda))_{j,k=1,\ldots,p}$. For a matrix A, A^* stands for its Hermitian transpose. Finally, we will use the symbol $\sim$ to indicate a limit, e.g. $a_n \sim a$ stands for

$\lim a_n = a$. In the last example, note that this is equivalent to $a_n/a \to 1$ when $a \neq 0$, and that there is no such interpretation when $a = 0$.

Definition 2.1. (*Time domain*) A multivariate (p-vector) second-order stationary time series is LRD if its autocovariance matrix function satisfies:

$$\gamma(n) = n^{D-(1/2)I} R(n) n^{D-(1/2)I} = \left(R_{jk}(n) n^{(d_j+d_k)-1} \right)_{j,k=1,\ldots,p}, \tag{2.15}$$

where $D = \mathrm{diag}(d_1, \ldots, d_p)$ with $d_j \in (0, 1/2), j = 1, \ldots, p$, and $R(u) = (R_{jk}(u))_{j,k=1,\ldots,p}$ is an $\mathbb{R}^{p\times p}$-valued function satisfying

$$R(u) \sim R = (R_{jk})_{j,k=1,\ldots,p}, \quad \text{as } u \to +\infty, \tag{2.16}$$

for some $p \times p$ matrix R, where $R_{jk} \in \mathbb{R}$ and $R_{jj} \neq 0, j = 1, \ldots, p$.

Definition 2.2. (*Spectral domain*) A multivariate (p-vector) second-order stationary time series is LRD if its spectral density matrix function satisfies

$$f(\lambda) = \lambda^{-D} G(\lambda) \lambda^{-D^*} = \left(G_{jk}(\lambda) \lambda^{-(d_j+d_k)} \right)_{j,k=1,\ldots,p}, \tag{2.17}$$

where $D = \mathrm{diag}(d_1, \ldots, d_p)$ with $d_j \in (0, 1/2), j = 1, \ldots, p$, and $G(\lambda) = (G_{jk}(\lambda))_{j,k=1,\ldots,p}$ is a $\mathbb{C}^{p\times p}$-valued, Hermitian symmetric, non-negative definite matrix function satisfying

$$G(\lambda) \sim G = (G_{jk})_{j,k=1,\ldots,p} = \left(g_{jk} e^{i\phi_{jk}} \right)_{j,k=1,\ldots,p}, \qquad \text{as } \lambda \to 0^+, \tag{2.18}$$

for some $p \times p$, Hermitian symmetric, non-negative definite matrix G, where $g_{jk} \in \mathbb{R}$, $g_{jj} \neq 0$ and $\phi_{jk} \in (-\pi, \pi]$.

A number of remarks regarding Definitions 2.1 and 2.2 are in place.

Remark 2.2.1. Note that the component series $\{X_n^j\}_{n\in\mathbb{Z}}$, $j = 1, \ldots, p$, of a multivariate LRD series are LRD with parameters d_j, $j = 1, \ldots, p$. (Note that $\phi_{jj} = 0$ since the matrix G is Hermitian symmetric and hence has real-valued entries on the diagonal.) Another possibility would be to require that at least one of the component series $\{X_n^j\}_{n\in\mathbb{Z}}$ is LRD. This could be achieved by assuming in (2.17) that $d_j \in [0, 1/2)$, $j = 1, \ldots, p$, and that at

least one $d_j > 0$. But note that such assumption is not appropriate in (2.15). See also Section 2.6 for a discussion on the so-called antipersistence case $d_j < 0$.

Remark 2.2.2. Note that the structure of (2.17) is such that $f(\lambda)$ is Hermitian and non-negative definite. D^* appearing in (2.17) can be replaced by D since it is diagonal and consists of real-valued entries. We write D^* to make the non-negative definiteness of $f(\lambda)$ more evident. The entries ϕ_{jk} in (2.18) are referred to as *phase parameters*. Their role will become more apparent in Proposition 2.2.1 below where Definitions 2.1 and 2.2 are compared.

There are two common but slightly different ways to represent the phase parameters. Note that a complex number $z = z_1 + iz_2 \in \mathbb{C}$, $z_1, z_2 \in \mathbb{R}$, can be represented as (assuming $z_1 \neq 0$ in the second relation below and denoting the principle value of the argument by $\mathrm{Arg}(z) \in (-\pi, \pi]$)

$$
\begin{aligned}
z &= |z| e^{i\mathrm{Arg}(z)} && (2.19)\\
&= \sqrt{z_1^2 + z_2^2}\, e^{i(\arctan(\frac{z_2}{z_1}) + \pi \mathrm{sign}(z_2) 1_{\{z_1<0\}})} = \mathrm{sign}(z_1)\sqrt{z_1^2 + z_2^2}\, e^{i \arctan(\frac{z_2}{z_1})} \\
&= z_1 \sqrt{1 + \frac{z_2^2}{z_1^2}}\, e^{i \arctan(\frac{z_2}{z_1})} = z_1 \sqrt{1 + \tan^2(\arctan \frac{z_2}{z_1})}\, e^{i \arctan(\frac{z_2}{z_1})} \\
&= \frac{z_1}{\cos\phi} e^{i\phi}, \quad \text{with} \quad \phi = \arctan(\frac{z_2}{z_1}). && (2.20)
\end{aligned}
$$

The two specifications of g_{jk} and the phase ϕ_{jk} correspond to (2.19) (e.g. Brockwell and Davis (2009), p. 422) and (2.20) (e.g. Hannan (1970), pp. 43–44):

$$g_{jk} = |G_{jk}|, \quad \phi_{jk} = \mathrm{Arg}(G_{jk}), \tag{2.21}$$

$$g_{jk} = \frac{\Re G_{jk}}{\cos\phi_{jk}}, \quad \phi_{jk} = \arctan \frac{\Im G_{jk}}{\Re G_{jk}}. \tag{2.22}$$

Note that, in the case (2.22), $\phi_{jk} \in (-\pi/2, \pi/2)$.

Remark 2.2.3. It should also be noted that the phase parameters are unique to the LRD case. Taking

$$\sum_{n=-\infty}^{\infty} ||\gamma(n)|| < \infty \tag{2.23}$$

for the definition of *short-range dependent* (SRD) series (with $||\cdot||$ denoting the Frobenius norm of a matrix), we have

$$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda}\gamma(n). \tag{2.24}$$

In particular, $f(0) = (2\pi)^{-1}\sum_{n=-\infty}^{\infty}\gamma(n)$ consists of real-valued entries, and

$$f(\lambda) \sim G \ (= f(0)), \quad \text{as } \lambda \to 0, \tag{2.25}$$

where G consists of real-valued entries. The relation (2.25) corresponds to (2.17)−(2.18) with $d_j = 0$, $j = 1, \ldots, p$, and all phase parameters $\phi_{jk} = 0$. Note that we view antipersistence corresponding to $G = (2\pi)^{-1}\sum_{n=-\infty}^{\infty}\gamma(n) = 0$ as a special case of short-range dependence. The antipersistence case, however, is expected to involve phase parameters similarly to the case of long-range dependence. See Section 2.6 for further discussion.

Remark 2.2.4. The squared coherence function

$$\mathcal{H}_{jk}^2(\lambda) = \frac{|f_{jk}(\lambda)|^2}{f_{jj}(\lambda)f_{kk}(\lambda)}$$

satisfies $0 \le \mathcal{H}_{jk}^2(\lambda) \le 1$. As $\lambda \to 0^+$, this translates into

$$0 \le \lim_{\lambda\to 0^+} \frac{|G_{jk}|^2\lambda^{-2(d_j+d_k)}}{G_{jj}\lambda^{-2d_j}G_{kk}\lambda^{-2d_k}} = \frac{|G_{jk}|^2}{G_{jj}G_{kk}} \le 1 \tag{2.26}$$

and also explains why the choice of $\lambda^{-(d_j+d_k)}$ is natural for the cross-spectral density $f_{jk}(\lambda)$.

Remark 2.2.5. Note also that (2.17) is considered for $\lambda \to 0^+$. Since f is Hermitian symmetric, (2.17)−(2.18) can be replaced by

$$f_{jk}(\lambda) = G_{jk}(\lambda)|\lambda|^{-(d_j+d_k)} \sim g_{jk}e^{i\phi_{jk}\text{sign}(\lambda)}|\lambda|^{-(d_j+d_k)}, \quad \text{as } \lambda \to 0, \tag{2.27}$$

where $G_{jk}(-\lambda) = G_{jk}(\lambda)^*$, if both positive and negative λ's are considered. (Note that since G is Hermitian symmetric, we have $g_{jk} = g_{kj}$.)

Proposition 2.2.1 below compares Definitions 2.1 and 2.2. It uses the notion of a quasi-monotone slowly varying function whose definition is recalled in Appendix A.1. The proof of the proposition can also be found in the appendix. As usual, $\Gamma(\cdot)$ denotes the gamma function.

Proposition 2.2.1. *(i) Suppose the component functions R_{jk} are quasi-monotone slowly varying. Then, Definition 2.1 implies Definition 2.2 with*

$$G_{jk} = \frac{\Gamma(d_j + d_k)}{2\pi}\Big\{(R_{jk} + R_{kj})\cos(\frac{\pi}{2}(d_j + d_k)) - i(R_{jk} - R_{kj})\sin(\frac{\pi}{2}(d_j + d_k))\Big\} \quad (2.28)$$

in the relation (2.18).

(ii) Suppose the component functions $\Re G_{jk}$, $\Im G_{jk}$ are quasi-monotone slowly varying. Then, Definition 2.2 implies Definition 2.1 with

$$R_{jk} = 2\Gamma(1 - (d_j + d_k))\Big\{\Re G_{jk}\sin(\frac{\pi}{2}(d_j + d_k)) - \Im G_{jk}\cos(\frac{\pi}{2}(d_j + d_k))\Big\} \quad (2.29)$$

in the relation (2.16).

Note that under the specification (2.22), with $G_{jk} = g_{jk}e^{i\phi_{jk}}$, the relation (2.28) yields

$$\begin{aligned} \phi_{jk} &= -\arctan\Big\{\frac{R_{jk} - R_{kj}}{R_{jk} + R_{kj}}\tan(\frac{\pi}{2}(d_j + d_k))\Big\}, & (2.30)\\ g_{jk} &= \frac{\Gamma(d_j + d_k)(R_{jk} + R_{kj})\cos(\frac{\pi}{2}(d_j + d_k))}{2\pi\cos(\phi_{jk})}. & (2.31) \end{aligned}$$

Similarly, under the specification (2.21),

$$\begin{aligned} \phi_{jk} &= -\arctan\Big\{\frac{R_{jk} - R_{kj}}{R_{jk} + R_{kj}}\tan(\frac{\pi}{2}(d_j + d_k))\Big\} - \pi\,\mathrm{sign}(R_{jk} - R_{kj})1_{\{R_{jk}+R_{kj}<0\}}, & (2.32)\\ g_{jk} &= \frac{\Gamma(d_j + d_k)}{2\pi}\left(R_{jk}^2 + R_{kj}^2 + 2R_{jk}R_{kj}\cos(\pi(d_j + d_k))\right)^{1/2}, & (2.33) \end{aligned}$$

where we used the first equality following (2.19).

The relation (2.30) sheds light on the phase parameters ϕ_{jk}. Note that $\phi_{jk} = 0$ if and only if $R_{jk} = R_{kj}$. In view of (2.15)–(2.16), the last property corresponds to $\gamma_{jk}(n)$ being

symmetric as $n \to \infty$ and $n \to -\infty$, that is, $\gamma_{jk}(-n) \sim \gamma_{jk}(n)$, as $n \to \infty$. (We used here the fact that $\gamma(-n) = \gamma(n)'$ and hence $\gamma_{jk}(-n) = \gamma_{kj}(n)$.)

2.3 Non-causal linear representations

We are interested here in linear representations (2.9) of multivariate LRD time series. In the next result, we show that linear time series (2.9) with power-law coefficients Ψ_k in the asymptotic sense as $k \to \infty$ and $k \to -\infty$, are multivariate LRD. We argue at the end of the section that such non-causal time series can lead to general phase parameters.

Proposition 2.3.1. *Let $\{\epsilon_n\}_{n\in\mathbb{Z}}$ be an $\mathbb{R}^p$-valued white noise, satisfying $\mathbb{E}\epsilon_n = 0$ and $\mathbb{E}\epsilon_n\epsilon_n' = I$. Let also $\{\Psi_m = (\psi_{jk,m})_{j,k=1,\dots,p}\}_{m\in\mathbb{Z}}$ be a sequence of real-valued matrices such that*

$$\psi_{jk,m} = L_{jk}(m)|m|^{d_j-1}, \quad m \in \mathbb{Z}, \tag{2.34}$$

where $d_j \in (0, 1/2)$ and $L(m) = (L_{jk}(m))_{j,k=1,\dots,p}$ is an $\mathbb{R}^{p\times p}$-valued function satisfying

$$L(m) \sim A^+, \quad as \quad m \to \infty, \quad and \quad L(m) \sim A^-, \quad as \quad m \to -\infty, \tag{2.35}$$

for some $p \times p$ real-valued matrices $A^+ = (\alpha^+_{jk})_{j,k=1,\dots,p}$, $A^- = (\alpha^-_{jk})_{j,k=1,\dots,p}$. Then, the autocovariance function of the time series X_n given by a linear representation

$$X_n = \sum_{m=-\infty}^{\infty} \Psi_m \epsilon_{n-m}, \tag{2.36}$$

satisfies (2.15)–(2.16) with

$$R_{jk} = \frac{\Gamma(d_j)\Gamma(d_k)}{\Gamma(d_j+d_k)} \left(c^1_{jk} \frac{\sin(\pi d_j)}{\sin(\pi(d_j+d_k))} + c^2_{jk} + c^3_{jk} \frac{\sin(\pi d_k)}{\sin(\pi(d_j+d_k))} \right), \tag{2.37}$$

where

$$\begin{array}{rcccl} c^1_{jk} & = & \sum_{t=1}^p \alpha^-_{jt}\alpha^-_{kt} & = & (A^-(A^-)^*)_{jk}, \\ c^2_{jk} & = & \sum_{t=1}^p \alpha^-_{jt}\alpha^+_{kt} & = & (A^-(A^+)^*)_{jk}, \\ c^3_{jk} & = & \sum_{t=1}^p \alpha^+_{jt}\alpha^+_{kt} & = & (A^+(A^+)^*)_{jk}. \end{array} \tag{2.38}$$

In particular, if $R_{jj} \neq 0,\ j = 1, \ldots, p$, then the series is LRD in the sense of Definition 2.1.

The proof of Proposition 2.3.1 can be found in Appendix A.1. Entries of the matrices A^+ and A^- in (2.35) are allowed to be zero. In particular, the case

$$\psi_{jk,m} = l_{jk}(m)|m|^{d_{jk}-1}, \quad m \in \mathbb{Z}, \tag{2.39}$$

with possibly different $d_{jk} \in (0, 1/2)$ across k for fixed j and $l_{jk}(m) \sim \beta_{jk}^{\pm}$ with $\beta_{jk}^{\pm} \in \mathbb{R}$, as $m \to \pm\infty$, can be expressed as (2.34) with $d_j = \max_k d_{jk}$.

Note also that the time series (2.36) is proved to be LRD in the sense of Definition 2.1. In view of Proposition 2.2.1, (i), the time series is also expected to be LRD in the sense of Definition 2.2 with G_{jk} given by (2.28). To calculate G_{jk}, note that

$$\begin{aligned} R_{jk} + R_{kj} &= \frac{\Gamma(d_j)\Gamma(d_k)}{\Gamma(d_j + d_k)} \left(c_{jk}^1 \frac{\sin(\pi d_j) + \sin(\pi d_k)}{\sin(\pi(d_j + d_k))} + c_{jk}^2 + c_{kj}^2 + c_{jk}^3 \frac{\sin(\pi d_k) + \sin(\pi d_j)}{\sin(\pi(d_j + d_k))} \right) \\ &= \frac{\Gamma(d_j)\Gamma(d_k)}{\Gamma(d_j + d_k)} \left(c_{jk}^1 \frac{\cos(\frac{\pi}{2}(d_j - d_k))}{\cos(\frac{\pi}{2}(d_j + d_k))} + c_{jk}^2 + c_{kj}^2 + c_{jk}^3 \frac{\cos(\frac{\pi}{2}(d_j - d_k))}{\cos(\frac{\pi}{2}(d_j + d_k))} \right), \end{aligned}$$

where we used basic trigonometric identities and the facts that $c_{jk}^1 = c_{kj}^1$, $c_{jk}^3 = c_{kj}^3$. Hence,

$$\begin{aligned} &\Gamma(d_j + d_k)(R_{jk} + R_{kj}) \cos(\frac{\pi}{2}(d_j + d_k)) \\ &\qquad = \Gamma(d_j)\Gamma(d_k) \left((c_{jk}^1 + c_{jk}^3) \cos(\frac{\pi}{2}(d_j - d_k)) + (c_{jk}^2 + c_{kj}^2) \cos(\frac{\pi}{2}(d_j + d_k)) \right). \end{aligned} \tag{2.40}$$

Similarly, one can show that

$$\begin{aligned} &\Gamma(d_j + d_k)(R_{jk} - R_{kj}) \sin(\frac{\pi}{2}(d_j + d_k)) \\ &\qquad = \Gamma(d_j)\Gamma(d_k) \left((c_{jk}^1 - c_{jk}^3) \sin(\frac{\pi}{2}(d_j - d_k)) + (c_{jk}^2 - c_{kj}^2) \sin(\frac{\pi}{2}(d_j + d_k)) \right). \end{aligned} \tag{2.41}$$

Combining (2.40) and (2.41), the relation (2.28) now yields

$$\begin{aligned} G_{jk} &= \frac{\Gamma(d_j)\Gamma(d_k)}{\Gamma(d_j+d_k)} \Big((c_{jk}^1 + c_{jk}^3)\cos(\frac{\pi}{2}(d_j-d_k)) + (c_{jk}^2 + c_{kj}^2)\cos(\frac{\pi}{2}(d_j+d_k)) \\ &\qquad + i(c_{jk}^1 - c_{jk}^3)\sin(\frac{\pi}{2}(d_j-d_k)) + i(c_{jk}^2 - c_{kj}^2)\sin(\frac{\pi}{2}(d_j+d_k)) \Big) \\ &= \frac{\Gamma(d_j)\Gamma(d_k)}{\Gamma(d_j+d_k)} \left(c_{jk}^1 e^{i\frac{\pi}{2}(d_j-d_k)} + c_{jk}^3 e^{-i\frac{\pi}{2}(d_j-d_k)} + c_{jk}^2 e^{i\frac{\pi}{2}(d_j+d_k)} + c_{kj}^2 e^{-i\frac{\pi}{2}(d_j+d_k)} \right). \end{aligned} \tag{2.42}$$

Setting

$$F = \mathrm{diag}\left(\Gamma(d_1) e^{i\frac{\pi}{2}d_1}, \ldots, \Gamma(d_p) e^{i\frac{\pi}{2}d_p} \right) \tag{2.43}$$

and using (2.38), the relation (2.42) can also be expressed as

$$\begin{aligned} G_{jk} &= \frac{\Gamma(d_j)\Gamma(d_k)}{\Gamma(d_j+d_k)} \Big((A^-(A^-)^*)_{jk} e^{i\frac{\pi}{2}(d_j-d_k)} + (A^+(A^+)^*)_{jk} e^{-i\frac{\pi}{2}(d_j-d_k)} \\ &\qquad + (A^-(A^+)^*)_{jk} e^{i\frac{\pi}{2}(d_j+d_k)} + (A^+(A^-)^*)_{jk} e^{-i\frac{\pi}{2}(d_j+d_k)} \Big) \\ &= \frac{1}{2\pi} \left((FA^-)(FA^-)^* + (F^*A^+)(F^*A^+)^* + (FA^-)(F^*A^+)^* + (F^*A^+)(FA^-)^* \right)_{jk} \\ &= \frac{1}{2\pi} \left((F^*A^+ + FA^-)(F^*A^+ + FA^-)^* \right)_{jk}. \end{aligned} \tag{2.44}$$

The relation (2.44) can also be derived informally as follows. Since $\sum_{m=-\infty}^{\infty} ||\Psi_m||^2 < \infty$, the relation (5.3) in Hannan (1970) implies that the time series (2.36) has the spectral density

$$\frac{1}{2\pi} \left(\sum_{m=-\infty}^{\infty} \Psi_m e^{-im\lambda} \right) \left(\sum_{m=-\infty}^{\infty} \Psi_m^* e^{im\lambda} \right). \tag{2.45}$$

Be Lemma A.1 in Appendix A.1, it is expected that

$$\sum_{m=-\infty}^{\infty} \Psi_m e^{-im\lambda} = \sum_{m=1}^{\infty} \Psi_m e^{-im\lambda} + \sum_{m=0}^{\infty} \Psi_{-m} e^{im\lambda} \sim \lambda^{-D} \left(F^*A^+ + FA^- \right), \tag{2.46}$$

which, when combined with (2.45), is consistent with (2.44).

Finally, note that, for fixed d_j's and $G = (G_{jk})_{j,k=1,\ldots,p}$, we can find matrices A^+, A^-, so that (2.44) holds. Indeed, since G is Hermitian symmetric and non-negative definite, we have $G = WW^*$ for some matrix W. The real matrices A^+, A^- can now be found by setting

$(2\pi)^{-1}(F^*A^+ + FA^-) = W$. (Note that, since $e^{-i\pi d/2}$ and $e^{i\pi d/2}$ are linearly independent, there are real α^+ and α^- such that $e^{-i\pi d/2}\alpha^+ + e^{i\pi d/2}\alpha^- = y$ for any $y \in \mathbb{C}$.) In particular, any phase ϕ_{jk} can be obtained with a suitable choice of A^+, A^-.

2.4 Causal linear representations

In this section, we focus on causal linear representations of multivariate LRD series, that is, the representations (2.9) with $\Psi_k = 0$ for $k < 0$. As shown in Section 2.4.1 below, causal representations with power-law coefficients can only have very special phase parameters. Causal representations with zero and more general phases, based on trigonometric power-law coefficients (2.10), are considered in Sections 2.4.2 and 2.4.3.

2.4.1 The special case of power-law coefficients

One consequence of the results of Section 2.3 is that the causal representations of multivariate LRD series with power-law coefficients can only have very special phase parameters. The next result restates Proposition 2.3.1 in the causal case.

Corollary 2.4.1. *Let $\{\epsilon_n\}_{n\in\mathbb{Z}}$ and $\{\Psi_m\}_{m\in\mathbb{Z}}$ be as in Proposition 2.3.1 but with $\Psi_m = 0$ for $m < 0$. Then, the autocovariance of the time series X_n given by a causal linear representation*

$$X_n = \sum_{m=0}^{\infty} \Psi_m \epsilon_{n-m}, \tag{2.47}$$

satisfies (2.15)–(2.16) with

$$R_{jk} = \frac{\Gamma(d_j)\Gamma(d_k)}{\Gamma(d_j + d_k)} \frac{\sin(\pi d_k)}{\sin(\pi(d_j + d_k))} (A^+(A^+)^*)_{jk}. \tag{2.48}$$

In particular, if $R_{jj} \neq 0$, $j = 1, \ldots, p$, then the series is LRD in the sense of Definition 2.1.

Arguing as for (2.44), the relation (2.48) yields

$$G_{jk} = \frac{\Gamma(d_j)\Gamma(d_k)}{2\pi} (A^+(A^+)^*)_{jk} e^{-i\frac{\pi}{2}(d_j - d_k)} \tag{2.49}$$

and hence the phase parameters

$$\phi_{jk} = -\frac{\pi}{2}(d_j - d_k). \tag{2.50}$$

This can also be deduced from (2.45)–(2.46) when $A^- = 0$.

2.4.2 The case of zero phases

The causal time series (2.47) with power-law coefficients leads necessarily to the phase parameters (2.50). What coefficient matrices Ψ_m could one take to obtain general phases? (Such coefficient matrices exist in theory by the multivariate version of the Paley-Wiener theorem.) It is instructive to begin the discussion with the case of zero phases (that is, the symmetric case), before moving to the general case.

Informally, in the case of zero phases, we are looking for coefficient matrices Ψ_m such that, as $\lambda \to 0^+$,

$$\left(\sum_{m=0}^{\infty} \Psi_m e^{-im\lambda}\right)\left(\sum_{m=0}^{\infty} \Psi_m^* e^{im\lambda}\right) \sim \lambda^{-D} G \lambda^{-D^*}, \tag{2.51}$$

where G is real-valued (and we included the factor $1/2\pi$ on the left-hand side of (2.45) into Ψ_m's). This relation would follow from

$$\left(\sum_{m=0}^{\infty} \psi_{jk,m} e^{-im\lambda}\right)\left(\sum_{m=0}^{\infty} \psi_{j'k',m} e^{im\lambda}\right) \sim c_{jk,j'k'} \lambda^{-(d_j + d_{j'})} \tag{2.52}$$

with real $c_{jk,j'k'}$. This in turn would suggest to look for coefficients ψ_m such that

$$\sum_{m=0}^{\infty} \psi_m e^{-im\lambda} \sim c\lambda^{-d} \tag{2.53}$$

with real c. Note, however, that (2.53) with real c is not plausible. Writing $c\lambda^{-d} = ci^d(i\lambda)^{-d}$ the behavior of $(i\lambda)^{-d}$ can be captured by taking power-law coefficients. It is, however, impossible to recover i^d through a Fourier series of real coefficients. In fact, (2.53) is only expected with complex $c = Ci^{-d}$, corresponding to power-law coefficients ψ_m. But again,

power-law coefficients lead to zero phase only when the component LRD parameters are all identical.

Instead of (2.53), another possibility is to look for coefficients ψ_m such that

$$\sum_{m=0}^{\infty} \psi_m e^{-im\lambda} \sim c\lambda^{-d} e^{ia(\lambda)}, \tag{2.54}$$

where $a(\lambda) \to \infty$, as $\lambda \to 0^+$. Moreover, $a(\lambda)$ should be flexible enough in its relation to d. The idea here is that the complex-valued terms $e^{ia(\lambda)}$ associated with the two Fourier series would cancel out on the left-hand side of (2.52). In fact, coefficients whose Fourier transform behaves as (2.54) have already been studied by Wainger (1965).

For example, adapting the arguments of Wainger (1965), we show in Appendix A.2 that

$$\sum_{n=0}^{\infty} n^{-b} \cos(2\pi n^a) e^{2\pi i n x} \sim c_{a,b} x^{-d} e^{-i(\xi_a x^{-\frac{a}{1-a}} + \psi)}, \tag{2.55}$$

as $x \to 0^+$, where $0 < a < 1$, $0 < b \leq 1 - \frac{a}{2}$, $\psi = -\frac{\pi}{4}$, $c_{a,b}$ and ξ_a are two non-zero real constants and

$$d = \frac{1 - b - \frac{a}{2}}{1 - a}. \tag{2.56}$$

The next proposition, Proposition 2.4.1, uses (2.55) to construct multivariate LRD series with zero phases. Before stating the proposition, we shed some light on (2.55)–(2.56) and one further assumption to be made.

Note that, viewing d in (2.56) as a LRD parameter, we need $d < 1/2$ which translates to

$$\frac{1}{2} < b. \tag{2.57}$$

This additional assumption will be made in the proposition below. Under (2.57), the coefficients $n^{-b}\cos(2\pi n^a)$ are also square-summable and thus can be used to define linear time series. When (2.57) holds, observe that

$$d = \frac{1 - b - \frac{a}{2}}{1 - a} < 1 - b =: d_0, \tag{2.58}$$

where d_0 corresponds to the value of d in (2.56) when formally setting $a = 0$ in the left-hand side of (2.55) (see Lemma A.1). Moreover, when $a_1 < a_2$ (and (2.57) holds),

$$\frac{1 - b - \frac{a_1}{2}}{1 - a_1} > \frac{1 - b - \frac{a_2}{2}}{1 - a_2}. \tag{2.59}$$

Thus, viewing d as a LRD parameter, it decreases from $d_0 = 1-b$ associated with power-law coefficients as a increases.

Proposition 2.4.1. *Let $\{\epsilon_n\}_{n\in\mathbb{Z}}$ be an $\mathbb{R}^p$-valued white noise, satisfying $\mathbb{E}\epsilon_n = 0$ and $\mathbb{E}\epsilon_n\epsilon_n' = I$. Let also $\{\Psi_m = (\psi_{jk,m})_{j,k=1,\ldots,p}\}_{m\geq 0}$ be a sequence of real-valued matrices such that*

$$\psi_{jk,m} = \alpha_{jk} m^{-b_j} \cos(2\pi m^a), \quad m \geq 0, \tag{2.60}$$

where $\alpha_{jk} \in \mathbb{R}$, $0 < a < 1$, $\frac{1}{2} < b_j \leq 1 - \frac{1}{2}a$, $j = 1, \ldots, p$. Then, the spectral density of the time series

$$X_n = \sum_{m=0}^{\infty} \Psi_m \epsilon_{n-m}, \tag{2.61}$$

satisfies (2.17)–(2.18) with

$$d_j = \frac{1 - b_j - \frac{a}{2}}{1 - a}, \tag{2.62}$$

$$G_{jk} = (2\pi)^{d_j + d_k - 1} c_{a,b_j} c_{a,b_k} (AA^*)_{jk}, \tag{2.63}$$

where $A = (\alpha_{jk})_{j,k=1,\ldots,p}$, $c_{a,b}$ is a non-zero real constant given in Theorem A.1, and hence the phase parameters

$$\phi_{jk} = 0.$$

When the cosine in (2.60) is replaced by the sine, the statements above continue to hold. In particular, when $G_{jj} \neq 0, j = 1, \ldots, p$, the series is LRD in the sense of Definition 2.2.

Proposition 2.4.1 is proved in Appendix A.1.

2.4.3 The case of general phases

We showed in Proposition 2.4.1 that trigonometric power-law coefficients lead to zero phases for multivariate LRD series. The next result shows that linear combinations of trigonometric power-law coefficients can lead to general phases (see also the discussion following Proposition 2.4.2).

Proposition 2.4.2. *Let* $\{\epsilon_n\}_{n\in\mathbb{Z}}$ *be as in Proposition 2.4.1. Let also* $\{\Psi_m = (\psi_{jk,m})_{j,k=1,\ldots,p}\}_{m\in\mathbb{Z}}$ *be a sequence of real-valued matrices such that*

$$\psi_{jk,m} = \alpha_{jk} m^{-b_j} \cos(2\pi m^a) + \beta_{jk} m^{-b_j} \sin(2\pi m^a), \quad m \geq 0, \tag{2.64}$$

where $\alpha_{jk}, \beta_{jk} \in \mathbb{R}$, $0 < a < 1$, $\frac{1}{2} < b_j \leq 1 - \frac{1}{2}a$, $j = 1, \ldots, p$. *Then, the spectral density of the time series*

$$X_n = \sum_{m=0}^{\infty} \Psi_m \epsilon_{n-m} \tag{2.65}$$

satisfies (2.17)–(2.18) with

$$d_j = \frac{1 - b_j - \frac{a}{2}}{1 - a}, \tag{2.66}$$

$$G_{jk} = (2\pi)^{d_j + d_k - 1} c_{a,b_j} c_{a,b_k} \sum_{t=1}^{p} \overline{z}_{jt} z_{kt}, \tag{2.67}$$

where $z_{jk} = \alpha_{jk} + i\beta_{jk}$, *and* $c_{a,b}$ *is a non-zero real constant given in Theorem A.1. In particular, when* $G_{jj} \neq 0, j = 1, \ldots, p$, *the series is LRD in the sense of Definition 2.2.*

Proposition 2.4.2 is proved in Appendix A.1.

Note also that the coefficients $\psi_{jk,m}$ in (2.64) can be expressed as (supposing, for example, $\alpha_{jk} \neq 0$)

$$\begin{aligned}
\psi_{jk,m} &= \alpha_{jk}\left(\cos(2\pi m^a) + \frac{\beta_{jk}}{\alpha_{jk}}\sin(2\pi m^a)\right) m^{-b_j} \\
&= a_{jk}\sqrt{1+\beta_{jk}^2/\alpha_{jk}^2}\left(\frac{1}{\sqrt{1+\beta_{jk}^2/\alpha_{jk}^2}}\cos(2\pi m^a) + \frac{\beta_{jk}/a_{jk}}{\sqrt{1+\beta_{jk}^2/\alpha_{jk}^2}}\sin(2\pi m^a)\right) m^{-b_j} \\
&= \tau_{jk}\left(\cos(\theta_{jk})\cos(2\pi m^a) - \sin(\theta_{jk})\sin(2\pi m^a)\right) m^{-b_j} \\
&= \tau_{jk} m^{-b_j}\cos(2\pi m^a + \theta_{jk}), \qquad (2.68)
\end{aligned}$$

where $\tau_{jk} = a_{jk}\sqrt{1+\beta_{jk}^2/\alpha_{jk}^2}$ and $\theta_{jk} \in (-\pi/2, \pi/2)$ is such that $\tan(\theta_{jk}) = -\beta_{jk}/\alpha_{jk}$.

Note that the matrix $G = (G_{jk})_{j,k=1,\ldots,p}$ with the entries (2.67) can be written as

$$G = ZZ^*,$$

where $Z = ((2\pi)^{d_j} c_{a,b_j}\overline{z}_{jk})_{j,k=1,\ldots,p}$. Any Hermitian symmetric, non-negative definite G can be written as $G = WW^*$ and α_{jk}, β_{jk} can be found by setting $Z = W$.

2.5 Multivariate FARIMA$(0, D, 0)$ series

We provide here a multivariate, non-causal extension of FARIMA series. Let $D = \mathrm{diag}(d_1, \ldots, d_p)$ with $d_j < 1/2$, $j = 1, \ldots, p$, and $Q_+ = (q_{jk}^+)$, $Q_- = (q_{jk}^-) \in \mathbb{R}^{p\times p}$. Let also $\{\epsilon_n\}_{n\in\mathbb{Z}}$ be an $\mathbb{R}^p$-valued white noise series, satisfying $\mathbb{E}\epsilon_n = 0$ and $\mathbb{E}\epsilon_n\epsilon_n' = I$. Define a *multivariate FARIMA*$(0, D, 0)$ *series* as

$$X_n = (I-B)^{-D}Q_+\epsilon_n \;+\; (I-B^{-1})^{-D}Q_-\epsilon_n, \qquad (2.69)$$

where B is the backshift operator. The series X_n is thus given by a non-causal linear representation (when Q_- is not identically zero). In the next result, we give the exact form of the autocovariance matrix function of the multivariate FARIMA$(0, D, 0)$ series in (2.69).

Proposition 2.5.1. *The (j,k) component $\gamma_{jk}(n)$ of the autocovariance matrix function $\gamma(n)$ of the multivariate FARIMA$(0,D,0)$ series in (2.69) is given by*

$$\gamma_{jk}(n) = \frac{1}{2\pi}\left(b^1_{jk}\gamma_{1,jk}(n) + b^2_{jk}\gamma_{2,jk}(n) + b^3_{jk}\gamma_{3,jk}(n) + b^4_{jk}\gamma_{4,jk}(n)\right), \tag{2.70}$$

where

$$\begin{aligned} b^1_{jk} &= \sum_{t=1}^{p} q^-_{jt}q^-_{jt} = (Q^-(Q^-)^*)_{jk}, \quad b^3_{jk} = \sum_{t=1}^{p} q^+_{jt}q^+_{jt} = (Q^+(Q^+)^*)_{jk}, \\ b^2_{jk} &= \sum_{t=1}^{p} q^-_{jt}q^+_{jt} = (Q^-(Q^+)^*)_{jk}, \quad b^4_{jk} = \sum_{t=1}^{p} q^+_{jt}q^-_{jt} = (Q^+(Q^-)^*)_{jk}, \end{aligned} \tag{2.71}$$

and

$$\begin{aligned} \gamma_{1,jk}(n) &= \gamma_{3,kj}(n) = 2\Gamma(1-d_j-d_k)\sin(\pi d_k)\tfrac{\Gamma(n+d_k)}{\Gamma(n+1-d_j)}, \\ \gamma_{4,jk}(n) &= \gamma_{2,jk}(-n) = \begin{cases} 2\pi\frac{1}{\Gamma(d_j+d_k)}\frac{\Gamma(d_j+d_k+n)}{\Gamma(1+n)} & , \quad n = 0,\ 1, 2, \ldots, \\ 0 & , \quad n = -1,\ -2, \ldots \end{cases} \end{aligned} \tag{2.72}$$

Proposition 2.5.1 is proved in Appendix A.1.

Remark 2.5.1. Since $(I-B)^{-d} = \sum_{j=0}^{\infty} b_j B^j$ with $b_j = j^{d-1}/\Gamma(d)$, as $j \to \infty$, observe that the FARIMA$(0,D,0)$ series satisfies (2.34) with

$$A^+ = \Gamma(D)^{-1}Q^+, \quad A^- = \Gamma(D)^{-1}Q^-, \tag{2.73}$$

where $\Gamma(D)^{-1} = \mathrm{diag}(\Gamma(d_1)^{-1}, \ldots, \Gamma(d_p)^{-1})$. By using $\Gamma(j+a)/\Gamma(j+b) \sim j^{a-b}$, as $j \to \infty$, observe also that, as $n \to \infty$,

$$\gamma_{1,jk}(n) \sim 2\Gamma(1-d_j-d_k)\sin(\pi d_j)\frac{\Gamma(n+d_j)}{\Gamma(n+1-d_k)} = \frac{2\pi\sin(\pi d_j)}{\Gamma(d_j+d_k)\sin(\pi(d_j+d_k))}n^{d_j+d_k-1}, \tag{2.74}$$

$$\gamma_{4,jk}(n) \sim \frac{2\pi}{\Gamma(d_j+d_k)}n^{d_j+d_k-1}. \tag{2.75}$$

The relations (2.73)–(2.75) show that (2.70)–(2.72) are consistent with (2.37)–(2.38).

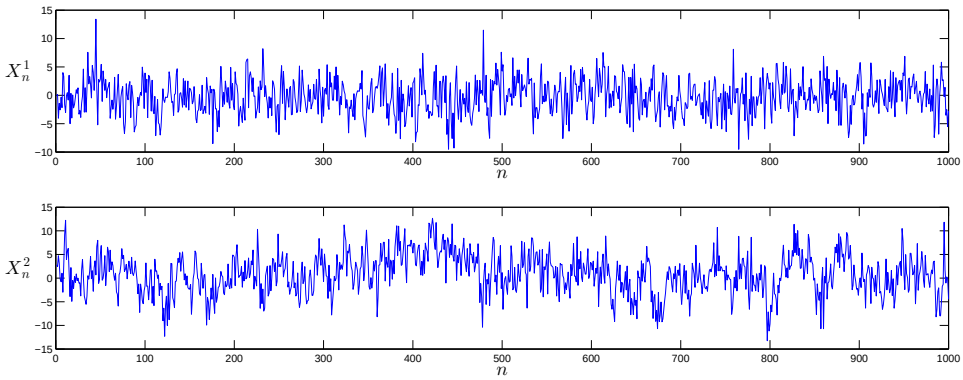

Figure 2.1: Components of bivariate FARIMA$(0, D, 0)$ series where $D = \text{diag}(0.2, 0.4)$, and $\phi = 1.4587$.

The exact form of the autovovariance function in (2.70)–(2.72) can be used, for example, in a fast generation of the Gaussian FARIMA$(0, D, 0)$ series by using a circulant matrix embedding method (see Helgason et al. (2011)). For example, Figure 1 presents the plots of a bivariate FARIMA$(0, D, 0)$ series with $\phi = 1.4587$ and

$$Q^+ = \begin{pmatrix} 0.50246 & 0 \\ 0 & 1.2436 \end{pmatrix}, \quad Q^- = \begin{pmatrix} 0 & -1.8878 \\ 3.4191 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.4 \end{pmatrix}.$$

The phase parameter ϕ above is computed using the relation (2.22), where the (j, k) entry of the matrix G is given by (2.44) with the matrices A^+ and A^- appearing in the relation (2.73).

2.6 Conclusions

In this chapter, we studied the definitions, representations and models of multivariate long-range dependent time series. Particular quantities of interest were phase parameters appearing in the cross spectra which control the (a)symmetry properties at large lags among the components of a multivariate LRD series. The major theoretical contribution was the construction of causal (one-sided) linear representations of multivariate LRD series with general phase parameters, based on the trigonometric power-law coefficients.

The results of Chapter 2 raise a number of questions left for future work. For one possible future direction, note that Definition 2.2 can readily be extended to consider the fundamental range $d_j \in (-1/2, 1/2)$, not just $d_j \in (0, 1/2)$. Examples of such models are FARIMA$(0, D, 0)$ series considered in Section 2.5, which even allow $d_j \in (-\infty, 1/2)$. The case $d_j \in (-1/2, 0)$ corresponds to antipersistent series. In particular, according to the extended Definition 2.2, the cross spectra of antipersistent series could involve phase parameters as in the case of LRD series.

Proposition 2.2.1 could likely be extended to cover the mixed case $d_j \in (-1/2, 0)$ or $(0, 1/2)$. But note that Definition 2.1 would have to be modified at least in part (i) of the proposition since (2.15)–(2.16) need not imply the condition $\sum_{n=-\infty}^{\infty} \gamma_{jj}(n) = 0$ if the jth component series is antipersistent. A more challenging open problem is to produce causal representations in the mixed case. It is yet to be seen whether the trigonometric power-law coefficients can be used in this regard. (Note that the key Theorem A.1 yields $d > 0$ only.)

In another direction, it would be interesting to study models with poles at other frequencies than the origin. In the univariate case, the well-known examples are the so-called Gegenbauer processes (e.g. Beran et al. (2013)). What are natural extensions of these processes in the multivariate case? What is an appropriate semiparametric specification for the cross spectra? Finally, estimation issues will be considered in Chapter 3 and also in future work.

CHAPTER 3

Inference and applications of a bivariate long-range dependent time series model with general phase

3.1 Introduction

In this chapter, we are interested in modeling bivariate ($\mathbb{R}^2$–vector) time series exhibiting long-range dependence (LRD, in short). In the univariate case, long-range dependent (LRD) time series models are stationary with the autocovariance function decaying slowly like a power-law function at large lags, or the spectral density diverging like a power-law function at the zero frequency. The univariate LRD is understood well in theory and used widely in applications. See, for example, Park and Willinger (2000), Robinson (2003), Doukhan et al. (2003), Palma (2007), Giraitis et al. (2012), Beran et al. (2013), Pipiras and Taqqu (2015).

Bivariate and, more generally, multivariate (vector-valued) LRD time series models have also been considered by a number of researchers. But theoretical foundations for a general class of such models were laid only recently in Kechagias and Pipiras (2015) or Chapter 2 of this thesis. In particular, Kechagias and Pipiras (2015) stressed the importance of the so-called phase parameters. Turning to the bivariate case which is the focus of this chapter, the phase ϕ appears in the cross spectrum of a bivariate LRD series around the zero frequency and controls the (a)symmetry of the series at large time lags. There are currently no parametric models of bivariate LRD with general phase that can be used in inference and applications. The goal of this chapter is to introduce such a class of models, and to examine it through a simulation study and an application to real data. In the rest of this section, we describe our contributions in greater detail.

A common parametric VARFIMA$(0, D, 0)$ (Vector Autoregressive Fractionally Integrated Moving Average) model for a bivariate LRD time series $\{X_n\}_{n\in\mathbb{Z}}=\{(X_n^1, X_n^2)'\}_{n\in\mathbb{Z}}$ is obtained as a natural extension of the univariate ARFIMA$(0, d, 0)$ model by fractionally

integrating the component series of a bivariate white noise series, namely,

$$X_n = \begin{pmatrix} X_n^1 \\ X_n^2 \end{pmatrix} = \begin{pmatrix} (I-B)^{-d_1} & 0 \\ 0 & (I-B)^{-d_2} \end{pmatrix} \begin{pmatrix} \eta_n^1 \\ \eta_n^2 \end{pmatrix} = (I-B)^{-D}\eta_n, \tag{3.1}$$

where B is the backshift operator, $d_1, d_2 \in (0, 1/2)$ are the LRD parameters of the component series $\{X_n^1\}_{n\in\mathbb{Z}}$ and $\{X_n^2\}_{n\in\mathbb{Z}}$, respectively, $D = \mathrm{diag}(d_1, d_2)$ and $\{\eta_n\}_{n\in\mathbb{Z}} = \{(\eta_n^1, \eta_n^2)'\}_{n\in\mathbb{Z}}$ is a bivariate white noise series with zero mean $\mathbb{E}\eta_n = 0$ and covariance $\mathbb{E}\eta_n\eta_n' = \Sigma$. If $\Sigma = QQ'$, note that the model (3.1) can also be written as

$$X_n = (I-B)^{-D}Q\epsilon_n \quad \text{or} \quad (I-B)^D X_n = \eta_n = Q\epsilon_n, \tag{3.2}$$

where $\{\epsilon_n\}_{n\in\mathbb{Z}}$ is a bivariate white noise with the identity covariance matrix $\mathbb{E}\epsilon_n\epsilon_n' = I$.

The model (3.1) admits a *causal* (one-sided) linear representation of the form

$$X_n = \sum_{k\in I} \widetilde{\Psi}_k \eta_{n-k} = \sum_{k\in I} \Psi_k \epsilon_{n-k}, \tag{3.3}$$

where $I = \{n \in \mathbb{Z} : n \geq 0\}$ and $\widetilde{\Psi}_k$, Ψ_k are real-valued 2×2 matrices whose entries decay as a power law. In the frequency domain, the matrix-valued spectral density function[1] $f(\lambda)$ of the series X_n defined in (3.1) satisfies

$$f(\lambda) \sim \begin{pmatrix} \omega_{11}|\lambda|^{-2d_1} & \omega_{12}|\lambda|^{-(d_1+d_2)}e^{-i\mathrm{sign}(\lambda)\phi} \\ \omega_{21}|\lambda|^{-(d_1+d_2)}e^{i\mathrm{sign}(\lambda)\phi} & \omega_{22}|\lambda|^{-2d_2} \end{pmatrix}, \quad \text{as } \lambda \to 0, \tag{3.4}$$

where $\omega_{11}, \omega_{12}, \omega_{21}, \omega_{22} \in \mathbb{R}$ and

$$\phi = (d_1 - d_2)\pi/2. \tag{3.5}$$

The asymptotic behavior (3.4) of the spectral density f with general $\phi \in (-\pi/2, \pi/2)$ is taken for the definition of bivariate LRD in Kechagias and Pipiras (2015). Note that $\phi \in (-\pi/2, \pi/2)$ is taken and the following polar coordinate representation $z = \frac{z_1}{\cos(\phi)}e^{i\phi}$

[1]The following conventions are used here. The autocovariance function γ is defined as $\gamma(n) = \mathbb{E}X_0X_n'$ and the spectral density $f(\lambda)$ satisfies $\gamma(n) = \int_{-\pi}^{\pi} e^{in\lambda}f(\lambda)d\lambda$.

with $\phi = \arctan(z_2/z_1)$ of $z = z_1 + iz_2 \in \mathbb{C}$ is used throughout. The special form of the phase parameter ϕ in (3.5) limits the type of bivariate LRD behavior that can be captured by the model (3.1). For example, in the case of time-reversible models satisfying $\gamma(n) = \gamma(-n), n \in \mathbb{Z}$, the spectral density matrix f has real-valued entries and hence $\phi = 0$. Under the model (3.1) and (3.4), however, $\phi = 0$ holds only when $d_1 = d_2$. These observations naturally raise the following question:

Can one define a bivariate parametric LRD model with general phase?

One solution to the question above is to consider *noncausal* (two-sided) linear representations with power-law decaying coefficients, that is, representations of the form (3.3) with the index set I now being the set of all integers $\mathbb{Z}$. Specifically, Kechagias and Pipiras (2015) constructed a noncausal VARFIMA$(0, D, 0)$ model with general phase by taking

$$X_n = (I - B)^{-D} Q_+ \epsilon_n + (I - B^{-1})^{-D} Q_- \epsilon_n, \tag{3.6}$$

where Q_+, Q_- are two real-valued 2×2 matrices. The reason we refer to (3.6) as noncausal is the presence of B^{-1} in the second term of the right-hand side of (3.6), which introduces dependence between X_n and the leads of the innovation process ϵ_n. Also, the positive and negative powers of the backshift operator motivate our notation for the subscripts of the matrices Q_+ and Q_-.

We shall use (3.6) in developing our parametric bivariate LRD model with general phase. A first issue that needs to be addressed is finding a suitable parametrization under which this model is identifiable, while still yielding a general phase parameter. We show in Section 3.2 that this two-fold goal can be achieved by taking Q_- as

$$Q_- = \begin{pmatrix} c & 0 \\ 0 & -c \end{pmatrix} Q_+ =: CQ_+, \tag{3.7}$$

for some real constant c. Under the relation (3.7) and letting $\{Z_n\}_{n\in\mathbb{Z}}$ be a zero mean bivariate white noise series with covariance matrix $\mathbb{E}Z_n Z_n^{'} = Q_+ Q_+^{'} =: \Sigma$, we can rewrite

(3.6) in the more compact form

$$X_n = \Delta_c(B)Z_n, \tag{3.8}$$

where the operator Δ_c is defined as

$$\Delta_c(B) = (I - B)^{-D} + (I - B^{-1})^{-D}C. \tag{3.9}$$

Note that when $c = 0$ (and $C = 0$), the filter $\Delta_c(B)$ becomes the causal (one-sided) fractional integration filter $\Delta_0(B) = (I - B)^{-D}$.

The focus of Chapter 3 will be on the model (3.8)–(3.9) and its extensions involving autoregressive and moving average parts, namely, a general phase VARFIMA(p, D, q) model

$$\Phi(B)\Delta_c(B)^{-1}X_n = \Theta(B)Z_n, \tag{3.10}$$

where $\Phi(B)$, $\Theta(B)$ are matrix polynomials of finite orders p and q satisfying the usual stationarity and invertibility conditions. In fact, for identifiability and estimation purposes, we shall work with diagonal AR filters $\Phi(B)$ in which the general phase VARFIMA(p, D, q) model (3.10) can also be expressed as

$$\Phi(B)X_n = \Delta_c(B)\Theta(B)Z_n, \tag{3.11}$$

since the diagonal filters $\Phi(B), \Delta_c(B)$ commute. The advantage of the model (3.11) is that the autocovariance function of the right-hand side of (3.11) can be computed efficiently. In inference, we can then employ a conditional likelihood approach where the Gaussian likelihood is written for $\Phi(B)X_n$ (though the maximum is still sought over all unknown parameters).

Our estimation follows the approach of Tsay (2010) who considered causal models (3.11) with $c = 0$. Still in the case $c = 0$, Sowell (1986) calculated (numerically) the autocovariance function of the model (3.10) and performed exact likelihood inference. For other approaches, (all in the case $c = 0$), see also Ravishanker and Ray (1997) who considered the Bayesian analysis and Pai and Ravishanker (2009*a*; 2009*b*) who employed the EM and

PCG algorithms as well as Dueker and Starz (1998), Martin and Wilkins (1999), Sela and Hurvich (2009) and Diongue (2010).

The rest of Chapter 3 is structured as follows. General phase VARFIMA$(0, D, 0)$ and VARFIMA(p, D, q) series are presented in Sections 3.2 and 3.3. Inference and other tasks are considered in Section 3.4. Section 3.5 contains a simulation study, and Section 3.6 contains an application to the U.S. inflation rates.

3.2 General phase VARFIMA$(0, D, 0)$ series

In this section, we consider the noncausal bivariate VARFIMA$(0, D, 0)$ model (3.8)–(3.9). Kechagias and Pipiras (2015) showed that any phase parameter ϕ in (3.4) can be obtained with the model (3.6) for an appropriate choice of Q_+ and Q_-. However, letting the entries of these matrices take any real value, causes identifiability issues around the zero frequency, as the same phase parameter can be obtained by more than one choice of Q_+ and Q_-. Indeed, from the following simple counting perspective note that the specification (3.4) has 6 parameters ($d_1, d_2, \omega_{11}, \omega_{12}, \omega_{22}$ and ϕ) whereas the model (3.6) has 10 (d_1, d_2 and the entries of Q_+ and Q_-). One might naturally expect identifiability up to Q_+Q_+' and Q_-Q_-' but this still leaves the number of parameters at 8 (d_1, d_2 and the 6 different entries of Q_+Q_+' and Q_-Q_-'). In Proposition 3.2.1 and Corollary 3.2.1 below (see also the discussion following the latter), we show that the parameterization (3.7) addresses the identifiability and general phase issues. For one, note that the model (3.8)–(3.9) has the required number of parameters (d_1, d_2, c and the three different entries of $\Sigma = Q_+Q_+^{'}$).

Proposition 3.2.1. *Let $d_1, d_2 \in (0, 1/2)$ and Q_+ be a 2×2 matrix with real-valued entries. Let also $\{X_n^{(c)}\}_{n\in\mathbb{Z}}$ be a time series defined by* (3.8)–(3.9) *where $D = \mathrm{diag}(d_1, d_2)$. For any $\phi_c \in (-\pi/2, \pi/2)$, there exists a unique constant $c \in (-1, 1)$ such that the series $\{X_n^{(c)}\}_{n\in\mathbb{Z}}$ has the phase parameter $\phi = \phi_c$ in* (3.4). *Moreover, the constant c has a closed form given by*

$$c = \frac{2(a_1 + a_2) - \sqrt{\Delta}}{2(a_1 - a_2 + \tan(\phi_c)(1 + a_1 a_2))}, \tag{3.12}$$

where

$$a_1 = \tan\left(\frac{\pi d_1}{2}\right),\ a_2 = \tan\left(\frac{\pi d_2}{2}\right) \quad \textit{and} \quad \Delta = 16a_1a_2 + 4(1+a_1a_2)^2\tan^2(\phi_c). \quad (3.13)$$

PROOF: By using Theorem 11.8.3 in Brockwell and Davis (2009), the VARFIMA$(0, D, 0)$ series in (3.8)–(3.9) has a spectral density matrix

$$f(\lambda) = \frac{1}{2\pi}\Delta_c(e^{-i\lambda})\Sigma\Delta_c(e^{-i\lambda})^*, \quad (3.14)$$

where the superscript $*$ denotes the complex conjugate operation. From (3.7) and by using the fact that $1 - e^{\pm i\lambda} \sim \mp i\lambda$, as $\lambda \to 0$, we have

$$f(\lambda) \sim \left((i\lambda)^{-D} + (-i\lambda)^{-D}C\right)\Sigma\left((-i\lambda)^{-D} + C(i\lambda)^{-D}\right), \quad \text{as} \quad \lambda \to 0. \quad (3.15)$$

Next, by denoting $\Sigma = (\sigma_{jk})_{j,k=1,2}$ and using the relation $\pm i = e^{\pm i\pi/2}$, we get that the (j,k) element of the spectral density $f(\lambda)$ satisfies

$$f_{jk}(\lambda) \sim g_{jk}\lambda^{-(d_j+d_k)}, \quad \text{as} \quad \lambda \to 0^+, \quad (3.16)$$

where the complex constant g_{jk} is given by

$$g_{jk} = \frac{\sigma_{jk}}{2\pi}(e^{-i\pi d_j/2} + (-1)^{j+1}ce^{i\pi d_j/2}) \cdot (e^{i\pi d_k/2} + (-1)^{k+1}ce^{-i\pi d_k/2}). \quad (3.17)$$

Focusing on the $(1,2)$ element, and by using the polar-coordinate representation $z = \frac{z_1}{\cos(\phi)}e^{i\phi}$ of $z = z_1 + iz_2 \in \mathbb{C}$ with $\phi = \arctan(z_2/z_1)$ (see (2.20) above or the relation (20) in Kechagias and Pipiras (2015)), we have

$$\begin{aligned} g_{12} &= \frac{\sigma_{jk}}{2\pi}\left(\cos(\frac{\pi d_1}{2})(1+c) + i\sin(\frac{\pi d_1}{2})(c-1)\right) \cdot \left(\cos(\frac{\pi d_2}{2})(1-c) + i\sin(\frac{\pi d_2}{2})(1+c)\right) \\ &= \frac{\sigma_{jk}}{2\pi}\frac{\cos(\frac{\pi d_1}{2})\cos(\frac{\pi d_2}{2})}{\cos(\phi_{c,1})\cos(\phi_{c,2})}(1-c^2)e^{-i\phi_c}, \end{aligned} \quad (3.18)$$

where

$$\phi_c = -(\phi_{c,1} + \phi_{c,2}), \quad \phi_{c,1} = \arctan\left(a_1 \frac{c-1}{1+c}\right) \quad \text{and} \quad \phi_{c,2} = \arctan\left(a_2 \frac{1+c}{1-c}\right) \tag{3.19}$$

with a_1 and a_2 given in (3.13).

By using the arctangent addition formula $\arctan(u) + \arctan(v) = \arctan(\frac{u+v}{1-uv})$ for $uv < 1$ (in our case $uv = -a_1 a_2 < 0$), we can rewrite ϕ_c as

$$\phi_c = -\arctan\left(\frac{a_1 \frac{c-1}{1+c} + a_2 \frac{1+c}{1-c}}{1 + a_1 a_2}\right) =: h(c). \tag{3.20}$$

For all $d_1, d_2 \in (0, 1/2)$, the function $h : (-1, 1) \to (-\pi/2, \pi/2)$ is strictly decreasing (and therefore 1-1) and also satisfies

$$\lim_{c \downarrow -1} h(c) = \frac{\pi}{2}, \quad \lim_{c \uparrow 1} h(c) = -\frac{\pi}{2}.$$

Since h is continuous, it is also onto its range which completes the existence and uniqueness part of the proof.

To obtain the formula (3.12), we invert the relation (3.20) to get the quadratic equation

$$(a_1 - a_2 + \tan(\phi_c)(1 + a_1 a_2))c^2 - 2(a_1 + a_2)c + a_1 - a_2 - \tan(\phi_c)(1 + a_1 a_2) = 0, \tag{3.21}$$

whose discriminant Δ_c is given by

$$\Delta = 16 a_1 a_2 + 4(1 + a_1 a_2)^2 \tan^2(\phi_c)$$

and is is always positive. The solutions of (3.21) are then given by

$$c_1 = \frac{2(a_1 + a_2) + \sqrt{\Delta_c}}{2(a_1 - a_2 + \tan(\phi_c)(1 + a_1 a_2))}, \quad c_2 = \frac{2(a_1 + a_2) - \sqrt{\Delta_c}}{2(a_1 - a_2 + \tan(\phi_c)(1 + a_1 a_2))}.$$

It can be checked that $c_1 \notin (-1, 1)$ and $c_2 \in (-1, 1)$. $\square$

The following result is a direct consequence of the proof of Proposition 3.2.1.

Corollary 3.2.1. *The spectral density of the time series* $\{X_n^{(c)}\}_{n\in\mathbb{Z}}$ *in Proposition 3.2.1 satisfies the asymptotic relation* (3.4) *with* $\phi = \phi_c$ *and*

$$\omega_{jj} = \frac{\sigma_{jj}}{2\pi}(1 + c^2 + 2c\cos(\pi d_j)), \quad j = 1, 2, \tag{3.22}$$

$$\omega_{12} = \frac{\sigma_{12}}{2\pi}\frac{\cos(\frac{\pi d_1}{2})\cos(\frac{\pi d_2}{2})}{\cos(\phi_{c,1})\cos(\phi_{c,2})}(1 - c^2), \tag{3.23}$$

where $\Sigma = Q_+Q_+' = (\sigma_{jk})_{j,k=1,2}$ *and* $\phi_{c,1}, \phi_{c,2}$ *are given in* (3.19).

PROOF: The relations (3.22)–(3.23) follow from (3.16)–(3.17) and (3.18)–(3.19). □

Corollary 3.2.1 shows that the bivariate LRD model (3.8)–(3.9) is identifiable around the zero frequency when parametrized by $d_1, d_2, \Sigma = Q_+Q_+'$ and c.

Remark 3.2.1. Proposition 3.2.1 relates the phase ϕ at the zero frequency and the constant c which appears in the full model (3.8)–(3.9). For this model, however, the phase function $\phi_{12}(\lambda)$ of the full cross spectral density $f_{12}(\lambda) = g_{12}(\lambda)e^{-i\phi(\lambda)}, \lambda \in (0, \pi)$, is not a constant function of the frequency λ. Instead, by using the identity $1 - e^{\mp i\lambda} = 2\sin(\frac{\lambda}{2})e^{\mp i(\lambda-\pi)}$ and arguing as for (3.18)–(3.20) above, it can be shown that

$$\phi_{12}(\lambda) = -\arctan\left(\frac{x_1(\lambda)\frac{c-1}{1+c} + x_2(\lambda)\frac{1+c}{1-c}}{1 + x_1(\lambda)x_2(\lambda)}\right), \tag{3.24}$$

where

$$x_1(\lambda) = \tan\left(\frac{d_1(\pi - \lambda)}{2}\right), \quad x_2(\lambda) = \tan\left(\frac{d_2(\pi - \lambda)}{2}\right). \tag{3.25}$$

Several plots of the phase function (3.24) are given in Figure 3.1. We also note that Sela (2010) considers LRD models with phase functions $\phi_{12}(\lambda)$ following special power laws, but we will not expand in this direction.

Remark 3.2.2. The autocovariance function of the model (3.6) has an explicit form given in Proposition 5.1 of Kechagias and Pipiras (2015). The same form can obviously be used for the model (3.8)–(3.9).

Remark 3.2.3. We have tried several choices for Q_- other than (3.7), for example $Q_- = cQ_+$, but they did not lead to a general phase parameter for the resulting bivariate LRD

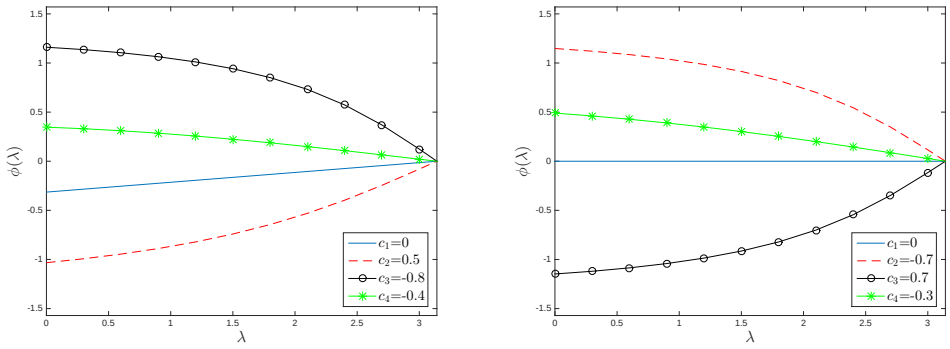

Figure 3.1: Phase functions $\phi(\lambda)$ for the model (3.8)–(3.9) for different parameter values. Left: $c = 0, 0.5, -0.8, -0.4$ and $d_1 = 0.2$, $d_2 = 0.4$. Right: $c = 0, -0.7, 0.7, -0.3$, $d_1 = d_2 = 0.3$.

models. A related note is that we presently do not have an explicit form for the inverse filter $\Delta_c(B)^{-1}$. But this observation also suggests that one could possibly work with the filter

$$\widetilde{\Delta}_c(B) = ((I - B)^D + C(I - B^{-1})^D)^{-1}, \tag{3.26}$$

if the goal is to have an explicit form of the filter $\widetilde{\Delta}_c(B)^{-1}$ applied to the series $\{X_n\}_{n\in\mathbb{Z}}$. Other filters than (3.9) and (3.26) with interesting properties might also exist and could also be considered. In using (3.9), we aimed to have a general phase and an explicit form of the autocovariance function.

Remark 3.2.4. We have assumed in Proposition 3.2.1 that both component series are LRD, that is, $d_1, d_2 \in (0, 1/2)$. The introduced model could also be used when one of the d's is 0, that is, one of the series is short-range dependent (SRD, for short).

Remark 3.2.5. We stress again that the case $c = 0$ corresponds to the phase $\phi = (d_1 - d_2)\pi/2$ (and in particular not necessarily $\phi = 0$.) Note also that if the two component series are interchanged (so that d_1 and d_2 are interchanged, and ϕ becomes $-\phi$), then the constant c in (3.12) changes to $-c$.

3.3 General phase VARFIMA(p, D, q) series

In this section, we generalize the model (3.8)–(3.9) by introducing autoregressive (AR, for short) and moving average (MA, for short) components to capture potential short-

range dependence effects. For the causal model (3.1), this extension has been achieved in a number of ways. Naturally, we focus on extensions that preserve the general phase and identifiability properties. We also consider the problem of computing (theoretically or numerically) the autocovariance functions of the introduced models, since these functions are used in estimation (see Section 3.4 below).

3.3.1 VARFIMA$(0, D, q)$ series

We begin with the case $p = 0$ (where there is no AR part). Define the *general phase VARFIMA*$(0, D, q)$ *series* as

$$Y_n = \Delta_c(B)\Theta(B)Z_n, \tag{3.27}$$

where $\Delta_c(B)$ is the operator given by (3.9) and

$$\Theta(B) = I_2 + \Theta_1 B + \ldots + \Theta_q B^q \tag{3.28}$$

is a matrix polynomial with 2×2 real-valued matrices $\Theta_s = (\theta_{jk,s})_{j,k=1,2}, s = 1, \ldots, q$. As throughout the chapter, $\{Z_n\}_{n\in\mathbb{Z}}$ is a white noise series with $\mathbb{E}Z_n Z_n^{'} = \Sigma = (\sigma_{jk})_{j,k=1,2}$. In the special case where $\Theta(B)$ is diagonal or when $d_1 = d_2$, the model (3.27) is equivalent to

$$Y_n = \Theta(B)\Delta_c(B)Z_n. \tag{3.29}$$

The two operators $\Delta_c(B)$ and $\Theta(B)$, however, do not commute in general. In fact, the two models in (3.27) and (3.29) are quite different. More specifically, if $\Theta(B)$ has at least one nonzero element on the off diagonal and if $d_1 \neq d_2$, the series $\{Y_n\}_{n\in\mathbb{Z}}$ in (3.29) can be thought to exhibit a form of fractional cointegration, by writing $\Theta(B)^{-1}Y_n = \Delta_c(B)Z_n$ where reduction of memory in one of the component series of $\{Y_n\}_{n\in\mathbb{Z}}$ could occur from linear combination of present and past variables of the two component series. On the other hand, fractional cointegration cannot occur under the model (3.27). In the rest of this chapter, we will restrict our attention to this simpler case, leaving the investigation of fractional cointegration for future work.

In the next proposition, we compute the autocovariance function of the series in (3.27). Tsay (2010) calculated the autocovariance function of the causal analogue of (3.27) using the properties of the hypergeometric function. Our approach, which we find less cumbersome for the multivariate case, is similar to the one used for the VARFIMA$(0, D, 0)$ series in Proposition 2.5.1 above or Proposition 5.1 of Kechagias and Pipiras (2015).

Proposition 3.3.1. *The (j,k) component $\gamma_{jk}(n)$ of the autocovariance matrix function $\gamma(n)$ of the bivariate VARFIMA$(0,D,q)$ series in (3.27) is given by*

$$\gamma_{jk}(n) = \frac{1}{2\pi} \sum_{u,v=1}^{2} \sum_{s,t=0}^{q} \theta_{ju,s}\theta_{kv,t}\sigma_{uv} \left(a_{1,jk}\gamma^{(1)}_{st,jk}(n) + a_{2,j}\gamma^{(2)}_{st,jk}(n) + \gamma^{(3)}_{st,jk}(n) + a_{4,k}\gamma^{(4)}_{st,jk}(n) \right), \tag{3.30}$$

where

$$a_{1,jk} = c^2(-1)^{j+k}, \quad a_{2,j} = c(-1)^{j+1}, \quad a_{4,k} = c(-1)^{k+1}, \tag{3.31}$$

and

$$\begin{array}{ccccc} \gamma^{(1)}_{st,jk}(n) & = & \gamma^{(3)}_{st,kj}(n) & = & 2\Gamma(1-d_j-d_k)\sin(\pi d_k)\frac{\Gamma(n+t-s+d_k)}{\Gamma(n+t-s+1-d_j)}, \\ \gamma^{(4)}_{st,jk}(n) & = & \gamma^{(2)}_{ts,jk}(-n) & = & \begin{cases} 2\pi\frac{1}{\Gamma(d_j+d_k)}\frac{\Gamma(d_j+d_k+n+t-s)}{\Gamma(1+n+t-s)} & , \quad n \geq s-t, \\ 0 & , \quad n < s-t. \end{cases} \end{array} \tag{3.32}$$

Proof: By using Theorem 11.8.3 in Brockwell and Davis (2009), the VARFIMA$(0,D,q)$ series in (3.27) has a spectral density matrix

$$f(\lambda) = \frac{1}{2\pi} G(\lambda)\Sigma G(\lambda)^*, \tag{3.33}$$

where $G(\lambda) = \Delta_c(e^{-i\lambda})\Theta(e^{-i\lambda})$. The (j,k) component of the spectral density is given by

$$f_{jk}(\lambda) = \frac{1}{2\pi} \sum_{u,v=1}^{2} \sum_{s,t=0}^{q} \theta_{ju,s}\theta_{kv,t}\sigma_{uv} e^{-i(s-t)\lambda} \left(f_{1,jk}(\lambda) + f_{2,jk}(\lambda) + f_{3,jk}(\lambda) + f_{4,jk}(\lambda) \right), \tag{3.34}$$

where

$$\begin{aligned} f_{1,jk}(\lambda) &= a_{1,jk}(1-e^{i\lambda})^{-d_j}(1-e^{-i\lambda})^{-d_k}, \quad & f_{2,jk}(\lambda) &= a_{2,j}(1-e^{i\lambda})^{-(d_j+d_k)}, \\ f_{3,jk}(\lambda) &= (1-e^{-i\lambda})^{-d_j}(1-e^{i\lambda})^{-d_k}, \quad & f_{4,jk}(\lambda) &= a_{4,k}(1-e^{-i\lambda})^{-(d_j+d_k)}. \end{aligned} \tag{3.35}$$

Consequently, the (j,k) component of the autocovariance matrix satisfies $\gamma_{jk}(n) = \int_0^{2\pi} e^{in\lambda} f_{jk}(\lambda) d\lambda$, which in view of the relations (3.34)–(3.35) implies (3.30)–(3.31) with

$$\gamma^{(1)}_{st,jk}(n) = \gamma^{(3)}_{st,kj}(n) = \int_0^{2\pi} e^{i(n-s+t)\lambda}(1-e^{i\lambda})^{-d_j}(1-e^{-i\lambda})^{-d_k} d\lambda,$$

$$\gamma^{(2)}_{st,jk}(n) = \int_0^{2\pi} e^{i(n-s+t)\lambda}(1-e^{i\lambda})^{-x_{jk}} d\lambda, \quad \gamma^{(4)}_{st,jk}(n) = \int_0^{2\pi} e^{i(n-s+t)\lambda}(1-e^{-i\lambda})^{-x_{jk}} d\lambda,$$

where $x_{jk} = d_j + d_k$. The relations (3.32) follow from the evaluation of the integrals above as in the proof of Proposition 2.5.1 above or Proposition 5.1 of Kechagias and Pipiras (2015). □

Remark 3.3.1. Since $\Theta(e^{-i\lambda}) \sim I_2 + \Theta_1 + \ldots + \Theta_q$ as $\lambda \to 0$, the VARFIMA$(0, D, q)$ model has general phase at the zero frequency, with the same relation (3.12) as in Proposition 3.2.1. The parameters of Θ are identifiable if and only if they are identifiable for the same VARMA$(0, q)$ model.

3.3.2 VARFIMA(p, D, q) series

We extend here the model (3.27) to a general phase fractionally integrated model containing both autoregressive and moving average components. As for the causal model (3.1), two possibilities can be considered for this extension. Let $\Phi(B) = I_2 - \Phi_1 B - \ldots - \Phi_p B^p$ be the AR polynomial, where $\Phi_r = (\phi_{jk,r})_{j,k=1,2}, r = 1, \ldots, p$, are 2×2 real-valued matrices. Following the terminology of Sela and Hurvich (2009), define the noncausal VARFIMA(p, D, q) series $\{X_n\}_{n\in\mathbb{Z}}$ as

$$\Phi(B) X_n = \Delta_c(B)\Theta(B) Z_n, \tag{3.36}$$

and the noncausal FIVARMA(p, D, q) series as

$$\Phi(B)\Delta_c(B)^{-1} X_n = \Theta(B) Z_n. \tag{3.37}$$

The causal FIVARMA(p, D, q) series have been more popular in the literature, with Lobato (1997), Sela and Hurvich (2009) and Tsay (2010) being notable exceptions. In particular, Sela and Hurvich (2009) investigated thoroughly the differences between the causal analogues of the models (3.36) and (3.37), focusing on models with no MA part. Excluding the MA part is natural as any VARMA(p, q) series can be written in a VAR(1) form. Such transformations, however, lead to higher dimensional models, where inference is not always computationally feasible.

As expected, the noncausal VARFIMA and FIVARMA series differ if $\Phi(B)$ is nondiagonal and if $d_1 \neq d_2$. Similarly to the discussion around the models (3.27) and (3.29), the VARFIMA model with nondiagonal $\Phi(B)$ allows for fractional cointegration in the sense discussed following the relation (3.29), which however cannot be produced by the FIVARMA model (3.37) (see Sela and Hurvich (2009) for more details in the causal case). As indicated earlier, the case of fractional cointegration will be pursued elsewhere (though we shall also briefly mention some numerical results in Section 3.5).

We will focus on the VARFIMA(p, D, q) series (3.36) with a *diagonal AR part*, in which case the two models (3.36) and (3.37) are equivalent. Besides the obvious computational and simplification advantages of this approach, our consideration is also justified by similar assumptions recently used in Dufour and Pelletier (2011) for the construction of identifiable multivariate short-range dependent time series models. More specifically, Dufour and Pelletier (2011) show that any VARMA(p, q) series can be transformed to have a diagonal AR (or MA) part with the cost of increasing the order of the MA (or AR) component. As a consequence, they construct identifiable representations of VARMA(p, q) series where either AR or MA is diagonal.

There is yet another reason for making our assumption of diagonal AR part $\Phi(B)$. By using the reparametrizations of Dufour and Pelletier (2011), the FIVARMA model (3.37) can take the form (3.36) with diagonal $\Phi(B)$. Indeed, inverting the fractional operator $\Delta_c(B)$, the model (3.37) becomes

$$\Delta_c(B)^{-1} X_n = \Phi(B)^{-1} \Theta(B) Z_n.$$

Next, by using the relation $\Phi(B)^{-1} = |\Phi(B)|^{-1}\text{adj}(\Phi(B))$, where $|\cdot|$ and $\text{adj}(\cdot)$ denote the determinant and adjoint of a matrix respectively, we can write

$$\Delta_c(B)^{-1}|\Phi(B)|X_n = \text{adj}(\Phi(B))\Theta(B)Z_n, \tag{3.38}$$

where the commutation of $\Delta_c(B)^{-1}$ and $|\Phi(B)|$ is possible since $|\Phi(B)|$ is scalar-valued. Letting $\widetilde{\Phi}(B) = \text{diag}(|\Phi(B)|)$ and $\widetilde{\Theta}(B) = \text{adj}(\Phi(B))\Theta(B)$, the relation (3.38) yields

$$\Delta_c(B)^{-1}\widetilde{\Phi}(B)X_n = \widetilde{\Theta}(B)Z_n. \tag{3.39}$$

Thus, a FIVARMA model with AR component of order p can indeed be written as a VARFIMA model with diagonal AR part whose order will not exceed $2p$ (maximum possible order of $|\Phi(B)|$).

The presence of the AR filter on the left-hand side of (3.36) makes it difficult to compute the autocovariance function of the process explicitly. Closed form formulas for the autocovariance function of the causal model (3.37) with $c = 0$ were provided by Sowell (1986), albeit their implementation is computationally inefficient as it requires multiple expensive evaluations of hypergeometric functions. The slow performance of Sowell's approach was also noted by Sela (2010), who proposed fast approximate algorithms for calculating the autocovariance functions of the causal models (3.36) and (3.37) with $c = 0$ when $p = 1$ and $q = 0$. Although not exact, Sela's algorithms are fast with negligible approximation errors. In fact, it is straightforward to extend these algorithms to calculate the autocovariance function of a noncausal VARFIMA$(1, D, q)$ series. For models with AR components of higher orders, however, this extension seems to require restrictive assumptions on the AR coefficients and therefore we do not pursue this approach.

3.4 Inference and other tasks

In this section, we discuss the statistical inference of the VARFIMA(p, D, q) model (3.36) introduced in Section 3.3.2. Estimation of the parameters of this model can be carried out by adapting suitably the CLDL (Conditional Likelihood Durbin Levinson) estimation of Tsay

(2010). Tsay's method is appealing in our case for a number of reasons. First, as discussed in Section 3.4.1 below, the method requires only the knowledge of the autocovariance function of the VARFIMA$(0, D, q)$ series (3.27) for which we have an explicit form. Second, Tsay's algorithm can be modified easily to yield multiple steps-ahead (finite sample) predictions of the series. Finally, Tsay's method has a mild computational cost, compared to most alternative estimation methods.

3.4.1 Estimation

The basic idea of Tsay's CLDL algorithm is to transform a VARFIMA(p, D, q) series to a VARFIMA$(0, D, q)$ series whose autocovariance function is known. Then, a straightforward implementation of the well-known Durbin-Levinson (DL, for short) algorithm allows one to replace the computationally expensive likelihood calculations of the determinant and the quadratic part with less time consuming operations. Below we give a brief review of the algorithm.

We start with some notation. Let $\{Y_n\}_{n=1,\ldots,N}$ be the VARFIMA$(0, D, q)$ series (3.27) and let $\Gamma(n) = \mathbb{E}Y_0Y_h^{'}$ denote its autocovariance function. Let also $\Theta = (\mathrm{vec}(\Theta_1)', \ldots, \mathrm{vec}(\Theta_q)')'$ be the vector containing the entries of the coefficient matrices of the MA polynomial $\Theta(B)$. Assuming that the bivariate white noise series $\{Z_n\}$ is Gaussian, we can express the likelihood function of $\{Y_n\}_{n=1,\ldots,N}$ with the aid of the multivariate DL algorithm (see Brockwell and Davis (2009), p. 422). More specifically, letting $\theta = (d_1, d_2, c, \sigma_{11}, \sigma_{12}, \sigma_{22}, \Theta')'$ be the $(6+4q)$–dimensional vector containing all the parameters of the model (3.27), we can write the likelihood function as

$$L(\theta; Y) = (2\pi)^{-N} \left(\prod_{j=0}^{N-1} V_j \right)^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{j=0}^{N-1} (Y_{j+1} - \widehat{Y}_{j+1})' V_j^{-1} (Y_{j+1} - \widehat{Y}_{j+1}) \right\}, \quad (3.40)$$

where $\widehat{Y}_{j+1} := \mathbb{E}(Y_{j+1}|Y_1, \ldots, Y_j)$ and $V_j, j = 0, \ldots, N-1$, are the one-step-ahead finite sample predictors and their corresponding error covariance matrices obtained by the multivariate DL algorithm. Using the fact that the series $\{Y_n\}_{n=1,\ldots,N}$ satisfies the relation

$$\Phi(B) X_n = Y_n, \quad (3.41)$$

where $\{X_n\}_{n=1,\ldots,N}$ is the VARFIMA(p, D, q) series (3.36), we can view $\{\Phi(B)X_n\}_{n=p+1,\ldots,N}$ as a VARFIMA$(0, D, q)$ process, whose likelihood function conditional on $X_1, \ldots, X_p$ and $\Phi = (\text{vec}(\Phi_1)', \ldots, \text{vec}(\Phi_p)')'$ is given by

$$L(\Phi, \theta; X_n | X_1, \ldots, X_p) \equiv L(\theta; \Phi(B)X_n), \quad n = p+1, \ldots, N. \tag{3.42}$$

The reason we do not absorb Φ in θ, is to emphasize the different roles that these two parameters have in calculating the likelihood function in (3.42). More specifically, Φ is used to transform the available data $\{X_n\}_{n=1,\ldots,N,}$ to a VARFIMA$(0, D, q)$ series $\{Y_n\}_{n=1,\ldots,N,}$ while θ is necessary to apply the DL algorithm.

The conditional likelihood estimators of Φ and θ are then given by

$$(\widehat{\Phi}, \widehat{\theta}) = \underset{\Phi, \theta \in S}{\text{argmax}}\ L(\Phi, \theta; X_n | X_1, \ldots, X_p), \tag{3.43}$$

where $S = \{\theta \in \mathbb{R}^{6+4q} : 0 < d_1, d_2 < 0.5,\ -1 < c < 1,\ |\Sigma| > 0,\ (\Sigma_{jj})_{j=1,2} \geq 0\}$ denotes the parameter space for θ. Although there is no closed form for the estimates $\widehat{\Phi}$ and $\widehat{\theta}$, they can be computed numerically using the quasi-Newton algorithm of Broyden, Fletcher, Goldfarb, and Shanno (BFGS).

Remark 3.4.1. The finite sample predictors $\widehat{Y}_{j+1}$ and their corresponding error covariance matrices $V_j, j = 1, \ldots, N$, in (3.42) can also be computed using the multivariate Innovations (IN, for short) algorithm as $\widehat{Y}_1 = 0$ by convention and

$$\widehat{Y}_{j+1} = \sum_{k=1}^{j} \Theta_{jk}(Y_{j+1-k} - \widehat{Y}_{j+1-k}), \quad j \geq 1, \tag{3.44}$$

$$V_j = \Gamma(0) - \sum_{k=0}^{j-1} \Theta_{j,j-k} V_j \Theta'_{j,j-k}, \tag{3.45}$$

where $\Gamma(n)$ is the autocovariance function of $\{Y_n\}_{n \in \mathbb{Z}}$ at lag n and the matrix-valued coefficients $\Theta_{jk}, k = 1, \ldots, j$, are given by the relation (11.4.23) in Brockwell and Davis (2009). The multivariate DL and IN algorithms have computational complexities $O(N^2)$ and $O(N^3)$ respectively. As argued in Section 3.4.2 below, however, the latter is more suitable for

forecasting purposes and hence we will use it instead of the DL algorithm for prediction purposes.

3.4.2 Forecasting

In contrast to the DL algorithm, the IN algorithm also provides an efficient way to compute the h–step-ahead finite sample predictors $\widehat{Y}_{n+h|n} := \mathbb{E}(Y_{n+h}|Y_1, \ldots, Y_n), h = 1, 2, \ldots,$ and the corresponding prediction error matrices $V_{n+h-1} = \mathbb{E}(Y_{n+h} - \widehat{Y}_{n+h|n})(Y_{n+h} - \widehat{Y}_{n+h|n})'$ of a process $\{Y_n\}_{n\in\mathbb{Z}}$ whose autocovariance function is known. More specifically, once the one-step-ahead predictors $\widehat{Y}_1, \ldots, \widehat{Y}_n$ and the matrix coefficients $\Theta_{n,n-k}, k = 0, \ldots, n-1,$ are computed by the IN algorithm, $\widehat{Y}_{n+h|n}, h = 1, 2, \ldots,$ and V_{n+h-1} can be calculated as

$$\widehat{Y}_{n+h|n} = \sum_{k=1}^{n} \Theta_{n+h-1,n+h-k}(Y_k - \widehat{Y}_k), \tag{3.46}$$

$$V_{n+h-1} = \Gamma(0) - \sum_{k=1}^{n} \Theta_{n+h-1,k+h-1} V_{n-k} \Theta'_{n+h-1,k+h-1} \tag{3.47}$$

(see Brockwell and Davis (2009), pp. 174-175, for the univariate analogues of the relations (3.46)–(3.47)).

We now turn our attention to the VARFIMA(p, D, q) series X_n defined in (3.41). As we do not have an explicit form of the autocovariance function of $\{X_n\}_{n\in\mathbb{Z}}$, it is not immediately clear how to calculate the h–step-ahead predictors $\widehat{X}_{n+h|n} := \mathbb{E}(X_{n+h}|X_1, \ldots, X_n), h = 1, 2, \ldots,$ and the corresponding error matrices $\widetilde{V}_{n+h-1} = \mathbb{E}(X_{n+h} - \widehat{X}_{n+h|n})(X_{n+h} - \widehat{X}_{n+h|n})'$. In the next proposition, we show that $\widehat{X}_{n+h|n}$ and $\widetilde{V}_{n+h-1}$ can be calculated recursively from $\widehat{Y}_{n+h|n}$ and V_{n+h-1}. For simplicity we focus on the case $p = 1$. However, the proposition can be extended for larger values of p.

Proposition 3.4.1. *Let $\{Y_n\}_{n\in\mathbb{Z}}$ be the VARFIMA$(0, D, q)$ series given in (3.27) with autocovariance function $\Gamma(n)$, and let $\Theta_{nk}, V_k, k = 1, \ldots, n$, be the coefficients and prediction error matrices obtained by the IN algorithm. Let also $\{X_n\}_{n\in\mathbb{Z}}$ be the VARFIMA(p, D, q) series defined in (3.41), where $\Phi(B)$ is an AR polynomial of order $p = 1$, and suppose for simplicity that X_1 and Y_1 are uncorrelated with $Y_2, \ldots, Y_n$. Then, the h–step-ahead*

predictors $\widehat{X}_{n+h|n} = \mathbb{E}(X_{n+h}|X_1, \ldots, X_n)$ *satisfy*

$$\widehat{X}_{n+h|n} = \Phi_1^h X_n + \sum_{s=0}^{h-1} \Phi_1^s \widehat{Y}_{n+h-s|n}, \tag{3.48}$$

where $\widehat{Y}_{n+h|n} = \mathbb{E}(Y_{n+h}|Y_1, \ldots, Y_n)$. *Moreover, the prediction error matrices* $\widetilde{V}_{n+h-1} = \mathbb{E}(X_{n+h} - \widehat{X}_{n+h|n})(X_{n+h} - \widehat{X}_{n+h|n})'$ *can be computed by*

$$\widetilde{V}_{n+h-1} = \sum_{s=0}^{h-1} \Phi_1^s V_{n+h-s-1} (\Phi_1^s)' + \sum_{\substack{s,t=0 \\ s \neq t}}^{h-1} \Phi_1^s A_{s,t}(n+h)(\Phi_1^t)', \tag{3.49}$$

where

$$A_{s,t}(u) = \Gamma(t-s) - \sum_{k=1}^{n} \Theta_{u-s-1,u-s-k} V_{k-1} \Theta'_{u-t-1,u-t-k}. \tag{3.50}$$

PROOF: By using the relation (3.41) recursively, we can write

$$X_{n+h} = \Phi_1^h X_n + \sum_{s=0}^{h-1} \Phi_1^s Y_{n+h-s}, \quad h = 1, 2, \ldots \tag{3.51}$$

which implies that

$$\mathbb{E}(X_{n+h}|X_1, \ldots, X_n) = \Phi_1^h X_n + \sum_{s=0}^{h-1} \Phi_1^s \mathbb{E}(Y_{n+h-s}|X_1, \ldots, X_n). \tag{3.52}$$

Note that $\mathbb{E}(Y_{n+h-s}|X_1, \ldots, X_n) = \mathbb{E}(Y_{n+h-s}|Y_2, \ldots, Y_n) = \mathbb{E}(Y_{n+h-s}|Y_1, \ldots, Y_n) = \widehat{Y}_{n+h-s|n}$, since we assumed for simplicity that X_1 and Y_1 are uncorrelated with $Y_2, \ldots, Y_n$. Then, the relation (3.52) yields (3.48).

Next, we subtract (3.48) from (3.51) to get

$$X_{n+h} - \widehat{X}_{n+h|n} = \sum_{s=0}^{h-1} \Phi^s (Y_{n+h-s} - \widehat{Y}_{n+h-s|n}).$$

The h–step-ahead error matrix $\widetilde{V}_{n+h-1}$ is then given by

$$\begin{aligned}\widetilde{V}_{n+h-1} &= \mathbb{E}\left(\sum_{s=0}^{h-1}\Phi_1^s(Y_{n+h-s}-\widehat{Y}_{n+h-s|n})\right)\left(\sum_{t=0}^{h-1}\Phi_1^t(Y_{n+h-t}-\widehat{Y}_{n+h-t|n})\right)' \\ &= \sum_{s=0}^{h-1}\Phi_1^s V_{n+h-s-1}(\Phi_1^s)' + \sum_{\substack{s,t=0\\ s\neq t}}^{h-1}\Phi_1^s A_{s,t}(n+h)(\Phi_1^t)', \end{aligned} \tag{3.53}$$

where $A_{s,t}(u) = \mathbb{E}(Y_{u-s}-\widehat{Y}_{u-s|n})(Y_{u-t}-\widehat{Y}_{u-t|n})'$. To show that $A_{s,t}(u)$ satisfies (3.50) note that for $s,t = 0,\ldots,u-n-1, s\neq t$, we have

$$\mathbb{E}\widehat{Y}_{u-s|n}Y_{u-t}^{'} = \mathbb{E}(\mathbb{E}(\widehat{Y}_{u-s|n}Y_{u-t}^{'}|Y_1,\ldots,Y_n)) = \mathbb{E}\widehat{Y}_{u-s|n}\mathbb{E}(Y_{u-t}|Y_1,\ldots,Y_n) = \mathbb{E}\widehat{Y}_{u-s|n}\widehat{Y}_{u-t|n}^{'}.$$

Hence,

$$\begin{aligned} A_{t,s}(u) &= \mathbb{E}Y_{u-s}Y_{u-t}^{'} - \mathbb{E}Y_{u-s}\widehat{Y}_{u-t|n}^{'} - \mathbb{E}\widehat{Y}_{u-s|n}Y_{u-t}^{'} + \mathbb{E}\widehat{Y}_{u-s|n}\widehat{Y}_{u-t|n}^{'} \\ &= \Gamma(t-s) - \mathbb{E}\widehat{Y}_{u-s|n}\widehat{Y}_{u-t|n}^{'} \\ &= \Gamma(t-s) - \mathbb{E}\left(\sum_{k=1}^{n}\Theta_{u-s-1,u-s-k}(Y_k-\widehat{Y}_k)\right)\left(\sum_{k=1}^{n}\Theta_{u-t-1,u-t-k}(Y_k-\widehat{Y}_k)\right)' \\ &= \Gamma(t-s) - \sum_{k=1}^{n}\Theta_{u-s-1,u-s-k}V_{k-1}\Theta_{u-t-1,u-t-k}^{'}, \end{aligned} \tag{3.54}$$

where the last relations comes from the orthogonality of $Y_k-\widehat{Y}_k$ and $Y_j-\widehat{Y}_j$, $k,j = 1,\ldots,n$, $k\neq j$. $\square$

It is worth noting here that typical forecasting methods based on the infinite past cannot be immediately extended to the noncausal case. For example, note that for a series $\{X_n\}_{n\in\mathbb{Z}}$ admitting a causal representation (3.2), we can write

$$X_{n+h} = \sum_{j=0}^{\infty}\Psi_j Z_{n+h-j}.$$

Then, by replacing the ϵ's that contribute to X_{n+h} but not to X_n with their expected values (which are zero), we get the infinite past forecast $\widehat{X}_{n+h|n,\infty}$ as

$$\widehat{X}_{n+h|n,\infty} = \sum_{j=h}^{\infty} \Psi_j Z_{n+h-j}, \tag{3.55}$$

with the forecast error

$$X_{n+h} - \widehat{X}_{n+h|n,\infty} = \sum_{j=0}^{h-1} \Psi_j Z_{n+h-j}. \tag{3.56}$$

A noncausal series X_n, however, satisfies $X_{n+h} = \sum_{j=-\infty}^{\infty} \Psi_j \epsilon_{n+h-j}$, where the expected values of the infinite future $\epsilon_k,\ k = n+1, n+2, \ldots$, are not necessarily zero and hence we cannot derive a relation similar to (3.55).

3.4.3 Model selection

Here, we briefly review two typical model selection methods, namely, the Akaike information criterion (AIC, for short) and the Bayesian information criterion (BIC, for short). The AIC was proposed by Akaike (1974; 1998) as a method to select a model that best fits a given dataset amongst a set of candidate models. It is defined as

$$\text{AIC} = -2\ln L + 2k, \tag{3.57}$$

where L is the maximized value of the model's likelihood function, and k is the number of its parameters. Models with smaller AIC values are preferred to those with larger AIC values. Note, however, that AIC is a relative measure of model quality in the sense that it gives no information as to whether or not a given model is a good choice for the data at hand. Instead, the AIC ranks the candidate models rewarding the ones with large likelihood functions while penalizing the ones with a large number of parameters.

A number of similar model selection criteria have appeared in the literature. The most notable is a version of AIC with a correction for finite sample sizes defined as

$$\text{AICc} = AIC + \frac{2k(k+1)}{n-k-1}, \tag{3.58}$$

where n is the sample size (see Hurvich and Tsai (1989)). In particular, Brockwell and Davis (2009) prefer AICc as a model selection criterion for short-range dependent time series models as it counteracts the overfitting tendency of AIC. The sample size of the dataset in Section 3.6 is much larger than the number of parameters for the models we will consider and so the AIC and AICc differ only by a small amount.

BIC was proposed by Schwarz (1978) as a competitor to AIC. It is defined as

$$\text{BIC} = -2\ln L + k\ln(n), \tag{3.59}$$

and tends to favor smaller models. In other words, BIC requires stronger evidence to allow for higher complexity. Informally, this is justified by the parameter uncertainty that is present in Bayesian analysis. Below we give a rule of thumb for the strength of evidence against the model with the higher BIC (see Kass and Raftery (1995)).

BIC Difference	**Evidence against higher BIC**
0 to 2	Weak
2 to 6	Positive
6 to 10	Strong
>10	Very strong

Table 3.1: Rule of thumb for assessing the strength of evidence against the model with the higher BIC value.

We will not expand more on model selection criteria here. For further details, see Burnham and Anderson (2002).

3.5 Simulation study

In this section, we perform a Monte Carlo simulation study to assess the performance of the CLDL algorithm for the VARFIMA(p, D, q) model (3.36) described in Section 3.4.1. We examine four different models with AR and MA components of orders $p, q = 0, 1$. As discussed in Section 3.3.2, the focus will be on models with diagonal AR components. For each model, we consider two sample sizes $N = 200, 400$. The Gaussian time series data are

generated using the fast and exact synthesis algorithm of Helgason et al. (2011), while the number of replications will be either 100 or 200.

To solve the maximization problem (3.43), we use the SAS procedure **nlpqn**, which implements the BFGS quasi-Newton method, a popular iterative optimization algorithm. For our optimization scheme, we follow an approach found in Tsay (2010). A first step is to eliminate the nonlinear inequality constraint $|\Sigma| \geq 0$ in the parameter space S, by letting $\Sigma = U^T U$, where $U = (U_{jk})_{j,k=1,2}$ is an upper triangular matrix (Σ is nonnegative definite and such a factorization always exists). Therefore, the parameter vector θ can be written as $\theta = (d_1, d_2, c, U_{11}, U_{12}, U_{22}, \Theta')'$ while the parameter space becomes $S = \{\theta \in \mathbb{R}^{6+kq} : 0 < d_1, d_2 < 0.5,\ -1 < c < 1\}$, where k is 2 for models with diagonal MA part and 4 for models with general MA part.

Next, we describe our strategy on selecting initial parameter values (Φ^I, θ^I) for the BFGS method. Let

$$\Phi_r^0 = (\phi_{jk,r}^0)_{j,k=1,2}, \quad \theta^0 = (d_1^0, d_2^0, c^0, U_{11}^0, U_{12}^0, U_{22}^0, (\Theta_s^0)')', \quad r, s = 1, 2, \tag{3.60}$$

where $\Theta_s^0 = (\theta_{jk,s}^0)_{j,k=1,2}$ be the true parameter values. We consider initial values

$$d_k^I = \frac{d_k^0}{1 + 2d_k^0}, \quad c^I = \frac{2c^0}{1 + |c^0|}, \quad U_{jk}^I = 1, \quad \theta_{jk,s}^I = \frac{e^{\theta_{jk,s}^0} - 1}{e^{\theta_{jk,s}^0} + 1}, \quad \phi_{jk,r}^I = \frac{e^{\phi_{jk,r}^0} - 1}{e^{\phi_{jk,r}^0} + 1}, \tag{3.61}$$

where $j, k, r, s = 1, 2$. Note that the transformations (3.61) are essentially perturbations of the true parameter values that also retain the range of the parameter space S. For example, the value of d_k^I will be zero (or 1/2) when d_k^0 is also zero (or 1/2). Moreover, even though the parameter space S does not include identifiability constraints for the elements of the AR and MA polynomials, we did not encounter any cases where the optimization algorithm considers such values.

Figures 3.2–3.12 show the histograms of the parameter estimates for 12 different models. For all simulations we take $d_1 = 0.2, d_2 = 0.4, c = 0.6, \Sigma_{11} = 3, \Sigma_{12} = 0.5, \Sigma_{22} = 3$, while the true SRD parameters along with the sample size and number of replications are given

in the caption of each plot.[2] We also performed simulations for several other values of these parameters and got similar results and therefore we omit them in favor of space economy. Moreover, each plot in Figures 3.2–3.12 has dashed and solid lines which indicate the median of all replications and the true parameter values respectively. Note that in some cases, the estimates of the SRD parameters and of the entries of the matrix U, have large absolute values for a few number of realizations. To preserve a more informative scale of the horizontal axis, we do not to include these estimates in histogram bins. Instead, we indicate the range and frequency of these "outlier" estimates inside a legend box.

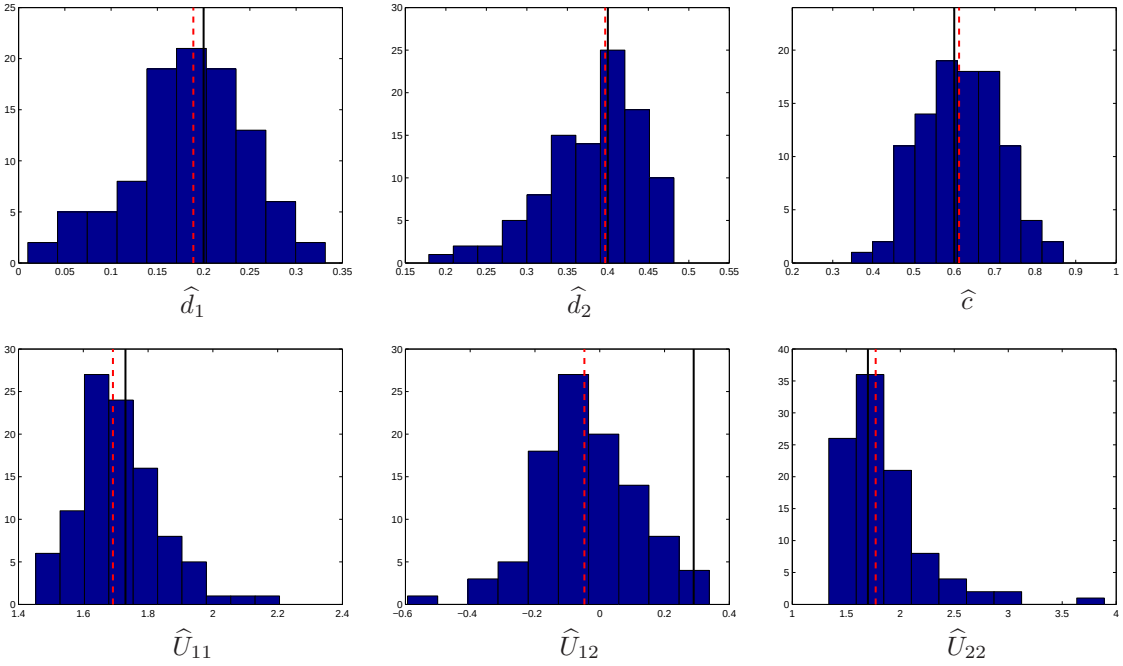

Figure 3.2: Estimated parameters for 100 replications of a VARFIMA$(0, D, 0)$ series with sample size $N = 200$. The dashed vertical lines indicate the median over all replications while the solid vertical lines indicate the true parameter values.

[2]For this choice of d_1, d_2, c, the phase parameter is equal to $\phi = 1.15$. Taking $c = -0.1985$ with the same d's would yield zero phase.

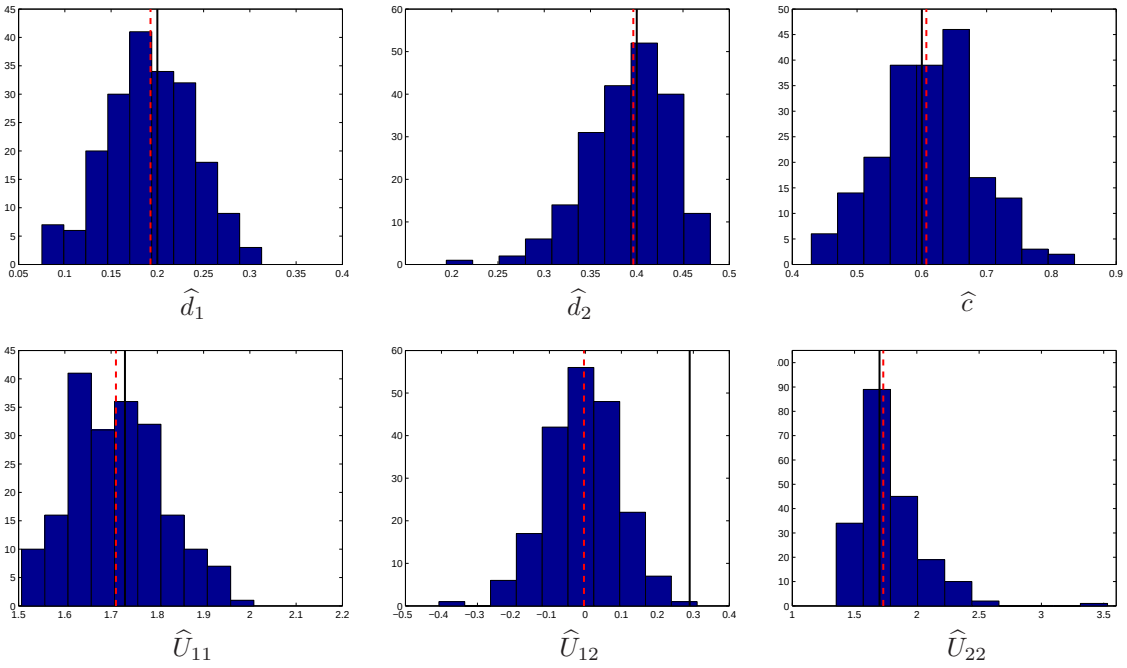

Figure 3.3: Estimated parameters for 200 replications of a VARFIMA$(0, D, 0)$ series with sample size $N = 400$. The dashed vertical lines indicate the median over all replications while the solid vertical lines indicate the true parameter values.

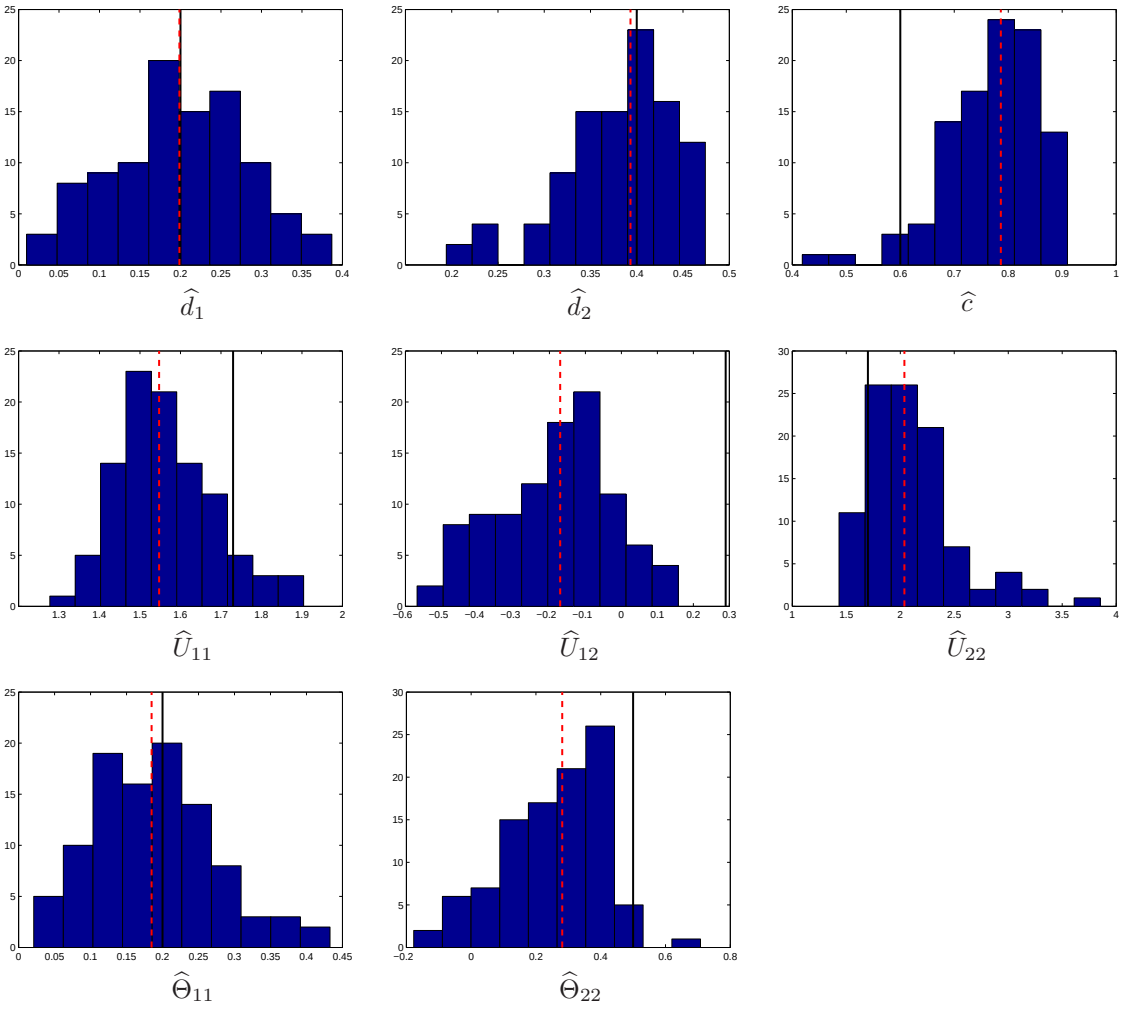

Figure 3.4: Estimated parameters for 100 replications of a VARFIMA$(0, D, 1)$ series with sample size $N = 200$ and diagonal MA component. The dashed vertical lines indicate the median over all replications while the solid vertical lines indicate the true parameter values.

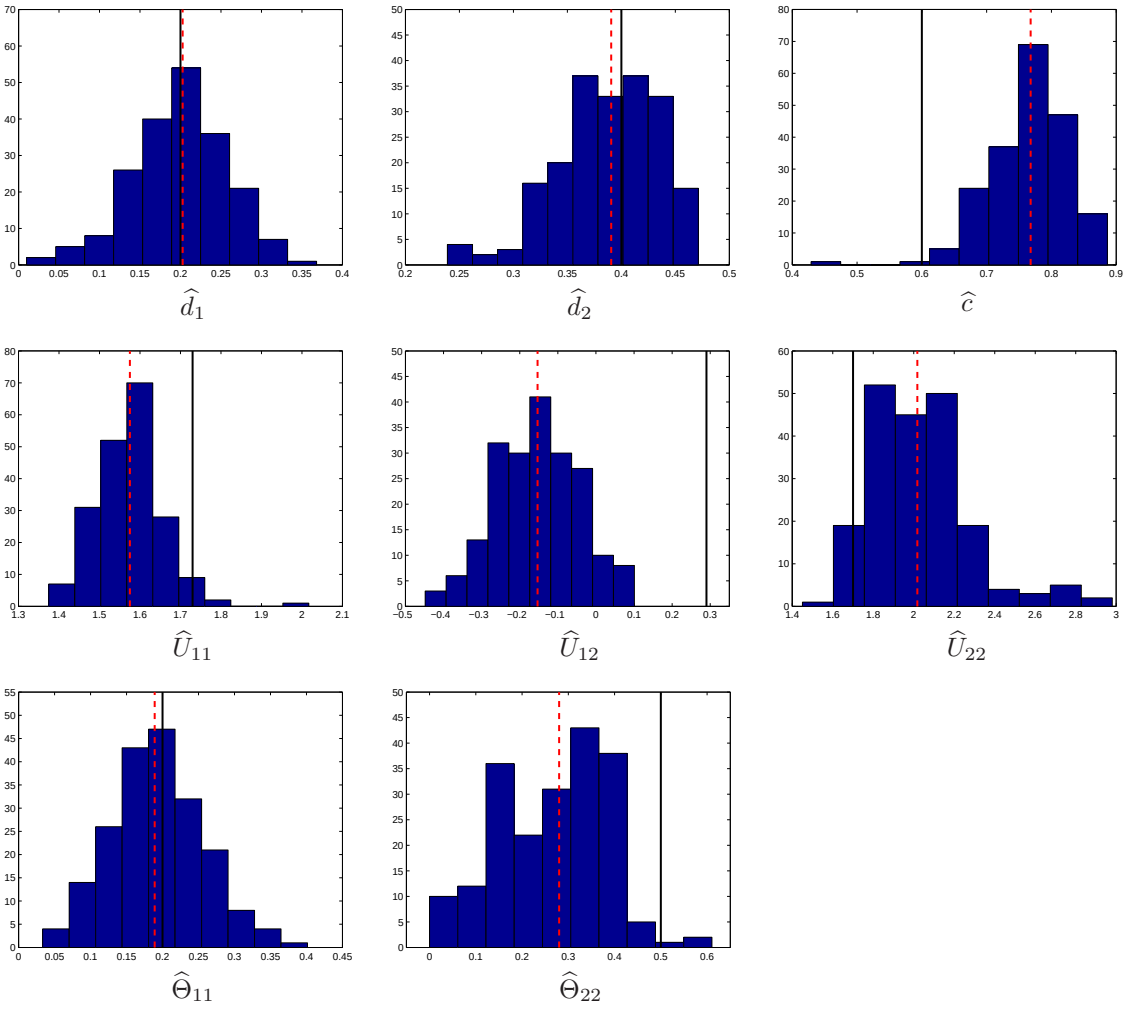

Figure 3.5: Estimated parameters for 200 replications of a VARFIMA$(0, D, 1)$ series with sample size $N = 400$ and diagonal MA component. The dashed vertical lines indicate the median over all replications while the solid vertical lines indicate the true parameter values.

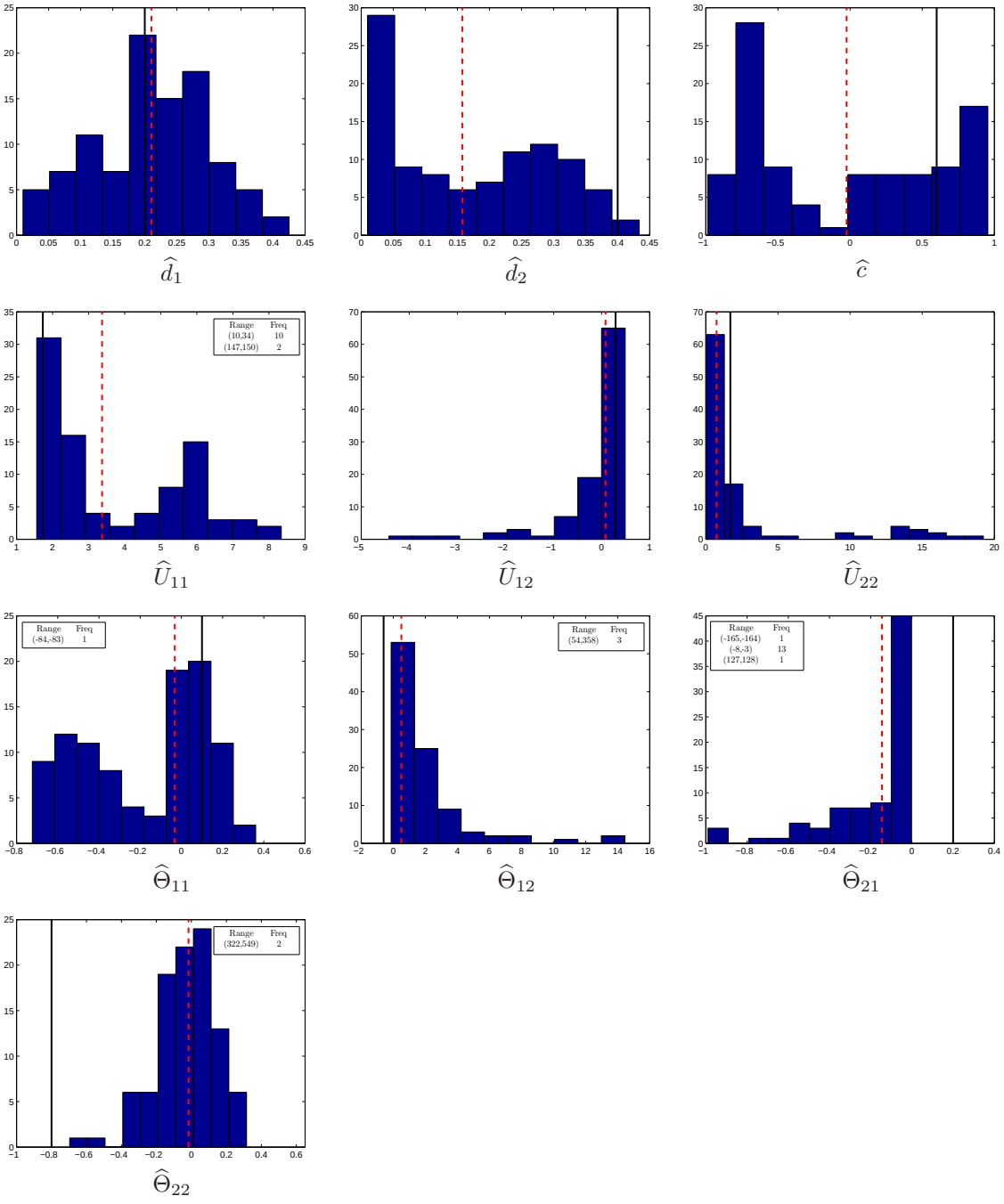

Figure 3.6: Estimated parameters for 100 replications of a VARFIMA$(0, D, 1)$ series with sample size $N = 200$ and nondiagonal MA component. The dashed vertical lines indicate the median over all replications while the solid vertical lines indicate the true parameter values.

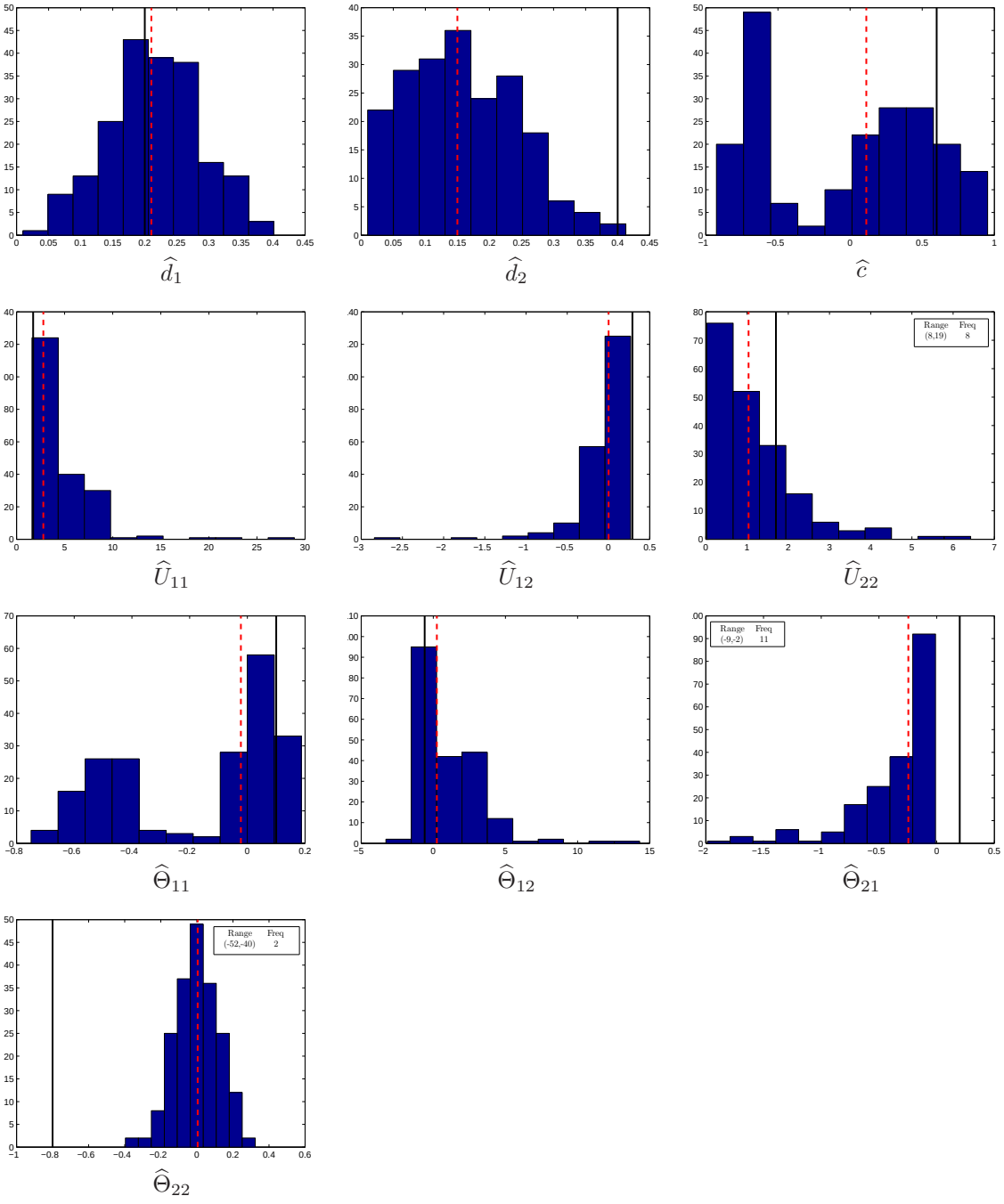

Figure 3.7: Estimated parameters for 200 replications of a VARFIMA$(0, D, 1)$ series with sample size $N = 400$ and nondiagonal MA component. The dashed vertical lines indicate the median over all replications while the solid vertical lines indicate the true parameter values.

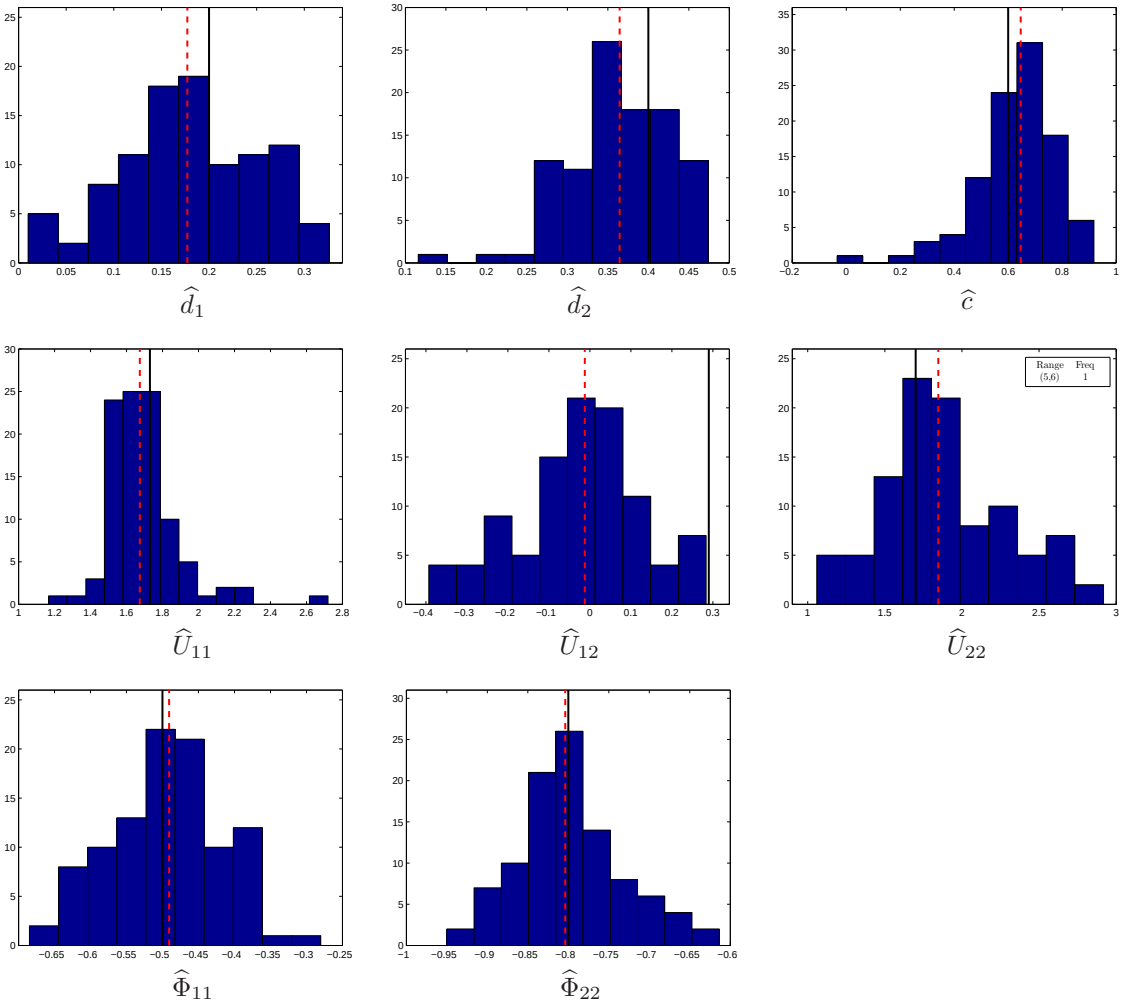

Figure 3.8: Estimated parameters for 100 replications of a VARFIMA$(1, D, 0)$ series with sample size $N = 200$ and diagonal AR component. The dashed vertical lines indicate the median over all replications while the solid vertical lines indicate the true parameter values.

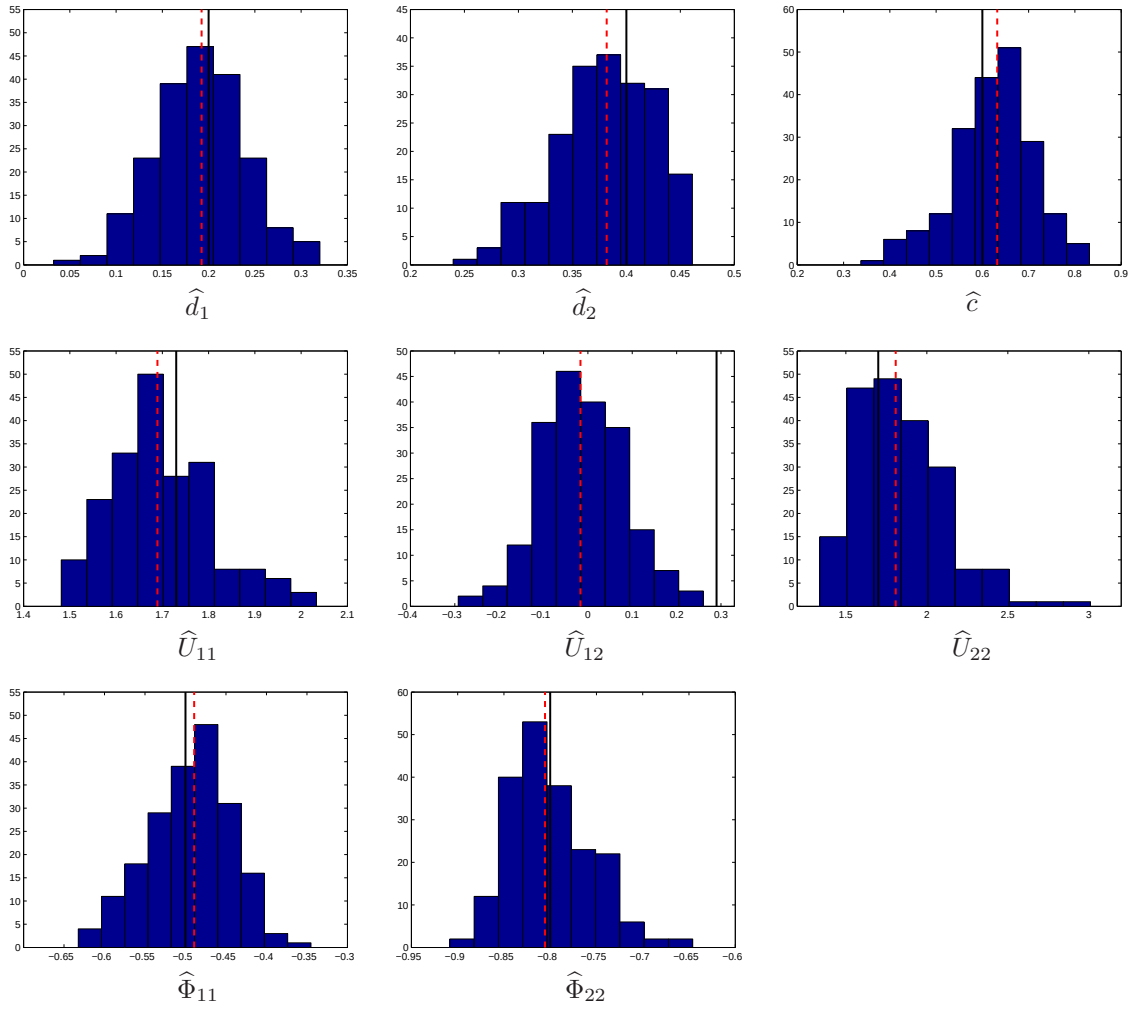

Figure 3.9: Estimated parameters for 200 replications of a VARFIMA$(1, D, 0)$ series with sample size $N = 400$ and diagonal AR component. The dashed vertical lines indicate the median over all replications while the solid vertical lines indicate the true parameter values.

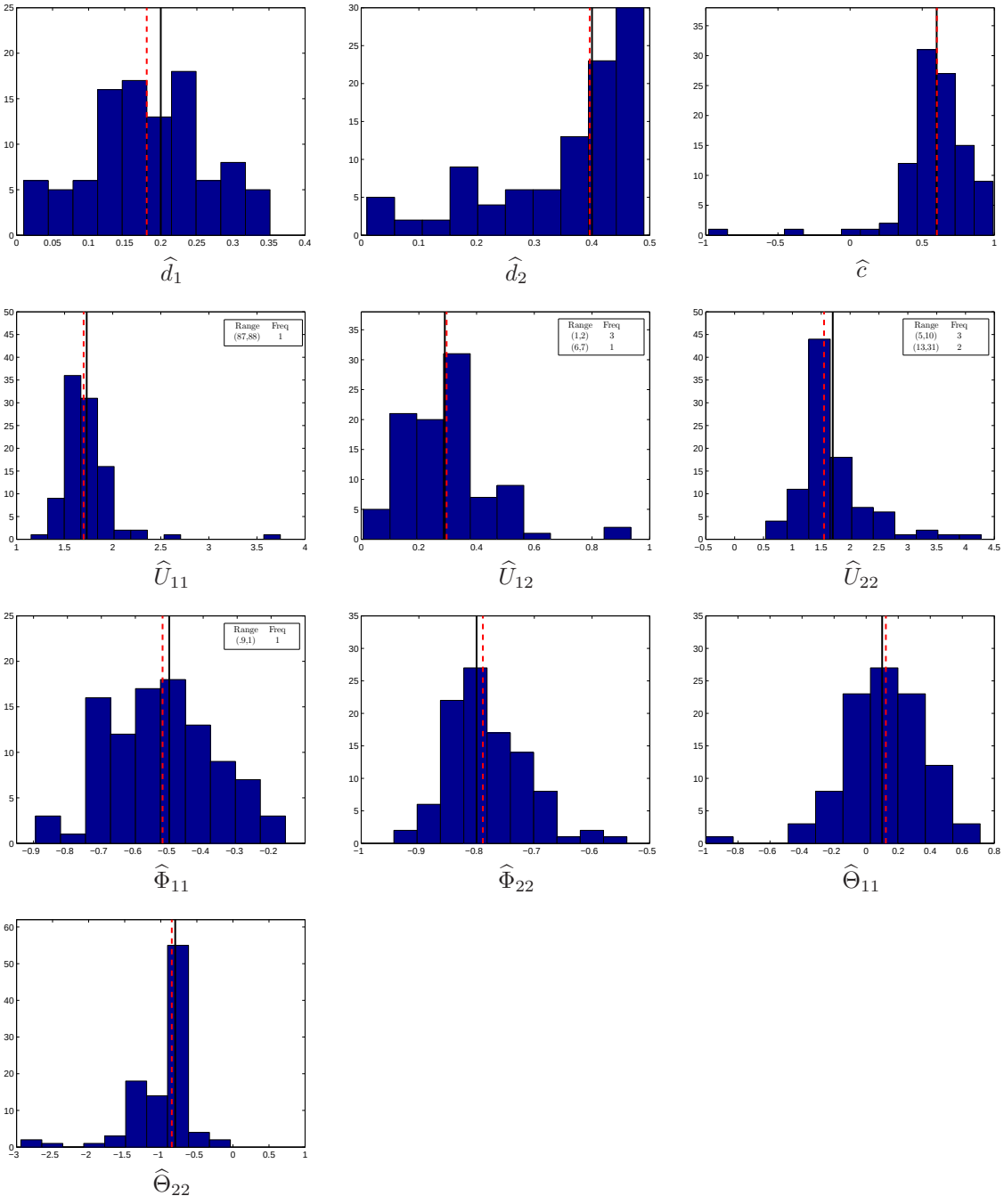

Figure 3.10: Estimated parameters for 100 replications of a VARFIMA$(1, D, 1)$ series with sample size $N = 200$ and diagonal AR and MA components. The dashed vertical lines indicate the median over all replications while the solid vertical lines indicate the true parameter values.

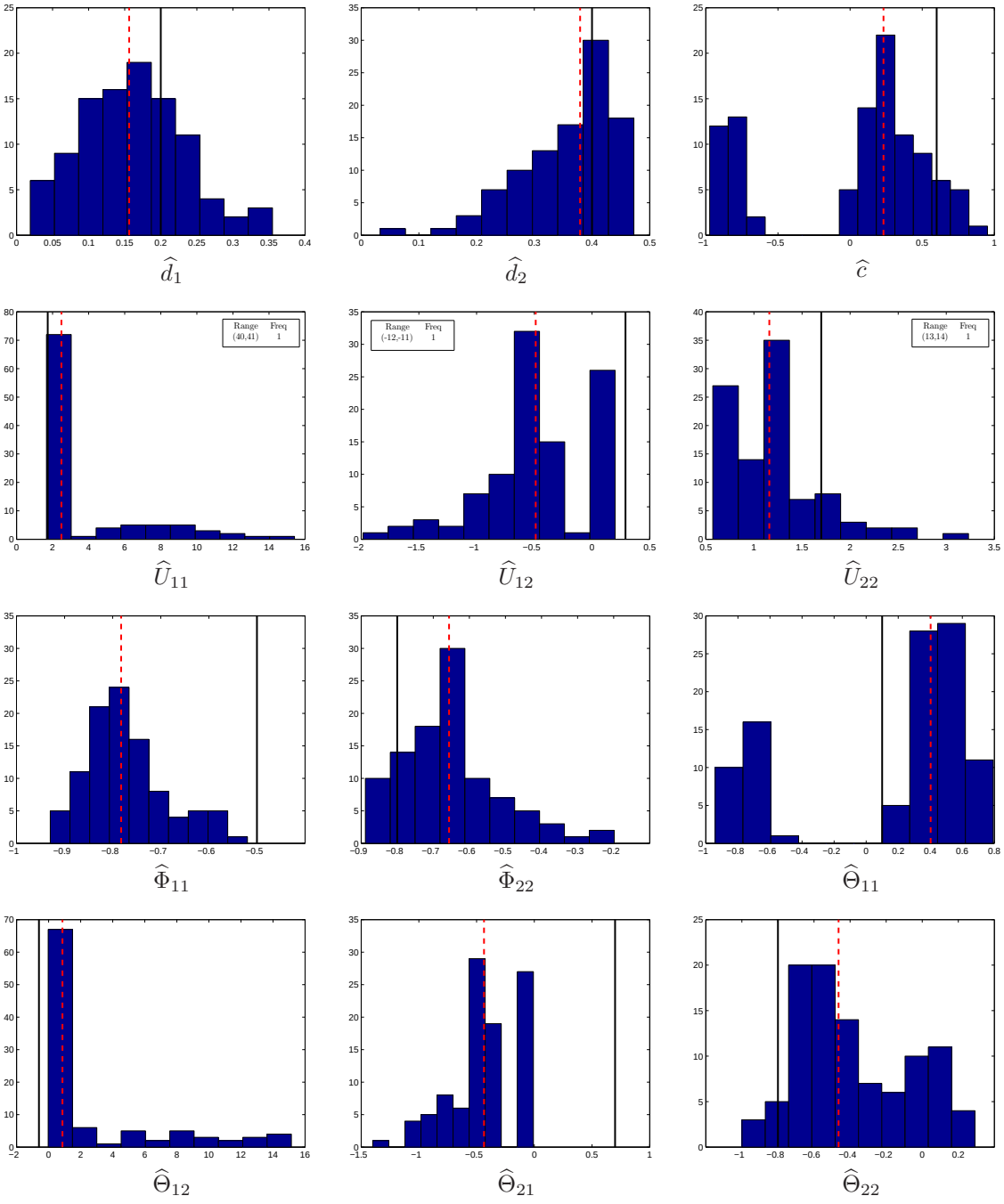

Figure 3.11: Estimated parameters for 100 replications of a VARFIMA$(1, D, 1)$ series with sample size $N = 200$ and diagonal AR and nondiagonal MA components. The dashed vertical lines indicate the median over all replications while the solid vertical lines indicate the true parameter values.

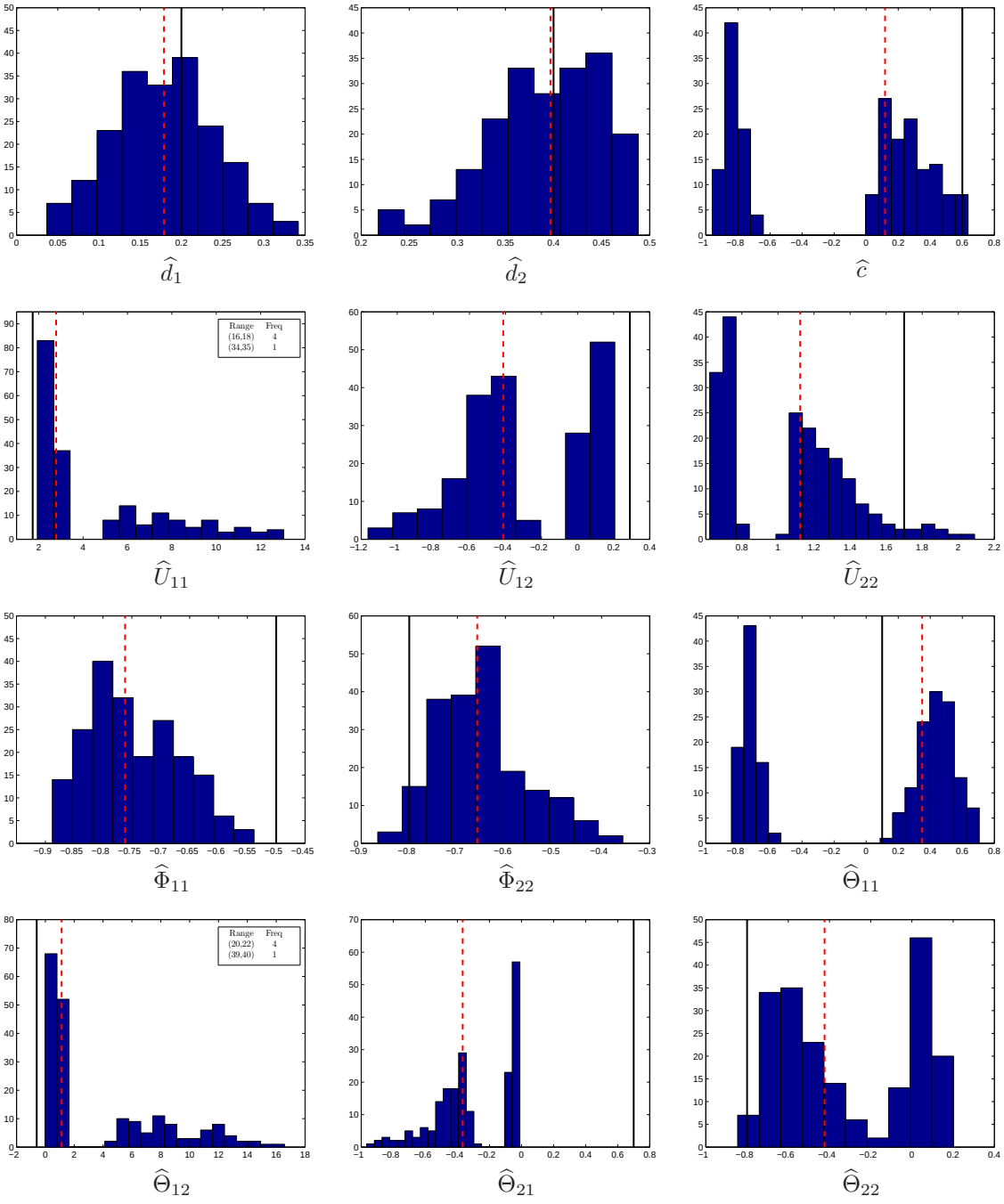

Figure 3.12: Estimated parameters for 200 replications of a VARFIMA$(1, D, 1)$ series with sample size $N = 400$ and diagonal AR and nondiagonal MA components. The dashed vertical lines indicate the median over all replications while the solid vertical lines indicate the true parameter values.

The results of the simulation study indicate a satisfactory performance of the CLDL algorithm for most cases considered. The small differences between the dashed and solid lines in Figures 3.2–3.3 indicate the small bias of the parameter estimates for the VARFIMA$(0, D, 0)$ model, with the exception of $\widehat{U}_{12}$. The reason $\widehat{U}_{12}$ is close to zero is because the value of the true parameter U_{12} is small (0.28), and not because the algorithm fails to estimate the dependence across the series. To check this, we tried higher values of U_{12}, which led to estimates $\widehat{U}_{12}$ far from zero but also, as expected, an increased bias of the other parameter estimates. Moreover, as the sample size grows from N =100 to 200, the bias measured through the median$(\widehat{U}_{12}) - U_{12}$ decreases from -0.33 to -0.29.

Moving on to higher order models with short-range dependent components, we notice bias in the estimates of other parameters besides U_{12}. The estimates of the LRD parameters, in particular, have the smallest bias across all models. We would also like to note the bimodality features of some histograms in Figures 3.6–3.7 (VARFIMA$(0, D, 1)$ with nondiagonal MA part) and 3.11–3.12 (VARFIMA$(1, D, 1)$ with nondiagonal MA and diagonal AR parts), especially those of $\widehat{c}$.[3] One explanation of the bimodality is that the optimization procedure is trapped at a local maximum of the likelihood function. Another reason could be that the graph of the likelihood (viewed as a function of c only, with all other parameters fixed at their estimated values) may be very flat around the maxima for a number of realizations. We shall explore these issues in greater depth elsewhere.

Finally, it is worth mentioning here an interesting pattern in the estimates of Figures 3.11 and 3.12. More specifically, the estimates of all parameters can be classified according to two types of realizations. For example, the estimates $\widehat{c}$ in Figure 3.11 divide into two clusters one to the left and one to the right of 0.5. Figure 3.13 depicts the histograms of $\widehat{U}_{11}, \widehat{\Phi}_{22}, \widehat{\Theta}_{11}$ from Figure 3.11, classified according to the values of $\widehat{c}$. The white bins correspond to the realizations for which $\widehat{c} > -0.1$, while the dark colored bins correspond to the realizations for which $\widehat{c} < -0.1$.

[3]Similar features were also observed in simulations where other values for c were taken (not reported here).

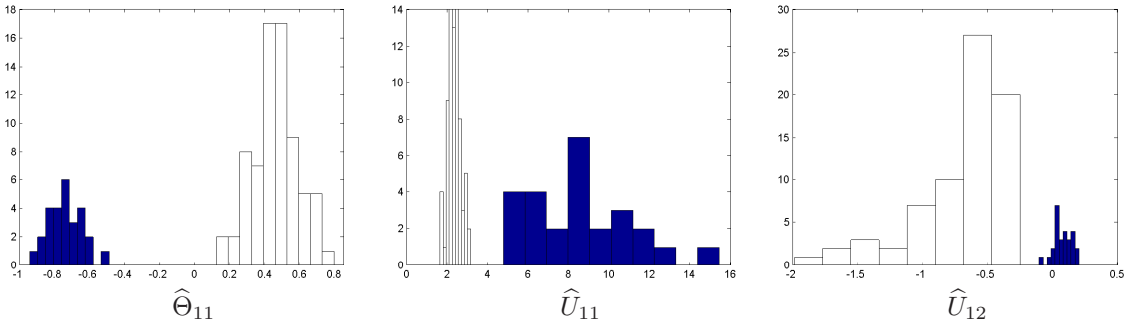

Figure 3.13: Estimated parameters for 100 replications of a VARFIMA$(1, D, 1)$ series with sample size $N = 200$ and diagonal AR and nondiagonal MA components. The white bins correspond to the realizations for which $\widehat{c} > -0.1$, while the dark colored bins correspond to the realizations for which $\widehat{c} < -0.1$

3.6 Application

In this section, we apply the CLDL algorithm to analyze inflation rates in the U.S. under the noncausal VARFIMA(p, D, q) model discussed in Section 3.3.2. Evidence of long-range dependence behavior in inflation rates has been found in a number of works (see, for example, Baillie et al. (1996), Doornik and Ooms (2004), Hurvich and Sela (2009), Baillie and Moreno (2012) and references therein). More specifically, Hurvich and Sela (2009) tested the fit of several long- and short-range dependent models on the annualized monthly inflation rates for goods and services in the U.S. during the period of February 1956–January 2008 ($N = 624$ months) and selected a causal VARFIMA model as the best choice. Besides their long memory features, however, the time series of inflation rates often exhibit asymmetric behavior, and therefore call for multivariate LRD models that allow for general phase.

Following the notation of Sela (2010),[4] we denote the Consumer Price Indices series for commodities as $\{CPI_t^c\}_{t=0,\ldots,N}$ and the corresponding series for services as $\{CPI_t^s\}_{t=0,\ldots,N}$. Then, we define the annualized monthly inflation rates for goods and services as

$$g_t = 1200\frac{CPI_t^c - CPI_{t-1}^c}{CPI_{t-1}^c} \quad \text{and} \quad s_t = 1200\frac{CPI_t^s - CPI_{t-1}^s}{CPI_{t-1}^s},$$

respectively. The two series $\{g_t\}_{t=1,\ldots,N}$, $\{s_t\}_{t=1,\ldots,N}$ are depicted in Figure 3.14.[5]

[4]See also the accompanying R code.

[5]The consumer price indices (raw) data are available online from the Bureau of Labor Statistics.

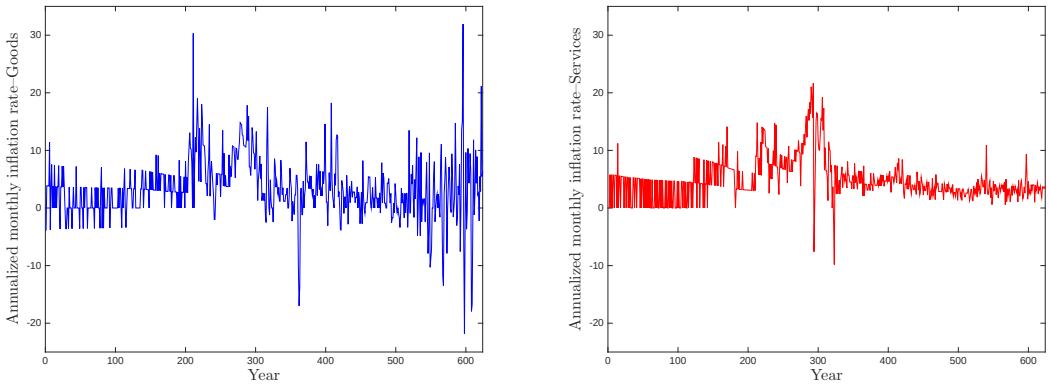

Figure 3.14: Annualized monthly inflation rates for goods (left) and services (right) from February 1956 to January 2008.

The two plots in Figure 3.15 provide some motivation for why a general phase model is needed for this dataset. More specifically, the left plot in Figure 3.15 depicts the sample crosscorrelation function $\widehat{\rho}_{12}(h)$ of the two series for all lags such that $|h| < 25$. Observe that for negative lags the sample crosscorrelation function decays faster than for positive lags suggesting time-non-reversibility of the series and hence non-zero phase (see (2.30)).

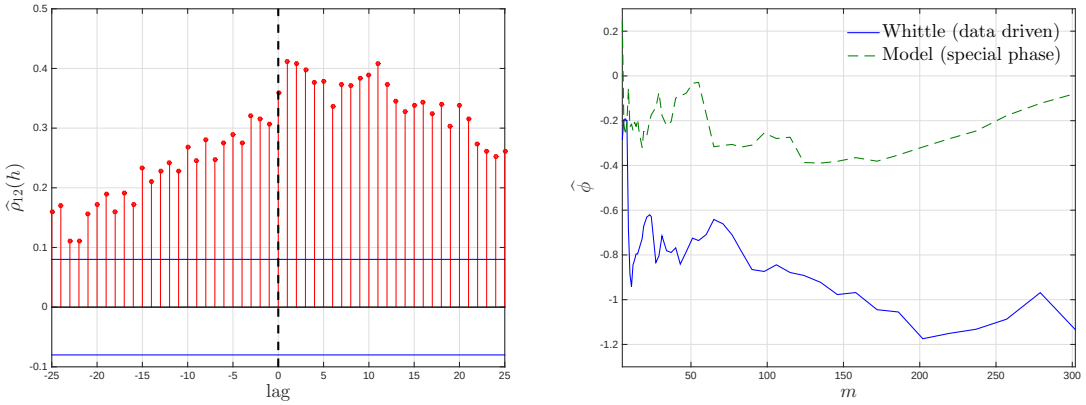

Figure 3.15: Left plot: Sample cross correlation $\widehat{\rho}_{12}(h)$ of the series $\{g_t\}_{t=1,\ldots,N}$ and $\{s_t\}_{t=1,\ldots,N}$ depicted in Figure 3.14 for $|h| \leq 25$. Right plot: Local Whittle phase estimates, one corresponding to the causal VARFIMA (dashed line) and one estimated directly from the data (solid line). Both estimates are plotted as functions of a tuning parameter $m = N^{0.25+0.0125k}$, $k = 1, \ldots, 51$, where $N = 624$ is the size of the series.

Further evidence for general phase can be obtained from the local Whittle estimation of Robinson (2008) which can be used to estimate the phase and the LRD parameters directly from the data. The estimation is semiparametric in the sense that it only requires

specification of the spectral density at low frequencies (see, for example, the relation (2.5)). The right plot in Figure 3.15 depicts two local Whittle estimates of the phase parameter ϕ as functions of m – a tuning parameter representing the number of lower frequencies used in the estimation. The dashed line corresponds to the special phase estimate $\widehat{\phi} = (\widehat{d}_1 - \widehat{d}_2)\pi/2$ of the causal VARFIMA model based on the local Whittle estimates of the two d's. On the other hand, the solid line shows the phase parameter estimated directly from the data. The two lines being visibly different suggest that the special phase parameter and the associated VARFIMA model are not appropriate.

In the analysis of Hurvich and Sela (2009), the causal VARFIMA$(1, D, 0)$ model was selected as the best choice (based on AIC), amongst vector autoregressive models of both low and high orders and also amongst causal VARFIMA$(p, D, 0)$ and FIVARMA$(p, D, 0)$ models with $p \leq 1$. Note that for the FIVARMA model, the authors also allowed for typical fractional cointegration, an extension that was not discussed in Section 3.3.2. The estimated VARFIMA$(1, D, 0)$ model, in particular, was

$$\begin{aligned} g_t &= 0.3027 g_{t-1} + 0.4245 s_{t-1} + \epsilon_{1t}, \\ s_t &= -0.0237 g_{t-1} - 0.3085 s_{t-1} + \epsilon_{2t}, \end{aligned} \tag{3.62}$$

where

$$\begin{pmatrix} \epsilon_{1t} \\ (I-B)^{0.4835}\epsilon_{2t} \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} 20.23 & 0.46 \\ 0.46 & 7.08 \end{pmatrix}\right). \tag{3.63}$$

We should note here that the SAS optimization algorithm we used produced estimates similar to those of Sela's algorithm (implemented in R), for all parameters except d_1 which Sela estimates to be zero while for this model we estimated it to be 0.17.

The parameter estimates in (3.62)–(3.63) reveal an interesting feature, noted by Sela (2010). In particular, while the lagged services inflation has a significant influence on goods inflation, the lagged goods inflation seems to have a small effect on services inflation. This behavior is potentially related to the so-called gap between the prices in services and the prices in goods which was studied by Peach et al. (2004). More specifically, the term gap

refers to the tendency of prices in services to increase faster than prices in goods. We plan to examine the relation of these two phenomena in greater depth in a future work.

Table 3.2 below contains the AIC and BIC values for 9 causal and 9 noncausal VARFIMA models fitted to the inflation rates series. The superscripts $+$ and $*$ indicate models with diagonal AR and MA components, respectively, while k denotes the number of parameters of each model.

(a) Noncausal models

Model	k	**AIC**	**BIC**
$(0, D, 0)$	6	4412	4421
$(0, D, 1)^*$	8	4385.7	4394.6
$(1, D, 0)^+$	8	4367	4376
$(0, D, 1)$	10	4361	4370
$(1, D, 0)$	10	4351	4360
$(1, D, 1)^{*+}$	10	4339	4348
$(1, D, 1)^*$	12	4314.8	4323.7
$(1, D, 1)^+$	12	4314.1	4322.9
$(1, D, 1)$	14	4303	4312

(b) Causal models

Model	k	**AIC**	**BIC**
$(0, D, 0)$	5	4420	4429
$(0, D, 1)^*$	7	4385.9	4394.8
$(1, D, 0)^+$	7	4376	4385
$(0, D, 1)$	9	4380	4389
$(1, D, 0)$	9	4358	4367
$(1, D, 1)^{*+}$	9	4372	4381
$(1, D, 1)^*$	11	4315.1	4323.9
$(1, D, 1)^+$	11	4315.3	4324.1
$(1, D, 1)$	13	4314	4323

Table 3.2: AIC and BIC values from fitting 9 noncausal (left table) and 9 causal (right table) VARFIMA(p, D, q) models to the annualized inflation rates for goods and services. The superscripts $+$ and $*$ indicate diagonal AR and MA components, respectively.

A number of remarks are in place regarding Table 3.2. First, note that all noncausal models have smaller AIC and BIC values compared to their causal counterparts. Second, amongst all the models considered the best choice according to both the AIC and BIC selection criteria is the VARFIMA$(1, D, 1)$ with general AR and MA components. This observation hints that we should extend out study to models with short-range dependent components of higher orders, which we plan to pursue in a future work. Considering higher order AR components will also allow us to consider FIVARMA(p, D, q) models (see the discussion under the relation (3.39)).

Next, we present the two estimated noncausal VARFIMA$(1, D, 0)$ models whose AIC and BIC values were calculated in Table 3.2. Even though other models have smaller AIC and BIC values, these models allow us to compare our results with Sela's. The estimated

noncausal VARFIMA$(1, D, 0)$ model with a diagonal AR component is

$$\begin{aligned} g_t &= 0.122 g_{t-1} + \epsilon_{1t}, \\ s_t &= -0.514 s_{t-1} + \epsilon_{2t}, \end{aligned} \tag{3.64}$$

where

$$\begin{pmatrix} ((I-B)^{-0.215} + 0.534(I-B^{-1})^{-0.215})^{-1}\epsilon_{1t} \\ ((I-B)^{-0.379} - 0.534(I-B^{-1})^{-0.379})^{-1}\epsilon_{2t} \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} 9.34 & 1.72 \\ 1.72 & 17.43 \end{pmatrix}\right). \tag{3.65}$$

Moreover, the estimated noncausal VARFIMA$(1, D, 0)$ model with general AR component is

$$\begin{aligned} g_t &= 0.176 g_{t-1} + 0.03 s_{t-1} + \epsilon_{1t}, \\ s_t &= 0.09 g_{t-1} - 0.49 s_{t-1} + \epsilon_{2t}, \end{aligned} \tag{3.66}$$

where

$$\begin{pmatrix} ((I-B)^{-0.175} + 0.528(I-B^{-1})^{-0.175})^{-1}\epsilon_{1t} \\ ((I-B)^{-0.363} - 0.528(I-B^{-1})^{-0.363})^{-1}\epsilon_{2t} \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} 9.34 & 1.48 \\ 1.48 & 17.23 \end{pmatrix}\right). \tag{3.67}$$

We have a number of interesting observations related to the estimated parameters in (3.64)–(3.67). First, notice that the off-diagonal elements of the AR component in (3.66) are close to 0. This suggests that the (a)symmetry behavior in the inflation rates is now captured by the parameter c. Indeed, plugging the estimates of d_1, d_2 and c from the models (3.64)–(3.65) and (3.66)–(3.67) in relation (3.20), we obtain phase estimates close to -1 which agree with the local Whittle estimate of the phase in the right plot of Figure 3.15. Note that the local Whittle estimates of d_1 and d_2 which we did not report here are close to 0.2 and 0.38 respectively, and hence they are also close to the estimated parameters of the two noncausal VARFIMA models.

Second, even though the parameter estimates in (3.64)–(3.65) and (3.66)–(3.67) are very close, the AIC and BIC values of these models differ a lot. We believe that this happens because there exists some type of fractional cointegration relationship between the two series

$\{g_t\}_{t=1,\ldots,N}$ and $\{s_t\}_{t=1,\ldots,N}$ which can be captured by the model (3.66)–(3.67) but not by the model (3.64)–(3.65).

Finally, observe that the diagonal elements of the variance matrices in (3.65) and (3.67) seem to be reversed when compared to the corresponding estimates from the causal model (3.63). At the moment, we are not certain what causes this behavior but we plan to further investigate this in a future work.

We conclude this section with a comment regarding forecasting. We are currently testing the performance of the multistep ahead forecasting method proposed in Section 3.4.2. Preliminary results indicate that for short and medium-range forecast horizons (1-2 years), the mean squared prediction error (MSPE) in the inflation data for noncausal models is slightly smaller than for causal ones. On the other hand, the noncausal models had significantly smaller MSPEs for long-range forecast horizons (more than 3 years).

CHAPTER 4

Convex optimization and feasible circulant matrix embeddings in synthesis of stationary Gaussian fields

4.1 Introduction

In this chapter, we are interested in the synthesis of stationary Gaussian two-dimensional random fields. One of the most popular methods for the exact generation of such fields is based on *circulant matrix embedding* (CME). Introduced by Davis and Harte (1987) for simulating stationary Gaussian univariate processes, the method was later extended to random fields by Dietrich and Newsam (1993), Wood and Chan (1994), Chan and Wood (1997), Gneiting et al. (2006), Stein (2012). See also Stein (2002) concerning the case of fractional Brownian surfaces; Chan and Wood (1999), Helgason et al. (2011) for extensions to the multivariate (vector) context; Percival (2006) for the case of complex-valued Gaussian processes; and Craigmile (2003) for some theoretical contributions.

The idea of the method is to use a suitable periodization to embed a covariance matrix of interest into a larger circulant matrix whose eigenvalues can be computed efficiently using the fast Fourier transform (FFT). If all the eigenvalues are nonnegative, a stationary Gaussian random field can be constructed with the embedding circulant matrix as its covariance matrix. Its suitable subfield is then the desired stationary Gaussian field. While in theory all eigenvalues can be shown to be nonnegative in the limit of the increasing sample size, some of the eigenvalues often remain negative for computationally feasible large sample sizes and many covariance structures of practical interest.

To have all the eigenvalues nonnegative, two approaches have been suggested in the literature, namely, the *cutoff circulant embedding* or CCE, for short (Stein (2002), Gneiting et al. (2006)) and the *smoothing windows circulant embedding* or SWCE, for short (Helgason et al. (2014)). In the CCE method, the initial covariance is extended suitably in a parametric

fashion, based on the model at hand, to a larger domain, leading to a covariance embedding with nonnegative eigenvalues. While such extensions have been found for several examples of stationary fields, their construction is often nontrivial.

The SWCE approach, on the other hand, does not depend on the model at hand and is easier to implement by suitably modifying the standard embedding. A possible reason that some eigenvalues are negative is that the covariance embedding is not "smooth" at the boundary of periodization. In the SWCE method, the covariance is extended over a transition region where it is smoothed at the boundary, using a smoothing kernel. Helgason et al. (2014) propose two types of smoothing: overlapping and nonoverlapping windows. Numerical studies show that both variants of SWCE work well for a range of covariance structures. Moreover, when using overlapping windows, the SWCE method greatly outperforms CCE in the sense of efficiency considered in Helgason et al. (2014). However, for some covariance structures (for example, as those in Section 4.4 below), a large transition region is needed for the SWCE method to work.

We propose and study here a new method to simulate stationary Gaussian fields using circulant matrix embedding, called *optimal circulant embedding* or OCE, for short. We find the method interesting and promising for several reasons. First, the method proposes a novel approach based on quadratic constrained optimization. This is quite fitting given the growing integration of optimization tools in statistics. Second, as seen below, several key components of the optimization procedure can be implemented more efficiently using FFT, making it natural to combine the circulant matrix embedding and optimization approaches. Third, we show numerically that the OCE method works better than the SWCE method (and hence the CCE method by the discussion above) for several covariance classes, especially when the transition region is small.

More specifically, the OCE method consists of formulating a quadratic programming problem with linear inequality constraints, to find a *feasible* covariance embedding, meaning

that its eigenvalues are nonnegative. The exact optimization problem has the form

$$\begin{aligned} \min_{\widetilde{r}(n_1,n_2)} \quad & \sum_{n_1,n_2=-N+1}^{N-1} \left(r(n_1,n_2)-\widetilde{r}(n_1,n_2)\right)^2, \\ \text{subject to} \quad & \lambda_{k_1,k_2}(\widetilde{r}) \geq 0, \quad -\widetilde{N}+1 \leq k_1,k_2 \leq \widetilde{N}-1, \end{aligned} \tag{4.1}$$

where $\{r(n_1,n_2),\ -N+1 \leq n_1,n_2 \leq N-1\}$ is a given covariance function of a random field $X_{(n_1,n_2)}$ on a square grid $\{(n_1,n_2):\ 0 \leq n_1,n_2 \leq N-1\}$, $\{\widetilde{r}(n_1,n_2),\ -\widetilde{N}+1 \leq n_1,n_2 \leq \widetilde{N}-1\}$ is the corresponding covariance embedding on a larger grid $\{(n_1,n_2),\ -\widetilde{N}+1 \leq n_1,n_2 \leq \widetilde{N}-1\}$ with $\widetilde{N} \geq N$, and $\lambda_{k_1,k_2}(\widetilde{r})$ are the eigenvalues of the covariance embedding $\widetilde{r}$. The difference between the larger and smaller square grids is the transition region, referred to above. When the minimum value of the objective function is equal to zero, then the method is exact, namely, the synthesized field will have the targeted covariance. If for some choice of $\widetilde{N}$ this is not true, then the solution of (4.1) can be thought as the best approximation to the targeted covariance, that has no negative eigenvalues. The focus of this chapter though is on the situations where the objective function is zero (up to some tolerance error) and the simulation is exact.

To solve (4.1) numerically, we use an interior-point algorithm called the *primal log-barrier method* (see Section 4.3 for more details). The basic idea of this algorithm is to solve a sequence of unconstrained problems of the form

$$\min_{\widetilde{r}(n_1,n_2)} \quad F_t(\widetilde{r}) := \sum_{n_1,n_2=-N+1}^{N-1} \left(r(n_1,n_2)-\widetilde{r}(n_1,n_2)\right)^2 - \frac{1}{t} \sum_{k_1,k_2=-\widetilde{N}+1}^{\widetilde{N}-1} \log(\lambda_{k_1,k_2}(\widetilde{r})), \tag{4.2}$$

for positive increasing values of t. The log term in (4.2) serves as a "barrier" not allowing the search algorithm starting from a strictly feasible point $\widetilde{r}$ (with all eigenvalues $\lambda_{k_1,k_2}(\widetilde{r})$ positive) to move into the region of non-feasible points $\widetilde{r}$ (with some eigenvalues $\lambda_{k_1,k_2}(\widetilde{r})$ zero or negative). On the other hand, a large t ensures that the minimized function $F_t(\widetilde{r})$ is close to that of interest in (4.1). One can show that the solutions $\widetilde{r}_t(n_1,n_2)$ of (4.2) approach a solution of (4.1) as t increases. Eliminating inequality constraints of (4.1) by inserting them in the objective function in (4.2) is a frequently used practice, since unconstrained problems are far easier to handle.

To solve the problem (4.2) (for a fixed t), we approximate F_t in (4.2) by a quadratic function $\widehat{F}_t$ (around a strictly feasible initial point), and calculate its minimizer $\widetilde{r}_{t,a}(n_1, n_2)$. This reduces the problem to solving a linear system of the form

$$Hv = b, \tag{4.3}$$

where H is the Hessian of F_t in (4.2) and v is the covariance embedding $\widetilde{r}$ indexed as a vector. The symmetry and the positive definiteness of the coefficient matrix H allows solving the systems (4.3) using a popular iterative algorithm called the *conjugate gradient* method. As we show below, key steps of the conjugate gradient method applied to our problem can be carried out more efficiently using FFT.

Though the various components of the optimization procedure discussed above might appear straightforward, there are in fact a number of technical issues that need to be addressed. For example, as indicated above, we need to show how FFT is used in solving (4.3). Among other issues arising in the procedure are the role and choice of the updating rule for t in (4.2), the effect that the approximation of F_t by $\widehat{F}_t$ has on the convergence of the approximate solutions $\widetilde{r}_{t,a}(n_1, n_2)$ to the solution of (4.1), and so on. No less important is to see how the suggested approach performs on covariance structures of interest and how it compares to other approaches. All these issues are addressed in this chapter.

The rest of Chapter 4 is organized as follows. In Section 4.2, we briefly review the existing circulant embedding schemes. Optimal circulant embedding is presented in Section 4.3. The performance of the method for several covariance structures is studied in Section 4.4. Section 4.5 contains all figures while technical proofs are moved to Appendix B.1.

4.2 Available circulant matrix embeddings

Let $X = \{X_n, n = (n_1, n_2) \in \mathbb{Z}^2\}$ be a zero mean, stationary Gaussian two-dimensional random field. The autocovariance function of X is defined as

$$r(n) = r(n_1, n_2) = \mathbb{E}X_0X_n, \tag{4.4}$$

and satisfies $r(n) = r(-n),\ n \in \mathbb{Z}^2$. We are interested here in the synthesis of the two dimensional field X on the square grid

$$G(N) = \{n = (n_1, n_2) \in \mathbb{Z}^2 : 0 \leq n_1, n_2 \leq N - 1\}, \tag{4.5}$$

though the methods described below can be extended easily to rectangular grids and likely to higher dimensions. For reference and comparison, the rest of Section 2 contains a short description of the existing circulant matrix embedding methods. We recall the standard CME (Circulant Matrix Embedding) method in Section 4.2.1. The CCE (Cutoff Circulant Embedding) and SWCE (Smoothing Windows Circulant Embedding) methods are discussed briefly in Sections 4.2.2 and 4.2.3, respectively.

4.2.1 Standard embedding

Set $M = 2N-1$ to be the (side) size of a larger embedding grid $G(M)$, where an embedding field will be generated. Let also $\widetilde{r}(n),\ n \in G(M)$, be the covariance embedding defined through the symmetry relation

$$\widetilde{r}(n) = r(\xi_N(n)), \quad n \in G(M), \tag{4.6}$$

where $\xi_L(u) = (\xi_{1,L}(u), \xi_{2,L}(u))$ is defined by

$$\xi_{1,L}(u) = \xi_{2,L}(u) = \begin{cases} u, & \text{if } 0 \leq u \leq L-1, \\ u - M, & \text{if } L \leq u \leq M-1. \end{cases} \tag{4.7}$$

(The subscript L in (4.7) differs from N in (4.6) since the function (4.7) will also be used for other indices than N.) Extend $\widetilde{r}$ periodically in both dimensions by period M. Note that $\widetilde{r}$ can also be defined as the function that is M–periodic in both dimensions, and satisfies

$$\widetilde{r}(n) = r(n), \quad n \in \widetilde{G}(N), \tag{4.8}$$

where the grid $\widetilde{G}(N)$ is defined as

$$\widetilde{G}(N) = \{n = (n_1, n_2) \in \mathbb{Z}^2 : -(N-1) \leq n_1, n_2 \leq N-1\} \tag{4.9}$$

(see also Remark 4.2.1 below).

Next, consider a circular convolution operator $\widetilde{\Sigma}$, whose action on a vector $u(n), n \in G(M)$, is defined by

$$\widetilde{\Sigma}u(n) = \sum_{m \in G(M)} \widetilde{r}(m-n)u(m), \quad n \in G(M). \tag{4.10}$$

The two-dimensional DFT basis $\{e^{-i2\pi k\cdot(n/M)},\ n \in G(M)\},\ k \in G(M)$, where $n/M = (n_1/M, n_2/M)$ and $\cdot$ is the usual inner product, diagonalizes the operator $\widetilde{\Sigma}$. Hence, the eigenvalues of $\widetilde{\Sigma}$ are given by

$$\lambda_k = \sum_{n \in G(M)} \widetilde{r}(n)e^{-i2\pi k\cdot(n/M)}, \quad k \in G(M), \tag{4.11}$$

and can be computed efficiently using the two-dimensional FFT. Assuming that

$$\lambda_k \geq 0, \quad k \in G(M), \tag{4.12}$$

consider the complex-valued random variables

$$\widetilde{X}_n = M^{-1} \sum_{k \in G(M)} \lambda_k^{1/2}(Z_k^0 + iZ_k^1)e^{-i2\pi n\cdot(k/M)}, \quad n \in G(M), \tag{4.13}$$

where $Z_k^0,\ Z_k^1,\ k \in G(M)$, are independent standard Gaussian random variables. One can show that $\{\Re(\widetilde{X}_n), n \in G(M)\}$ and $\{\Im(\widetilde{X}_n), n \in G(M)\}$, that is, the real and imaginary parts of $\widetilde{X}_n$, are independent random fields with the covariance structure

$$\mathbb{E}\Re(\widetilde{X}_n)\Re(\widetilde{X}_{n+h}) = \mathbb{E}\Im(\widetilde{X}_n)\Im(\widetilde{X}_{n+h}) = \widetilde{r}(h), \quad n, n+h \in G(M). \tag{4.14}$$

By using (4.8) and (4.14), a subfield $X_n = \Re(\widetilde{X}_n)$ or $X_n = \Im(\widetilde{X}_n)$, $n \in G(N)$, is a Gaussian random field with the desired covariance structure r.

The construction of the variables $\widetilde{X}_n$ is possible only under the condition (4.12). This condition, however, often does not hold for some $k \in G(M)$, and hence the above standard circulant matrix embedding (CME) method fails. A common way to make it work is to increase N to some $\widetilde{N}$. (As noted in Section 4.1, the condition (4.12) holds in the limit $N \to \infty$, under mild assumptions.) It is convenient to think of $\widetilde{G}(\widetilde{N})\backslash\widetilde{G}(N)$ (or its periodization in $G(M)$ with $M = 2\widetilde{N} - 1$) as a *transition region*. Increasing N to $\widetilde{N}$ can thus be thought as extending the covariance function $r(n)$ over the transition region. Since increasing N to any computationally feasible $\widetilde{N}$ often does not ensure (4.12), it is natural to consider other ways to extend $r(n)$ over the transition region $\widetilde{G}(\widetilde{N})\backslash\widetilde{G}(N)$.

Remark 4.2.1. The reader may find the use of both domains $\widetilde{G}(N)$ and $G(M)$ excessive, and wonder why not use only one of them. In particular, for example in view of (4.6) and (4.8), the domain $\widetilde{G}(N)$ may seem to be preferred for simplicity. There are several reasons why we, and others, work on $G(M)$. First, it is natural to think of extending $G(N)$ to $G(M)$, where a subfield is ultimately selected. (In one dimension, a longer series is generated at times $0, 1, \ldots, M-1$, and a shorter series is selected of size N.) Second, the indexing by $G(M)$ is more common in scientific software, e.g. MATLAB. Third and more importantly in our context, we will refer below to the discontinuities of the covariance embedding at the boundary of periodization (see the beginning of Section 4.2.3). The discontinuities are easier to visualize on $G(M)$, not $\widetilde{G}(N)$.

4.2.2 Cutoff embedding

The CCE (Cutoff Circulant Embedding) method considers two extension schemes over the transition region, and is used in models with isotropic covariances $r(n) = \psi(||n||_2)$, where ψ is a function and $||\cdot||_2$ denotes the usual Euclidean distance. In one CCE extension (Gneiting et al. (2006)), $\widetilde{r}$ is defined on $\widetilde{G}(\widetilde{N})$ as

$$\widetilde{r}(n) = \begin{cases} r(n), & \text{if } 0 \le ||n||_2 \le \sqrt{2}N, \\ b_1(a_1 - ||n||_2^{1/2}), & \text{if } \sqrt{2}N \le ||n||_2 \le \widetilde{N} - 1, \end{cases} \tag{4.15}$$

and then periodically extended in both dimensions. The constants a_1, b_1 and $\widetilde{N}$ are chosen as

$$a_1 = (\sqrt{2}N)^{1/2} - (\sqrt{2}N)^{-1/2}\frac{1}{2}\frac{\psi(\sqrt{2}N)}{\psi'\sqrt{2}N}, \quad b_1 = -2(\sqrt{2}N)^{1/2}\psi'(\sqrt{2}N), \quad \widetilde{N} = [a_1^2], \tag{4.16}$$

where $[x]$ denotes the integer part of x. The choices (4.16) ensure that $\widetilde{r}(n)$ is once continuously differentiable at $t = \sqrt{2}N$, while the choice of $\sqrt{2}N$ in (4.15) is to have $\widetilde{r}(n) = r(n)$ for $n \in \widetilde{G}(N)$.

In another CCE extension, the embedding $\widetilde{r}$ is defined as

$$\widetilde{r}(n) = \begin{cases} r(n), & \text{if } 0 \le ||n||_2 \le \sqrt{2}N, \\ b_2(a_2 - ||n||_2)^2, & \text{if } \sqrt{2}N \le ||n||_2 \le \widetilde{N} - 1, \end{cases} \tag{4.17}$$

where the constants a_2, b_2 and $\widetilde{N}$ satisfy

$$a_1 = \sqrt{2}N - 2\frac{\psi(\sqrt{2}N)}{\psi'(\sqrt{2}N)}, \qquad b_1 = \frac{(\psi'(\sqrt{2}N))^2}{4\psi(\sqrt{2}N)}, \qquad \widetilde{N} = [a_2^2]. \tag{4.18}$$

After extending r into $\widetilde{r}$ as in (4.15) or (4.17), and periodization, the rest of CCE is the same as the algorithm in (4.10)–(4.13) with $M = 2\widetilde{N} - 1$. As shown in in Theorems 1 and 2 of Gneiting et al. (2006), under the extension schemes (4.15) and (4.17), the condition (4.12) always holds for several classes of covariance models.

4.2.3 Smoothing windows

The SWCE (Smoothing Windows Circulant Embedding) method is another way of extending the covariance function over the transition region. In the standard CME, the cross sections $\{(n_1, n_2) : n_1 = N-1,\ 0 \le n_2 \le M-1\}$ and $\{(n_1, n_2) : n_2 = N-1,\ 0 \le n_1 \le M-1\}$ are the boundary of periodization, when extending r periodically to $\widetilde{r}$ from $\widetilde{G}(N)$ to $\mathbb{Z}^2$ (and, in particular, to $\widetilde{r}$ on $G(M)$). Along this boundary, the covariance embedding $\widetilde{r}$ is often not smooth (see, for example, the top left plots in Figures 4.1–4.2). The idea of the SWCE method is to smooth the discontinuities at the boundary of periodization in the standard CME, since the discontinuities can be thought to be the source of negative

eigenvalues (of FFT). Helgason et al. (2014) showed that the SWCE method outperforms the CCE and standard CME for a range of covariance functions. In this section, we review briefly the two variants of the SWCE method, called the *overlapping* and *nonoverlapping windows*. For more details, see Helgason et al. (2014).

We begin with the definition of a smoothing window. Let $0 < P < Q$. A smoothing window is defined as

$$w(x) = \begin{cases} 1, & \text{if } x \in \widetilde{G}(P), \\ \rho(x), & \text{if } x \in \widetilde{G}(Q) \setminus \widetilde{G}(P), \\ 0, & \text{if } x \in \mathbb{R}^2 \setminus \widetilde{G}(Q), \end{cases} \tag{4.19}$$

where $\rho(x), x \in \mathbb{R}^2$, is a real-valued function that decays smoothly from 1 to 0 when moving from the boundary of $\widetilde{G}(P)$ toward that of $\widetilde{G}(Q)$.

In the *nonoverlapping* SWCE, one considers a transition region of side length $\widetilde{N} - N$ and an embedding size $M = 2\widetilde{N} - 1$. The covariance function is then extended through the transition region as

$$\widetilde{r}(n) = r(n)w(n), \quad n \in \widetilde{G}(\widetilde{N}), \tag{4.20}$$

where $w(n)$ is the smoothing window in (4.19) with

$$P = N, \quad \text{and} \quad Q = \widetilde{N}. \tag{4.21}$$

Since $w(n) = 1$ for $n \in \widetilde{G}(N)$, we have $\widetilde{r}(n) = r(n)$ for $n \in \widetilde{G}(N)$. The rest of the algorithm (after periodizing $\widetilde{r}$) is the same as CME in Section 4.2.1. The basic idea behind (4.20) is that $w(n)$ smooths $\widetilde{r}(n)$ at the boundary of periodization by forcing it to zero.

Note that (4.20) is the embedding on $\widetilde{G}(\widetilde{N})$, which is then periodized in both dimensions. The covariance embedding $\widetilde{r}(n)$ can equivalently be defined on $G(M)$ and then periodized by setting

$$\begin{aligned} \widetilde{r}(n) =& r(n_1, n_2)w_1(n) + r(n_1, n_2 - M)w_2(n) \\ &+ r(n_1 - M, n_2 - M)w_3(n) + r(n_1 - M, n_2)w_4(n), \quad n \in G(M), \end{aligned} \tag{4.22}$$

where $w_k(n) = w_k(n_1, n_2)$, $k = 1, 2, 3, 4$ and

$$\begin{aligned} w_1(n_1, n_2) &= w(n_1, n_2), & w_2(n_1, n_2) &= w(n_1, n_2 - M), \\ w_3(n_1, n_2) &= w(n_1 - M, n_2 - M), & w_4(n_1, n_2) &= w(n_1 - M, n_2). \end{aligned} \tag{4.23}$$

To see the equivalence, note that only one of the $w_i(n)$'s can be different from zero. See Section 3.2 in Helgason et al. (2014) for more details.

In the *overlapping* SWCE, the covariance embedding $\widetilde{r}(n)$ is smoothed without forcing it to be zero as in the nonoverlapping SWCE. The basic idea is to use the embedding given by (4.22)–(4.23) but choosing a smoothing window (4.19) with

$$P = N, \quad \text{and} \quad Q = 2\widetilde{N} - N. \tag{4.24}$$

The fact that $Q > \widetilde{N}$ in (4.24) ensures the overlap of $w_i(n)$'s. One can show that under (4.24), $\widetilde{r}(n) = r(n)$ for $n \in \widetilde{G}(N)$. The rest of the algorithm (after periodizing $\widetilde{r}$) is the same as CME in Section 4.2.1 using the embedding (4.22)–(4.23) with (4.24). The overlapping SWCE was found to outperform the nonoverlapping SWCE in Helgason et al. (2014).

4.3 Optimal circulant embedding

We propose here a new type of circulant matrix embedding method to generate a zero mean, stationary Gaussian random field X on the grid $G(N)$. Take $\widetilde{N} > N$ and set $M = 2\widetilde{N} - 1$. We shall also assume that the covariance embedding $\widetilde{r} = \{\widetilde{r}(n),\ n \in G(M)\}$ has the property $\widetilde{r}(n) = \widetilde{r}(-n)$, and satisfies the symmetry condition

$$\widetilde{r}(n) = \widetilde{r}(\xi_{\widetilde{N}}(n)), \quad n \in G(M), \tag{4.25}$$

where the function $\xi_{\widetilde{N}}(u)$ is given in (4.7). The symmetry condition (4.25) ensures, in particular, that the eigenvalues of the covariance embedding $\widetilde{r}$ are real-valued (see also the discussion following Lemma 4.1 below).

4.3.1 Formulation of the constrained optimization problem

The basic idea of the OCE (Optimal Circulant Embedding) method is to obtain the covariance embedding $\widetilde{r} = \{\widetilde{r}(n),\ n \in G(M)\}$, or by periodization $\widetilde{r} = \{\widetilde{r}(n),\ n \in \widetilde{G}(\widetilde{N})\}$, by solving the following optimization problem:

$$\begin{aligned} \min_{\widetilde{r}} \quad & f^*(\widetilde{r}) = \sum_{n \in \widetilde{G}(N)} \big(r(n) - \widetilde{r}(n)\big)^2, \\ \text{subject to} \quad & g_k(\widetilde{r}) \geq 0, \quad k \in \widetilde{G}(\widetilde{N}), \end{aligned} \tag{4.26}$$

where $r = \{r(n), n \in \widetilde{G}(N)\}$ is a given covariance structure and $g_k(\widetilde{r})$ are the eigenvalues of the covariance embedding $\widetilde{r}$, given by (after replacing $G(M)$ by $\widetilde{G}(\widetilde{N})$ in (4.11) for convenience and using $\widetilde{G}(\widetilde{N})$ for indexing)

$$g_k(\widetilde{r}) = \sum_{n \in \widetilde{G}(\widetilde{N})} \widetilde{r}(n) e^{-i2\pi k \cdot (n/M)}, \quad k \in \widetilde{G}(\widetilde{N}). \tag{4.27}$$

Note that the sum in the first relation in (4.26) is taken over a smaller grid $\widetilde{G}(N)$ than $\widetilde{G}(\widetilde{N})$ of the sum in (4.27). Minimization of f forces the covariance embedding $\widetilde{r}$ to be as close as possible to the targeted covariance r over the domain of interest $\widetilde{G}(N)$. Hence, when the minimum value of f is zero, the solution of the problem can be used for the exact synthesis of the desired field. The inequality constraints in (4.26) rule out candidate (non-feasible) minimizers that lead to covariance embeddings with negative eigenvalues.

To write the problem (4.26) on the domain $G(M)$, introduce the weights

$$\beta(n) = \begin{cases} 1, & n \in \widetilde{G}(N), \\ 0, & n \in \widetilde{G}(\widetilde{N}) \backslash \widetilde{G}(N), \end{cases} \tag{4.28}$$

on $\widetilde{G}(\widetilde{N})$ and extend them periodically in both dimensions. With r extended periodically in the same way, the problem (4.26) can be written as

$$\begin{aligned} \min_{\widetilde{r}} \quad & f^*(\widetilde{r}) = \sum_{n \in G(M)} \beta(n) \big(r(n) - \widetilde{r}(n)\big)^2, \\ \text{subject to} \quad & g_k(\widetilde{r}) \geq 0, \quad k \in G(M). \end{aligned} \tag{4.29}$$

The optimization problem (4.29) is a quadratic programming problem with linear inequality constraints. In the optimization literature, this problem is treated by regarding the underlying variable as a vector. In the case considered here, the underlying variable $\widetilde{r}$ is indexed by the square grid $G(M)$. Indexing is just a question of perspective and we will continue using parameters indexed by grids (though also sometimes switch to vectors, as for example, in one of the proofs).

An issue requiring immediate attention, however, is the symmetry of the covariance embedding $\widetilde{r}$, expressed through the relation (4.25). We need to keep track only of the non-symmetric half of the covariance, which we choose to be over the smaller rectangular grid

$$G_+(M) = \{n \in \mathbb{Z}^2,\ 0 \le n_1 \le \widetilde{N} - 1,\ 0 \le n_2 \le M - 1\}. \tag{4.30}$$

Note that the symmetry relation (4.25) implies that, for $(n_1, n_2) \in G(M) \backslash G_+(M)$, $\widetilde{r}(n_1, n_2) = \widetilde{r}(M - n_1, M - n_2)$, so that the values of $\widetilde{r}(n)$ for $n \in G(M) \backslash G_+(M)$ are indeed determined from those on $G_+(M)$.

When working on the smaller rectangle $G_+(M)$, we need to have a representation of the eigenvalues $g_k(\widetilde{r})$ where the sum $\sum_{n \in G(M)}$ is replaced by $\sum_{n \in G_+(M)}$, that is, a representation of the form

$$g_k(\widetilde{r}) = \sum_{n \in G_+(M)} c_k(n) \widetilde{r}(n) =: [A\widetilde{r}](k), \tag{4.31}$$

where $c_k(n)$ are suitable weights and A is viewed as a linear operator acting on a field defined on $G_+(M)$. The next lemma provides an expression for the weights $c_k(n)$. The proof of the lemma can be found in Appendix B.1.

Lemma 4.1. *Suppose that (4.25) holds, and the eigenvalues $g_k(\widetilde{r})$ are given by* (4.11). *Then, the relation* (4.31) *holds with the weights $c_k(n),\ k \in G(M),\ n \in G_+(M)$, given by*

$$c_k(n) = c_{(k_1,k_2)}(n_1, n_2) = \begin{cases} 2\cos(2\pi k \cdot (n/M)), & \text{if } n_1 \neq 0, \\ \cos(2\pi k_2 n_2 / M), & \text{if } n_1 = 0. \end{cases} \tag{4.32}$$

A number of comments are in place concerning Lemma 4.1. First, note that the relation (4.32) implies that $c_{(M,M)-k}(n) = c_k(n)$. Therefore,

$$g_k(\widetilde{r}) \geq 0, \quad k \in G_+(M), \tag{4.33}$$

yields nonnegative $g_k(\widetilde{r})$ for general k in the larger grid $G(M)$. Thus, the number of inequality constraints in (4.29) can be reduced by half. Second, it also follows from (4.32) that g_k are real-valued.

We will not use Lemma 4.1 to calculate the eigenvalues $g_k(\widetilde{r}) = [A\widetilde{r}](k)$ – these can be computed efficiently by using the two-dimensional FFT. The relation (4.31), however, plays a significant role in the practical implementation of the OCE method in the following sense. As can be seen below in Section 4.3.2, the algorithm we use to solve the problem (4.29) requires multiple evaluations of both $[A\widetilde{r}](k)$ and $[A^T\widetilde{r}](k)$, where A^T refers to the adjoint operator of A (if A is viewed as a matrix, A^T is its transpose). We show in the next lemma that $[A^T\widetilde{r}](k)$ can be computed using $[A\widetilde{r}](k)$ and hence FFT. The proof of the lemma is based on (4.31) and can be found in Appendix B.1.

Lemma 4.2. *The operators A and A^T are related as*

$$[A^T y](k) = [Ay](k) + E_k, \quad k \in G_+(M), \tag{4.34}$$

where

$$E_k = \begin{cases} -\sum_{n_2=0}^{M-1} \cos(2\pi k_2 n_2/M) \left[\sum_{n_1=1}^{\widetilde{N}-1} y(n_1, n_2)\right], & \text{if } k_1 = 0, \\ \sum_{n_2=0}^{M-1} \cos(2\pi k_2 n_2/M) y(0, n_2), & \text{if } k_1 \neq 0. \end{cases}$$

Note that E_k in the lemma above can also be calculated efficiently using FFT. We are now ready to present the method to solve the optimization problem (4.29), which by the discussion around (4.33) becomes

$$\begin{aligned} \min_{\widetilde{r}} \quad & f(\widetilde{r}) = \sum_{n \in G_+(M)} \beta(n)\big(r(n) - \widetilde{r}(n)\big)^2, \\ \text{subject to} \quad & g_k(\widetilde{r}) \geq 0, \quad k \in G_+(M). \end{aligned} \tag{4.35}$$

4.3.2 Primal log-barrier method

To solve (4.35), we use the *primal log-barrier* (PLB) method. We outline the method below in order to refer to its parts that will be specialized in our problem. For a more detailed account, see Chapter 11 in Boyd and Vandenberghe (2004), and Section 19.6 in Nocedal and Wright (2006).[1]

We start by recalling some convex optimization notions from Boyd and Vandenberghe (2004), p. 128, adjusting the notation suitably for two-dimensional fields. The *optimal value* z^* of the problem (4.35) is defined as $z^* = \inf\{f(\widetilde{r}) \mid g_k(\widetilde{r}) \geq 0,\ k \in G_+(M)\}$. We will say that a (field) point $\widetilde{r}$ is *feasible* if $g_k(\widetilde{r}) \geq 0,\ k \in G_+(M)$ (if $g_k(\widetilde{r}) > 0,\ k \in G_+(M)$, the point $\widetilde{r}$ is called strictly feasible). The set of all feasible points will be called *feasible region*. A (field) point $\widetilde{r}^*$ is said to be an *optimal point*, or to solve the problem (4.35) if it is feasible and $f(\widetilde{r}^*) = z^*$. Moreover, a feasible point $\widetilde{r}$ with $f(\widetilde{r}) \leq z^* + \epsilon$ ($\epsilon > 0$) is called $\epsilon-$*suboptimal*.

To ease the presentation of the PLB method, we break it down into 3 parts described next. The actual steps of the PLB algorithm are outlined at the end of this section.

Part 1: Eliminating the constraints

Define the *logarithmic barrier* function

$$\phi(\widetilde{r}) = -\sum_{k \in G_+(M)} \log(g_k(\widetilde{r})), \tag{4.36}$$

with the domain $\{\widetilde{r} = \widetilde{r}(n) \in G_+(M) \mid g_k(\widetilde{r}) > 0, k \in G_+(M)\}$. To eliminate the constraints in (4.35), the logarithmic barrier ϕ is introduced as a penalty term in the objective function of (4.35). More specifically, let $t > 0$ and consider the unconstrained problem

$$\min_{\widetilde{r}} \quad f_t(\widetilde{r}) := tf(\widetilde{r}) + \phi(\widetilde{r}). \tag{4.37}$$

If $g_k(\widetilde{r}) < 0$, then by convention the value of f_t in (4.37) is ∞.

[1]The reader familiar with optimization methods may skip to the end of the section where the use of FFT in the CG method is discussed.

One can show that the minimizers of (4.37), which we denote by $\widetilde{r}_t^*(n)$, approach a solution of (4.35) as t grows, under certain conditions (see, for example, Theorem 3.10 in Forsgren et al. (2002), p. 548, for a detailed proof and Boyd and Vandenberghe (2004), pp. 562-563, for an intuitive discussion). The points $\widetilde{r}_t^*(n)$ form the so-called *central path* $\{\widetilde{r}_t^*(n), t > 0\}$, and are at most m/t*–suboptimal* for the problem (4.35), i.e.

$$f(\widetilde{r}_t^*(n)) - z^* \leq m/t, \tag{4.38}$$

where $m = \widetilde{N}M$ is the number of inequality constraints in (4.35). See Boyd and Vandenberghe (2004), pp. 565-566, for a proof.

Part 2: Quadratic approximation

We now turn to solving the unconstrained optimization problem (4.37). It is convenient here to view each f_t as a function whose argument is a vector $x \in \mathbb{R}^m$, where m is the number of points in the grid $G_+(M)$. Consider the second-order Taylor approximation $\widehat{f_t}$ of f_t around some given $x = x_0$ (for a fixed t)

$$\widehat{f_t}(x+v) = f_t(x) + \nabla f_t(x)v + \frac{1}{2}v^T\nabla^2 f_t(x)v, \tag{4.39}$$

where ∇f_t and $\nabla^2 f_t$ denote the gradient and Hessian of f_t, respectively. Instead of minimizing the function $f(x)$, we will choose a point x_0 and find the direction v that minimizes the Taylor approximation of f_t around x_0. In other words, we will solve the quadratic problem

$$\min_v \quad \nabla f_t(x)v + \tfrac{1}{2}v^T\nabla^2 f_t(x)v \tag{4.40}$$

for $x = x_0$. In the optimization literature, the vector v is called the *Newton direction* and the process of calculating v is called the *Newton step*. Carrying out multiple Newton steps yields a sequence of points that converges to the solution of the exact problem (4.37).

For our problem, however, since t is increasing, it is not necessary to find exact minimizers of the problem (4.37). In fact, it is sufficient to calculate *only one* Newton direction for each value of t and initial value $x = x_0$. This fact is argued in Bertsimas and Tsitsiklis

(1997) for linear objective functions (see the discussion under relation (9.19) on p. 424 and figure 9.6 on p. 425) and in Boyd and Vandenberghe (2004), p. 570, for general convex functions.

As expected, the approximate problem (4.40) is much easier to solve. The first-order condition linear system is

$$Hv = b, \tag{4.41}$$

where H and b are given by

$$H = \nabla^2 f_t(x), \qquad b = -\nabla f_t(x). \tag{4.42}$$

Thus, the minimizer of $\widehat{f}_t$, for a fixed t and initial point $x = x_0$, is given by

$$\widehat{x}^* = x_0 + v^*, \tag{4.43}$$

where v^* is the solution of (4.41).

Part 3: Conjugate gradient algorithm

The *conjugate gradient* (CG) algorithm is an iterative procedure for solving linear systems of the form (4.41) with symmetric and positive definite matrices H. It is particularly appropriate for large problems, which is the case with (4.35) even for moderate M. Moreover, in our case, the key steps of the algorithm can be implemented more efficiently through FFT. We give next a description of the CG method following Nocedal and Wright (2006), p. 112.

Outline of the CG algorithm

Given v_0; Set $\epsilon_0 = Hv_0 - b$, $p_0 = \epsilon_0$, $k = 0$;

while $\epsilon_k \neq 0$

$$\begin{aligned}
\alpha_k &\leftarrow \frac{\epsilon_k^T \epsilon_k}{p_k^T H p_k}; \\
v_{k+1} &\leftarrow v_k + \alpha_k p_k; \\
\epsilon_{k+1} &\leftarrow \epsilon_k + \alpha_k H p_k; \\
s_{k+1} &\leftarrow \frac{\epsilon_{k+1}^T \epsilon_{k+1}}{\epsilon_k^T \epsilon_k}; \\
p_{k+1} &\leftarrow -\epsilon_{k+1} + s_{k+1} p_k; \\
k &\leftarrow k+1;
\end{aligned}$$

end (**while**)

The major computational tasks to be performed at each iteration of the CG algorithm, are calculating the products Hp_k, $p_k^T H p_k$ and $\epsilon_{k+1}^T \epsilon_k$. In our case, however, these calculations can be done more efficiently using FFT. This is possible because, as shown in Lemma 4.3 below, both the Hessian matrix H and the negative gradient b can be expressed as functions of the linear operators A and A^T, which can be implemented using FFT by Lemmas 4.1 and 4.2. Indeed, note that a direct matrix-vector calculation of Hp is of the order m^2 for an $m \times m$ matrix H and an $m \times 1$ vector p. In our problem, m is of the order N^2 and hence m^2 of the order N^4. On the other hand, if H refers to FFT and p to a field, the order of calculation becomes $N^2 \log(N)$.

In the practical implementation of the CG algorithm, we considered the stopping rule $\epsilon_k < \text{TOL}$ for a tolerance $\text{TOL} = 0.1$. For all covariance functions considered in Section 4.4, we found that calculating the Newton step with higher precision (taking smaller tolerance) led to the same optimal value for f, requiring however a larger number of steps.

Before we state Lemma 4.3, some notation is necessary. Given a field $r = \{r(n), n \in G_+(M)\}$, define $\{d(k), k \in G_+(M)\}$ to be the field whose kth element is given by

$$d(k) = -([Ar](k))^{-1} = -1/([Ar](k)). \tag{4.44}$$

Next, let $D := \mathrm{diag}(d)$ be a linear operator whose action on the field r is defined as

$$[Dr](k) = d(k) \cdot r(k), \quad k \in G_+(M). \tag{4.45}$$

We are now ready to state Lemma 4.3. The proof of the lemma can be found in Appendix B.1.

Lemma 4.3. *Let $\beta(n)$ be the weights defined in (4.28). Then, the Hessian matrix H and the negative gradient b in (4.41) can be viewed as a linear operator and a two-dimensional field given by, respectively,*

$$H = tW + A^T D^2 A, \quad \textit{and} \quad b(n) = t(W\widetilde{r}(n) - s(n)) + [A^T d](n), \tag{4.46}$$

where $W = \mathrm{diag}(\beta(n))$ and $s(n) = \beta(n)r(n)$.

We conclude with a general outline of the primal log-barrier method. Further comments can be found in Section 4.3.3. Let $\widetilde{r}^*_{t,a}$ denote the field analogue of the approximate minimizer $\widehat{x}^*$ in (4.43), i.e. the solution of the optimization problem

$$\min_{\widetilde{r}} \widehat{f}_t(\widetilde{r}). \tag{4.47}$$

Steps of the PLB method

1. Find an initial strictly feasible point $\widetilde{r}^0$ for the problem (4.35). Pick constants $t_0 > 0$, $\mu > 1$, and accuracy $\epsilon > 0$. Set $t = t_0$ and $\widetilde{r} = \widetilde{r}^0$.

2. Compute $\widetilde{r}^*_{t,a}$ by solving the problem (4.47) or equivalently (4.41) with initial point $\widetilde{r}$, using the CG algorithm.

3. Update $\widetilde{r} = \widetilde{r}^*_{t,a}$. If $m/t < \epsilon$, stop and return $\widetilde{r}$; since $f(\widetilde{r}^*_{t,a})$ $(\approx f(\widetilde{r}^*_t))$ does not differ from the objective at the optimum by more than m/t, we can ensure that $f(\widetilde{r})$ is less than ϵ away from the optimal value.

4. Increase t to μt and start again from Step 2.

4.3.3 Further discussion on the PLB method

In this section, we discuss the choice of the parameters $\widetilde{r}_0$, t_0 and μ above. The initial covariance embedding $\widetilde{r}_0$ needs to be strictly feasible (i.e. $[A\widetilde{r}_0](k) > 0$ for all k). We describe below a straightforward way to choose such a point. The roles of t_0 and μ, on the other hand, are less straightforward. Note that the domain of the function $\widehat{f}_t$ is different from the feasible region of the constrained problem (4.35). This means that the CG algorithm used to find approximate minimizers of f_t will possibly result in minimizers of $\widehat{f}_t$ which are not feasible points for (4.35). The parameters t_0 and μ are selected to mitigate this effect.

Choice of $\widetilde{r}_0$: To find a strictly feasible initial point $\widetilde{r}^0$ in Step 1 of the PLB method, we first calculate the standard covariance embedding $\widetilde{r}(n),\ n \in G(M)$, as defined through the relation (4.6). Let $F(k) := [A\widetilde{r}](k),\ k \in G(M)$, be the corresponding spectrum (which can be calculated efficiently through FFT). We then shift the spectrum by a positive constant c so that all of its elements become positive (taking $c = 2|\min_{k\in G(M)} F(k)|$ will ensure that). Denote the resulting spectrum by $F_+(k)$. Finally, let $y(n), n \in G(M)$, denote the inverse FFT of $F_+(k)$. Taking $\widetilde{r}^0(n) = r(0)y(n)/y(0)$ yields a strictly feasible initial point with the targeted scale $\widetilde{r}^0(0) = r(0)$.

Choice of t_0: Observe that $\widehat{f}_t$ can be written as

$$\widehat{f}_t(\widetilde{r}) = tf(\widetilde{r}) + \widehat{\phi}(\widetilde{r}), \tag{4.48}$$

where $\widehat{\phi}$ denotes the second-order Taylor approximation of the logarithmic barrier ϕ (defined in a similar manner as $\widehat{f}_t$ in (4.39)). Selecting a small t_0 essentially diminishes the contribution of f to $\widehat{f}_t$ and enhances the role of $\widehat{\phi}$, in the early steps of the PLB method. For points outside the feasible region, the value of ϕ is (by our convention) infinite, and thus we expect the value of $\widehat{\phi}$ to be very large. As a result, these points will not be favored by the CG algorithm in the initial iterations of the PLB method. For our simulations (see Section 4.4), we used a small $t_0 = 0.0001$, which proved to work well for all the covariance

structures considered. See Boyd and Vandenberghe (2004), pp. 570-571, for more elaborate choices for t_0.

Choice of μ: When μ is small (near 1), two consecutive t's in the PLB method are not very different. As a result, the problem (4.47) is perturbed only slightly from one iteration to another. Moreover, the update rule in Step 3 of the algorithm implies that the initial point used when solving the problem (4.47) is the solution of the previous iteration. Since the two problems are only slightly different, we expect this choice of initial point to be a good one, and therefore the conjugate gradient algorithm will require less time to find the solution. The downside of this strategy is that a large number of iterations will be required to reach the suboptimality property $m/t < \epsilon$ in Step 3 of the method.

On the other hand, when μ is large, t is increased rapidly at each PLB iteration and thus the convergence to the suboptimality property is much faster. This aggressive update of t, however, means that there will be a large difference between the problems (4.47) at two consecutive iterations, and therefore the solution of one iteration might not be a good initial point for the next. This might decrease the quality of the Taylor approximation, potentially leading to a solution outside the feasible region.

When t is updated aggressively (large μ), note also that the first term in (4.48) dominates the objective function and thus the CG algorithm focuses more on iterates that minimize f in (4.35) and less on iterates that satisfy the constraints in (4.35). In other words, large values of μ are more likely to lead to an exact circulant embedding, which however might have some negative eigenvalues. Taking smaller values of μ will ensure that the eigenvalues are nonnegative, however leading to approximate embeddings. We discuss these situations further in Section 4.4.

4.4 Simulations

In this section, we provide a Monte Carlo study comparing the OCE and SWCE methods, as well as numerical illustrations of all the circulant embedding methods discussed above. For our comparisons, we consider the *anisotropic* covariance function of the powered exponential

form

$$r(n) = e^{-(\theta||n||_W)^\alpha}, \quad 0 < \alpha \leq 2, \quad \theta > 0, \tag{4.49}$$

where $||t||_W := \sqrt{t'Wt}$ for column vectors t and W is a symmetric positive definite 2×2 matrix, and the *isotropic* Cauchy covariance function of the form

$$r(n) = (1 + (\theta||n||_2)^\alpha)^{-\beta/\alpha}, \quad 0 < \alpha \leq 2, \quad \beta > 0, \quad \theta > 0. \tag{4.50}$$

For the powered exponential covariance structure (4.49), we consider two configurations, with $\alpha = 0.5,\ \theta = 0.01$ and $W = W_1$ or W_2, where

$$W_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad W_2 = \begin{pmatrix} 1.6388 & -1.489 \\ -1.489 & 1.3712 \end{pmatrix}. \tag{4.51}$$

(While there is nothing particularly special about the choice of W_1, the matrix W_2 is chosen to be nearly singular as discussed in Section 4.4.1 below.) For the Cauchy covariance structure (4.50), we consider the following parameter values

$$\alpha = 1.3, \quad \beta = 0.01, \quad \theta = 0.01. \tag{4.52}$$

For the SWCE method, overlapping windows are used throughout.

4.4.1 Powered exponential covariance

Figure 4.1 shows the covariance embeddings resulting from the standard CME, SWCE and OCE procedures for $N = 200$ and $\widetilde{N} = 240$ (transition region length 40) when the matrix W_1 is used in (4.49). In the standard CME (top left), the covariance embedding is not smooth at the boundary of periodization resulting in many negative eigenvalues (77,300). The SWCE method (top right), as expected, smooths the covariance embedding at the boundary. However, for this transition region length, some of the eigenvalues (50) are still negative. Finally, the covariance embeddings of the OCE method (middle left and middle right plots) have no negative eigenvalues, while their values within the corner areas marked

with broken lines are those of the desired covariance (up to the indicated error $\min f$). These plots visualize the goal of the optimization problem (4.26). The two bottom plots show 2 cross sections of the covariance embeddings from all 3 methods.

Figure 4.2 shows the covariance embeddings resulting from the standard CME, SWCE and OCE procedures for $N = 100$ and $\widetilde{N} = 359$ (transition region length 80) when the matrix W_2 is used in (4.49). The matrix W_2 is nearly singular having eigenvalues 0.01 and 3. This singularity causes more pronounced discontinuities at the boundary of periodization – c.f. the standard CME plots in Figures 4.1–4.2 (upper left). The SWCE method smooths the discontinuities. However, the resulting embedding still has some negative eigenvalues (116). Moreover, even though the length of the transition region here is larger compared to Figure 4.1, the number of negative eigenvalues has increased. This is due to the ill behavior of the weight matrix W. On the other hand, the two middle plots of Figure 4.2 depict the covariance embedding of the OCE method with no negative eigenvalues. The two bottom plots show 2 cross sections of the covariance embeddings from all 3 methods.

4.4.2 Cauchy covariance

Figure 4.3 shows the covariance embeddings resulting from the CME, SWCE and OCE procedures for the Cauchy covariance function (4.50) with parameters (4.52) and $N = 200$. The transition region length for the upper right (SWCE) and middle left (OCE) covariance embeddings is 40, while for the middle left (SWCE) is 400. As in Figures 4.1–4.2, the covariance embedding resulting from the standard CME method (upper left) is not smooth at the boundary of periodization (2,836 negative eigenvalues). Note that non-smoothness here is in the derivative (and not in the covariance itself as in Figures 4.1–4.2, top left) as indicated through a highlighted contour plot which is spiky at the boundary.

In the top-right plot where the transition region length is 40, the SWCE method results in an embedding with 24,720 negative eigenvalues. One possible explanation for this is a very small range of the values of the given covariance function (the range of approximate size 0.05 as seen from the vertical scale). That is, when overlapping windows are used in the SWCE, the values of the covariance function are superimposed according to (4.22) and as in the case considered here, can result in larger values over the transition region (the

white region in the middle of the top-right plot). In fact, the SWCE method still fails to produce a covariance embedding with nonnegative eigenvalues even when $\widetilde{N} = 600$ (for example, the smoothing covariance embedding with $\widetilde{N} = 400$ shown in the middle right plot has 28 negative eigenvalues). The OCE method, on the other hand, finds a feasible covariance embedding even when $\widetilde{N} = 240$. The two bottom plots show 2 cross sections of the covariance embeddings from all 3 methods.

4.4.3 Efficiency and related issues

We compare here the performance of the SWCE and OCE methods. Following Helgason et al. (2014), define the efficiency of the embedding as

$$\gamma = \gamma(M) = \frac{2N^* - 1}{M}, \tag{4.53}$$

where $M = 2\widetilde{N} - 1$ is the embedding size. For the SWCE method, $N^*(\leq N)$ is the largest size for which the covariance embedding has nonnegative eigenvalues (and hence for which the field can be generated exactly on the grid $G(N^*)$). The efficiency γ satisfies $0 < \gamma \leq 1$ and the closer γ is to 1, the larger the grid $G(N^*)$ one can simulate the field on.

For the OCE, the value N^* can be defined in several ways. One possibility is to take N^* as the largest value for which the covariance embedding yields the objective function smaller than some small value, for example, 10^{-5}. Another possibility which we find more informative, is to take N^* as the largest value for which the OCE method results in a feasible embedding (i.e. an embedding with nonnegative eigenvalues) – see Section 4.3.3 for a related discussion. The value of N^*, however, needs to be supplemented by the value of the objective function since the latter is not necessarily small. This choice of N^* is used below, though we also comment on what N^* is expected under the first possibility above.

In Figures 4.4–4.5, we compare the embedding efficiency of the OCE (solid line) and the SWCE (dashed line) methods. Figure 4.4 compares the embedding efficiency for the powered exponential covariance (4.49) as a function of $M\theta$ (this choice, rather than θ, allows a nicer scale on the $x-$axis). The comparisons were done for 3 different values of $\alpha = 0.5, 1, 1.5$ (corresponding to the three rows), $\theta = 0.05k,\ k = 1, \ldots, 10$, and embedding size $M = 101$.

The OCE method was implemented for $\mu = 2$ (left column) and $\mu = 3$ (right column), and the optimum values of f in (4.35) are given below each point on the solid lines. The OCE method performs significantly better than the SWCE in all cases considered (the solid lines are located above the dashed ones). Note also that as α decreases, the performance of the SWCE method gets worse, while the OCE method is only slightly affected.

Focusing on the OCE method, the plots illustrate nicely the tradeoff in the choice of μ discussed in Section 4.3. Observe, for example, in the second (middle) row of Figure 4.4, that the embedding efficiency γ has higher values for $\mu = 2$ (left) than for $\mu = 3$ (right). This means that, for the given covariance $r(n), n \in \widetilde{G}(N)$, and the given size M, taking $\mu = 2$ allows the synthesis of the desired field on larger grids, than taking $\mu = 3$. The synthesized fields, however, will not necessarily be exact, since the optimum value of the objective function is often non-negligible (e.g. for $\mu = 2$ and $M\theta = 20$ the optimum value is 1). On the other hand, the optimum values of the objective function f in (4.35) are lower for $\mu = 3$ (right plots) illustrating the fact that larger values of μ are more likely to yield exact embeddings.

Note also that, for example, all the values of the objective function in the top-right plot ($\mu = 3$) of Figure 4.4 are smaller than 10^{-5}. If we defined the efficiency for the OCE method requiring the objective function to be smaller than 10^{-5} at N^*, the resulting top-left plot ($\mu = 2$) of Figure 4.4 would naturally have a lower efficiency for the OCE method. We find, however, that this efficiency would still be higher than the one for $\mu = 3$ in the top-right plot. The same observation applies to other plots concerning the cases $\mu = 2$ and $\mu = 3$.

Remark 4.4.1. A field $X = \{X_n, n \in G(N)\}$ on grids of small size ($N < 100$) can be synthesized exactly using Cholesky decomposition. However, even a fast variant of the method developed by Dietrich (1993) has complexity $O(N^5)$, and thus is not suitable for the synthesis of fields on grids of size larger than $N = 100$ (see Table 2 in Gneiting et al. (2006), p. 485, for a detailed comparison of the computational requirements of Cholesky decomposition and standard CME methods). On the other hand, the OCE method works well for grids of size up to $N = 1,000$. Note also that the speed of the OCE method varies depending on how strong the discontinuities of the covariance embedding are. For

the Cauchy covariance of Figure 4.3 for example, where $N = 200$ and $M = 479$, the OCE method for $\mu = 2.5$ needs 150 seconds to produce a feasible covariance embedding.[2] On the other hand, for the powered exponential covariance function as in Figure 4.2 but for $N = 60$ and $M = 215$, the time needed is almost the same (140 secs) even though the grid size is much smaller.

Remark 4.4.2. Regarding the computational requirements of the SWCE and OCE methods, at first glance, the SWCE method seems to be much faster. As discussed above, however, the SWCE method can break down for small transition regions. In fact, the minimum transition region length needed for the method to work is not known in advance. This means that to produce a feasible covariance embedding, it may be necessary to perform the SWCE method multiple times, each time increasing the transition region length until a feasible covariance embedding is produced. Although this trial and error approach can be optimized, it is still likely that a large number of transition region lengths will need to be tested. In other words, the OCE and SWCE methods should not be compared in speed solely for fixed transition regions leading to feasible embeddings, where the SWCE method is likely to be faster (even if the corresponding transition region is larger). One should also take into account the uncertainty in the size of the transition region, where the OCE method is the favorite (in allowing for smaller regions).

Finally, note again that the approach taken in this chapter stresses nonnegative eigenvalues of covariance embeddings and focuses on the values of the objective function. Related to this, it is natural to ask how our approach compares to the following simple procedure. If the standard CME fails, a straightforward way to obtain a feasible covariance embedding $\widetilde{r}_N$ is by setting all negative eigenvalues of the CME covariance $\widetilde{r}_I$ equal to 0. Variations of this are suggested, for example, in Chan and Wood (1997), p. 167. Moreover, by using Parseval's identity, note that

$$\underset{\widetilde{r}}{\operatorname{argmin}} \, ||\widetilde{r}_I - \widetilde{r}||_2 = \underset{\widetilde{r}}{\operatorname{argmin}} \, ||g(\widetilde{r}_I) - g(\widetilde{r})||_2 = \widetilde{r}_N,$$

[2] All simulations were performed in an Intel Core i5 processor with 4GB of RAM.

where $||\cdot||_2$ denotes the Frobenius norm, $g(\widetilde{r}) = \{g_k(\widetilde{r}), k \in \widetilde{G}(N)\}$ are the eigenvalues of the covariance embedding $\widetilde{r}$ and the minimum is taken over all covariance embeddings $\widetilde{r}$ having nonnegative eigenvalues. In other words, $\widetilde{r}_N$ is the solution to our optimization problem (4.26) when $\widetilde{N} = N$.

Indeed, when the transition region is empty ($\widetilde{N} = N$), the OCE method results in the covariance embedding $\widetilde{r}_O$ very close to $\widetilde{r}_N$. On the other hand, when $\widetilde{N} > N$, the OCE method can result in significantly smaller optimal values of the objective function (which moreover can be made smaller if desired). For example, in the case of the powered exponential covariance function of Figure 4.1, the optimal value in the OCE method is less than 10^{-6}, while the above modification of the CME method leads to the Frobenius distance between $\widetilde{r}_I$ and $\widetilde{r}_N$ equal to 24.

4.5 Figures

We gather below the figures referred to in Section 4.4.

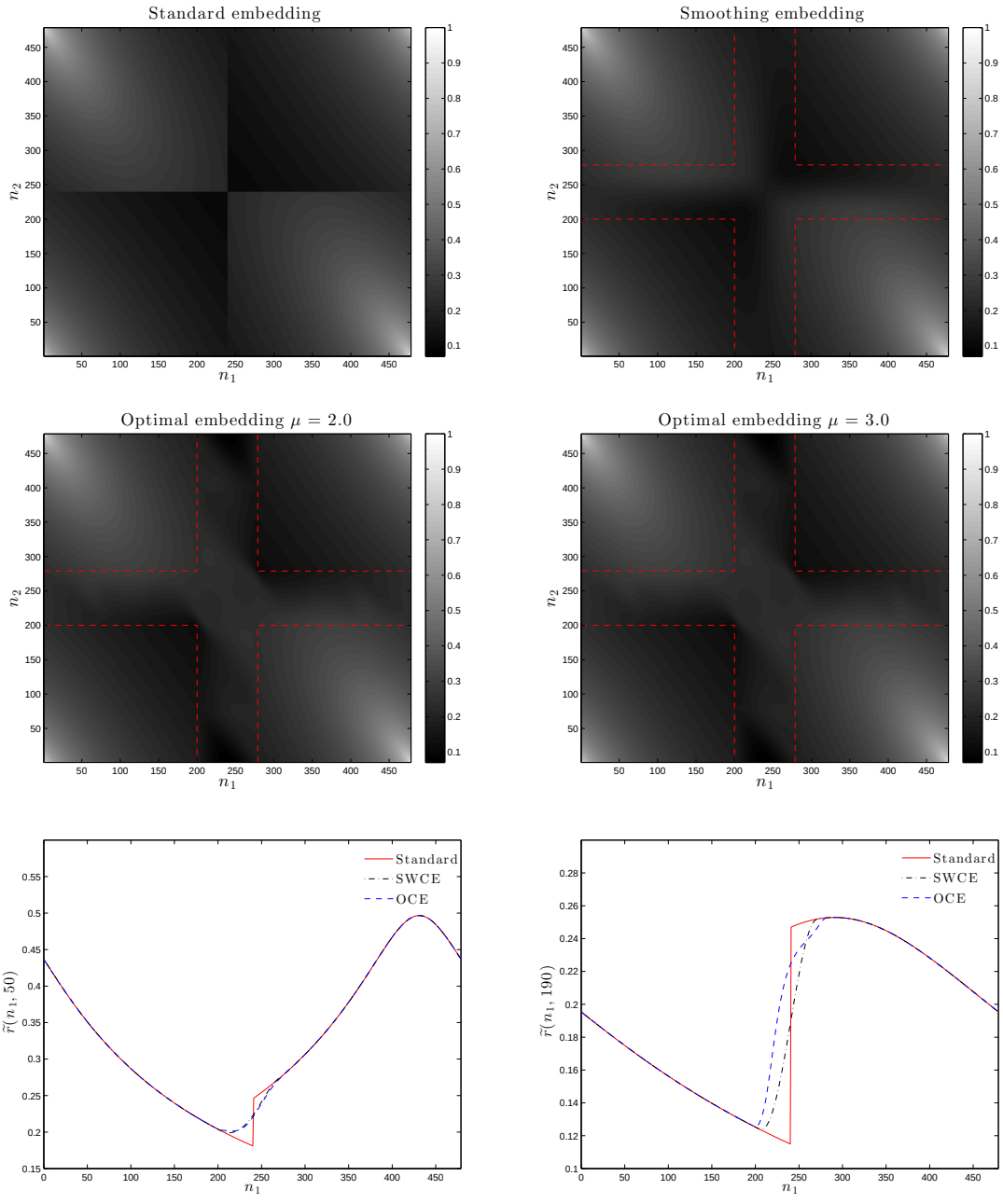

Figure 4.1: Plots for the anisotropic covariance of the powered exponential form in (4.49) with $W = W_1$ in (4.51), $N = 200$ and $M = 479$. Top left: Standard embedding. Top right: Smoothing windows embedding with 50 negative eigenvalues. Middle left: Optimal embedding with $\mu = 3$ and $\min f = 6.7 \cdot 10^{-7}$. Middle right: Optimal embedding with $\mu = 1.5$ and $\min f = 8 \cdot 10^{-7}$. Bottom plots: Cross sections of standard CME, SWCE and OCE ($\mu = 1.5$) for $n_2 = 50$, 190.

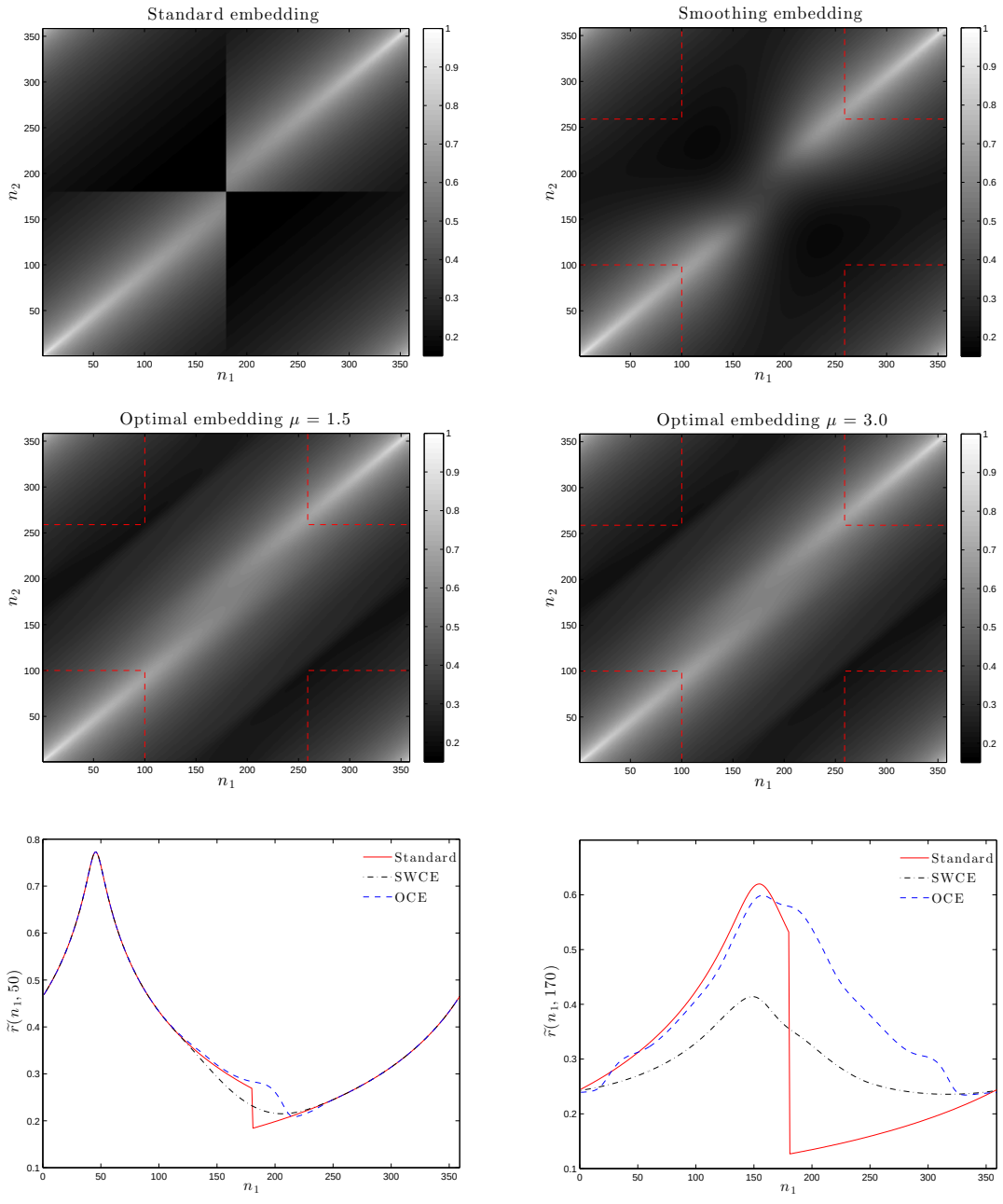

Figure 4.2: Plots for the anisotropic covariance of the powered exponential form in (4.49) with $W = W_2$ in (4.51) and $N = 100$ and $M = 359$. Top left: Standard embedding. Top right: Smoothing windows embedding with 116 negative eigenvalues. Middle left: Optimal embedding with $\mu = 1.5$, $\min f = 5 \cdot 10^{-6}$, and no negative eigenvalues. Middle right: Optimal embedding with $\mu = 3$, $\min f = 4 \cdot 10^{-6}$, and no negative eigenvalues. Bottom plots: Cross sections of standard CME, SWCE and OCE ($\mu = 1.5$) for $n_2 = 50$, 170.

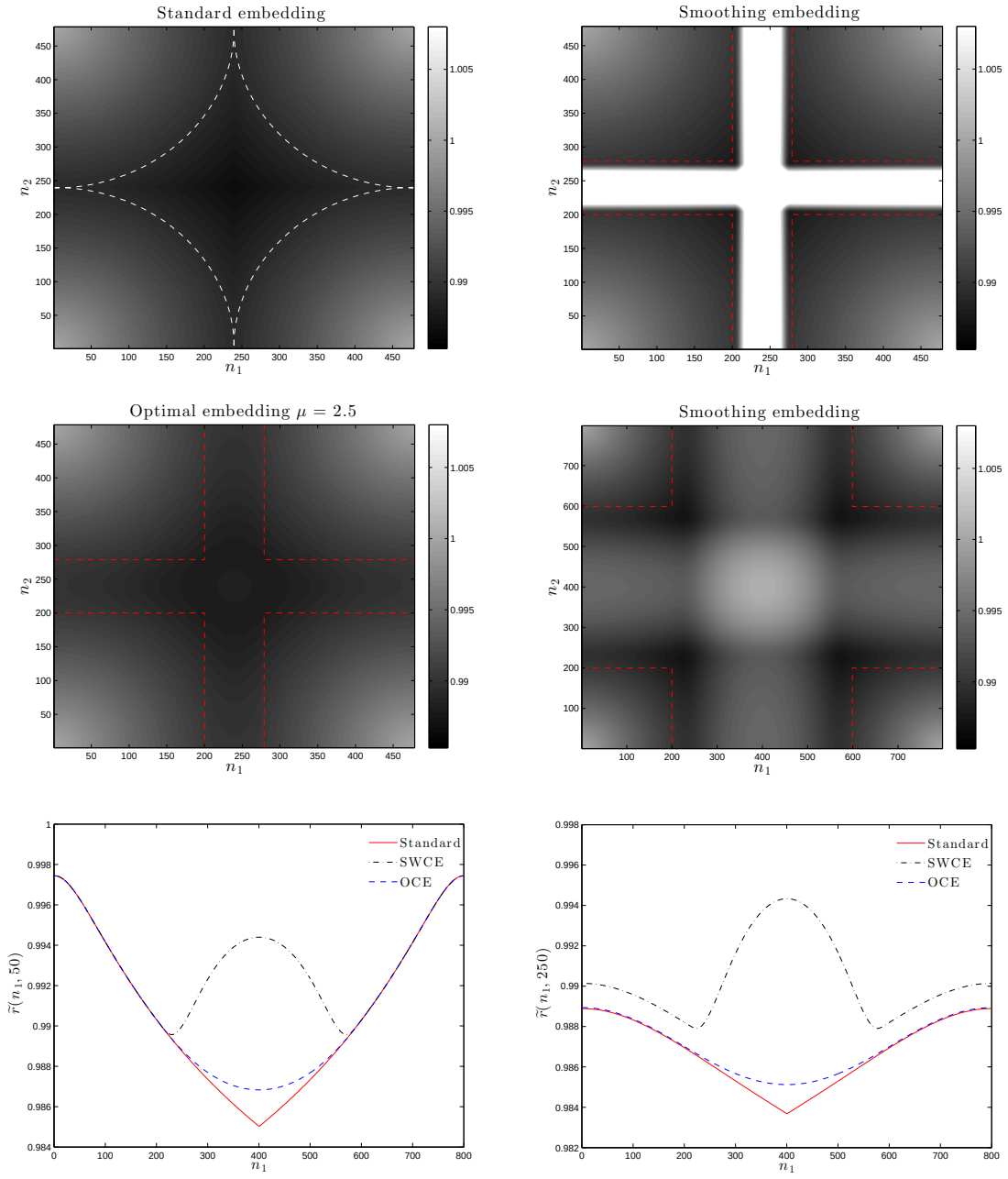

Figure 4.3: Plots for the isotropic covariance of the Cauchy form (4.50) with parameters (4.52) $N = 200$. Top left: Standard embedding (2,836 negative eigenvalues). Top right: Smoothing windows embedding with transition region $\widetilde{N} - N = 40$ and 24,720 negative eigenvalues. Middle left: Optimal embedding with $\mu = 2.5$, $\min f = 7{\cdot}10^{-5}$, and no negative eigenvalues. Middle right: Smoothing embedding with transition region $\widetilde{N} - N = 200$ and 136 negative eigenvalues. Bottom plots: Cross sections of standard CME, SWCE and OCE ($\mu = 1.5$) for $n_2 = 50,\ 250$ and $\widetilde{N} - N = 200$.

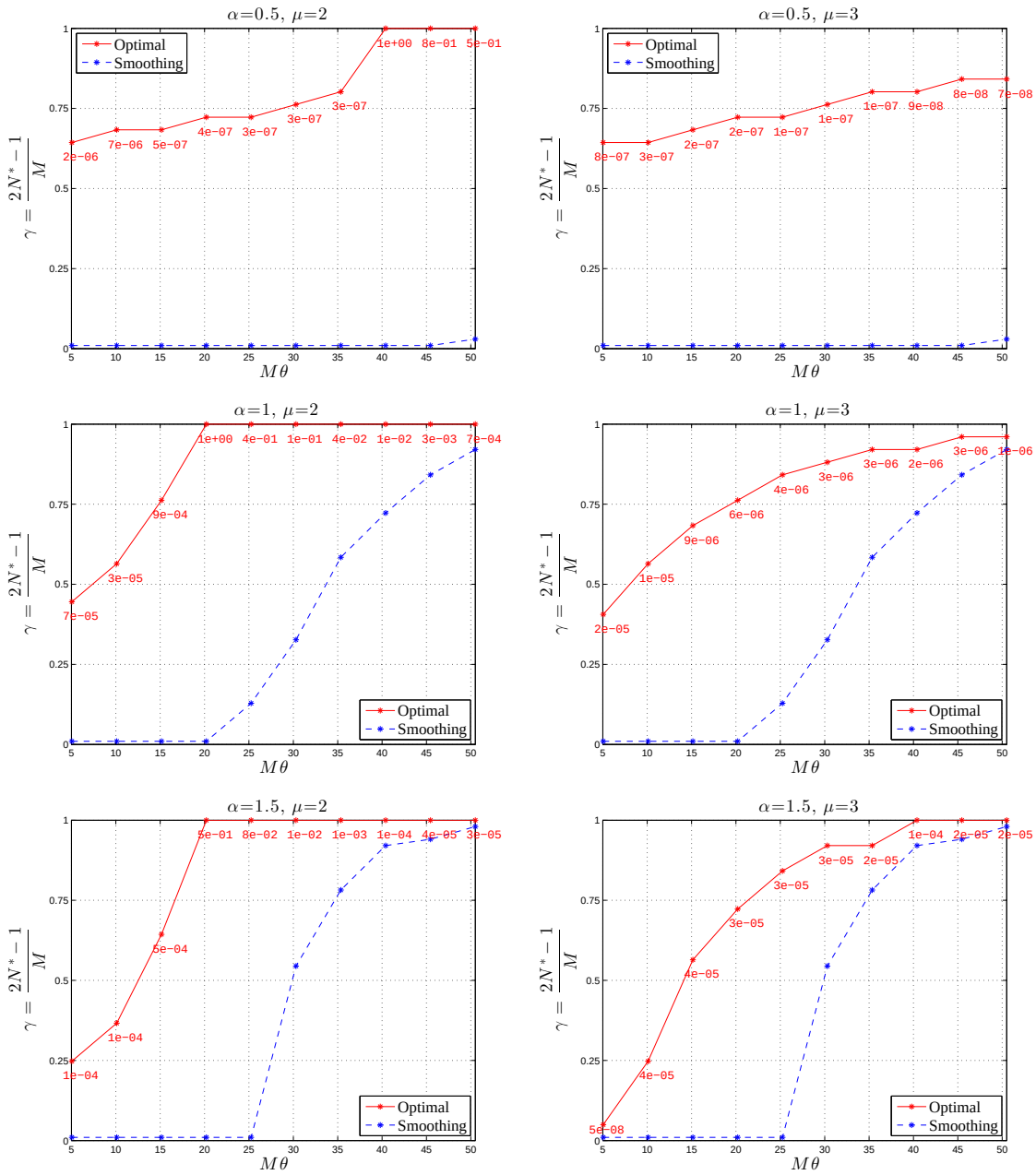

Figure 4.4: Efficiency values of SWCE (dashed lines) and OCE (solid lines) methods for the powered exponential covariance function (4.49) with $\theta = 0.05k$, $k = 1\ldots,10$, and $\alpha = 0.5, 1, 1.5$. Left: OCE implemented with $\mu = 2$. Right: OCE implemented with $\mu = 3$. The tolerance ϵ in the PLB method is equal to 10^{-5} in all cases.

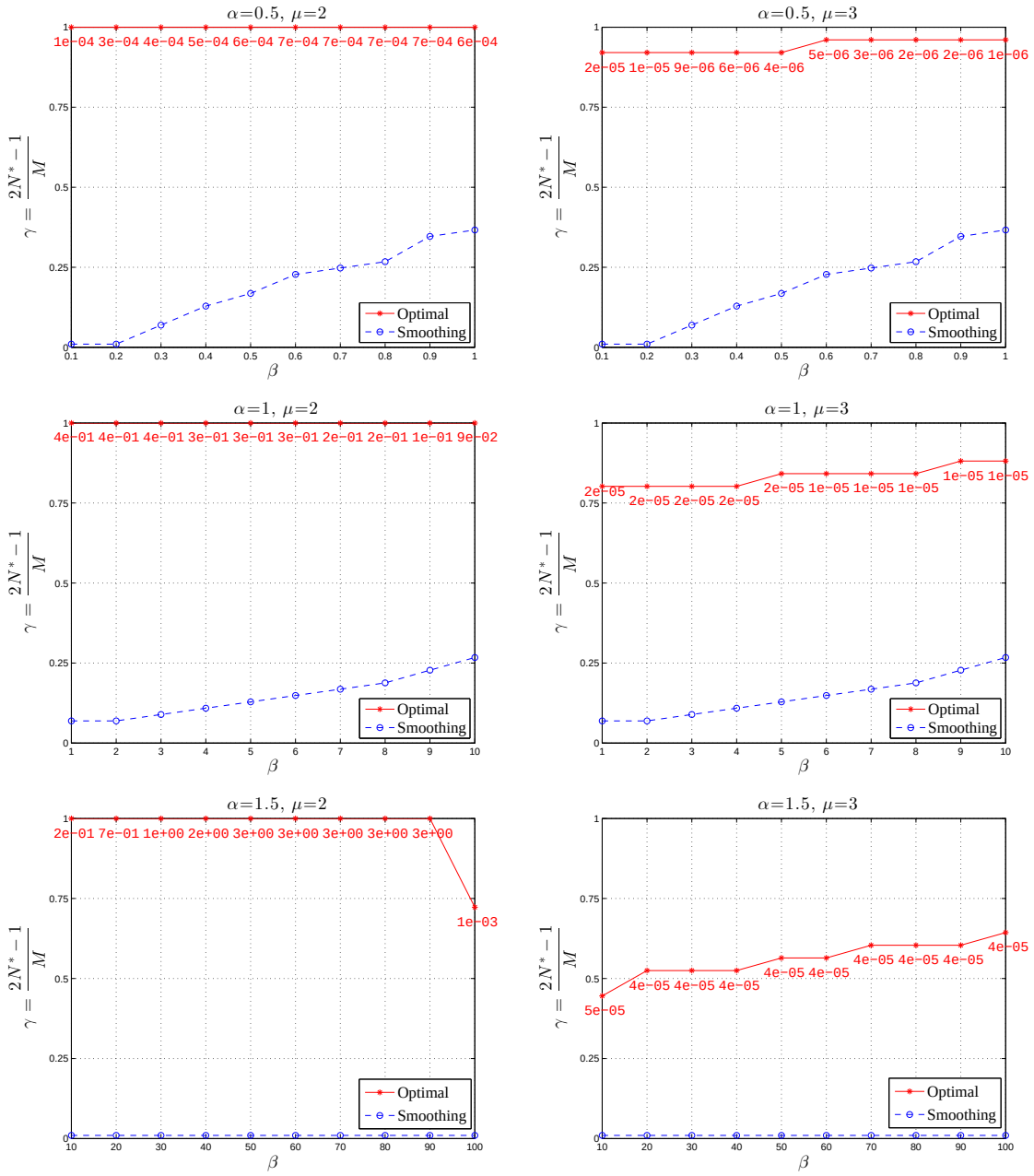

Figure 4.5: Efficiency values of SWCE (dashed lines) and OCE (solid lines) methods for the Cauchy covariance (4.50) with $\theta = 0.001$ and $\alpha = 0.5, 1, 1.5$. For the first row $\beta = 0.1, 0.2, \ldots, 1$, for the second row $\beta = 1, 2, \ldots, 10$ and for the third row $\beta = 10, 20, \ldots, 100$. Left: OCE implemented with $\mu = 2$. Right: OCE implemented with $\mu = 3$. The tolerance ϵ in the PLB method is equal to 10^{-5} in all cases.

APPENDIX A

A.1 Technical proofs for Chapter 2

We gather here the proofs of several results given in Chapter 2, starting with Proposition 2.2.1. Recall that a slowly varying function L is *quasi-monotone* if it is of bounded variation on any compact interval of $[0, \infty)$ and, if for some $\delta > 0$,

$$\int_0^x u^\delta |dL(u)| = O\left(x^\delta L(x)\right), \quad \text{as} \quad x \to \infty \tag{A.1}$$

(Bingham, Goldie and Teugels (1989)). One interest in quasi-monotone slowly varying functions lies in the following classical result (see, for example, Theorem 4.3.2 in Bingham et al. (1989)).

Lemma A.1. *Suppose L is a quasi-monotone slowly varying function. Let $g(u)$ stand for $\cos(u)$, $\sin(u)$ or e^{iu}, and $0 < p < 1$. Then, the following series converges conditionally for all $\lambda \in (0, \pi]$, and*

$$\sum_{k=0}^{\infty} g(k\lambda)\frac{L(k)}{k^p} \sim \lambda^{p-1} L\left(\frac{1}{\lambda}\right)\Gamma(1-p)g\left(\frac{\pi}{2}(1-p)\right), \quad \textit{as} \quad \lambda \to 0^+. \tag{A.2}$$

A converse of Lemma A.1, allowing one to go from the spectral domain to the time domain, is also available (see (4.3.8) in Bingham et al. (1989)).

Lemma A.2. *Suppose $l(1/u)$ is a quasi-monotone slowly varying function on $(1/\pi, \infty)$, and $0 < p < 1$. Then,*

$$\int_0^\pi e^{in\lambda}\lambda^{p-1} l(\lambda) d\lambda \sim n^{-p} l\left(\frac{1}{n}\right)\Gamma(p) e^{\frac{i\pi p}{2}}, \quad \textit{as} \quad n \to \infty.$$

PROOF OF PROPOSITION 2.2.1: (i) One consequence of Lemma A.1 (omitted here for the shortness sake) is that we can write

$$f_{jk}(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \gamma_{jk}(n). \tag{A.3}$$

(Proving (A.3) amounts to showing that $f_{jk}(\lambda)$ given by the right-hand side of (A.3) satisfies $\int_{-\pi}^{\pi} e^{in\lambda} f_{jk}(\lambda) d\lambda = \gamma_{jk}(n)$.) Then, by using $\gamma_{jk}(-n) = \gamma_{kj}(n)$,

$$f_{jk}(\lambda) = \frac{1}{2\pi}\Big\{ \sum_{n=-\infty}^{\infty} \cos(n\lambda)\gamma_{jk}(n) - i \sum_{n=-\infty}^{\infty} \sin(n\lambda)\gamma_{jk}(n)\Big\}$$

$$= \frac{1}{2\pi}\Big\{\gamma_{jk}(0) + \sum_{n=1}^{\infty} \cos(n\lambda)(\gamma_{jk}(n) + \gamma_{kj}(n))\Big\} - \frac{i}{2\pi}\Big\{ \sum_{n=1}^{\infty} \sin(n\lambda)(\gamma_{jk}(n) - \gamma_{kj}(n))\Big\}$$

$$= \frac{1}{2\pi}\Big\{\gamma_{jk}(0) + \sum_{n=1}^{\infty} \cos(n\lambda)\frac{R_{jk}(n) + R_{kj}(n)}{n^{1-(d_j+d_k)}}\Big\} - \frac{i}{2\pi}\Big\{ \sum_{n=1}^{\infty} \sin(n\lambda)\frac{R_{jk}(n) - R_{kj}(n)}{n^{1-(d_j+d_k)}}\Big\}.$$

It follows from Lemma A.1 that

$$f_{jk}(\lambda) \sim \frac{\Gamma(d_j + d_k)}{2\pi} \lambda^{-(d_j+d_k)} \Big\{ (R_{jk} + R_{kj}) \cos(\frac{\pi}{2}(d_j + d_k)) - i(R_{jk} - R_{kj}) \sin(\frac{\pi}{2}(d_j + d_k))\Big\},$$

showing (2.28). This also immediately yields (2.30)–(2.31) by using (2.22).

(ii) Note that

$$\gamma_{jk}(n) = \int_{-\pi}^{\pi} e^{in\lambda} f_{jk}(\lambda) d\lambda = \int_{-\pi}^{\pi} e^{in\lambda} G_{jk}(\lambda) |\lambda|^{-(d_j+d_k)} d\lambda$$

$$= \int_0^{\pi} e^{in\lambda} (\Re G_{jk}(\lambda) + i\Im G_{jk}(\lambda)) \lambda^{-(d_j+d_k)} d\lambda + \int_0^{\pi} e^{-in\lambda} (\Re G_{jk}(-\lambda) + i\Im G_{jk}(-\lambda)) \lambda^{-(d_j+d_k)} d\lambda$$

$$= \int_0^{\pi} e^{in\lambda} (\Re G_{jk}(\lambda) + i\Im G_{jk}(\lambda)) \lambda^{-(d_j+d_k)} d\lambda + \int_0^{\pi} e^{-in\lambda} (\Re G_{kj}(\lambda) - i\Im G_{kj}(\lambda)) \lambda^{-(d_j+d_k)} d\lambda$$

$$= 2\int_0^{\pi} \cos(n\lambda) \Re G_{jk}(\lambda) \lambda^{-(d_j+d_k)} d\lambda - \int_0^{\pi} \sin(n\lambda) \Im G_{jk}(\lambda) \lambda^{-(d_j+d_k)} d\lambda,$$

where we used $\Re G_{jk}(\lambda) = \Re G_{kj}(\lambda)$ and $\Im G_{jk}(\lambda) = -\Im G_{kj}(\lambda)$. By Lemma A.2 we get that

$$\gamma_{jk}(n) \sim 2\Gamma(1-(d_j+d_k)) n^{(d_j+d_k)-1} \Big\{ \Re G_{jk} \cos(\frac{\pi}{2}(1-(d_j+d_k))) - \Im G_{jk} \sin(\frac{\pi}{2}(1-(d_j+d_k)))\Big\},$$

which yields (2.29). $\square$

We next turn to Proposition 2.3.1.

PROOF OF PROPOSITION 2.3.1: Write the autocovariance function $\gamma(n)$ of X_n as

$$\gamma(n) = \sum_{m=-\infty}^{\infty} \Psi_m \Psi'_{m+n} = \sum_{m=-\infty}^{-n-1} \Psi_m \Psi'_{m+n} + \sum_{m=-n}^{0} \Psi_m \Psi'_{m+n} + \sum_{m=0}^{\infty} \Psi_m \Psi'_{m+n}$$
$$=: \gamma_1(n) + \gamma_2(n) + \gamma_3(n). \quad \text{(A.4)}$$

Denote by $\gamma_{i,jk}(n)$ the (j,k) component of $\gamma_i(n)$, $i = 1, 2, 3$. Then, by using (2.34), we have

$$\begin{aligned}
\gamma_{1,jk}(n) &= \sum_{m=-\infty}^{-n-1} \sum_{t=1}^{p} L_{jt}(m) L_{kt}(m+n) |m|^{d_j - 1} |m+n|^{d_k - 1} \\
&= \sum_{t=1}^{p} \sum_{m=n+1}^{\infty} L_{jt}(-m) L_{kt}(n-m) m^{d_j - 1} (m-n)^{d_k - 1} \\
&= n^{d_j + d_k - 1} \sum_{t=1}^{p} \sum_{m=n+1}^{\infty} L_{jt}(-m) L_{kt}(n-m) \left(\frac{m}{n}\right)^{d_j - 1} \left(\frac{m}{n} - 1\right)^{d_k - 1} \frac{1}{n} \\
&\sim n^{d_j + d_k - 1} \sum_{t=1}^{p} \alpha^-_{jt} \alpha^-_{kt} \int_1^{\infty} x^{d_j - 1} (x-1)^{d_k - 1} dx,
\end{aligned}$$

where the last asymptotic relation follows by the dominated convergence theorem and (2.35). By using Formula 3.191.2 in Gradshteyn and Ryzhik (2007), p. 315, we have

$$\gamma_{1,jk}(n) \sim R^1_{jk} n^{d_j + d_k - 1}, \quad \text{as} \quad n \to \infty, \quad \text{(A.5)}$$

where $R^1_{jk} = \sum_{t=1}^{p} \alpha^-_{jt} \alpha^-_{kt} \frac{\Gamma(d_k)\Gamma(1-d_j-d_k)}{\Gamma(1-d_j)}$. Similarly for $\gamma_{2,jk}(n)$ and $\gamma_{3,jk}(n)$, as $n \to \infty$, we have

$$\gamma_{2,jk}(n) \sim n^{d_j + d_k - 1} \sum_{t=1}^{p} \alpha^-_{jt} \alpha^+_{kt} \int_0^1 x^{d_j - 1} (1-x)^{d_k - 1} dx = R^2_{jk} n^{d_j + d_k - 1}, \quad \text{(A.6)}$$

$$\gamma_{3,jk}(n) \sim n^{d_j + d_k - 1} \sum_{t=1}^{p} \alpha^+_{jt} a^+_{kt} \int_0^{\infty} x^{d_j - 1} (x+1)^{d_k - 1} dx = R^3_{jk} n^{d_j + d_k - 1}, \quad \text{(A.7)}$$

where $R^2_{jk} = \sum_{t=1}^p \alpha^-_{jt}\alpha^+_{kt}\frac{\Gamma(d_j)\Gamma(d_k)}{\Gamma(d_j+d_k)}$ and $R^3_{jk} = \sum_{t=1}^p \alpha^+_{jt}\alpha^+_{kt}\frac{\Gamma(d_j)\Gamma(1-d_j-d_k)}{\Gamma(1-d_k)}$. Combining (A.5), (A.6) and (A.7), we get (2.15)–(2.16) with

$$R_{jk} = c^1_{jk}\frac{\Gamma(d_k)\Gamma(1-d_j-d_k)}{\Gamma(1-d_j)} + c^2_{jk}\frac{\Gamma(d_j)\Gamma(d_k)}{\Gamma(d_j+d_k)} + c^3_{jk}\frac{\Gamma(d_j)\Gamma(1-d_j-d_k)}{\Gamma(1-d_k)}, \tag{A.8}$$

where $c^1_{jk} = \sum_{t=1}^p \alpha^-_{jt}\alpha^-_{kt}$, $c^2_{jk} = \sum_{t=1}^p \alpha^-_{jt}\alpha^+_{kt}$, $c^3_{jk} = \sum_{t=1}^p \alpha^+_{jt}\alpha^+_{kt}$. The coefficients R_{jk} can be expressed as in (2.37) by using the identity $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$, $0 < z < 1$. □

Next, we will prove Propositions 2.4.1 and 2.4.2.

Proof of Proposition 2.4.1: The series X_n in (2.61) is well defined (in the $L^2(\Omega)$–sense) since $b_j > 1/2$ and hence

$$\sum_{m=0}^{\infty} |\psi_{m,jk}|^2 = \sum_{m=0}^{\infty} |\alpha_{jk} m^{-b_j}\cos(2\pi m^a)|^2 \le \alpha^2_{jk}\sum_{m=0}^{\infty} m^{-2b_j} < \infty,$$

for $j,k = 1,\ldots,p$. Moreover, from (5.3) in Hannan (1970), p. 61, the spectral density matrix $f(\lambda)$ of X_n is given by

$$f(\lambda) = \frac{1}{2\pi}\left(\sum_{m=0}^{\infty}\Psi_m e^{-im\lambda}\right)\left(\sum_{m=0}^{\infty}\Psi_m e^{-im\lambda}\right)^*, \tag{A.9}$$

where the series $\sum_{m=0}^{\infty}\Psi_m e^{-im\lambda}$ is defined a.e. in the $L^2(-\pi,\pi]$-sense. The (j,k) entry of the series $\sum_{m=0}^{\infty}\Psi_m e^{-im\lambda}$ is given by $\sum_{m=0}^{\infty}\psi_{m,jk}e^{-im\lambda} = \alpha_{jk}\sum_{m=0}^{\infty} m^{-b_j}\cos(2\pi m^a)e^{-im\lambda}$. By Lemma A.5, this entry is equal to $\alpha_{jk}\overline{f_1(\lambda/2\pi)}$, where $f_1(x)$ is defined in Theorem A.1. Hence, with this interpretation and by using Theorem A.1 with $x = \lambda/2\pi$,

$$\sum_{m=0}^{\infty}\psi_{m,jk}e^{-im\lambda} \sim a_{jk}(2\pi)^{d_j}c_{a,b_j}\lambda^{-d_j}e^{i(\xi_a(2\pi)^{\frac{a}{1-a}}\lambda^{-\frac{a}{1-a}} - \frac{\pi}{4})}, \quad \text{as} \quad \lambda \to 0^+. \tag{A.10}$$

By using (A.9) and (A.10), we conclude that

$$f_{jk}(\lambda) \sim (2\pi)^{d_j+d_k-1}c_{a,b_j}c_{a,b_k}\sum_{t=1}^{p} a_{jt}a_{kt}\lambda^{-(d_j+d_k)}. \tag{A.11}$$

Finally, the statement of the proposition concerning the sine can be deduced similarly. □

PROOF OF PROPOSITION 2.4.2: As in the proof above, the series X_n in (2.65) is well defined in the $L^2(\Omega)$–sense and its spectral density is given by (A.9), where the series $\sum_{m=0}^{\infty} \Psi_m e^{-im\lambda}$ is defined a.e. in the $L^2(-\pi,\pi]$–sense. The (j,k) entry of the series $\sum_{m=0}^{\infty} \Psi_m e^{-im\lambda}$ is given by $\sum_{m=0}^{\infty} \psi_{m,jk} e^{-im\lambda} = \alpha_{jk} \sum_{m=0}^{\infty} m^{-b_j} \cos(2\pi m^a) e^{-im\lambda} + \beta_{jk} \sum_{m=0}^{\infty} m^{-b_j} \sin(2\pi m^a) e^{-im\lambda}$. By Lemma A.5, this entry is equal to $\alpha_{jk}\overline{f_1(\lambda/2\pi)} + \beta_{jk}\overline{f_2(\lambda/2\pi)}$, where $f_1(x), f_2(x)$ are defined in Theorem A.1. Then, by using Theorem A.1 with $x = \lambda/2\pi$,

$$\sum_{m=0}^{\infty} \psi_{m,jk} e^{-im\lambda} \sim \overline{z}_{jk} (2\pi)^{d_j} c_{a,b_j} \lambda^{-d_j} e^{i(\xi_a (2\pi)^{\frac{a}{1-a}} \lambda^{-\frac{a}{1-a}} - \psi)}, \quad \text{as} \quad \lambda \to 0^+, \tag{A.12}$$

where $z_{jk} = \alpha_{jk} + i\beta_{jk}$. By using (A.9) and (A.12), we conclude that

$$f_{jk}(\lambda) \sim (2\pi)^{d_j + d_k - 1} c_{a,b_j} c_{a,b_k} \sum_{t=1}^{p} \overline{z}_{jt} z_{kt} \lambda^{-(d_j + d_k)}. \tag{A.13}$$

□

Finally, we prove Proposition 2.5.1.

PROOF OF PROPOSITION 2.5.1: By using the relation (5.3) in Hannan (1970), p. 61, the FARIMA$(0, D, 0)$ series in (2.69) has the spectral density matrix

$$f(\lambda) = \frac{1}{2\pi} G(\lambda) G(\lambda)^*, \tag{A.14}$$

where $G(\lambda) = (1 - e^{-i\lambda})^{-D} Q_+ + (1 - e^{i\lambda})^{-D} Q_-$. This can be expressed component-wise as

$$f_{jk}(\lambda) = \frac{1}{2\pi} g_j(\lambda) g_k(\lambda)^*, \tag{A.15}$$

where g_j is the jth row of G. Then, the (j,k) component of the autocovariance matrix is

$$\begin{aligned}\gamma_{jk}(n) &= \int_0^{2\pi} e^{in\lambda} f_{jk}(\lambda) d\lambda = \frac{1}{2\pi}\int_0^{2\pi} e^{in\lambda} g_j(\lambda) g_k(\lambda)^* d\lambda \\ &= \frac{1}{2\pi}\left(b_{jk}^1\gamma_{1,jk}(n) + b_{jk}^2\gamma_{2,jk}(n) + b_{jk}^3\gamma_{3,jk}(n) + b_{jk}^4\gamma_{4,jk}(n)\right), \qquad \text{(A.16)}\end{aligned}$$

where $b_{jk}^1, b_{jk}^2, b_{jk}^3, b_{jk}^4$ are given in (2.71), and

$$\gamma_{1,jk}(n) = \gamma_{3,kj}(n) = \int_0^{2\pi} e^{in\lambda}(1-e^{i\lambda})^{-d_j}(1-e^{-i\lambda})^{-d_k} d\lambda,$$

$$\gamma_{2,jk}(n) = \int_0^{2\pi} e^{in\lambda}(1-e^{i\lambda})^{-(d_j+d_k)} d\lambda, \qquad \gamma_{4,jk}(n) = \int_0^{2\pi} e^{in\lambda}(1-e^{-i\lambda})^{-(d_j+d_k)} d\lambda.$$

By writing $1 - e^{\pm i\lambda} = 2\sin(\frac{\lambda}{2})e^{\pm i(\lambda-\pi)/2}$, we have

$$\gamma_{1,jk}(n) = \frac{e^{i\pi(d_j-d_k)/2}}{2^{d_j+d_k}} \int_0^{2\pi} e^{in\lambda} \sin^{-d_j-d_k}(\frac{\lambda}{2}) e^{i\lambda(d_k-d_j)/2} d\lambda$$

$$= \frac{2e^{i\pi(d_j-d_k)/2}}{2^{d_j+d_k}} \int_0^{\pi} e^{i\omega(2n+d_k-d_j)} \sin^{-d_j-d_k}(\omega) d\omega.$$

By using Formula 3.892.1 in Gradshteyn and Ryzhik (2007), p. 485, we deduce that

$$\gamma_{1,jk}(n) = \frac{2e^{i\pi(d_j-d_k)/2}}{2^{d_j+d_k}} \frac{\pi e^{i\beta\pi/2}}{2^{\nu-1}\nu B(\frac{\nu+\beta+1}{2}, \frac{\nu-\beta+1}{2})},$$

where $\beta = 2n + d_k - d_j$ and $\nu = 1 - d_k - d_j$. Then,

$$\gamma_{1,jk}(n) = 2\pi(-1)^n \frac{\Gamma(1-d_j-d_k)}{\Gamma(1-d_j+n)\Gamma(1-d_k-n)}. \qquad \text{(A.17)}$$

Similar calculations yield

$$\gamma_{2,jk}(n) = 2\pi(-1)^n \frac{\Gamma(1-d_j-d_k)}{\Gamma(1-n)\Gamma(1+n-d_j-d_k)}. \qquad \text{(A.18)}$$

The relations (2.72) can now be deduced from (A.17) and (A.18) by using the identities $\Gamma(z)\Gamma(1-z) = \pi/\sin(\pi z)$ and $\Gamma(z)\Gamma(1-z) = (-1)^n\Gamma(n+z)\Gamma(1-n-z)$, $0 < z < 1$. □

A.2 Fourier series of trigonometric power-law coefficients

In the next result, we establish the asymptotic behavior of the Fourier series of the trigonometric power-law coefficients (2.10). The proof is based on the work of Wainger (1965) who obtained a similar result for double-sided trigonometric power-law coefficients (Theorem 10 in Wainger (1965), p. 53). For shortness sake, we shall abbreviate the work of Wainger (1965) by WA.

Theorem A.1. *Let* $0 < a < 1$ *and* $0 < b \leq 1 - \frac{1}{2}a$. *For* $\epsilon > 0$, *define*

$$f_{\epsilon,1}(x) = \sum_{n=0}^{\infty} n^{-b} \cos(2\pi n^a) e^{2\pi i n x - \epsilon n}, \qquad f_{\epsilon,2}(x) = \sum_{n=0}^{\infty} n^{-b} \sin(2\pi n^a) e^{2\pi i n x - \epsilon n}.$$

Then, the limits $f_j(x) = \lim_{\epsilon \to 0^+} f_{\epsilon,j}(x)$, $j = 1, 2$, *exist in the pointwise sense for* $x \neq 0$. *Moreover,* $f_j(x)$ *are continuous for* $x \neq 0$, $j = 1, 2$, *and*

$$f_1(x) = |x|^{-d} e^{-\operatorname{sign}(x) i (\xi_a |x|^{-\frac{a}{1-a}} + \psi)} \left(c_{a,b} + O(|x|^{\frac{a}{1-a}}) \right) + C_1(x), \tag{A.19}$$

$$f_2(x) = |x|^{-d} \operatorname{sign}(x) i e^{-\operatorname{sign}(x) i (\xi_a |x|^{-\frac{a}{1-a}} + \psi)} \left(c_{a,b} + O(|x|^{\frac{a}{1-a}}) \right) + C_2(x), \tag{A.20}$$

where $d = \frac{1-b-\frac{a}{2}}{1-a}$, $c_{a,b} = \frac{1}{2} a^{-\frac{b-1/2}{1-a}} (1-a)^{-1/2}$, $\xi_a = 2\pi(a^{\frac{a}{1-a}} - a^{\frac{1}{1-a}})$, $\psi = -\frac{\pi}{4}$ *and* $C_1(x)$, $C_2(x)$ *are continuous functions.*

Proof. We follow to the extent possible the notation of Wainger (1965), abbreviated WA below. Consider the functions

$$\Phi_1(u) = \psi(u) |u|^{-b} \cos(2\pi |u|^a), \qquad \Phi_2(u) = \psi(u) |u|^{-b} \sin(2\pi |u|^a),$$

where $\psi(u) \in C^\infty(-\infty, \infty)$, $\psi(u) = 0$ for $u \leq 1/2$, $\psi(u) = 1$ for $u \geq 1$ and $0 \leq \psi(u) \leq 1$. Let $\epsilon > 0$ and $D_{\epsilon,j}(x) = \{\Phi_j(u) e^{-\epsilon u}\}^\vee(x) = \int_{\mathbb{R}} \Phi_j(u) e^{2\pi i x u - \epsilon u} du$ be the inverse Fourier transforms of $e^{-\epsilon u} \Phi_j(u)$, $j = 1, 2$. Observe that the functions $f_{\epsilon,j}(x) = \sum_{n=0}^{\infty} \Phi_j(n) e^{2\pi i n x - \epsilon n}$ are discrete counterparts of $D_{\epsilon,j}(x)$. The proof below will show that

$$D_j(x) = \lim_{\epsilon \to 0^+} D_{\epsilon,j}(x), \quad j = 1, 2, \tag{A.21}$$

exist and are continuous at $x \neq 0$, that $f_j(x)$ and $D_j(x)$ are equal up to a continuous function, and that $D_j(x)$, $j = 1, 2$, have the asymptotic behavior of the first terms on the right-hand sides in (A.19)–(A.20). We will use Lemma 11 (p. 37) and Theorem 9 (p. 41) in WA.

Consider the function $F_\epsilon^{k,a,b}(x)$ appearing in 2.4 of WA, p. 44,

$$F_\epsilon^{k,a,b}(x) = 2\pi|x|^{\frac{1}{2}(2-k)} \int_0^\infty \psi(u) u^{-b+\frac{1}{2}k} e^{2\pi i u^a - \epsilon u} J_{\frac{1}{2}(k-2)}(2\pi|x|u) du, \tag{A.22}$$

where $J_\mu(x)$ is a Bessel function of the first kind. (See, for example, Korenev (2002) for more information on Bessel functions.) The function (A.22) is denoted by $F_\epsilon(x)$ in WA. We added the superscripts k, a and b to avoid confusion regarding the values of these parameters. By Lemma A.3 below,

$$D_{\epsilon,1}(x) = \frac{1}{2}\mathrm{Re}(F_\epsilon^{1,a,b}(x)) + \mathrm{sign}(x) i \frac{1}{2}|x| \mathrm{Re}(F_\epsilon^{3,a,b+1}(x)), \tag{A.23}$$

$$D_{\epsilon,2}(x) = \frac{1}{2}\mathrm{Im}(F_\epsilon^{1,a,b}(x)) + \mathrm{sign}(x) i \frac{1}{2}|x| \mathrm{Im}(F_\epsilon^{3,a,b+1}(x)). \tag{A.24}$$

By Theorem 9 in WA (see again 2.4 in WA, p. 44), $F^{k,a,b}(x) = \lim_{\epsilon \to 0^+} F_\epsilon^{k,a,b}(x)$ exists in the pointwise sense for $x \neq 0$, and $F^{k,a,b}(x)$ is continuous for $x \neq 0$. Thus, in view of (A.23) and (A.24), the same holds for $D_{\epsilon,j}(x)$ and $D_j(x)$, $j = 1, 2$.

We now want to use Lemma 11 in WA, p. 37, with $\Phi = \Phi_j$, $F_\epsilon = D_{\epsilon,j}$ and $F = D_j$ in the lemma. The established relation (A.21) is one of the assumptions of the lemma. The other assumptions are $|\Phi_j(u)| = O(e^{\epsilon|x|})$ as $u \to \infty$, and $|D_{\epsilon,j}(x)| = O(|x|^{-1-\mu})$ uniformly in ϵ for some $\mu > 0$, as $x \to \infty$. The first of these assumptions certainly holds. The second assumption can be verified by using Theorem 9 in WA. Thus, Lemma 11 of WA yields that the limits $f_j(x)$ exist, are continuous at $x \neq 0$, and are equal to $D_j(x)$ up to continuous functions.

It remains to show that the functions $D_j(x)$ behave as the first terms on the right-hand sides of (A.19)–(A.20). Theorem 9, ii), in WA shows that, for $b \leq k(1 - \frac{1}{2}a)$ and $m_0 = 0$,

$$F^{k,a,b}(x) = |x|^{-\frac{k-b-\frac{ka}{2}}{1-a}} e^{i\xi_a |x|^{-\frac{a}{1-a}}} \left(\alpha_0 + O(|x|^{\frac{a}{1-a}})\right) + \widetilde{C}_k(|x|), \tag{A.25}$$

where $\widetilde{C}_k(|x|)$ is a continuous function and $\alpha_0 \in \mathbb{C}\backslash\{0\}$ depends on a, b and k. For the asymptotic behavior of (A.19)–(A.20), we need an exact form of the constant α_0 when $k = 1$ and $k = 3$. By using a version of the saddle point method, which is finer than the one used in Lemma 13 of WA, we show in Lemma A.4 below that

$$\alpha_0 = 2c_{a,b}e^{i\psi} \tag{A.26}$$

when $k = 1$, and

$$\alpha_0 = 2c_{a,b}ie^{i\psi} \tag{A.27}$$

when $k = 3$. This yields

$$F^{1,a,b}(x) = |x|^{-d}e^{i(\xi_a|x|^{-\frac{a}{1-a}}+\psi)}\left(2c_{a,b} + O(|x|^{\frac{a}{1-a}})\right) + \widetilde{C}_1(|x|), \tag{A.28}$$

$$F^{3,a,b+1}(x) = |x|^{-d-1}ie^{i(\xi_a|x|^{-\frac{a}{1-a}}+\psi)}\left(2c_{a,b} + O(|x|^{\frac{a}{1-a}})\right) + \widetilde{C}_3(|x|). \tag{A.29}$$

Letting $\epsilon \to 0^+$ in (A.23), (A.24), and using (A.28) and (A.29), we conclude that

$$D_1(x) = |x|^{-d}e^{-\operatorname{sign}(x)i(\xi_a|x|^{-\frac{a}{1-a}}+\psi)}\left(c_{a,b} + O(|x|^{\frac{a}{1-a}})\right) + \frac{1}{2}\left(\operatorname{Re}(\widetilde{C}_1(|x|)) + i|x|\operatorname{Re}(\widetilde{C}_3(|x|))\right),$$

$$D_2(x) = |x|^{-d}\operatorname{sign}(x)ie^{-\operatorname{sign}(x)i(\xi_a|x|^{-\frac{a}{1-a}}+\psi)}\left(c_{a,b} + O(|x|^{\frac{a}{1-a}})\right) + \frac{1}{2}\left(\operatorname{Im}(\widetilde{C}_1(|x|)) + i|x|\operatorname{Im}(\widetilde{C}_3(|x|))\right), \tag{A.30}$$

where we also used the identities

$$\cos(y) - \operatorname{sign}(x)i\sin(y) = e^{-\operatorname{sign}(x)iy}, \quad \sin(y) + \operatorname{sign}(x)i\cos(y) = \operatorname{sign}(x)ie^{-\operatorname{sign}(x)iy}.$$

This completes the proof. □

The next two auxiliary lemmas were used in the proof of Theorem A.1 above.

Lemma A.3. *The functions $D_{\epsilon,j}(x)$, $j = 1, 2$, $F_\epsilon^{k,a,b}(x)$, defined in (A.21) and (A.22), satisfy the relations (A.23) and (A.24).*

Proof. By using the Bessel function properties $J_{-\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi x}}\cos(x)$ and $J_{\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi x}}\sin(x)$ (Korenev (2002), p. 16) and the facts that $\cos(|x|) = \cos(x)$ and $\sin(|x|) = \text{sign}(x)\sin(x)$, we have

$$F_\epsilon^{1,a,b}(x) = 2\int_0^\infty \psi(u)u^{-b}e^{2\pi i u^a - \epsilon u}\cos(2\pi x u)du,$$

$$F_\epsilon^{3,a,b+1}(x) = 2\text{sign}(x)|x|^{-1}\int_0^\infty \psi(u)u^{-b}e^{2\pi i u^a - \epsilon u}\sin(2\pi x u)du.$$

Since $\psi(u) = 0$ for $u < 0$, we can also rewrite the inverse Fourier transforms $D_{\epsilon,j}(x)$ as

$$D_{\epsilon,1}(x) = \int_0^\infty \psi(u)u^{-b}\cos(2\pi u^a)e^{2\pi i x u - \epsilon u}du,$$

$$D_{\epsilon,2}(x) = \int_0^\infty \psi(u)u^{-b}\sin(2\pi u^a)e^{2\pi i x u - \epsilon u}du.$$

Since

$$\text{Re}(D_{\epsilon,1}(x)) = \frac{1}{2}\text{Re}(F_\epsilon^{1,a,b}(x)), \quad \text{Im}(D_{\epsilon,1}(x)) = \frac{1}{2}\text{sign}(x)|x|\text{Re}(F_\epsilon^{3,a,b+1}(x)),$$

$$\text{Re}(D_{\epsilon,2}(x)) = \frac{1}{2}\text{Im}(F_\epsilon^{1,a,b}(x)), \quad \text{Im}(D_{\epsilon,2}(x)) = \frac{1}{2}\text{sign}(x)|x|\text{Im}(F_\epsilon^{3,a,b+1}(x)),$$

we conclude that (A.23) and (A.24) hold. □

The second auxiliary lemma uses a saddle point method. The saddle point method allows computing asymptotic expansions of integrals of the form

$$I(t) = \int_C f(z)e^{th(z)}dz, \quad \text{as} \quad t \to \infty,$$

where C is a contour in the complex plane and the functions $f(z)$, $h(z)$ are holomorphic in a neighborhood of this contour. According to the method, if z_0 is an interior point of C

and a saddle point of $h(z)$, that is, $h'(z_0) = 0$, $h''(z_0) \neq 0$, then

$$I(t) = \sqrt{\frac{2\pi}{-h''(z_0)}} t^{-1/2} e^{th(z_0)} \Big(f(z_0) + O(t^{-1}) \Big), \quad \text{as} \quad t \to \infty. \tag{A.31}$$

See, for example, Fedoryuk (2011). The version of the saddle point result (A.31) used by WA, Lemma 13, pp. 42-43, provides only the absolute value of the constant at $t^{1/2} e^{th(z_0)}$ in (A.31), that is, the value $(2\pi)^{1/2} |h''(z_0)|^{-1/2} |f(z_0)|$. (This is also after correcting the typo in WA, p. 43, where the exponent $1/2$ of $|h''(\xi)|$ should be replaced by $-1/2$.) The finer version (A.31) allows us to identify the constant α_0 in (A.25) as stated in the next lemma.

Lemma A.4. *The coefficient α_0 appearing in (A.25) is given by (A.26) and (A.27) when $k = 1$ and 3, respectively, and the relations (A.28) and (A.29) hold.*

Proof. To prove that $F^{k,a,b}(x)$ in (A.25) is the limit of $F_\epsilon^{k,a,b}(x)$, Wainger (1965) decomposes $F_\epsilon^{k,a,b}(x)$ into several integrals. The main contribution to $F^{k,a,b}(x)$ comes from the integral given in 2.20 on p. 49 in WA,

$$H_{II}^{k,a,b}(x) = 2|x|^{\frac{b-k}{1-a}} t^{\frac{1}{2}(1-k)} \int_{II} s^{-b+\frac{1}{2}(k-1)} e^{th_1(s)} S_1^{\frac{1}{2}(k-2)}(2\pi st) ds,$$

where $h_1(s) = 2\pi i s^a - 2\pi i s$, $t = |x|^{-\frac{a}{1-a}}$, $S_1^\mu(z)$ is an analytic function given in Lemma 12 of WA, and II is the contour described on p. 47 of WA. Therefore, it is enough to show that $H_{II}^{1,a,b}(x)$ and $H_{II}^{3,a,b+1}(x)$ are equal to the right-hand sides of (A.28) and (A.29), respectively, up to a continuous function. When $k = 1$ and $k = 3$, we get from Lemma 12 of WA that $S_1^{-1/2}(z) = 1/2$, $S_1^{1/2}(z) = i/2$. Then,

$$H_{II}^{1,a,b}(x) = |x|^{\frac{b-1}{1-a}} \int_{II} s^{-b} e^{th_1(s)} ds, \qquad H_{II}^{3,a,b+1}(x) = |x|^{\frac{b-2+a}{1-a}} i \int_{II} s^{-b} e^{th_1(s)} ds.$$

(Note that the second term is with $b+1$ to correspond to $F^{3,a,b+1}(x)$ in (A.29)). Next, consider the integral $I(t) = \int_{II} s^{-b} e^{th_1(s)} ds$. Let $\xi = a^{\frac{1}{1-a}}$ and observe that the point $(\xi, 0)$ is a saddle point of h_1 that lies in the interior of the contour II (as seen in figure 1 of WA,

p. 46). Then, from (A.31), $I(t) = t^{-\frac{1}{2}} e^{i(t\xi_a+\psi)} \left(c_{a,b} + O(t^{-1})\right)$, as $t \to \infty$. This yields

$$H^{1,a,b}(x) = |x|^{-d} e^{i(\xi_a |x|^{-\frac{a}{1-a}}+\psi)} \left(2c_{a,b} + O(|x|^{\frac{a}{1-a}})\right),$$

$$H^{3,a,b+1}(x) = |x|^{-d-1} i e^{i(\xi_a |x|^{-\frac{a}{1-a}}+\psi)} \left(2c_{a,b} + O(|x|^{\frac{a}{1-a}})\right).$$

□

Finally, we include the following elementary lemma which is used in the proofs of Propositions 2.4.1 and 2.4.2.

Lemma A.5. *Let a, b and f_1, f_2 be as in Theorem A.1. If*

$$b > \frac{1}{2},$$

then the trigonometric power-law coefficients (2.10) are in $l^2(\mathbb{Z})$, and their Fourier series (defined in the $L^2(-1,1]$–sense) satisfy

$$f_1(x) = \sum_{n=0}^{\infty} \cos(2\pi n^a) n^{-b} e^{2\pi i n x}, \qquad f_2(x) = \sum_{n=0}^{\infty} \sin(2\pi n^a) n^{-b} e^{2\pi i n x}, \qquad a.e.\ dx. \tag{A.32}$$

Proof. Consider the functions $f_{\epsilon,j}(x)$, $j = 1, 2$, defined in Theorem A.1. Since

$$\int_{-1/2}^{1/2} \left| f_{\epsilon,1}(x) - \sum_{n=0}^{\infty} \cos(2\pi n^a) n^{-b} e^{2\pi i n x} \right|^2 dx = \sum_{n=0}^{\infty} n^{-2b} \cos^2(2\pi n^a)(e^{-\epsilon n} - 1)^2 \to 0,$$

as $\epsilon \to 0^+$, we have $f_{\epsilon,1}(x)$ converging to the Fourier series $\sum_{n=0}^{\infty} \cos(2\pi n^a) n^{-b} e^{2\pi i n x}$ in the $L^2(-1/2, 1/2]-$sense. By Theorem A.1, $f_{\epsilon,1}(x)$ to $f_1(x)$ pointwise. The uniqueness of the limit yields the first relation in (A.32). The proof of the second relation in (A.32) is analogous. □

APPENDIX B

B.1 Technical proofs for Chapter 4

We first prove Lemma 4.1.

Proof of Lemma 4.1: First we write (4.25) as

$$\widetilde{r}(n_1,n_2)=\begin{cases}\widetilde{r}(n_1,n_2), & \text{if} \quad (n_1,n_2)\in Q_1,\\ \widetilde{r}(n_1-M,n_2-M), & \text{if} \quad (n_1,n_2)\in Q_3,\\ \widetilde{r}(n_1,n_2-M), & \text{if} \quad (n_1,n_2)\in Q_2,\\ \widetilde{r}(n_1-M,n_2), & \text{if} \quad (n_1,n_2)\in Q_4,\end{cases} \tag{B.1}$$

where the regions Q_1, Q_2, Q_3 and Q_4 are given by

$$\begin{aligned} Q_1 &= \{(n_1,n_2): 0\le n_1\le \widetilde{N}-1,\ 0\le n_2\le \widetilde{N}-1\},\\ Q_2 &= \{(n_1,n_2): 0\le n_1\le \widetilde{N}-1,\ \widetilde{N}\le n_2\le M-1\},\\ Q_3 &= \{(n_1,n_2): \widetilde{N}\le n_1\le M-1,\ \widetilde{N}\le n_2\le M-1\},\\ Q_4 &= \{(n_1,n_2): \widetilde{N}\le n_1\le M-1,\ 0\le n_2\le \widetilde{N}-1\} \end{aligned} \tag{B.2}$$

and partition the grid $G(M)$. By using this partition, we rewrite the eigenvalues $g_k(\widetilde{r})$ in (4.11) as

$$g_k(\widetilde{r})=\left(\sum_{n\in Q_1}+\sum_{n\in Q_2}+\sum_{n\in Q_3}+\sum_{n\in Q_4}\right)\widetilde{r}(n)e^{-i2\pi k\cdot n/M}. \tag{B.3}$$

The first sum of (B.3) can be written as

$$\sum_{n\in Q_1} \widetilde{r}(n_1,n_2)e^{-i2\pi k\cdot n/M} = \sum_{n_1=1}^{\widetilde{N}-1}\sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(n_1,n_2)e^{-i2\pi k\cdot n/M} + \sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(0,n_2)e^{-i2\pi k_2 n_2/M} + \sum_{n_1=1}^{\widetilde{N}-1} \widetilde{r}(n_1,0)e^{-i2\pi k_1 n_1/M} + \widetilde{r}(0,0). \quad \text{(B.4)}$$

Using the symmetry condition (4.25) and the property $\widetilde{r}(n) = \widetilde{r}(-n)$, we have for the third sum of (B.3),

$$\begin{aligned}
\sum_{n\in Q_3} \widetilde{r}(n_1,n_2)e^{-i2\pi k\cdot n/M} &= \sum_{n_1=\widetilde{N}}^{M-1}\sum_{n_2=\widetilde{N}}^{M-1} \widetilde{r}(n_1,n_2)e^{-i2\pi k\cdot n/M} \\
&= \sum_{n_1=-\widetilde{N}+1}^{-1}\sum_{n_2=-\widetilde{N}+1}^{-1} \widetilde{r}(n_1+M,n_2+M)e^{-i2\pi k\cdot (n+M)/M} \\
&= \sum_{n_1=1}^{\widetilde{N}-1}\sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(M-n_1,M-n_2)e^{i2\pi k\cdot n/M} \\
&= \sum_{n_1=1}^{\widetilde{N}-1}\sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(n_1-M,n_2-M)e^{i2\pi k\cdot n/M} \\
&= \sum_{n_1=1}^{\widetilde{N}-1}\sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(n_1,n_2)e^{i2\pi k\cdot n/M}.
\end{aligned} \quad \text{(B.5)}$$

By combining the relations (B.4) and (B.5), we get

$$\left(\sum_{n\in Q_1} + \sum_{n\in Q_3}\right) \widetilde{r}(n)e^{-i2\pi k\cdot (n/M)} = 2\sum_{n_1=1}^{\widetilde{N}-1}\sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(n_1,n_2)\cos\left(2\pi\left(\frac{k_1 n_1}{M}+\frac{k_2 n_2}{M}\right)\right) + \sum_{n_1=1}^{\widetilde{N}-1} \widetilde{r}(n_1,0)e^{-i2\pi k_1 n_1/M} + \sum_{n_2=1}^{\widetilde{N}-1} \widetilde{r}(0,n_2)e^{-i2\pi k_2 n_2/M} + \widetilde{r}(0,0). \quad \text{(B.6)}$$

Similar calculations for the second and fourth sums of (B.3) yield

$$\left(\sum_{n\in Q_2}+\sum_{n\in Q_4}\right)\widetilde{r}(n)e^{-i2\pi k\cdot(n/M)}=2\sum_{n_1=1}^{\widetilde{N}-1}\sum_{n_2=\widetilde{N}}^{M-1}\widetilde{r}(n_1,n_2)\cos\left(2\pi\left(\frac{k_1n_1}{M}+\frac{k_2n_2}{M}\right)\right)$$
$$+\sum_{n_1=1}^{\widetilde{N}-1}\widetilde{r}(n_1,0)e^{i2\pi k_1n_1/M}+\sum_{n_2=1}^{\widetilde{N}-1}\widetilde{r}(0,n_2)e^{i2\pi k_2n_2/M}. \quad \text{(B.7)}$$

Combining (B.6) and (B.7) yields

$$\begin{aligned} g_k(\widetilde{r}) &= 2\sum_{n_1=1}^{\widetilde{N}-1}\sum_{n_2=1}^{M-1}\widetilde{r}(n_1,n_2)\cos\left(2\pi\left(\frac{k_1n_1}{M}+\frac{k_2n_2}{M}\right)\right) \\ &\quad +2\sum_{n_1=1}^{\widetilde{N}-1}\widetilde{r}(n_1,0)\cos\left(\frac{2\pi k_1n_1}{M}\right)+2\sum_{n_2=1}^{\widetilde{N}-1}\widetilde{r}(0,n_2)\cos\left(\frac{2\pi k_2n_2}{M}\right)+\widetilde{r}(0,0). \quad \text{(B.8)}\end{aligned}$$

The first two terms on the right-hand side of (B.8) give $c_k(n)$ in (4.32) for $n_1\neq 0$, and the last two terms give $c_k(n)$ in (4.32) for $n_1=0$. $\square$

For the proofs of Lemmas 4.2 and 4.3 we first obtain an expression for the adjoint operator A^T of A similar to (4.31). More specifically, we will show that A^T satisfies

$$[A^T\widetilde{r}](k)=\sum_{n\in G_+(M)}c_n(k)\widetilde{r}(n),\quad k\in G_+(M). \quad \text{(B.9)}$$

Note that the only difference between the two operators A and A^T is that $c_k(n)$ in the expression (4.31) of A is interchanged with $c_n(k)$ in the expression (B.9) of A^T.

We will show (B.9) by viewing A and A^T as matrices. To make the transition to a matrix point of view, let m be the number of points in the grid $G_+(M)$ and consider two arbitrary fixed bijective mappings $\phi_i(u):\{1,\ldots,m\}\to G_+(M),\ i=1,2$, for rearranging the values of the fields $\widetilde{r}(n)$ and $c_k(n)$ in the relation (4.31) into vectors. More specifically, let

$$n=\phi_1(j)\quad\text{and}\quad k=\phi_2(l),\quad j,l=1,\ldots,m,$$

for $n,k\in G_+(M)$. Then, we can interpret the two-dimensional field $\widetilde{r}(n), n\in G_+(M)$, as a column vector $\widetilde{r}_v$ whose jth entry is $\widetilde{r}(\phi_1(j))$. Similarly, for each k, the coefficients

$c_k(n), n \in G_+(M)$, can be viewed as a row vector a_l^T whose jth entry is $c_{\phi_2(l)}(\phi_1(j))$. This allows us to rewrite the relation (4.31) as

$$[A\widetilde{r}](k) = \sum_{n \in G_+(M)} c_k(n)\widetilde{r}(n) = \sum_{j=1}^{m} a_l^T(j)\widetilde{r}(\phi_1(j)) = a_l^T\widetilde{r}_v = (A\widetilde{r}_v)_l, \tag{B.10}$$

where A in the last equation is viewed as a matrix with rows $a_l^T,\ l = 1, \ldots, m$, and $(\cdot)_l$ denotes the lth element of a vector. The subscript v in $\widetilde{r}_v$ is to avoid a possible confusion regarding which point of view is adopted, as $A\widetilde{r}$ will denote the action of the linear operator A on a field $\widetilde{r}$, whereas $A\widetilde{r}_v$ is the usual matrix-vector product.

Next, let $b_l^T,\ l = 1, \ldots, m$, denote the rows of the transpose matrix A^T of A. The jth entry of b_l^T satisfies

$$b_l^T(j) = a_j^T(l) = c_n(k), \quad \text{for } n = \phi_1(j) \text{ and } k = \phi_2(l).$$

Then, arguing as for (B.10) but in reverse order, we have

$$(A^T\widetilde{r}_v)_l = b_l^T\widetilde{r}_v = \sum_{j=1}^{m} b_l^T(j)\widetilde{r}(\phi_1(j)) = \sum_{n \in G_+(M)} c_n(k)\widetilde{r}(n), \tag{B.11}$$

which yields (B.9).

We are now ready to prove Lemmas 4.2 and 4.3.

Proof of Lemma 4.2: In view of the relations (4.31) and (B.9), it is enough to show that the weights $c_n(k)$ satisfy

$$c_n(k) = c_k(n) - \cos(2\pi k_2 n_2/M)1_{\{k_1=0, n_1\neq 0\}} + \cos(2\pi k_2 n_2/M)1_{\{k_1\neq 0, n_1=0\}}. \tag{B.12}$$

To show that (B.12) holds, we simply use (4.32) and compare the values of $c_k(n)$ and $c_n(k)$ for the four cases of the values of n and k shown in Tables B.1–B.2. □

Proof of Lemma 4.3: Recall from the relation (4.42), that H and b are the Hessian and negative gradient of the function $f(x) = tf(x) + \phi(x)$, where f and ϕ are given in (4.35)

$c_k(n)$	$n_1 \neq 0$	$n_1 = 0$
$k_1 \neq 0$	$2\cos(2\pi k \cdot (n/M))$	$\cos(2\pi k_2 n_2/M)$
$k_1 = 0$	$2\cos(2\pi k_2 n_2/M)$	$\cos(2\pi k_2 n_2/M)$

Table B.1: The values of $c_k(n)$

$c_n(k)$	$n_1 \neq 0$	$n_1 = 0$
$k_1 \neq 0$	$2\cos(2\pi k \cdot (n/M))$	$2\cos(2\pi k_2 n_2/M)$
$k_1 = 0$	$\cos(2\pi k_2 n_2/M)$	$\cos(2\pi k_2 n_2/M)$

Table B.2: The values of $c_n(k)$

and (4.36), respectively. To show that H and b satisfy the relations (4.46), we will consider the functions f and ϕ separately.

By using (4.31) and (B.10), we can express the function $\phi(\widetilde{r})$ in (4.36) from the vector perspective as

$$\phi(\widetilde{r}_v) = -\sum_{l=1}^{m} \log(a_l^T \widetilde{r}_v). \tag{B.13}$$

By using (B.10), we can also write the field $d(k)$ in (4.44) as a vector d_v whose lth entry $d_v(l)$ is $d(\phi_2(l)) = -(a_l^T \widetilde{r}_v)^{-1}$. Then, the gradient and Hessian of ϕ are given by

$$\nabla\phi(\widetilde{r}_v) = \sum_{l=1}^{m} \frac{1}{-a_l^T \widetilde{r}_v} a_l = A^T d_v, \tag{B.14}$$

$$\nabla^2\phi(\widetilde{r}_v) = \sum_{l=1}^{m} \frac{1}{(a_l^T \widetilde{r}_v)^2} a_l a_l^T = A^T D^2 A, \tag{B.15}$$

where $D = \mathrm{diag}(d_v)$. As in the case of the operator/matrix A, the diagonal matrix $D = \mathrm{diag}(d_v)$ is the matrix analogue of the operator D defined in (4.45). Indeed, let $y = [Au](k)$, for some two dimensional field $u = \{u(n), n \in G_+(M)\}$. Let also u_v and y_v be the vectors whose jth elements are $u(\phi_1(j))$ and $a_j^T u_v$, respectively. Then, the action of D on y yields

$$\begin{aligned} [Dy](k) &:= d(k) \cdot y(k) \\ &= d(\phi_2(l)) \cdot [Au](\phi_2(l)) \\ &= d_v(l) a_l^T u_v, \end{aligned}$$

where the last term is the lth element of the matrix Dy_v.

Next, we calculate the gradient and Hessian of f in (4.35). Letting s_v be a vector whose jth entry is $s(\phi_1(j))$, we can write f in a quadratic form as

$$f(\widetilde{r}_v) = \widetilde{r}_v^T W \widetilde{r}_v - 2 s_v^T \widetilde{r}_v + r_v^T W r_v. \tag{B.16}$$

Since the last term in the relation (B.16) is a constant, minimizing f is equivalent to minimizing the function

$$\tilde{f}_0(\widetilde{r}_v) = \frac{1}{2} \widetilde{r}_v^T W \widetilde{r}_v - s_v^T \widetilde{r}_v. \tag{B.17}$$

The gradient and Hessian of $\tilde{f}_0$ are given by

$$\nabla \tilde{f}_0(\widetilde{r}_v) = W \widetilde{r}_v - s_v, \tag{B.18}$$

$$\nabla^2 \tilde{f}_0(\widetilde{r}_v) = W. \tag{B.19}$$

Finally, by combining the relations (B.14)–(B.15) and (B.18)–(B.19), we get

$$\nabla f(\widetilde{r}_v) = t(W \widetilde{r}_v - s_v) + A^T d_v,$$

$$\nabla^2 f(\widetilde{r}_v) = tW + A^T D^2 A,$$

which are the vector equivalents of the relations (4.46). $\square$

BIBLIOGRAPHY

Achard, S., Bassett, D. S., Meyer-Lindenberg, A. & Bullmore, E. (2008), 'Fractal connectivity of long-memory networks', *Physical Review E* **77**, 036104.

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Akaike, H. (1998), Information theory and an extension of the maximum likelihood principle, *in* 'Selected Papers of Hirotugu Akaike', Springer, pp. 199–213.

Baillie, R. T. & Morana, C. (2012), 'Adaptive ARFIMA models with applications to inflation', *Economic Modelling* **29**(6), 2451–2459.

Baillie, R. T., Chung, C.-F. & Tieslau, M. A. (1996), 'Analysing inflation by the fractionally integrated ARFIMA-GARCH model', *Journal of applied econometrics* **11**(1), 23–40.

Beran, J., Feng, Y., Ghosh, S. & Kulik, R. (2013), *Long-Memory Processes*, Springer, Heidelberg.

Bertsimas, D. & Tsitsiklis, J. N. (1997), *Introduction to Linear Optimization*, Vol. 6, Athena Scientific Belmont, MA.

Bingham, N. H., Goldie, C. M. & Teugels, J. L. (1989), *Regular Variation*, Vol. 27 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge.

Boyd, S. & Vandenberghe, L. (2004), *Convex Optimization*, Cambridge University Press, Cambridge.

Brockwell, P. J. & Davis, R. A. (2009), *Time Series: Theory and Methods*, Springer Series in Statistics, Springer, New York. Reprint of the second (1991) edition.

Burnham, K. P. & Anderson, D. R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer Science & Business Media.

Chan, G. & Wood, A. T. (1997), 'An algorithm for simulating stationary Gaussian random fields', *Applied Statistics, Algorithm Section* **46**(1), 171–181.

Chan, G. & Wood, A. T. (1999), 'Simulation of stationary Gaussian vector fields', *Statistics and computing* **9**(4), 265–268.

Chan, N. H. & Terrin, N. (1995), 'Inference for unstable long-memory processes with applications to fractional unit root autoregressions', *The Annals of Statistics* **23**(5), 1662–1683.

Chen, W. W. & Hurvich, C. M. (2003), 'Semiparametric estimation of multivariate fractional cointegration', *Journal of the American Statistical Association* **98**(463), 629–642.

Chen, W. W. & Hurvich, C. M. (2006), 'Semiparametric estimation of fractional cointegrating subspaces', *The Annals of Statistics* **34**(6), 2939–2979.

Christensen, B. J. & Nielsen, M. Ø. (2006), 'Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting', *Journal of Econometrics* **133**(1), 343–371.

Craigmile, P. F. (2003), 'Simulating a class of stationary Gaussian processes using the Davies-Harte algorithm, with application to long memory processes', *Journal of Time Series Analysis* **24**(5), 505–511.

Davies, R. B. & Harte, D. S. (1987), 'Tests for Hurst effect', *Biometrika* **74**(1), 95–101.

Dietrich, C. & Newsam, G. (1993), 'A fast and exact method for multidimensional Gaussian stochastic simulations', *Water Resources Research* **29**(8), 2861–2869.

Dietrich, C. R. (1993), 'Computationally efficient Cholesky factorization of a covariance matrix with block Toeplitz structure', *Journal of Statistical Computation and Simulation* **45**(3-4), 203–218.

Diongue, A. K. (2010), 'A multivariate generalized long memory model', *Comptes Rendus Mathematique* **348**(5), 327–330.

Doornik, J. A. & Ooms, M. (2004), 'Inference and forecasting for ARFIMA models with an application to US and UK inflation', *Studies in Nonlinear Dynamics & Econometrics*.

Doukhan, P., Oppenheim, G. & Taqqu, M. S. (2003), *Theory and Applications of Long-Range Dependence*, Birkhäuser Boston Inc., Boston, MA.

Dueker, M. & Startz, R. (1998), 'Maximum-likelihood estimation of fractional cointegration with an application to US and Canadian bond rates', *Review of Economics and Statistics* **80**(3), 420–426.

Dufour, J.-M. & Pelletier, D. (2011), 'Practical methods for modeling weak VARMA processes: identification, estimation and specification with a macroeconomic application', *Preprint*.

Fedoryuk, M. V. (2011), Saddle point method, *Encyclopedia of Mathematics*, `http://www.encyclopediaofmath.org/index.php/Saddle_point_method`.

Forsgren, A., Gill, P. E. & Wright, M. H. (2002), 'Interior methods for nonlinear optimization', *SIAM Review* **44**(4), 525–597 (2003).

Giraitis, L., Koul, H. L. & Surgailis, D. (2012), *Large Sample Inference for Long Memory Processes*, Imperial College Press, London.

Gneiting, T., Ševčíková, H., Percival, D. B., Schlather, M. & Jiang, Y. (2006), 'Fast and exact simulation of large Gaussian lattice systems in $\mathbb{R}^2$: exploring the limits', *Journal of Computational and Graphical Statistics* **15**(3), 483–501.

Gradshteyn, I. S. & Ryzhik, I. M. (2007), *Table of Integrals, Series, and Products*, seventh edn, Elsevier/Academic Press, Amsterdam.

Hannan, E. J. (1970), *Multiple Time Series*, John Wiley and Sons, Inc., New York-London-Sydney.

Helgason, H., Pipiras, V. & Abry, P. (2011), 'Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding', *Signal Processing* **91**(5), 1123–1133.

Helgason, H., Pipiras, V. & Abry, P. (2014), 'Smoothing windows for the synthesis of Gaussian stationary random fields using circulant matrix embedding', *Journal of Computational and Graphical Statistics* **23**(3), 616–635.

Hurvich, C. M. & Tsai, C.-L. (1989), 'Regression and time series model selection in small samples', *Biometrika* **76**(2), 297–307.

Kass, R. E. & Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(430), 773–795.

Kechagias, S. & Pipiras, V. (2015), 'Definitions and representations of multivariate long-range dependent time series', *Journal of Time Series Analysis* **36**(1), 1–25.

Korenev, B. G. (2002), *Bessel Functions and Their Applications*, Vol. 8 of *Analytical Methods and Special Functions*, Taylor & Francis Ltd., London.

Lobato, I. N. (1997), 'Consistency of the averaged cross-periodogram in long memory series', *Journal of Time Series Analysis* **18**(2), 137–155.

Lobato, I. N. (1999), 'A semiparametric two-step estimator in a multivariate long memory model', *Journal of Econometrics* **90**(1), 129–153.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Marinucci, D. & Robinson, P. M. (2001), 'Semiparametric fractional cointegration analysis', *Journal of Econometrics* **105**(1), 225–247.

Martin, V. L. & Wilkins, N. P. (1999), 'Indirect estimation of ARFIMA and VARFIMA models', *Journal of Econometrics* **93**(1), 149–175.

Meerschaert, M. M. & Scheffler, H.-P. (2013), 'Tauberian theorems for matrix regular variation', *Transactions of the American Mathematical Society* **365**(4), 2207–2221.

Nielsen, F. S. (2011), 'Local Whittle estimation of multi-variate fractionally integrated processes', *Journal of Time Series Analysis* **32**(3), 317–335.

Nielsen, M. Ø. (2004), 'Local empirical spectral measure of multivariate processes with long range dependence', *Stochastic Processes and their Applications* **109**(1), 145–166.

Nielsen, M. Ø. (2007), 'Local Whittle analysis of stationary fractional cointegration and the implied-realized volatility relation', *J. Bus. Econom. Statist.* **25**(4), 427–446.

Nielsen, M. Ø. & Frederiksen, P. (2011), 'Fully modified narrow-band least squares estimation of weak fractional cointegration', *The Econometrics Journal* **14**(1), 77–120.

Nocedal, J. & Wright, S. J. (2006), *Numerical Optimization*, second edn, Springer, New York.

Pai, J. & Ravishanker, N. (2009*a*), 'Maximum likelihood estimation in vector long memory processes via em algorithm', *Computational Statistics & Data Analysis* **53**(12), 4133–4142.

Pai, J. & Ravishanker, N. (2009*b*), 'A multivariate preconditioned conjugate gradient approach for maximum likelihood estimation in vector long memory processes', *Statistics & Probability Letters* **79**(9), 1282–1289.

Palma, W. (2007), *Long-Memory Time Series*, John Wiley & Sons, Inc., Hoboken, New Jersey, USA.

Park, K. & Willinger, W. (2000), *Self-Similar Network Traffic and Performance Evaluation*, Wiley Online Library.

Peach, R.W., R. R. & Antoniades, A. (2004), 'The historical and recent behaviour of goods and services inflation', *Economic Policy Review* pp. 19–31.

Percival, D. B. (2006), 'Exact simulation of complex-valued Gaussian stationary processes via circulant embedding', *Signal Processing* **86**(7), 1470–1476.

Pipiras, V. & Taqqu, M. S. (2015), *Long-Range Dependence and Self-Similarity*, Cambridge University Press, Cambridge.

Ravishanker, N. & Ray, B. K. (1997), 'Bayesian analysis of vector ARFIMA processes', *Australian Journal of Statistics* **39**(3), 295–311.

Reinsel, G. C. (1997), *Elements of Multivariate Time Series Analysis*, Springer Series in Statistics, second edn, Springer-Verlag, New York.

Robinson, P. M. (1994), 'Semiparametric analysis of long-memory time series', *The Annals of Statistics* **22**(1), 515–539.

Robinson, P. M. (1995), 'Log-periodogram regression of time series with long range dependence', *The Annals of Statistics* **23**(3), 1048–1072.

Robinson, P. M. (2003), *Time Series with Long Memory*, Advanced Texts in Econometrics, Oxford University Press, Oxford.

Robinson, P. M. (2008), 'Multiple local Whittle estimation in stationary systems', *The Annals of Statistics* **36**(5), 2508–2530.

Robinson, P. M. & Marinucci, D. (2003), Semiparametric frequency domain analysis of fractional cointegration, *in* P. M. Robinson, ed., 'Time Series with Long Memory', Oxford University Press, Oxford.

Robinson, P. M. & Yajima, Y. (2002), 'Determination of cointegrating rank in fractional systems', *Journal of Econometrics* **106**(2), 217–241.

Schwarz, G. et al. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**(2), 461–464.

Sela, R. J. (2010), 'Three essays in econometrics: multivariate long memory time series and applying regression trees to longitudinal data', *Ph.D. Dissertation, New York University*.

Sela, R. J. & Hurvich, C. M. (2009), 'Computationally efficient methods for two multivariate fractionally integrated models', *Journal of Time Series Analysis* **30**(6), 631–651.

Shimotsu, K. (2007), 'Gaussian semiparametric estimation of multivariate fractionally integrated processes', *Journal of Econometrics* **137**(2), 277–310.

Sowell, F. (1986), 'Fractionally integrated vector time series', *Ph.D. Dissertation, Duke University*.

Stein, M. L. (2002), 'Fast and exact simulation of fractional Brownian surfaces', *Journal of Computational and Graphical Statistics* **11**(3), 587–599.

Stein, M. L. (2012), 'Simulation of Gaussian random fields with one derivative', *Journal of Computational and Graphical Statistics* **21**(1), 155–173.

Tsay, W.-J. (2010), 'Maximum likelihood estimation of stationary multivariate ARFIMA processes', *Journal of Statistical Computation and Simulation* **80**(7), 729–745.

Velasco, C. (2003), 'Gaussian semi-parametric estimation of fractional cointegration', *Journal of Time Series Analysis* **24**(3), 345–378.

Wainger, S. (1965), 'Special trigonometric series in k-dimensions', *Memoirs of the American Mathematical Society* **59**, 102.

Wendt, H., Scherrer, A., Abry, P. & Achard, S. (2009), Testing fractal connectivity in multivariate long memory processes, *in* 'Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on', IEEE, pp. 2913–2916.

Wood, A. T. A. & Chan, G. (1994), 'Simulation of stationary Gaussian processes in $[0, 1]^d$', *Journal of Computational and Graphical Statistics* **3**(4), 409–432.