Quasi-likelihood inference for modulated non-stationary time series

Arthur P. Guillaumin

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Statistical Science

University College London

March 1, 2018

I, Arthur P. Guillaumin, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

In this thesis we propose a new class of non-stationary time series models and a quasi-likelihood inference method that is computationally efficient and consistent for that class of processes. A standard class of non-stationary processes is that of locally-stationary processes, where a smooth time-varying spectral representation extends the spectral representation of stationary time series. This allows us to apply stationary estimation methods when analysing slowly-varying non-stationary processes. However, stationary inference methods may lead to large biases for more rapidly-varying non-stationary processes. We present a class of such processes based on the framework of modulated processes. A modulated process is formed by pointwise multiplying a stationary process, called the latent process, by a sequence, called the modulation sequence. Our interest lies in estimating a parametric model for the latent stationary process from observing the modulated process in parallel with the modulation sequence. Very often exact likelihood is not computationally viable when analysing large time series datasets. The Whittle likelihood is a standard quasi-likelihood for stationary time series. Our inference method adapts this function by specifying the expected periodogram of the modulated process for a given parameter vector of the latent time series model, and then fits this quantity to the sample periodogram. We prove that this approach conserves the computational efficiency and convergence rate of the Whittle likelihood under increasing sample size. Finally, our real-data application is concerned with the study of ocean surface currents. We analyse bivariate non-stationary velocities obtained from instruments following the ocean surface currents, and infer key physical quantities from this dataset. Our simulations show the benefit of our modelling and estimation method.

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Many phenomena observed over time can be modelled as stochastic processes, that is to say families of random variables indexed by time. Common real-world phenomena described as stochastic processes range from applications in econometrics, to applications in seismology and oceanography. The need for non-deterministic models may come from an actual belief that an observed phenomenon is "purely" random. Equally one may have a deterministic model for the observed phenomenon but may not be able to directly measure all inputs at play. Stochastic modelling is a way to account for the uncertainty resulting from this indetermination.

Inference from time series requires to posit a model for the generating process. The two most common modelling assumptions in time series analysis are Gaussianity and stationarity. Gaussianity of a process implies that its distribution is fully described by its first and second-order moments. Gaussian processes naturally arise in many observed phenomena, for instance as the combination of independent effects, according to the central limit theorem. A stochastic process is weakly stationary if its first and second-order moments are finite and shift-invariant over time. This allows us to reduce the degrees of freedom and makes averaging over time meaningful. Another key point to the assumption of stationarity is the associated spectral representation theorem, where any stationary stochastic process can be represented as a sum of complex exponentials with uncorrelated random amplitudes and phases. The variances of these amplitudes define an energy spectrum, characterizing the distribution of energy (or variance) over frequencies, which is the

Fourier transform of the autocovariance function.

Real-world processes are, however, non-stationary in their vast majority. A commonly observed departure from stationarity is a non-constant expectation over time. A trend can be defined as a smooth variation over time while seasonality refers to periodic patterns in the expectation. Second-order moments may as well vary over time, which will be the focus of this thesis. Defining models of non-stationary stochastic processes with time-varying covariance structures is difficult in practice. Firstly, the models' autocovariance matrices must satisfy some specific structure, as they must be non-negative definite. Secondly, even the class of all Gaussian processes with time-varying covariances is too large as it allows for covariance structures where no meaningful averaging from a single realization is permitted. A useful non-stationary time series model is one where the information in the covariance structure is somehow redundant. One approach is to model a covariance structure that varies with time but which contains repeated patterns. For instance, cyclostationary processes, equally called periodically-correlated processes, are such that their autocovariance functions are periodic (Gladyshev, 1963). Another class of processes constructed on the idea of repeated patterns is that of asymptotically stationary processes (Parzen, 1961, Dunsmuir and Robinson, 1981c, Toloi and Morettin, 1989), where it is required that the sample autocovariance sequence converges in a probabilistic way to a fixed function of the lags. Alternatively, a class of non-stationary processes can be constructed by specifying a bi-spectral representation of the autocovariance function. The class of harmonizable processes introduced by Loève (1945) considers autocovariance functions that can be expanded on a basis of separable exponential functions. A contrasting approach to define a non-stationary process with a redundant covariance structure is to assume that the time-variation of the covariance structure is "smooth" in some way, such that local averages can be carried out in a meaningful way. In that sense, the contributions of Priestley (1988) and Dahlhaus (1996) on defining stochastic processes with time-varying spectral densities have been key to the field. The framework of *infill asymptotics*, where the time-variation is made slower and slower, offers a

meaningful tool to compare statistics applied to these processes. We review the stationarity assumption in Chapter 2 as well as the classes of non-stationary processes cited above. Note that other general classes of non-stationary processes have been established, such as the Karhunen class or the Cramér class (Rao, 1985), but we do not focus on these approaches in this thesis, as it is usually quite difficult in practice to specify a model within those classes for real-world problems.

Real-world observations also tend to come as multivariate processes, adding to the challenge of defining models with valid covariance matrices. One simple method to define valid classes of multivariate non-stationary processes is via the framework of modulated processes, first introduced by (Parzen, 1963). A modulated process is obtained by multiplying pointwise a stationary process. The class of modulated processes was also considered by Priestley (1988), who required the modulation to be a smooth function, such that a time-varying spectrum can be defined. Parzen (1963) however considered the case where the modulation function is observed, but no smoothness assumptions are made. In particular, an application to periodically missing data was presented, where the modulation function takes the values zero for missing and one for observed. Not any form of modulation will lead to a stochastic process that may be inferred from a single realization. If the modulation function is periodic, then the modulated process is obviously cyclo-stationary. A more general assumption in the literature is that the modulation makes the modulated process asymptotically stationary. This requires that the sample autocovariance function of the modulation sequence converges to a non-zero value at all lags. However, this requirement is still too stringent for some practical applications. In Chapter 3 we review modulated processes and we propose a larger class of modulated processes, called *modulated processes with a significant correlation contribution*, where instead it is required that the sample autocovariance of the modulation sequence be asymptotically bounded below for certain lags that characterize the parametric model of the stationary latent process.

Parametric point inference within a parameter set can give some insight on the generating mechanism at work behind an observed time se-

ries. For instance, in our real-world application, our stochastic model is adapted from a deterministic model from oceanography, whose parameters can be interpreted in a meaningful way. The ubiquitous large data sets met in practice require fast estimation methods. The Global Drifter Program (`http://www.aoml.noaa.gov/phod/dac/index.php`) considered in our application to oceanography involves more than ten thousand drifters sending positional coordinates on a regular basis, with an average 1.4 hour interval since 2005, resulting in a data set with millions of points (Elipot et al., 2016). One solution to this computational burden is to use approximations of the likelihood function, commonly termed as *quasi-likelihoods* or *pseudo-likelihoods* (McCullagh, 1983). One such common approximation for stationary stochastic processes is the Whittle likelihood (Whittle, 1953). The Fourier transforms at Fourier frequencies of a stationary time series are known to be asymptotically independently distributed, with variances given by the spectral density. Thus the Gaussian maximum likelihood estimate is approximated by computing the periodogram of the time series and fitting it to the spectral density function. In Chapter 4 we review some inference methods for stationary processes, such as the method of moments, the maximum likelihood, the Whittle likelihood, and more generally for locally-stationary processes.

In Chapter 5 we consider a quasi-likelihood estimation method for modulated processes with a significant correlation contribution. The quasi-likelihood we introduce is based on an adaptation of the Whittle likelihood, involving the periodogram of the time series, and its theoretical expectation under the distribution specified by a vector of parameters at which we evaluate the quasi-likelihood. We establish the consistency of our quasi-likelihood estimator in Chapter 6, and show that it converges with rate $\mathcal{O}_P\left(N^{-1/2}\right)$ for a length-N time series.

We present various simulation studies in Chapter 7, including for missing data. In Chapter 8 we present a real-world application of our model and inference method, where we analyse the velocities of ocean surface drifting instruments from the Global Drifter Program. Some further simulation studies are included to show the adequacy of our method with the oceanographic model. In Chapter 9 we

present our conclusions and ideas for future work.

At the beginning of each chapter, we clearly state which section contains reviewed material versus new material. All proofs in the main body of the thesis are by the author, other proofs being left in the appendix. The work presented in this thesis is published in Guillaumin et al. (2017) and also contributed to Sykulski et al. (2016a).

Chapter 2

Stationarity and other time series models

In this chapter we formally define stochastic processes on the set of integers, and observations from stochastic processes, which are called time series. A common practical approach to defining a stochastic process usually consists in specifying finite joint distributions. In Section 2.1 we review the Kolmogorov extension theorem which provides mild conditions under which this methodology defines a valid stochastic process. The basics of time series analysis usually reduces to describing first and second-order moments, assuming they are finite. In the case of a Gaussian process, those moments completely specify the stochastic process. We first consider the analysis of first-order moments in Section 2.2, where we study the common notions of trend and seasonality. We then consider the analysis of second-order moments in Section 2.3 and study the standard stationarity assumption and the notion of a spectrum associated with a stationary time series. Following this we present different classes of non-stationary time series. The classes of asymptotically stationary time series and periodically correlated time series can be viewed as extending the concept of stationarity, in the sense that averaging still makes sense and allows for consistent estimators within the framework of increasing sample size asymptotics. Hence in Section 2.4 we present these two classes in a section called "almost stationary stochastic processes". We then review the class of harmonizable processes in Section 2.5, and we conclude this chapter with the class of locally-stationary pro-

cesses in Section 2.6, which are based on time-varying spectral representations. All the material presented in this chapter is reviewed material.

2.1 Stochastic processes

We start with the formal definition of a stochastic process and quickly review the inherent issue of existence given specified joint distributions, which is solved by the Kolmogorov extension theorem.

Definition 1 (Stochastic process). *Let T be a set. A stochastic process indexed by T is a family of random variables $\{X_t : t \in T\}$ from a probability space $(\Omega, \mathscr{A}, \mathscr{P})$ onto a measurable space $(E, \mathscr{E})$. Here Ω is any non-empty set, $\mathscr{A}$ is a σ-field on Ω (i.e. a class of subsets of Ω containing Ω, and closed under countable unions and the complementary operation), and $\mathscr{P}$ is a probability on $(\Omega, \mathscr{A})$, that is to say a measure with total measure $\mathscr{P}(\Omega)$ one. E is a non-empty set and $\mathscr{E}$ is a σ-field on E.*

In time series analysis, we consider the case $T = \mathbb{R}$ for *continuous-time* processes and $T = \mathbb{Z}$ or $T = \mathbb{N}$ for *discrete-time* processes, which implies that the set T possesses a natural ordering. Note that this choice for discrete-time processes implies that we only consider regularly sampled processes, and that we take the sampling step to be one without loss of generality. We usually consider the case where the stochastic process takes values in $E = \mathbb{R}^d$ (real-valued process) or $E = \mathbb{C}^d$ (complex-valued process), for some positive integer d. Complex-valued variables and time series will be reviewed in Section 3.5.1, and we will justify their use to analyse certain classes of bivariate real-valued time series. In this chapter, we consider a real-valued stochastic process $\{X_t\}$, where we use the notation of Definition 1. A *realization* of the so-defined process $\{X_t\}$ is given by $\{X_t(\omega) : t \in T\}$ for a particular value of $\omega \in \Omega$. Equivalently the term *path* is commonly used in the literature. For continuous-time processes, continuity of the paths, which corresponds to $\mathscr{P}(\{\omega : (t \to X_t(\omega)) \text{ is continuous}\}) = 1$, is called almost sure continuity.

The question of defining a valid stochastic process arises quickly. A simple

example of a stochastic process would be,

$$X_t = A\cos(2\pi ft + \phi), \quad \forall t \in \mathbb{Z}, \tag{2.1}$$

where f and ϕ are fixed, and A is a random variable on some probability space $(\Omega, \mathscr{A}, \mathscr{P})$ with a uniform distribution over $[-1, 1]$. In that case, for every $t \in \mathbb{Z}$, we can easily see that (2.1) amounts to setting $X_t(\omega) = A(\omega)\cos(2\pi ft + \phi), \forall \omega \in \Omega, \forall t \in \mathbb{Z}$. In more complicated situations finding a proper probability space $(\Omega, \mathscr{A}, \mathscr{P})$ to define a stochastic process may not be obvious at all. The Kolmogorov extension theorem, of which we give a version for time series below, solves this issue by ensuring that, under some mild conditions on specifying the joint distributions of finite sub-samples of the process, there exists a corresponding stochastic process on some probability space.

Theorem 1 (Kolmogorov extension theorem). *Let $T = \mathbb{R}$ or $T = \mathbb{Z}$. For any integer $n \geq 1$, for any times $t_1, \ldots, t_n \in T$, let $F_{t_1,\ldots,t_n}(x_1,\ldots,x_n)$ be a cumulative distribution function. Assume the following two natural requirements are met,*

- *for all integer $n \geq 1$, for all $t_1, \ldots, t_n \in T$, and for all $x_1, \ldots, x_n \in \mathbb{R}$, given any permutation $\pi(\cdot)$ of the set $\{1, \ldots, n\}$,*

$$F_{\pi(t_1),\ldots,\pi(t_n)}\big(x_{\pi(t_1)},\ldots,x_{\pi(t_2)}\big) = F_{t_1,\ldots,t_n}(x_1,\ldots,x_n),$$

- *for all integers $n \geq 1$, $m \geq 1$, for all $t_1, \ldots, t_{n+m} \in T$, and for all $x_1, \ldots, x_n \in \mathbb{R}$,*

$$F_{t_1,\ldots,t_n}(x_1,\ldots,x_n) = \lim_{y_1 \to \infty, \ldots, y_m \to \infty} F_{t_1,\ldots,t_n,t_{n+1},\ldots,t_{n+m}}(x_1,\ldots,x_n,y_1,\ldots,y_m).$$

If these conditions hold there exists a probability space $(\Omega, \mathscr{A}, \mathscr{P})$ and a stochastic process $\{X_t : t \in T\}$ on that probability space such that the joint cumulative distribution functions of the finite sub-samples $X_{t_1}, \ldots, X_{t_n}$ are given by $F_{t_1,\ldots,t_n}(x_1,\ldots,x_n)$.

Therefore one can define a valid stochastic process only by specifying the joint probabilities of any finite sub-sample from that process. For instance, a white noise

process W_t with parameter σ is a process made of uncorrelated random variables with mean zero and variance σ^2. If we assume all W_t have a standard normal distribution function and denote $\Phi(x)$ the corresponding cumulative distribution function, then for any integer $n \geq 1$, and any $t_1, \ldots, t_n \in T$, and for any $x_1, \ldots, x_n \in \mathbb{R}$, we define the following joint cumulative distribution functions,

$$F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) = \prod_{k=1}^{n} \Phi(x_k),$$

and the consistency conditions of Theorem 1 are easily verified based on that definition. This proves the existence of the discrete-time standard Gaussian white noise process. In practice, many models are based on a white noise process, like autoregressive and moving average models (Brockwell and Davis, 1991). Additionally, the Kolmogorov extension theorem shows that we can define entirely a Gaussian process by only specifying its expectations and its finite covariances, since in the Gaussian case those determine entirely the cumulative distribution functions. In this thesis we only consider Gaussian processes, unless stated otherwise. Additionally, we shall only consider discrete-time processes, i.e. the case $T = \mathbb{Z}$ or $T = \mathbb{N}$, unless stated otherwise.

2.2 Description of first-order moments

The most basic description of a discrete time series $\{X_t : t \in \mathbb{N}\}$ is given by its first-order moments, that is to say its expectations $\mu_t = \mathrm{E}\{X_t\}$. Consider a length-$N$ sample from that process, denoted $X_0, \ldots, X_{N-1}$. Different assumptions can be made as concerns these expectations. If the expectations are constant over time, say μ, and if the variances $\mathrm{var}\{X_t\}$ are bounded, according to the central limit theorem one can simply estimate μ by averaging, $\hat{\mu} = \frac{1}{N} \sum_{t=0}^{N-1} X_t$. In practice however, most time series do not satisfy this assumption. As an example, we retrieved the number of Google search requests (www.trends.google.com) for the term *Machine learning* in France from Jul 2012 to Jul 2017. Those are shown in Fig 2.1. It appears that the observations tend to increase in a polynomial, or maybe in an exponential way. The term used for such behaviour is usually that of trend. Another frequently observed
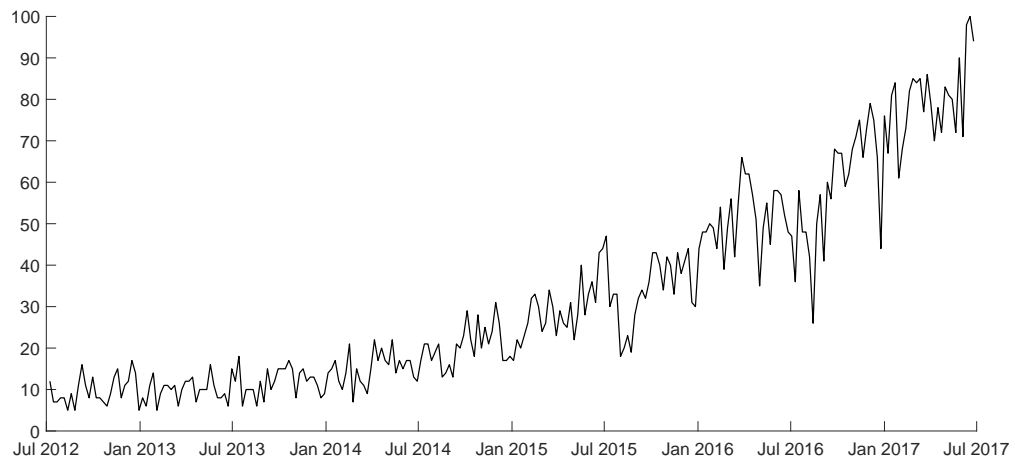
Figure 2.1: Number of Google searches for the term "Machine learning" in France from July 2012 to July 2017. The observations are scaled from 0 to 100 where 100 corresponds to the maximal peak attained (undisclosed). Source: `https://trends.google.fr/`.
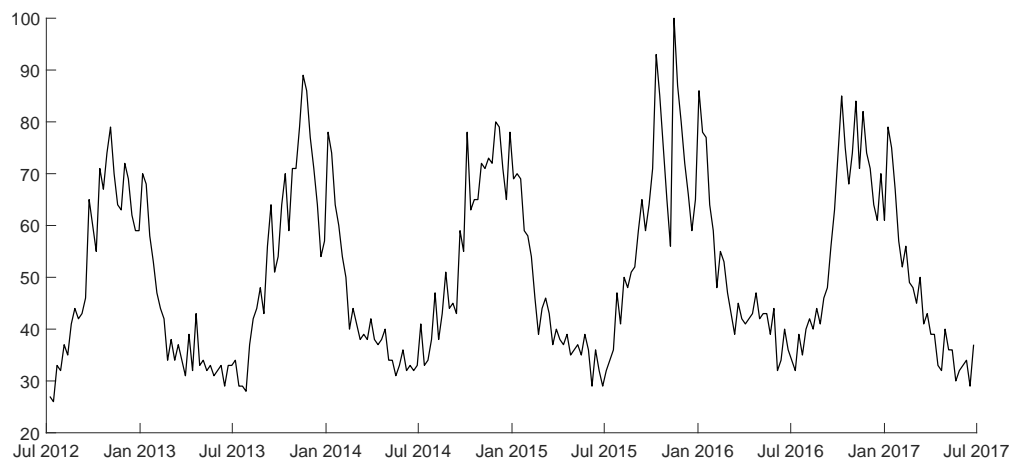
Figure 2.2: Number of Google searches for the term "Coat" in France from July 2012 to July 2017. The observations are scaled from 0 to 100 where 100 corresponds to the maximal peak attained (undisclosed). Source: `https://trends.google.fr/`.

behaviour is that of seasonality. Consider the example of Fig 2.2, which shows the number of Google search requests for the term *coat* in France from Jul 2012 to Jul 2017. Not surprisingly, this time series will be strongly related to weather conditions, explaining the yearly pattern that we can observe. In this thesis we do not focus on the problem of estimation of first-order moments, and assume all throughout this thesis that the processes have mean zero, unless stated otherwise.

2.3 Description of second-order moments

A more advanced description of a time series, once a model has been posited for the first-order moments, is to specify the second-order moments. Again some assumptions are required so that some form of averaging, and therefore estimation, is possible. In this section we first recall the definition of a stationary process and the corresponding spectral representation. We then consider two cases where the assumption of stationarity is relaxed: asymptotically stationary time series and periodically stationary time series. In both cases, the assumptions are such that consistent estimators can still be obtained under the asymptotic framework of increasing sample size. We then consider the class of locally-stationary time series that was introduced by Priestley (1988) and were given a proper asymptotic framework by Dahlhaus (1997).

2.3.1 Stationary processes

A key feature of stationarity is that is allows for estimation to performed by averaging over time. There are two main definitions for stationarity. The first one, which we define below, is very restrictive and difficult to verify in practice.

Definition 2 (Strictly-stationary processes)**.** *Let $\{X_t : t \in \mathbb{Z}\}$ be a discrete time stochastic process. For any $n \geq 1$ and any set of n time indices $t_1, \ldots, t_n$, we denote by $F_{t_1, \ldots, t_n}(x_1, \ldots, x_n)$ the cumulative distribution function of the vector $[X_{t_1}, \ldots, X_{t_n}]^T$. The process $\{X_t\}$ is strictly-stationary if, for any $\tau \in \mathbb{Z}$, we have the equality,*

$$F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) = F_{t_1 + \tau, \ldots, t_n + \tau}(x_1, \ldots, x_n), \quad \forall x_1, \ldots, x_n \in \mathbb{R}.$$

This definition is a very strong requirement. For that reason the concept of weak-stationarity, which we define below, is much more common in the time series literature, and we shall refer to it simply as "stationarity" since we will not make any use of Definition 2 in this thesis.

Definition 3 (Weakly-stationary processes)**.** *Let $\{X_t : t \in \mathbb{N}\}$ be a discrete time series. The process $\{X_t\}$ is said to be (weakly-)stationary if,*

- *its means* $\mathrm{E}\{X_t\}$ *are all finite and equal to a constant independent from t, say* μ_X,

- *its second-order joint moments are all finite and there exists a sequence* $\{c_X(\tau) : \tau \in \mathbb{Z}\}$ *such that,*

$$\mathrm{E}\{(X_t - \mu_X)(X_{t+\tau} - \mu_X)\} = c_X(\tau).$$

Stationarity therefore requires that first-order moments be independent from time, and second-order joint moments be only a function of the lag τ. Note that strictly-stationary processes are not always weakly-stationary, as a strictly-stationary process does not necessarily have finite second-order moments. Weakly-stationary processes are of course not strongly-stationary in general, except in some specific cases such as that of Gaussian processes where the two definitions are equivalent.

The quantity $\{c_X(\tau)\}$ that appears in the definition of a weakly-stationary process is called the autocovariance function of the process. A natural question that arises is, what sequences are valid autocovariance functions? It is clear that not all sequences are, since we must have that $c_X(0) \geq c_X(\tau), \forall \tau \in \mathbb{Z}$ according to the Cauchy-Schwarz inequality. This also means that if we define the autocorrelation function as $\gamma_X(\tau) = c_X(\tau)/c_X(0), \tau \in \mathbb{Z}$, it must be bounded in absolute value by one. To investigate more precisely the characterization of autocovariance sequences we need the following definition.

Definition 4 (Non-negative definite sequence). *Let* $(a_i)_{i \in \mathbb{Z}}$ *be a sequence indexed by the set of natural integers. It is said to be non-negative definite if for any* $n \geq 1$ *the following condition is satisfied,*

$$\sum_{i,j=1}^{n} a_{i-j} \lambda_i \lambda_j \geq 0, \quad \forall \lambda_1, \ldots, \lambda_n \in \mathbb{R}.$$

It is symmetric if $a_i = a_{-i}, \forall i \in \mathbb{Z}.$

The following lemma states that autocovariance sequences are exactly sym-

metric non-negative definite sequences.

Lemma 1 (Characterization of autocovariance sequences). *Let $(a_i)_{i \in \mathbb{Z}}$ be a sequence indexed by the set of natural integers. The sequence $(a_i)_{i \in \mathbb{Z}}$ is the autocovariance function of a stationary discrete-time process if and only if it is symmetric and non-negative definite.*

Proof. See Appendix A.1. □

The property of Lemma 1 is key in specifying valid models for stationary time series and in time series inference, where a desirable property of the estimated autocovariance of a stationary discrete time series is that it is non-negative definite. In particular, this will be an important consideration in specifying valid parametric models for stationary autocovariance functions. Moreover this property leads to the definition of the spectrum of a stationary time series, which is the topic of the next section.

2.3.2   Spectral distribution function and spectral representation

In this section we recall the definition of a spectrum for a stationary time series, which decomposes the variance of a stationary process across frequencies. We then review a fundamental result of stationary time series analysis, where a stationary stochastic process can be represented as a sum of sinusoids with random and uncorrelated amplitudes and phases.

The idea behind the spectrum of a stationary process is a one-to-one transformation between the set of autocovariance functions of stationary processes and the set of distribution functions defined on $[-\pi, \pi]$.

Proposition 1 (Spectral distribution function of a stationary process). *Let $\{X_t : t \in \mathbb{Z}\}$ be a stationary process with autocovariance function $c_X(\tau)$. There exists a distribution function $F : [-\pi, \pi) \to \mathbb{R}$ (distribution function here meaning non-decreasing, with value 0 at $-\pi$ and bounded) such that,*

$$c_X(\tau) = \int_{-\pi}^{\pi} e^{i\omega\tau} dF(\omega).$$

Here the integral is a Lebesgue-Stieltjes integral. In the case where the function
$F(\omega)$ admits a derivative everywhere, denoted $f(\omega)$, we have,

$$c_X(\tau) = \int_{-\pi}^{\pi} f(\omega)e^{i\omega\tau}d\omega.$$

The function $F(\omega)$ always exists and is called the spectral distribution function of
the process. When it exists, the function $f(\omega)$ is called the spectral density function
of the process, and is then related to the autocovariance function by,

$$f(\omega) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} c_X(\tau)e^{-i\omega\tau}, \tag{2.2}$$

i.e. $f(\omega)$ is the Fourier series with coefficients $c_X(\tau), \tau \in \mathbb{Z}$.

Proof. See Appendix A.2. □

The spectral distribution function describes how the variance (*energy* in the
signal-processing literature) of a stationary process spreads across frequencies. It
can be easier or more relevant to study the spectrum rather than the autocovari-
ance function. The following lemma gives a general case where a spectral density
function is well defined.

Lemma 2 (Spectral density for processes with short memory). *Let $\{X_t : t \in \mathbb{Z}\}$ be*
a stationary process with autocovariance sequence $c_X(\tau)$. Assume that the autoco-
variance sequence is absolutely summable, that is to say $\sum_{\tau=-\infty}^{\infty} |c_X(\tau)| < \infty$. Then
the process admits a spectral density function $f(\omega) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} c_X(\tau)e^{-i\omega\tau}$.

Proof. See Appendix A.3. □

As a simple example, the process $\{W_t : t \in \mathbb{Z}\}$ is called a white noise process
if it is stationary and its autocovariance function satisfies $c_W(0) > 0$ and $c_W(\tau) =$
$0, \tau > 0$. Its spectral density is constant and equal to $f(\omega) = c_W(0)/(2\pi)$.

Similarly to the spectral distribution function as an equivalent of the autoco-
variance function, there exists an equivalent representation of a stationary stochastic
process as a sum of sinusoids with uncorrelated random amplitudes, which we now
define.

Proposition 2 (Spectral representation theorem). *Let $\{X_t : t \in \mathbb{Z}\}$ be a stationary process with mean zero and with spectral distribution function $F(\omega)$. There exists a stochastic process with orthogonal increments $\psi(\omega)$ defined on $[-\pi, \pi]$ such that,*

$$X_t =_{m.s} \int_{-\pi}^{\pi} e^{i\omega t} d\psi(\omega), \tag{2.3}$$

where $=_{m.s}$ means that the equality is to be understood in the mean square sense. Moreover, $F(\omega)$ is the distribution function of $\psi(\omega)$, and in particular,

$$\mathrm{E}\left\{ [\psi(\omega_2) - \psi(\omega_1)]^2 \right\} = |F(\omega_2) - F(\omega_1)|.$$

This property is very often summed-up in the literature by the following approximate notation,

$$\mathrm{E}\{|d\psi(\omega)|^2\} = dF(\omega), \tag{2.4}$$

or $\mathrm{E}\{|d\psi(\omega)|^2\} = f(\omega)d\omega$ if the spectral density function exists.

Note that for the process $\{X_t\}$ to be real-valued we must have $\psi(-\omega) = \psi(\omega)^*$, where the symbol $*$ denotes complex-conjugation. The fact that equality (2.3) holds in terms of mean square error means that for any $t \in \mathbb{Z}$ we have,

$$\mathrm{var}\left\{ X_t - \int_{-\pi}^{\pi} e^{i\omega t} d\psi(\omega) \right\} = 0. \tag{2.5}$$

In the case where the spectral distribution function admits a discontinuity at $\omega_0 \in (-\pi, \pi]$, it is possible to write (Brockwell and Davis, 1991),

$$X_t = \int_{(-\pi,\pi)\backslash \omega_0} e^{i\omega t} d\psi(\omega) + \left[\psi(\omega_0) - \psi(\omega_0^-) \right] e^{i\omega_0 t},$$

where $\mathrm{var}\left\{ \psi(\omega_0) - \psi(\omega_0^-) \right\} = F(\omega_0) - F(\omega_0^-)$. Therefore discontinuities in the spectral distribution function account for the *deterministic* part of the process. Note that if the process $\{X_t\}$ is real-valued and its spectral distribution function admits a discontinuity at ω_0 then it must also admit a discontinuity at $-\omega_0$.

2.4 Almost stationary stochastic processes

Stationary processes present the advantage of a non-evolving first and second-order structure, making averaging possible and therefore allowing for a wide range of estimation methods, such as the method of moments or likelihood inference methods, which we shall review in Chapter 4. In this section we consider two classes of time series models where averaging remains feasible, despite relaxing the assumption of stationarity. For that reason we call this section "almost stationary stochastic processes". The first class of models we present here is that of asymptotically stationary processes (Parzen, 1961), for which it is not required that second-order moments be constant over time but instead the assumption is that the sample autocovariance sequence converges in some probabilistic way. The second class of models is that of periodically correlated processes (Gladyshev, 1963, Hurd and Miamee, 2007), where the assumption is that the autocovariance function is periodic.

2.4.1 Asymptotic stationarity

Asymptotically stationary time series were introduced by Parzen (1961) to account for the fact that despite the autocovariance of a stochastic process may not satisfy the assumption of stationarity, the sample autocovariance sequence may still converge in some probabilistic way.

Definition 5 (Asymptotically stationary process)**.** *Let* $\{X_t\}$ *be a discrete-time random process. We say that* $\{X_t\}$ *is an asymptotically stationary process if there exists a fixed positive function* $\{\gamma(\tau) : \tau \in \mathbb{N}\}$ *such that for all* $\tau \in \mathbb{N}$,

$$\lim_{N \to \infty} \mathrm{E}\left\{ \frac{1}{N} \sum_{t=0}^{N-\tau-1} X_t X_{t+\tau} \right\} = \gamma(\tau). \tag{2.6}$$

A particularly enticing subclass of asymptotically stationary processes is that of asymptotically stationary modulated processes which we review in Section 3.2.

2.4.2 Periodically correlated time series

A stationary process is one where the first and second-order structure are invariant under shifts. The class of periodically correlated time series (Gladyshev, 1963,

Gardner et al., 2006, Hurd and Miamee, 2007) extends that of stationary time series by allowing for the second-order structure to be time-invariant under shifts of multiples of a given length denoted d. With this assumption all features of the covariance structure can still be observed infinitely many times and averaging is still feasible. We give a formal definition of periodically correlated (equivalently called cyclostationary) processes below.

Definition 6 (Cyclostationary process). *Let $\{X_t : t \in \mathbb{Z}\}$ be a discrete-time process. This is said to be a periodically correlated process, or a cyclostationary process, if there exists $T \in \mathbb{N}$, $T > 1$, such that,*

1. $\mathrm{E}\{X_t\} = \mathrm{E}\{X_{t+T}\}$, $\forall t \in \mathbb{Z}$,

2. $\mathrm{cov}\{X_t, X_{t+\tau}\} = \mathrm{cov}\{X_{t+T}, X_{t+T+\tau}\}$, $\forall t, \tau \in \mathbb{Z}$.

Cyclostationary processes naturally occur in many instances in the real-world. In econometrics, the periodicity in the first and second-order moments is generated by the natural latent seasonalities of the processes.

2.5 Harmonizable processes

A large class of non-stationary processes is obtained by allowing for correlation between increments of the process $\psi(\omega)$ in the spectral representation (2.3). This corresponds to the class of *harmonizable* processes (Loève, 1945).

Definition 7 (Harmonizable process). *Let $\{X_t\}$ be a stochastic process with mean zero. The process $\{X_t\}$ is called harmonizable if there exists a complex-valued stochastic random measure of bounded variation $\psi(d\omega)$ over $([-\pi, \pi], \mathscr{B}([-\pi, \pi]))$, not necessarily orthogonally scattered, such that,*

$$X_t = \int_{-\pi}^{\pi} e^{i\omega t} \psi(d\omega).$$

There exists a scalar measure $F(d\omega_1, d\omega_2)$ such that, for all Borel sets $A, B \in \mathscr{B}([-\pi, \pi])$, we have,

$$\mathrm{cov}\{\psi(A), \psi(B)\} = F(A, B).$$

If the measure $F(d\omega_1, d\omega_2)$ admits a density $f(\omega_1, \omega_2)$, then it is called the bi-spectral density function. $F(d\omega_1, d\omega_2)$ is called the bi-spectral distribution function, and ψ the bi-spectral random distribution function.

The covariance function of the harmonizable process $\{X_t\}$ of Definition 7 is given by,

$$\text{cov}\{X_s, X_t\} = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i(\omega_1 t - \omega_2 s)} F(d\omega_1, d\omega_2).$$

Periodically correlated processes, presented in Section 2.4.2, are an example of harmonizable processes, where the bi-spectral distribution $F(d\omega_1, d\omega_2)$ is concentrated on lines parallel to the identity (Hurd and Miamee, 2007).

2.6 Locally stationary processes

Except for some specific cases, defining a valid class of non-stationary autocovariance functions is a complicated problem. This is because the covariance matrix of any sub-sample must be non-negative definite, as a generalization of Lemma 1 for stationary processes. Another approach is to allow for some variation in the spectral representation of a stationary process given in (2.3). We recall that a stationary process $\{X_t : t \in \mathbb{Z}\}$ with mean zero can be represented in the following way, according to the spectral representation theorem,

$$X_t = \int_{-\pi}^{\pi} e^{i\omega t} d\psi(\omega),$$

where ψ is a stochastic process with orthogonal increments and such that $\psi(-\omega) = \psi(\omega)^*$ as X_t is real-valued. Based on this, one way to define a non-stationary process with mean zero consists in specifying (Priestley, 1965, Dahlhaus, 1997, Mélard and Herteleer-De Schutter, 1989, Grenier, 1983),

$$X_t = \int_{-\pi}^{\pi} A(t, \omega) e^{i\omega t} d\psi(\omega), \tag{2.7}$$

where $\psi(\omega)$ is a stochastic process on $[-\pi, \pi)$ with orthogonal increments and spectral distribution function $F(\omega)$, and where $A(t, \omega)$ is a function from $\mathbb{Z} \times$

$[-\pi, \pi]$ to $\mathbb{C}$. This class of processes contains that of stationary processes with a spectral density function $f(\omega)$, as we can then choose $A(t, \omega) = f(\omega)$. The auto-covariance function of the process specified by (2.7) is given by, for any time $t \in \mathbb{Z}$ and lag $\tau \in \mathbb{Z}$,

$$
\begin{aligned}
\mathrm{E}\{X_t X_{t+\tau}\} &= \mathrm{E}\left\{ \left(\int_{-\pi}^{\pi} A(t, \omega) e^{i\omega t} d\psi(\omega) \right)^* \left(\int_{-\pi}^{\pi} A(t+\tau, \omega) e^{i\omega(t+\tau)} d\psi(\omega) \right) \right\} \\
&= \int_{-\pi}^{\pi} A(t, \omega)^* A(t+\tau, \omega) e^{i\omega\tau} dF(\omega),
\end{aligned}
$$

according to the properties of integration with respect to a stochastic process with orthogonal increments. Equation (2.7) always define a valid autocovariance form as we have, for any positive integer $n \in \mathbb{N}$, any $t_1, \dots, t_n \in \mathbb{Z}$, any $\lambda_1, \dots, \lambda_n \in \mathbb{C}$,

$$
\begin{aligned}
\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i^* \lambda_j \mathrm{cov}\{X_{t_i}, X_{t_j}\} &= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i^* \lambda_j \int_{-\pi}^{\pi} A(t_i, \omega)^* A(t_j, \omega) e^{i\omega(t_j - t_i)} dF(\omega) \\
&= \int_{-\pi}^{\pi} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i^* \lambda_j A(t_i, \omega)^* A(t_j, \omega) e^{i\omega(t_j - t_i)} dF(\omega) \\
&= \int_{-\pi}^{\pi} \left| \sum_{i=1}^{n} \lambda_i A(t_i, \omega) e^{i\omega t_i} \right|^2 dF(\omega) \geq 0.
\end{aligned}
$$

The non-negativeness of that last quantity results from the fact that F is non-decreasing. It implies that equation (2.7) always defines a valid stochastic process, in the sense that the resulting process has a valid, i.e. non-negative definite, auto-covariance function. A stationary process has a single representation in the form of equation (2.3) up to a multiplicative constant. However, the representation (2.7) may not be unique, and we may need some constraints so that one can interpret the quantity $A(t, \omega)$ as the spectral component of the process X_t at frequency ω and time t. Priestley (1965) introduces the notion of oscillatory functions, in the sense that $t \to A(t, \omega)$ is required to have a Fourier transform with a maximal absolute value at frequency zero, so that $t \to A(t, \omega) e^{i\omega}$ can be interpreted as an oscillatory function with frequency ω, and $A(t, \omega)$ can be interpreted as the *envelope*. With such a definition, a time-varying spectral distribution function $F_t(\omega)$ with respect to

$F(\omega)$ can be defined by,

$$dF_t(\omega) = |A(t,\omega)|^2 dF(\omega).$$

Different normalization constraints can be posited for the functions $A(t,\omega)$, but independently of those choices we get,

$$\text{var}\{X_t\} = \int_{-\pi}^{\pi} dF_t(\omega),$$

which provides us with some understanding of the frequency content of the process X_t at time t.

Building on this idea, Dahlhaus (1997) introduced the class of *locally stationary* processes, providing with a meaningful asymptotic framework. This is accomplished by defining a triangular array of processes according to,

$$X_{t,T} = \int_{-\pi}^{\pi} A_{t,T}(\omega) e^{i\omega t} d\psi(\omega), \quad T \in \mathbb{N}, \ t = 0,\ldots,T. \tag{2.8}$$

Such a process is called *locally stationary* if,

1. $\psi(\omega)$ is a stochastic process with orthogonal increments such that $\psi(-\omega) = \psi(\omega)^*$, and with spectral distribution function $F_\psi(\omega) = \omega + C$ (where C is any real-valued constant), i.e. such that $\text{E}\{|\psi(\omega_2) - \psi(\omega_1)|^2\} = |\omega_2 - \omega_1|$.

2. There exists a function $A(u,\omega)$ from $[0,1] \times [-\pi,\pi]$ to $\mathbb{C}$, which is continuous in u, and a positive constant K such that,

$$\left| A_{t,T}(\omega) - A\left(\frac{t}{T},\omega\right) \right| \leq \frac{K}{T}, \quad \forall T \in \mathbb{N}, \ t = 0,\ldots,T.$$

The distinction between $A_{t,T}(\omega)$ and $A(u,\omega)$ is required so that this definition can account for autoregressive models with time-varying parameters for instance (Dahlhaus, 1996). Such models are defined by,

$$\sum_{j=0}^{p} \phi_j(t/T) X_{t-j,T} = \sigma(t/T)\varepsilon_{t,T}, \tag{2.9}$$

with $\phi_0(u) = 1$ and the functions $\phi_j(u)$ are assumed continuous.

The process defined by (2.8) is such that the quantity $f(u, \omega) = |A(u, \omega)|^2$ is uniquely defined as the limit, for T going to infinity, of the Wigner-Ville spectrum,

$$f_T(u, \omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \text{cov}\{X_{\lfloor uT \rfloor - \tau/2, T}, X_{\lfloor uT \rfloor + \tau/2, T}\} e^{-i\omega\tau}. \qquad (2.10)$$

This as such allows to define a unique evolving spectral density $f(u, \omega) = |A(u, \omega)|^2$ associated with the non-stationary process $\{X_t\}$. The asymptotic framework corresponding to T going to infinity is called *infill asymptotics* (Dahlhaus, 1997) and does not correspond to the increasing sample size framework usually considered in stationary time series analysis. However it allows to derive asymptotic quantities, such as the asymptotic Kullback-Leibler information divergence between two locally stationary processes, which will depend on the respective evolving spectral densities. More precisely, assume Gaussianity, and let $f_1(u, \lambda)$ and $f_2(u, \lambda)$ be two time-varying spectral densities. Denote $\overline{f}_1, \overline{f}_2$ the respective probability density functions of samples $X_{1,T}, \ldots, X_{T,T}$. We then have, up to a constant,

$$\lim_{T \to \infty} \frac{1}{T} \text{E}_{\overline{f}_1} \left\{ \log \frac{\overline{f}_1(X_{1,T}, \ldots, X_{T,T})}{\overline{f}_2(X_{1,T}, \ldots, X_{T,T})} \right\} = \int_0^1 \int_{-\pi}^{\pi} \left\{ \log f_2(u, \lambda) + \frac{f_1(u, \lambda)}{f_2(u, \lambda)} \right\} d\lambda \, du.$$

In Section 4.4 we will review some parametric inference methods for locally stationary processes. The general idea behind those methods is to estimate the time-varying spectral density and minimize the approximate Kullback-Leibler information divergence between this estimate and a parametric form of the time-varying spectral density. The minimization is conducted within a parametric family of time-varying spectral densities that constitutes the model, and which needs not contain the true time-varying spectral density.

Chapter 3

Modulated stochastic processes

In Chapter 2 we considered numerous ways of departing from the assumption of stationarity. This is necessary to account for a wide range of real-world time series that cannot realistically be modelled as realizations of stationary stochastic processes. The class of locally-stationary processes introduced by Dahlhaus (1996) provides tools to understand the impact of fitting a stationary model to a non-stationary stochastic process that possesses a smooth time-varying spectral representation. Local Fourier transforms can be used to estimate a local spectrum (Dahlhaus, 1997). We shall quickly revisit this topic later in Section 4.4 with regards to parameter inference. The limitations of the locally-stationary framework however, is that in general we do not acquire more information about the behaviour of the process at present as we observe more points in the future. To hope for such, we need to impose on the model that part of the information repeats itself over time an infinite number of times. This is the idea behind cyclostationary processes, for example, as introduced in Section 2.4.2. Another way to construct a non-stationary stochastic process where part of the information is repeated over time is modulation. A modulated stochastic process is a stationary stochastic process multiplied pointwise by a continuous function (for time-continuous processes) or a sequence (for discrete-time processes). We will see that conditions on the modulating sequence will determine how much information about the latent stationary process is kept after modulation. In this chapter we first recall the formal definition of a modulated process in Section 3.1 and review the class of asymptotically stationary modulated

processes (Parzen, 1963) in Section 3.2. We then present our extended class of *modulated processes with a significant correlation contribution* in Section 3.3. In Section 3.4 we consider the application of modulated processes to the problem of missing data in the analysis of stationary processes, and show the use of our class of modulated processes with a significant correlation contribution to study larger classes of such problems than has been done in the past literature of modulated processes. We then propose an extension of modulation to a class of bivariate processes in Section 3.5, where the definition of significant correlation contribution is easily extended from scalar processes to bivariate processes. In this same section we introduce a bivariate non-stationary autoregressive model, show that it belongs to our class of bivariate modulated processes, and give sufficient conditions ensuring significant correlation contribution. In Chapter 8 we shall use this non-stationary model to study the velocity time series of drifting instruments which follow the ocean surface currents. In this chapter, Sections 3.1 and 3.2 constitute reviewed material. Section 3.3 is novel. Section 3.4 constitutes mostly revised material, except for the comments about significant correlation contribution for the scheme of (k, l)-periodically missing observations, and Example 3 of that section. Sections 3.5 and 3.6 are novel.

3.1 Modulated processes

Modulation is a natural and simple method of producing a non-stationary process (Parzen, 1963). A univariate Gaussian modulated process is defined as follows.

Definition 8 (Gaussian modulated process). *Let $\{X_t : t \in \mathbb{N}\}$ be a Gaussian, real-valued, zero-mean stationary process. Let $\{g_t : t \in \mathbb{N}\}$ be a given bounded real-valued deterministic sequence. Then a modulated process is defined pointwise as one taking the form*

$$\widetilde{X}_t = g_t X_t \qquad (3.1)$$

at all time points $t \in \mathbb{N}$.

Herein we treat $\{g_t\}$ as a known deterministic signal, unless stated otherwise. In our setting the process $\{X_t\}$, which is referred to as the *latent* process, is modelled

through a finite set of parameters $\theta \in \Theta \subset \mathbb{R}^d$, where d is a positive integer and Θ is the parameter space. Usually our object of interest is θ, the particular values of parameters that generated the observed realization. For example, if the latent process is an autoregressive process of order $p \geq 1$, we then have $d = p + 1$ if the mean is known (p autoregressive parameters and the variance of the innovations). We denote the autocovariance function of the stationary zero-mean process $\{X_t\}$ by $c_X(\tau)$, or $c_X(\tau; \theta)$ when we want to make the dependence on θ explicit. Its Fourier transform, the spectral density, is denoted $S_X(\omega)$ or $S_X(\omega; \theta)$, respectively.

The modulation of the latent process X_t is a convenient mechanism to account for a wide range of non-stationary processes. In particular this mechanism has been widely used as a modelling tool for missing data problems in time series, where g_t is assigned values 0 or 1 when respectively missing or observing a data point in time (Jones, 1962).

To understand when we can recover the parameters controlling the latent process X_t from observing $\widetilde{X}_t$, we need to put further conditions in place on g_t. The time series $\widetilde{X}_t / g_t$ cannot always be formed as g_t may be zero for some time indices, corresponding to missing observations. Another reason is that we may not directly observe $\widetilde{X}_t$, but instead we may observe an aggregated process $\widetilde{X}_t + Z_t$, where Z_t is a stationary process (or more generally another modulated process, see Section 3.6) independent from $\widetilde{X}_t$, thus preventing us from recovering the stationary latent process X_t by division. The latter situation occurs for instance in our real-data application to oceanography presented in Section 8. We discuss briefly the addition of two modulated processes in Section 3.6.

We assume that $\widetilde{X}_t$ satisfies (3.1) for a Gaussian, real-valued, zero-mean stationary X_t with absolutely summable autocovariance sequence. Then $\mathrm{E}\{\widetilde{X}_t\} = g_t \mathrm{E}\{X_t\} = 0$ and the time-varying autocovariance sequence is given by,

$$c_{\widetilde{X}}(t, t + \tau; \theta) = \mathrm{E}\left\{\widetilde{X}_t \widetilde{X}_{t+\tau}\right\} = g_t g_{t+\tau} c_X(\tau; \theta).$$

Given a single length-N realization $\widetilde{X}_0, \cdots, \widetilde{X}_{N-1}$, we start by computing the usual

method of moments estimator according to,

$$\hat{c}_{\widetilde{X}}^{(N)}(\tau) = \frac{1}{N} \sum_{t=0}^{N-\tau-1} \widetilde{X}_t \widetilde{X}_{t+\tau}, \tag{3.2}$$

for $\tau = 0, 1, ..., N-1$, such that τ is within the range of time offsets that is permissible given the length-N sample. Equation (3.2) is the biased sample autocovariance sequence of the modulated time series, which we define even though the process is non-stationary, as this object will become pivotal in our estimation procedure. The expectation of this object, which we denote $\bar{c}_{\widetilde{X}}^{(N)}(\tau; \theta)$ or simply $\bar{c}_{\widetilde{X}}^{(N)}(\tau)$, takes the following form,

$$
\begin{aligned}
\bar{c}_{\widetilde{X}}^{(N)}(\tau) &= \mathrm{E}\{\hat{c}_{\widetilde{X}}^{(N)}(\tau)\} = \mathrm{E}\left\{\frac{1}{N} \sum_{t=0}^{N-\tau-1} \widetilde{X}_t \widetilde{X}_{t+\tau}\right\} \\
&= c_X(\tau)\frac{1}{N} \sum_{t=0}^{N-\tau-1} g_t g_{t+\tau} = c_g^{(N)}(\tau) \cdot c_X(\tau),
\end{aligned}
\tag{3.3}
$$

where we have introduced the (deterministic) sample autocovariance of the modulating sequence,

$$c_g^{(N)}(\tau) = \frac{1}{N} \sum_{t=0}^{N-\tau-1} g_t g_{t+\tau}. \tag{3.4}$$

In the specific case where the modulating sequence $\{g_t\}$ is constant and equal to unity everywhere, which would correspond to observing the latent stationary process directly, we recover the expectation of the biased sample autocovariance for stationary time series, $(1 - \tau/N)c_X(\tau)$, for $\tau = 0, \cdots, N-1$. More generally, a standard assumption is to say that the modulated process $\widetilde{X}_t$ is an asymptotically stationary process (Parzen, 1961, 1963), which arises if for all lags τ, the quantity $c_g^{(N)}(\tau)$ in (3.4) converges as N tends to infinity. This is the subject of the following section.

3.2 Asymptotically stationary modulated processes

The class of asymptotically stationary modulated processes is due to Parzen (1961) and corresponds to modulated processes which satisfy the definition of asymptotic

stationarity as given in Section 2.4.1. More precisely, with the notations of the former section, $\{\widetilde{X}_t\}$ is an asymptotically stationary modulated process if there exists a fixed function $\{\gamma_g(\tau) : \tau \in \mathbb{N}\}$ such that for all $\tau \in \mathbb{N}$,

$$\lim_{N \to \infty} c_g^{(N)}(\tau) = \gamma_g(\tau), \tag{3.5}$$

where $c_g^{(N)}(\tau)$ is defined by (3.4). Indeed we then note that $\bar{c}_{\widetilde{X}}^{(N)}(\tau) \to \gamma_g(\tau)c_X(\tau)$ as $N \longrightarrow \infty$, so we could estimate $c_X(\tau)$ by defining,

$$\hat{c}_X^{(N)}(\tau) = \frac{\hat{c}_{\widetilde{X}}^{(N)}(\tau)}{\gamma_g(\tau)}, \tag{3.6}$$

assuming $\gamma_g(\tau) \neq 0$ for all $\tau \in \mathbb{N}$, and is known. It is shown in Parzen (1963) that (3.6) is a consistent estimator of the autocovariance sequence $c_X(\tau)$ of the latent stationary process under mild conditions. Further results are found in Dunsmuir and Robinson (1981a).

An example of a non-stationary but asymptotically stationary process is given by Parzen (1963), where a stationary process is observed according to a (k, l)-periodically missing data pattern, such that the first k values are observed, the next l values are missed, the next k values are observed, and so on, where k and l are two strictly positive integers.

The key feature in Definition 5 is that we average the time-varying autocovariance sequence $c_{\widetilde{X}}(t, t+\tau) = \mathrm{E}\left\{\widetilde{X}_t \widetilde{X}_{t+\tau}\right\}$ across a time period N to produce an average autocovariance across the time period, written as $\bar{c}_{\widetilde{X}}^{(N)}(\tau)$. If this converges (in N) to a function of τ, then by observing the modulated process over a suitably long time interval we can recover the second-order properties of the stationary latent process.

3.3 Significant correlation contribution

We now wish to explore a more general assumption than that of asymptotic stationarity for modulated processes. Specifically, we seek a larger class of models where consistent inference is still achievable. This will be smaller than the full class of

models for g_t, as using a trivial example, if $g_t \equiv 0$ always then we would not be able to infer properties of the generating mechanism of X_t. For consistent inference we propose the following class of modulated processes, which forms the key contribution of Guillaumin et al. (2017).

Definition 9 (Modulated process with a significant correlation contribution). *Assume that $\widetilde{X}_t$ is specified by* (3.1). *We say that $\widetilde{X}_t$ is a* modulated process with a significant correlation contribution *if there exists a finite subset of non-negative lags $\Gamma \subset \mathbb{N}$ such that,*

1. *The mapping $\theta \mapsto \{c_X(\tau; \theta) : \tau \in \Gamma\}$ is one-to-one (injective).*

2. *For all lags $\tau \in \Gamma$,*

$$\liminf_{N \to \infty} |c_g^{(N)}(\tau)| > 0, \tag{3.7}$$

   where $\liminf\limits_{N \to \infty}$ is the limit inferior.

Because of the symmetry of autocovariance sequences we do not need to consider $\tau < 0$ in this definition. Point 1 of Definition 9 means that for any two distinct parameter vectors $\theta, \theta' \in \Theta$, there exists at least one lag τ in the finite set Γ such that $c_X(\tau; \theta) \neq c_X(\tau; \theta')$. It is therefore an assumption about the latent process model. The quantity $|c_g^{(N)}(\tau)|$ is bounded above since the modulating sequence is assumed to be bounded above. Therefore the limit inferior in (3.7) is always finite. We observe that, for $\tau \in \Gamma$, (3.7) is equivalent to,

$$\exists \alpha_\tau > 0, \ \exists N_\tau \in \mathbb{N}, \ \forall N \in \mathbb{N}, \ N \geq N_\tau \ \Rightarrow \ |c_g^{(N)}(\tau)| \geq \alpha_\tau, \tag{3.8}$$

which we interpret as the fact that the sequence $\{|c_g^{(N)}(\tau) : N \in \mathbb{N}\}|$ is bounded below by a positive constant for N large enough. For further understanding of Point 1 in Definition 9 we provide the following two simple examples.

1. Let the latent process $\{X_t\}$ be an autoregressive process of order p, denoted AR(p), with known mean zero and unknown innovation variance, and with parameter set $\Theta \subset \mathbb{R}^{p+1}$. If the parameter set Θ is chosen appropriately, i.e.

such that the roots of the characteristic equation all lie outside the unit circle so that the process is causal, the Yule-Walker equations (Brockwell and Davis, 1991) show that $\theta \mapsto \{c_X(\tau; \theta) : \tau \in \Gamma\}$, where $\Gamma = \{0, \cdots, p\}$, is a one-to-one mapping. Similarly if $\{X_t\}$ is a moving average process of order q, denoted MA(q), with known mean an unknown innovation variance and if the parameter set Θ is chosen appropriately (Dzhaparidze and Yaglom, 1983), then the mapping $\theta \mapsto \{c_X(\tau; \theta) : \tau \in \Gamma\}$, where $\Gamma = \{0, \cdots, q\}$, is one-to-one.

2. Let the latent process $\{X_t\}$ be the MA(2) process defined by,

$$X_t = \sigma \left(\varepsilon_t + \theta_2 \varepsilon_{t-2} \right),$$

where the innovations ε_t are i.i.d. and have a standard normal distribution and $\sigma > 0$. The parameters of the model are (θ_2, σ), and the parameter set $\Theta = \mathbb{R} \times \mathbb{R} \backslash \{0\}$ ensures that the mapping $\theta \mapsto \{c_X(\tau; \theta) : \tau \in \Gamma\}$, where $\Gamma = \{0, 2\}$, is one-to-one. Note that observing lag-1 is not required here as we have assumed $\theta_1 = 0$ in the model.

The definition of a *significant correlation contribution* constrains how much "energy" adds up for any fixed lag $\tau \in \Gamma$. We see directly from (3.3) that if we assume a significant correlation contribution, the expectation of the estimated autocovariance of $\widetilde{X}_t$ does not vanish with the length of the observation N, at least for lags in Γ. This allows for consistent estimation of the parameter θ as we will prove in Section 6. As a trivial counterexample, assume for instance that $c_g^{(N)}(\tau)$ goes to zero when N goes to infinity. Then $\hat{c}_{\widetilde{X}}^{(N)}(\tau)$ in (3.2) goes to zero as well, independently of the parameter vector θ, resulting either in infeasible estimation or requiring a change of estimation approach.

Our class of modulated processes extends the concept of modulated processes with a significant correlation contribution. Specifically, for the class of asymptotically stationary modulated processes it is required that $c_g^{(N)}(\tau)$ converges to the non-zero quantity $R_g(\tau)$ as N grows, which is a stronger requirement than (3.7)

where we only require an asymptotic positive lower bound rather than convergence.

In this thesis we will also consider two situations that will require an extended definition of significant correlation contribution. In Section 3.6 we will consider a linear combination of modulated processes, and in Section 5.4 we will consider situations where the modulation sequence is not observed. These two additional contributions will require the following definition.

Definition 10 (Asymptotic injectivity). *For all $N \in \mathbb{N}$, let $h^{(N)} : \Theta \mapsto \mathbb{R}$ be a sequence of functions, from $\Theta \subset \mathbb{R}^d$ to $\mathbb{R}^m$, where d and m are positive integers. The sequence $h^{(N)}(\cdot)$ is said to be* asymptotically injective *if, for all distinct $\theta, \theta' \in \Theta$,*

$$\liminf_{N \to \infty} \|h^{(N)}(\theta) - h^{(N)}(\theta')\| > 0,$$

where $\| \cdot \|$ represents any norm on $\mathbb{R}^m$.

3.4 Missing observations

A compelling use of modulated processes is to account for missing observations in stationary time series. Let $\{X_t : t \in \mathbb{N}\}$ be a stationary Gaussian process. For each time point $t \in \mathbb{N}$, we set (Parzen, 1963),

$$g_t = \begin{cases} 0 & \text{if } X_t \text{ is missing} \\ 1 & \text{if } X_t \text{ is observed} \end{cases}. \tag{3.9}$$

The process $\widetilde{X}_t = g_t X_t$ is formed at all time points $t \in \mathbb{N}$, forming a Gaussian modulated process in the sense of Definition 8.

An example where the missing observation pattern is deterministic and leads to an asymptotically stationary modulated process is the case of (k, l)-periodically missing data treated by Jones (1962) and Parzen (1963). This corresponds to observing the k first values, missing the l next values, observing the k next values, and so on. Note that Parzen (1963) requires $k > l$ for non-parametric estimation of the spectral density of X_t based on (3.6). Our model of modulated processes with significant correlation contribution allows for $k \leq l$, as long as we observe the lags

in the set Γ used in Definition 9. A generalization of the (k,l)-periodically missing data scheme was introduced by Clinger and Ness (1976) with an application to oceanography.

Missing observations can also occur according to a random mechanism. This can be modelled by a random modulation sequence taking values zero and one (Scheinok, 1965, Bloomfield, 1970), when the random mechanism according to which missing points occur is independent from the observed process, which we shall assume. Conditioning on the observed modulation function, we then return to the deterministic modulating sequence described in this thesis. Most works, to our knowledge, have assumed some sort of stationarity for the random modulation sequence, i.e. that the sample autocovariance of the modulation sequence converges almost surely to a non-zero value at all lags (Dunsmuir and Robinson, 1981b,c). Some authors do not require such an assumption but have treated only specific models, usually autoregressive models (Jones, 1980, Broersen et al., 2004). The definition of a modulated process with a significant correlation contribution in such a situation needs to be understood in a probabilistic fashion, i.e. we require that Property 2 of Definition 9 be satisfied with probability one. Indeed, if one sees the general random experiment as a two-step experiment, where first the random modulating sequence $\{g_t\}$ is generated and observed and then a stationary process $\{X_t\}$ is modulated by this modulating sequence to produce $\{\widetilde{X}_t\}$, then with probability one the modulating sequence $\{g_t\}$ in the first step makes $\{\widetilde{X}_t\}$ a modulated process with a significant correlation contribution. Such a situation may be described by saying that $\{\widetilde{X}_t\}$ is a modulated process with an almost surely significant correlation contribution. We shall now give a few examples of cases satisfying the stated conditions.

1. Let X_t be an AR(p) Gaussian process with mean zero. If we set $\Gamma = \{0, \cdots, p\}$, and if the missing data occurs deterministically according to a (k,l)-periodic pattern, $k \geq p$ is a sufficient condition for the resulting modulated process to have a significant correlation contribution. This is because we are able to observe an infinite number of time the lags in Γ. We do not

require any additional condition on l.

2. Let X_t be an AR(p) process, and consider the missing data scheme treated by Scheinok (1965), where the random mechanism is a sequence of Bernoulli i.i.d. trials with identical probability of success (to be understood as *observation* here) $0 < p \leq 1$. According to the strong central limit theorem, for all lag $\tau \in \mathbb{N}$, $c_g^{(N)}(\tau)$ converges a.s. to $p^2 > 0$ and therefore $\liminf\limits_{N \to \infty} \left| c_g^{(N)}(\tau) \right| > 0$ a.s. Therefore the observed process is a modulated process with an almost surely significant correlation contribution.

3. Consider the random mechanism where the sequence $\{g_t\}$ is generated according to

$$g_t \sim \mathscr{B}(p_t), \tag{3.10}$$

where $\mathscr{B}(p)$ represents the Bernoulli distribution with parameter p, and where we set

$$p_t = \mathscr{P} + A_p \cos(\omega_p t), \tag{3.11}$$

with $0 < \mathscr{P} < 1$, $0 \leq A_p < \min(\mathscr{P}, 1 - \mathscr{P})$ (which ensures $0 < \mathscr{P} - A_p \leq p_t \leq 1, \forall t \in \mathbb{N}$), and $\omega_p \in [-\pi, \pi]$. The Bernoulli parameters p_t as given by (3.11) will oscillate around their mean value $\mathscr{P}$. This leads to $\liminf\limits_{N \to \infty} \left| c_g^{(N)}(\tau) \right| > 0$ a.s., using the fact that p_t is bounded below by $\mathscr{P} - A_p > 0$.

In section 7.1.2 we will provide a simulation study based on the third example. This is novel in comparison of previously studied missing observation schemes as we do not make an assumption of stationarity for the process g_t, as is for instance required by Dunsmuir and Robinson (1981a).

3.5 Extension to bivariate modulated processes

In this section we review complex-valued stochastic processes, and in particular the class of proper complex-valued processes. We introduce a class of non-stationary bivariate modulated processes in Section 3.5.2, and then consider a bivariate AR(1) stochastic process model which can be modelled via this class in Section 3.5.3.

3.5.1 Complex-valued time series

Up to this point we have only treated the case of real-valued stochastic processes, that is to say processes taking values in the set of real numbers. In this section, we review the basics of complex-valued time series analysis (Schreier and Scharf, 2010). We will use complex-valued time series analysis methods to analyse bivariate oceanographic time series in Section 8.4.

Firstly, we consider two complex-valued random variables Z_1 and Z_2 from some probability space $(\Omega, \mathscr{A}, \mathscr{P})$ to $\mathbb{C}$, and denote $Z_j = X_j + iY_j$, $j = 1, 2$ where X_j and Y_j are real-valued random variables, and i denotes the element of $\mathbb{C}$ that satisfies $i^2 = -1$. The expectations of the two random variables Z_1 and Z_2 are simply defined by,

$$\mathrm{E}\{Z_j\} = \mathrm{E}\{X_j\} + i\mathrm{E}\{Y_j\}, \quad j = 1, 2.$$

The covariance structure of those two complex-valued random variables possesses four degrees of freedom, and is equivalently represented by the two complex-valued quantities,

$$c_{Z_1, Z_2} = \mathrm{E}\{Z_1^* Z_2\}, \quad r_{Z_1, Z_2} = \mathrm{E}\{Z_1 Z_2\},$$

assuming zero-mean for both variables and recalling that * denotes the conjugation operator. The quantity c_{Z_1, Z_2} is the *covariance* of the two complex-valued random variables Z_1 and Z_2. The quantity r_{Z_1, Z_2} is called the *relation* or *complementary covariance*. To better understand those two quantities, observe that c_{Z_1, Z_2} is rotation-invariant. This means that the quantity c_{Z_1, Z_2} is unchanged if Z_1 and Z_2 are rotated by the same angle, i.e. $c_{e^{i\phi}Z_1, e^{i\phi}Z_2} = c_{Z_1, Z_2}, \forall \phi \in [-\pi, \pi)$. On the contrary, r_{Z_1, Z_2} is not rotation-invariant.

As a simple example let $Z = \rho e^{i\phi}$, where ρ is a non-negative real-valued random variable, and ϕ is a real-valued random variable taking values in $[-\pi, \pi)$. Assume that ϕ has a uniform distribution over $[-\pi, \pi)$, and is independent from ρ. It is not difficult to show that in that case $r_{Z,Z} = 0$ and $c_{Z,Z} = \mathrm{E}\{\rho^2\}$. A complex-valued random variable Z with relation $r_{Z,Z}$ equal to zero is called *proper*. We are now in position to extend the definition of stationarity to complex-valued processes, which

is the subject of the following definition.

Definition 11 (Stationary complex-valued stochastic process). *Let $\{Z_t : t \in \mathbb{Z}\}$ be a complex-valued stochastic process, and denote X_t and Y_t its real and imaginary parts respectively. The process $\{Z_t\}$ is called stationary if there exists two sequences $c_Z(\tau)$ and $r_Z(\tau)$ such that, for all $t, \tau \in \mathbb{Z}$,*

$$c_{Z_t, Z_{t+\tau}} = c_Z(\tau), \quad r_{Z_t, Z_{t+\tau}} = r_Z(\tau).$$

The sequences $c_Z(\tau)$ and $r_Z(\tau)$ appearing in Definition 11 are respectively called the *autocovariance* and *relation* (or *complementary autocovariance*) *sequences* of the process. If the relation sequence is zero everywhere, the process is said to be *proper* (Schreier and Scharf, 2010). Proper processes are common in the signal processing literature and in applications, and the complex-valued framework presented above offers a compact representation of such processes as compared to bivariate representations. Finally, note that complex-valued processes, unlike real-valued, no longer have a spectrum that needs to satisfy Hermitian symmetry, and if the series represents motion in the plane, the positive and negative frequencies represent anti-clockwise and clockwise rotations respectively.

3.5.2 A class of bivariate modulated processes

It is common in practical applications to observe more than one time series at any time, and to analyse a set together. Often the series in the set are related via phase-shifts and other small temporal inhomogeneities, see e.g. Allen and Robertson (1996), Rünstler (2004), Allefeld et al. (2009), Lilly and Olhede (2012). Bivariate non-stationary processes can be challenging to model, as they may not be representable in the same non-stationary oscillatory family (Tong, 1973, 1974). To explore the nature of multivariate modulation, we shall investigate the representation of bivariate processes. For ease of exposition we shall represent such series using complex-valued time series, see Walker (1993). We shall continue to assume that the latent process, now denoted Z_t for complex-valued processes, is Gaussian and zero-mean, leaving only the second-order structure to be modelled. Follow-

ing the classical modelling framework (Miller, 1969) for complex-valued processes we shall assume that the relation sequence takes the value zero for all lags. The complex-valued process is therefore proper, which is equivalent to the isotropy of the corresponding bivariate real-valued process. The assumption of propriety has the consequence of directly extending equation (3.1) to the complex-valued case from the real-valued case. Specifically, let Z_t be a complex-valued Gaussian proper zero-mean process, a complex-valued modulated process is defined as one taking the form,

$$\widetilde{Z}_t = g_t Z_t, \tag{3.12}$$

at all times $t \in \mathbb{N}$, where $g_t = \rho_t e^{i\phi_t}$ is a bounded modulation sequence. We note that for complex-valued time series the modulation sequence is complex-valued. With this definition, the modulation series g_t accomplishes a time-dependent rescaling or expansion/dilation, from ρ_t, together with a time-dependent rotation, from $e^{i\phi_t}$.

The autocovariance of the complex-valued modulated process $\widetilde{Z}_t$ at times t_1 and t_2 is given by the conveniently simple form,

$$c_{\widetilde{Z}}(t_1, t_2; \theta) = \mathrm{E}\left\{\widetilde{Z}_{t_1}^* \widetilde{Z}_{t_2}; \theta\right\} = g_{t_1}^* g_{t_2} c_Z(t_2 - t_1; \theta) = \rho_{t_1} \rho_{t_2} e^{i(\phi_{t_2} - \phi_{t_1})} c_Z(t_2 - t_1; \theta),$$

and $c_{\widetilde{Z}}(t_1, t_2; \theta)$ fully characterizes the process. Note that this quantity is not only a function of the lag $t_2 - t_1$ as the process is no longer stationary. Similarly to the univariate case cf. (3.4), let N be any positive integer, we define for $\tau = 0, \cdots, N-1$,

$$c_g^{(N)}(\tau) = \frac{1}{N} \sum_{t=0}^{N-\tau-1} g_t^* g_{t+\tau}. \tag{3.13}$$

Note that when g_t is real-valued (3.13) and (3.4) are the same. We also extend the notion of a significant correlation contribution for complex-valued modulated processes, which naturally mimics Definition 9.

A univariate real-valued modulated process is stationary if and only if the modulating sequence is a constant. A necessary and sufficient condition on the modulating sequence for the complex-valued modulated process (3.12) to be stationary is

more complicated to obtain, and is determined in the following proposition.

Proposition 3 (Stationary bivariate modulated processes). *Let $\widetilde{Z}_t$ be the complex-valued modulated process defined in (3.12). First, assume the latent process $\{Z_t\}$ is a white noise process. Then the modulated process $\{\widetilde{Z}_t\}$ is stationary if and only if the modulating sequence $g_t = \rho_t e^{i\phi_t}$ is of constant modulus, i.e. $\rho_t = a \geq 0$. In such case the modulated process is a white noise process with variance $a^2 \mathrm{E}\{|Z_0|^2\}$.*

More generally, assume the stationary latent process $\{Z_t\}$ is not a white noise process, and let $\mu = \gcd\{\tau \neq 0 \in \mathbb{N} : |c_Z(\tau;\theta)| > 0\}$ where gcd denotes the greatest common divisor. Then the modulated process is stationary if and only if $\{g_t\}$ is zero everywhere or if there exists two constants $a > 0$ and $\gamma \in [-\pi, \pi)$ such that for all $t \in \mathbb{N}$, letting $r = t \bmod \mu$ be the remainder of t divided by μ,

$$
\begin{aligned}
\rho_t &= a \\
\phi_t &= \phi_r + \gamma \left\lfloor \frac{t}{\mu} \right\rfloor \quad \bmod 2\pi,
\end{aligned}
$$

where $\left\lfloor \frac{t}{\mu} \right\rfloor$ denotes the floor of $\frac{t}{\mu}$ and $\bmod 2\pi$ indicates that the equality is true up to an additive multiple of 2π. In this case the spectral density of the modulated process $\{\widetilde{Z}_t\}$ is

$$
S_{\widetilde{Z}}(\omega) = a^2 S_Z \left(\omega - \frac{\gamma}{\mu} \right).
$$

Proof. We distinguish the case where $\{Z_t\}$ is a white noise process (the covariance is zero everywhere except for lag zero) from the case where $\{Z_t\}$ is not a white noise process.

1. Assume $\{Z_t\}$ is a white noise process.

   $\rightarrow$ Assume $\{\widetilde{Z}_t\}$ is a stationary process. Being stationary, it has a constant variance and therefore the modulating sequence must have a constant modulus.

   $\leftarrow$ Conversely, if $\{g_t\}$ has a constant modulus $\{\widetilde{Z}_t\}$ is stationary and is a white noise process.

2. Assume $\{Z_t\}$ is not a white noise process. The set $\{\tau \in \mathbb{N}^* : |c_Z(\tau; \theta)| > 0\}$ is therefore not empty (where $\mathbb{N}^*$ denotes the set of positive integers), so $\mu = \gcd\{\tau \in \mathbb{N}^* : |c_Z(\tau; \theta)| > 0\}$ is well defined.

$\rightarrow$ Assume $\{\widetilde{Z}_t\}$ is stationary. Then it must have a constant variance, so there must exists a real number $a \geq 0$ such that $\rho_t = a \ \forall t \in \mathbb{N}$. Leaving aside the trivial case in which $\{g_t\}$ is zero everywhere, let t_1, t_2 be two natural integers. We have,

$$c_{\widetilde{Z}}(t_1, t_2; \theta) = g_{t_1}^* g_{t_2} c_Z(t_2 - t_1; \theta) = a^2 e^{i(\phi_{t_2} - \phi_{t_1})} c_Z(t_2 - t_1; \theta).$$

If $c_Z(t_2 - t_1; \theta) \neq 0$ then

$$e^{i(\phi_{t_2} - \phi_{t_1})} = \frac{c_{\widetilde{Z}}(s, t; \theta)}{a^2 c_Z(t_2 - t_1; \theta)},$$

which leads to

$$\phi_{t_2} - \phi_{t_1} = \arg\left\{\frac{c_{\widetilde{Z}}(t_1, t_2; \theta)}{a^2 c_Z(t_2 - t_1; \theta)}\right\} \mod 2\pi,$$

where the equality is true up to a multiple of 2π, which we indicate by the use of the notation $\mod 2\pi$. Since $\{\widetilde{Z}_t\}$ is assumed stationary, there exists a function ζ, defined on $\{\tau \in \mathbb{N} : c_Z(\tau; \theta) \neq 0\}$, such that

$$\arg\left\{\frac{c_{\widetilde{Z}}(t_1, t_2; \theta)}{a^2 c_Z(t_2 - t_1; \theta)}\right\} = \zeta(t_2 - t_1) \mod 2\pi, \ \forall t_1, t_2 \in \mathbb{N}.$$

Therefore

$$\phi_{t_2} - \phi_{t_1} = \zeta(t_2 - t_1) \mod 2\pi.$$

Now let $t \in \mathbb{N}$ be any natural integer and write $t = \mu q + r$ where $0 \leq r < \mu$ and $q \in \mathbb{N}$ are uniquely defined as the remainder and quotient of the Euclidean division of t by μ.

$$\phi_t = \sum_{k=0}^{q-1} (\phi_{r+(k+1)\mu} - \phi_{r+k\mu}) + \phi_r$$

$$= \sum_{k=0}^{q-1} \zeta(\mu) + \phi_r \quad \mathrm{mod}\ 2\pi$$

$$= q\zeta(\mu) + \phi_r \quad \mathrm{mod}\ 2\pi.$$

Letting $\gamma = \zeta(\mu)$ we obtain,

$$\phi_t = \gamma \left\lfloor \frac{t}{\mu} \right\rfloor + \phi_{t \bmod \mu} \quad \mathrm{mod}\ 2\pi.$$

$\leftarrow$ Conversely assume there exists two constants $\gamma \in \mathbb{R}$ and $a \geq 0$ such that for all $t \in \mathbb{N}$,

$$\rho_t = a,$$

$$\phi_t = \phi_{t \bmod \mu} + \gamma \left\lfloor \frac{t}{\mu} \right\rfloor \quad \mathrm{mod}\ 2\pi.$$

Let t, τ be two natural integers. We have:

$$c_{\widetilde{Z}}(t, t+\tau; \theta) = g_t^* g_{t+\tau} c_Z(\tau; \theta) = a^2 e^{i(\phi_{t+\tau} - \phi_t)} c_Z(\tau; \theta).$$

If $c_Z(\tau; \theta) = 0$ then $c_{\widetilde{Z}} = (t, t+\tau; \theta) = 0$ which does not depend on t but only on τ. Otherwise, τ is a multiple of μ by definition of μ. Therefore there exists an integer q such that $\tau = q\mu$, and $(t+\tau) \bmod \mu = t \bmod \mu$. Finally,

$$\phi_{t+\tau} - \phi_t = \phi_{(t+\tau) \bmod \mu} + \gamma \left\lfloor \frac{t+\tau}{\mu} \right\rfloor - \phi_{t \bmod \mu} - \gamma \left\lfloor \frac{t}{\mu} \right\rfloor \quad \mathrm{mod}\ 2\pi$$

$$= \gamma \left\lfloor \frac{t}{\mu} + q \right\rfloor - \gamma \left\lfloor \frac{t}{\mu} \right\rfloor \quad \mathrm{mod}\ 2\pi$$

$$= \gamma \left(\left\lfloor \frac{t}{\mu} \right\rfloor + q - \left\lfloor \frac{t}{\mu} \right\rfloor \right) \quad \mathrm{mod}\ 2\pi$$

$$= \gamma q \quad \mathrm{mod}\ 2\pi,$$

where we have used the fact that $\left\lfloor \frac{t}{\mu} + q \right\rfloor = \left\lfloor \frac{t}{\mu} \right\rfloor + q$ as q is an integer. Again the obtained quantity does not depend on t. Therefore $c_{\widetilde{Z}}(t, t+$

$\tau; \theta)$ does not depend on t but only on the lag τ. This proves that $\{\widetilde{Z}_t\}$ is stationary with autocovariance sequence $c_{\widetilde{Z}}(\tau; \theta) = a^2 e^{i\gamma\frac{\tau}{\mu}} c_Z(\tau)$. As for the spectral density of the resulting stationary modulated process $\{\widetilde{Z}_t\}$ we have,

$$
\begin{aligned}
S_{\widetilde{Z}}(\omega; \theta) &= \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} c_{\widetilde{Z}}(\tau; \theta) e^{-i\omega\tau} \\
&= \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} a^2 c_Z(\tau; \theta) e^{-i(\omega\tau - \gamma\frac{\tau}{\mu})} \\
&= \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} a^2 c_Z(\tau; \theta) e^{-i(\omega - \frac{\gamma}{\mu})\tau} \\
&= \frac{a^2}{2\pi} \sum_{\tau=-\infty}^{\infty} c_Z(\tau; \theta) e^{-i(\omega - \frac{\gamma}{\mu})\tau} = a^2 S_Z\left(\omega - \frac{\gamma}{\mu}\right).
\end{aligned}
$$

This concludes the proof. Note that for a real-valued process this shift would be impossible as the spectral density has to retain symmetry. As both Z_t and $\widetilde{Z}_t$ are complex-valued, this is not a concern. $\qquad\square$

The value of μ in Proposition 3 depends on the location of zeros in the covariance sequence of the latent process. In particular, if $|c_Z(1; \theta)| > 0$ then $\mu = 1$ and $\widetilde{Z}_t$ is stationary only if there exists a constant $\gamma \in \mathbb{R}$ such that for all $t \in \mathbb{N}$, $\phi_t = \phi_0 + \gamma t$ mod 2π. If $|c_Z(2; \theta)| > 0$ but $|c_Z(\tau; \theta)| = 0$ for all $\tau \in \mathbb{N}, \tau \neq 0, 2$, then $\mu = 2$ (this can occur with a second-order moving average process for instance). In that case the modulated process $\widetilde{Z}_t$ is stationary if and only if there exists a constant $\gamma \in [-\pi, \pi)$ such that for all $t \in \mathbb{N}$, $\phi_t = \phi_0 + \gamma\frac{t}{2}$ mod 2π if t is even, or $\phi_t = \phi_1 + \gamma\frac{t-1}{2}$ mod 2π if t is odd.

3.5.3 A time-varying bivariate autoregressive process

We now introduce the specific non-stationary bivariate autoregressive model that will be used in our real-world data application. We consider the discrete-time complex-valued process $\{\widetilde{Z}_t : t \in \mathbb{N}\}$, defined by

$$
\begin{aligned}
\widetilde{Z}_t &= re^{i\beta_t}\widetilde{Z}_{t-1} + \varepsilon_t, \qquad t \geq 1, \ 0 \leq r < 1, \ \beta_t \in \mathbb{R}, \qquad (3.14) \\
\widetilde{Z}_0 &\sim \mathcal{N}_C\left(0, \frac{\sigma^2}{1 - r^2}\right), \qquad \sigma > 0,
\end{aligned}
$$

$$\varepsilon_t \quad \sim \quad \mathcal{N}_C\left(0, \sigma^2\right),$$

where $\mathcal{N}_C\left(0, \sigma^2\right)$ denotes the complex-valued normal distribution with mean 0 and variance σ^2, and with i.i.d. real and imaginary parts. Note that the real and imaginary parts of ε_t then each have variance $\sigma^2/2$. Here $0 \leq r < 1$ is commonly known as either the autoregressive or the damping parameter, ensuring the mean-reversion of the process. By mean-reversion we mean that, given any time $t \in \mathbb{N}$, we have $\lim_{\tau \to \infty} \mathrm{E}\left\{\widetilde{Z}_{t+\tau} | \widetilde{Z}_t\right\} = 0$, i.e. irrespective of the size of the perturbation ε_t at time t, the process is expected to return to its mean. This is seen from the following inductive relationship,

$$\tilde{Z}_{t+\tau} = r^\tau e^{i\Sigma_{j=1}^\tau \beta_{t+j}}\tilde{Z}_t + \sum_{j=1}^\tau r^{\tau-j} e^{i\Sigma_{k=j+1}^\tau \beta_k}\varepsilon_{t+j}, \quad \tau \geq 0,$$

which leads to

$$\mathrm{E}\left\{\tilde{Z}_{t+\tau} | \tilde{Z}_t\right\} = r^\tau e^{i\Sigma_{j=1}^\tau \beta_{t+j}}\tilde{Z}_t,$$

which goes to zero exponentially as τ goes to infinity, since $0 \leq r < 1$. A damping parameter r close to 1 will lead to a slowly-decaying autocorrelation sequence. A value of r close to 0 will lead to a process with very short memory, with the limiting behaviour of a white noise process as $r \to 0$, and that of a random walk as $r \to 1$. The parameter β_t is a known, dimensionless time-varying frequency, which we shall take within the interval $[-\pi, \pi)$ without loss of generality.

The process (3.14) is a non-stationary version of the complex-valued first-order autoregressive process (Sykulski et al., 2016b) introduced by Le Breton (1988), and also a discrete-time analogue of the complex-valued Ornstein-Ulhenbeck (OU) process (Arató et al., 1962) with time-varying oscillation frequency. We now prove in Proposition 4 that the model defined in (3.14) belongs to our class of bivariate modulated processes.

Proposition 4 (Modulated process representation). *Let $\{\widetilde{Z}_t\}$ be the process defined in (3.14). There exists a unit-magnitude complex-valued modulating sequence g_t, and a stationary complex-valued proper process $\{Z_t\}$ such that $\{\widetilde{Z}_t\}$ is the modula-*

tion of $\{Z_t\}$ by the non-random sequence $\{g_t\}$. More explicitly, we have $\widetilde{Z}_t = g_t Z_t$, for all $t \in \mathbb{N}$, where,

$$
\begin{aligned}
g_t &= e^{i\sum_{u=1}^{t}\beta_u}, & (3.15) \\
Z_t &= rZ_{t-1} + \varepsilon'_t, \ \ t \ge 1,
\end{aligned}
$$

and $Z_0 = \widetilde{Z}_0$. The process ε'_t is a Gaussian white noise process with the same properties (zero-mean, variance σ^2 and independence of real and imaginary parts) as those of ε_t. Defined as such, the latent complex-valued process Z_t is stationary and proper.

Proof. Let us define the complex-valued stochastic process $\{Z_t\}$ according to

$$
Z_t = e^{-i\sum_{u=1}^{t}\beta_u}\widetilde{Z}_t, \ \ t = 0, 1, 2, \cdots .
$$

By applying the definition of the process $\{\widetilde{Z}_t\}$ one can determine the following relationship, for all $t \ge 1$,

$$
\begin{aligned}
Z_t &= e^{-i\sum_{u=1}^{t}\beta_u}\widetilde{Z}_t \\
&= e^{-i\beta_t}e^{-i\sum_{u=1}^{t-1}\beta_u}\widetilde{Z}_t \\
&= e^{-i\beta_t}e^{-i\sum_{u=1}^{t-1}\beta_u}\left(re^{i\beta_t}\widetilde{Z}_{t-1} + \varepsilon_t\right) \\
&= re^{-i\sum_{u=1}^{t-1}\beta_u}\widetilde{Z}_{t-1} + \varepsilon'_t,
\end{aligned}
$$

and finally $Z_t = rZ_{t-1} + \varepsilon'_t, \ \ t \ge 1$, where $\varepsilon'_t = e^{-i\sum_{u=1}^{t}\beta_u}\varepsilon_t, \forall t \in \mathbb{N}$ has the same distribution as ε_t, as we have assumed that the complex-valued white noise process ε_t has variance σ^2 and independent real and imaginary parts. Therefore the process Z_t is a first-order complex-valued autoregressive process with constant stationary parameters. It is stationary if and only if $\text{var}\{Z_0\} = \frac{\sigma^2}{1-r^2}$. Since $Z_0 = \widetilde{Z}_0$, it follows that $\text{var}\{Z_0\} = \text{var}\{\widetilde{Z}_0\}$. Since $\text{var}\{\widetilde{Z}_0\} = \frac{\sigma^2}{1-r^2}$, the process $\{Z_t\}$ is stationary. The fact that the process $\{Z_t\}$ is proper stems from the fact that the innovations $\{\varepsilon_t\}$ as well as the random variable $\widetilde{Z}_0$ are proper, as using the following relation, $Z_t = r^t Z_0 + \sum_{j=1}^{t} r^{t-j}\varepsilon'_j$, we obtain for all $t, \tau \in \mathbb{N}$, $\text{E}\{Z_t Z_{t+\tau}\} = 0$. This shows how the

proposed process is generated by the stated mechanism of modulation as claimed in the proposition. □

The stationary latent process Z_t defined in (3.15) is a stationary complex-valued first-order autoregressive process, and is Gaussian. Its autocovariance sequence is given by,

$$c_Z(\tau) = \frac{\sigma^2}{1-r^2} r^{|\tau|}, \quad \tau \in \mathbb{Z}.$$

It is easy to verify that the mapping $(r, \sigma) \mapsto (c_Z(0), c_Z(1))$ is a one-to-one mapping. In the following proposition, we stipulate a sufficient condition on the frequencies β_t so that the process defined in (3.14) satisfies our assumption of significant correlation contribution, when represented as a modulated process as defined in Proposition 4.

Proposition 5 (Significant correlation). *Let $\tilde{Z}_t$ be the process defined by (3.14). Assume that there exists $\Xi \in [-\pi, \pi)$ and $0 \leq \kappa < \frac{\pi}{2}$ such that for all $t \in \mathbb{N}$, $|\beta_t - \Xi| \leq \kappa$. Then $\tilde{Z}_t$ is a modulated process with significant correlation contribution.*

Proof. Let $\Gamma = \{0, 1\}$. We show that conditions 1 and 2 given in Definition 9 are verified.

1. The function $(r, \sigma) \mapsto \{c_Z[\tau; (r, \sigma)] : \tau \in \Gamma\}$ is one-to-one.

2. According to Proposition 4, there exists a stationary proper complex-valued process Z_t such that $\tilde{Z}_t = g_t Z_t$, where $g_t = e^{i\sum_{u=1}^{t} \beta_u}$, i.e $\tilde{Z}_t$ is a modulated process. The autocovariance sequence of the process Z_t is given by

$$c_Z(\tau) = \frac{\sigma^2}{1-r^2} r^{|\tau|}, \tau \in \mathbb{Z},$$

   and we observe that the function $(r, \sigma) \mapsto (c_X(0), c_X(1))$ is one-to-one. Let L be the largest positive (i.e. greater than or equal to 1) integer such that $\kappa \leq \frac{\pi}{2L}$. This is well defined as we have assumed $0 \leq \kappa < \frac{\pi}{2}$. Fix an integer lag value

$0 \leq \tau \leq L$. We have

$$
\begin{aligned}
\left| \frac{1}{N} \sum_{t=0}^{N-1-\tau} g_t^* g_{t+\tau} \right| &= \left| \frac{1}{N} \sum_{t=0}^{N-1-\tau} e^{i \sum_{u=t+1}^{t+\tau} \beta_u} \right| = \left| \frac{1}{N} \sum_{t=0}^{N-1-\tau} e^{i \sum_{u=t+1}^{t+\tau} (\Xi + \beta_u - \Xi)} \right| \\
&= \left| \frac{1}{N} e^{i\tau\Xi} \sum_{t=0}^{N-1-\tau} e^{i \sum_{u=t+1}^{t+\tau} (\beta_u - \Xi)} \right| \\
&\geq \frac{1}{N} \left| \Re \left\{ \sum_{t=0}^{N-1-\tau} e^{i \sum_{u=t+1}^{t+\tau} (\beta_u - \Xi)} \right\} \right| \\
&= \frac{1}{N} \left| \sum_{t=0}^{N-1-\tau} \cos \left\{ \sum_{u=t+1}^{t+\tau} (\beta_u - \Xi) \right\} \right|.
\end{aligned}
$$

Using the triangle inequality it follows $\left| \sum_{u=t+1}^{t+\tau} \beta_u - \Xi \right| \leq \sum_{u=t+1}^{t+\tau} |\beta_u - \Xi| \leq \tau\kappa$. With the fact that $\tau\kappa < \frac{\pi}{2}$ by assumption, and that the cosine function is decreasing on the interval $[0, \frac{\pi}{2}]$ we obtain

$$
\left| \frac{1}{N} \sum_{t=0}^{N-1-\tau} g_t^* g_{t+\tau} \right| \geq \frac{1}{N} \sum_{t=0}^{N-1-\tau} \cos(\tau\kappa) = \left(1 - \frac{\tau}{N}\right) \cos(\tau\kappa) \overset{N \to \infty}{\to} \cos(\tau\kappa) > 0,
$$

as $0 \leq \tau\beta < \tau\frac{\pi}{2L} \leq \frac{\pi}{2}$. The quantity $c_g^{(N)}(\tau)$ is bounded below by a non-zero value as N goes to infinity, so that $\liminf_{N \to \infty} \left| c_g^{(N)}(\tau) \right| > 0$. It is true in particular for $\tau \in \Gamma$, as $L \geq 1$.

This shows that the process $\widetilde{Z}_t$ is a modulated process with a significant correlation contribution. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Hence the complex-valued autoregressive process defined by (3.14) belongs to the class of processes with a significant correlation contribution.

3.6 Linear combinations of modulated processes

The pointwise addition of two modulated processes is not a modulated process itself except under some particular conditions, for instance if two modulated processes have the same latent stationary process or if two modulated processes have the same modulation sequence. We can however extend the definition of significant correlation to the sum of two modulated processes (and more generally to any finite

linear combination of modulated processes). This is the subject of the following definition.

Definition 12 (Significant correlation contribution for a linear combination of modulated processes). *Let $\{\widetilde{X}_{1,t}\}, \ldots, \{\widetilde{X}_{p,t}\}$ be a family of $p \geq 1$ modulated processes with $\{X_{1,t}\}, \ldots, \{X_{p,t}\}$ as their respective latent processes, and $\{g_{1,t}\}, \ldots, \{g_{p,t}\}$ as their respective known modulation sequences. Let $\lambda_1, \ldots, \lambda_p \in \mathbb{R}^p$ be a family of real numbers. Assume that the latent processes are independent. Define the process $Z_t = \sum_{i=1}^p \lambda_i \widetilde{X}_{i,t}$. The process $\{Z_t\}$ is not a modulated process in the general case. We say that $\{Z_t\}$ has a significant correlation contribution if it satisfies both of the following,*

1. *Each process $\{\widetilde{X}_{i,t}\}$ is a modulated process with a significant correlation contribution. We denote Γ_i the set of lags that determine the parameter vector $\theta_i \in \Theta_i$ in the definition of significant correlation.*

2. *There exists $\Gamma \supset \cup_{i=1}^p \Gamma_i$ such that the map (denoting $\tau_1, \ldots, \tau_m$ the elements of Γ),*

$$\begin{cases} \Theta_1 \times \ldots \times \Theta_p & \longrightarrow & \mathbb{R}^m \\ \theta_1, \ldots, \theta_p & \longrightarrow & \left[\sum_{i=1}^p \lambda_i^2 \overline{c}_{X_i}^{(N)}(\tau_1; \theta_i), \ldots, \sum_{i=1}^p \lambda_i^2 \overline{c}_{X_i}^{(N)}(\tau_m; \theta_i) \right] \end{cases}$$

is asymptotically injective, as defined in Definition 10.

The expected autocovariance sequence of the linear combination of modulated processes is shown to be, with the assumption of independence of the latent processes,

$$\overline{c}_Z(\tau) = \mathrm{E}\{\hat{c}_X(\tau)\} = \sum_{i=1}^p \lambda_i^2 c_{X_i}(\tau; \theta_i) c_{g_i}^{(N)}(\tau), \ \tau \in \mathbb{N}. \tag{3.16}$$

We give a simple example of a linear combination of modulated processes with a significant correlation contribution. Let $\{X_{1,t}\}$ be a white-noise process with variance σ_1^2 and let $\{X_{2,t}\}$ be an AR(1) process with autoregressive parameter r and innovation variance σ_2^2. Let $\{g_{1,t}\}$ and $\{g_{2,t}\}$ be two modulation sequences such that the processes $\widetilde{X}_{1,t} = g_{1,t} X_{1,t}$ and $\widetilde{X}_{2,t} = g_{2,t} X_{2,t}$ are modulated processes with a

significant correlation contribution. Let λ_1, λ_2 be two known positive real-numbers, and define,

$$Z_t = \lambda_1 \widetilde{X}_{1,t} + \lambda_2 \widetilde{X}_{2,t}. \tag{3.17}$$

It can easily be verified that such a process satisfies the assumption of significant correlation contribution given above. We note the importance in the above definition to allow for Γ be larger than $\cup_{i=1}^{p} \Gamma_i$. Indeed, for any positive-definite γ_0, γ_1, there exists an infinite number of combinations of the parameters σ_1, r, σ_2 such that the autocovariance of the process $\widetilde{X}_{1,t} + \widetilde{X}_{2,t}$ has autocovariance γ_0 and γ_1 at lags zero and one, such that $\Gamma = \{0, 1\}$ would be too small a set to determine uniquely the set of parameters of the latent processes. However, we can verify that taking $\Gamma = \{0, 1, 2\}$ is enough. In Section 7.2, we simulate the process (3.17) and estimate its parameters by generalizing the inference method that will be described in Chapter 5.

3.7 Summary

The definition of significant correlation contribution allows us to relax the constraints on the modulation sequence which are necessary for asymptotically consistent estimation, in comparison to past literature. More precisely, in Chapter 5, we propose a computationally efficient estimation procedure for this class of processes. We prove the consistency of that procedure in Chapter 6. In terms of applications, we simulate and estimate modulated processes with a significant correlation contribution in Chapter 7, and the processes considered in our real-world application will be shown to belong to our class of processes making use of Proposition 5 in Section 8.2.2.

Chapter 4

Inference for stationary and locally stationary processes

The inference of stochastic processes from observed time series is key to understanding the underlying generating mechanisms behind many observed phenomena. In Chapter 8 we will present a stochastic process representation of velocity time series obtained from drifting instruments designed to follow the ocean surface currents. The stochastic modelling proposed is a parametric model adapted from a deterministic model from oceanography. The parameters of our stochastic model therefore have a physical interpretation and are worth investigating. In this chapter we review some inference methods for stationary and locally stationary processes. We review estimation of the mean, autocovariance sequence and spectral density of a stationary process. We then review some parametric estimation methods. The method of moments fits the parametric model whose first and second order moments equal those estimated from the sample. Very often this method, albeit consistent, is not optimal in terms of efficiency. Maximum likelihood estimators consist in maximizing the likelihood function. In some situations, these estimators are asymptotically equivalent to least square estimators, which are obtained by minimizing the sum of squares of errors from the one-step optimal linear predictor, weighted by the inverse of the variances of those predictors. For Gaussian processes the exact likelihood can be derived from the parametric mean and covariance matrix. More generally, it is common to compute and maximize the Gaussian likelihood for linear

processes even if they are not assumed to be Gaussian (Hannan, 1973). However, approximations to the likelihood are often required when the time series length N is large. This is because the exact Gaussian likelihood usually requires inverting large covariance matrices and computing their determinant. This can be achieved in order $\mathcal{O}(N^2)$ elementary operations for a regularly sampled stationary process, in which case the covariance matrix is Toeplitz (Jain, 1979), but will usually be even higher order for irregularly sampled or non-stationary processes. Approximations to the exact likelihood are referred to as *quasi-likelihoods* in this thesis (Wedderburn, 1974, McCullagh, 1983), although other authors use the term *pseudo-likelihood*. One such commonly favoured quasi-likelihood in time series analysis is the Whittle likelihood (Whittle, 1953), which has a computation cost of order $\mathcal{O}(N \log N)$ and has been shown to be asymptotically equivalent to the Gaussian likelihood (Dzhaparidze and Yaglom, 1983). Finally, we review how the Whittle likelihood can be extended to locally-stationary time series (Dahlhaus, 1997), and discuss succinctly another approach for the Gaussian scenario. Throughout this chapter, we consider $\mathbf{X} = X_0, \ldots, X_{N-1}$ a length-N sample from a stationary Gaussian process with mean zero (unless stated otherwise), $\{X_t\}$. We shall denote Ω_N the set of Fourier frequencies $\frac{2\pi}{N} \cdot \left(-\lceil \frac{N}{2} \rceil + 1, \cdots, -1, 0, 1, \cdots, \lfloor \frac{N}{2} \rfloor\right)$ in the rest of this thesis. All the material presented in this chapter is reviewed material.

4.1 Estimation of the mean, autocovariance sequence, and spectral density

We first review non-parametric estimators of the mean and autocovariance sequence of a stationary process. We then review the periodogram, a commonly used estimator of the spectral density defined in (2.2). We review the common assumptions that are needed for the mean and autocovariance sequence estimators to be consistent, and their asymptotic distributions. However, we will see that the variance of the periodogram does not decrease to zero. Finally, we review the use of such estimators in method of moments, a parametric estimation procedure where we fit the model that corresponds to the estimated mean and autocovariance.

4.1.1 Estimation of the mean

In this section we do not assume that the mean of the process $\{X_t\}$ is zero. We first recall the weak law of large numbers for i.i.d. random variables.

Theorem 2 (Weak law of large numbers for i.i.d. random variables). *Let $\{X_t : t \in \mathbb{N}\}$ be a family of i.i.d. random variables with mean μ. According to the weak law of large numbers, we have that,*

$$\frac{1}{N} \sum_{i=0}^{N-1} X_t \xrightarrow{p} \mu, \quad (N \longrightarrow \infty).$$

In the situation of a stochastic process, the random variables are not independent in the general case and therefore the above does not apply. A law of large numbers is therefore required for such kind of data. This is provided in the next theorem for stationary stochastic process, with some assumption on the autocovariance of the process. More specifically, it is required that the correlation between variables decreases to zero when the time lag goes to infinity.

Theorem 3 (Weak-law of large numbers for stationary time series). *Let $\{X_t : t \in \mathbb{N}\}$ be a stationary stochastic process. Assume that its autocovariance function, denoted $c_X(\tau)$, converges to zero as τ goes to infinity. Then,*

$$\frac{1}{N} \sum_{i=0}^{N-1} X_t \xrightarrow{p} \mu.$$

Moreover, if the series $\sum_{\tau \in \mathbb{Z}} |c_X(\tau)|$ converges, we have,

$$N \mathrm{var} \left\{ \frac{1}{N} \sum_{i=0}^{N-1} X_t \right\} \longrightarrow \sum_{\tau \in \mathbb{Z}} c_X(\tau).$$

Proof. See Appendix A.4. $\qquad\square$

For Gaussian linear processes, under mild conditions, the sample mean is normally distributed (Brockwell and Davis, 1991), allowing for establishing confidence intervals for the true mean value of the process.

4.1.2 Estimation of the autocovariance sequence

With the assumption of stationarity, the autocovariance function of a stochastic process can be estimated from a single realization. As we assume zero mean, an unbiased natural estimator would be (Guyon, 1982),

$$\hat{c}_X^u(\tau) = \frac{1}{N - |\tau|} \sum_{t=0}^{N-|\tau|-1} X_t X_{t+|\tau|}, \ \tau = 0, \ldots, N-1.$$

However, several reasons (Percival and Walden, 1993) have drawn many statisticians to replace the multiplicative factor $\frac{1}{N-|\tau|}$ by $\frac{1}{N}$, leading to the following biased sample autocovariance estimator,

$$\hat{c}_X(\tau) = \frac{1}{N} \sum_{t=0}^{N-|\tau|-1} X_t X_{t+|\tau|}, \ \tau = 0, \ldots, N-1.$$

One first reason to prefer the above biased estimator is that the unbiased estimator has a very large variance for lags close to the sample size. The other reason would be that the biased estimator satisfies the property of being non-negative definite, which is a basic characterization of autocovariance sequences. Because of these considerations, we shall only use the biased sample autocovariance estimator, which we will simply refer to as the autocovariance estimator or sample autocovariance. The asymptotic distribution of the sample autocovariance sequence can be determined for particular time series models, allowing for confidence intervals and goodness of fit tests (Brockwell and Davis, 1991). Note also that the sample autocovariance sequence can be computed in order $\mathcal{O}(N \log N)$ computations via a Fast Fourier Transform.

4.1.3 The periodogram

Recall that the spectral density of a stationary process $\{X_t\}$ with absolutely summable autocovariance sequence $c_X(\tau)$ is given by,

$$S_X(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} c_X(\tau) e^{-i\omega\tau}.$$

The squared modulus of the Fourier transform of the time series **X**, known as the *periodogram*, is a common statistic in stationary time series analysis (Percival and Walden, 1993), and is given by

$$\hat{S}_X^{(N)}(\omega) = \frac{1}{N} \left| \sum_{t=0}^{N-1} X_t e^{-i\omega t} \right|^2, \quad \omega \in \mathbb{R}. \tag{4.1}$$

It is easily seen that we have,

$$\hat{S}_X^{(N)}(\omega) = \sum_{\tau=-(N-1)}^{N-1} \hat{c}_X(\tau) e^{-i\omega \tau}. \tag{4.2}$$

Note that this quantity is 2π-periodic, i.e. $\hat{S}_X^{(N)}(\omega + 2\pi) = \hat{S}_X^{(N)}(\omega)$, $\omega \in \mathbb{R}$. In practice we compute the periodogram only for Fourier frequencies via a Fast Fourier Transform. The periodogram is an asymptotically unbiased estimator of the spectral density of the stationary process $\{X_t\}$, i.e. $\lim_{N\to\infty} \mathrm{E}\{\hat{S}_X^{(N)}(\omega); \theta\} = 2\pi S_X(\omega; \theta)$ for all $\omega \in [-\pi, \pi)$ (Brockwell and Davis, 1991). However the variance of the periodogram does not decrease to zero as the sample size increases. The asymptotic distribution of the periodogram at a Fourier frequency ω is shown to be exponentially distributed with mean $2\pi S_X(\omega; \theta)$. A consistent nonparametric estimator of a smooth spectral density $S_X(\omega; \theta)$ of the process $\{X_t\}$ is obtained by smoothing the periodogram across frequencies (Percival and Walden, 1993, p. 235–253), as long as $S_X(\omega; \theta)$ is continuous. More precisely, we form the following estimator, for some odd integer m,

$$\hat{S}_X^{(N),m}(\omega) = \frac{1}{m} \sum_{j=-(m-1)}^{m-1} \hat{S}_X^{(N)}\left(\omega + \frac{2j\pi}{N} \right). \tag{4.3}$$

If m goes to infinity while satisfying $m = o(N)$, and under some mild conditions on $\{X_t\}$, uniform mean square convergence of (4.3) to the true spectral density $S_X(\omega)$ follows.

4.1.4 Method of moments

The method of moments is a parametric inference method which consists in equating the time series sample moments with theoretical moments from a parametric class of models (Van der Vaart, 1998). Applied to inference for a stationary time series $\{X_t\}$, if for the estimated autocovariance sequence $\{\hat{c}_X(\tau) : \tau = 0, \ldots, N-1\}$, there exists a unique parameter vector θ within the parameter set Θ such that $\hat{c}_X(\tau) = c_X(\tau; \theta), \forall \tau = 0, \ldots, N-1$, then the moment estimator is given by θ.

We consider the use of method of moments for the class of auto-regressive models. Let $\{X_t\}$ be a causal AR(p) process,

$$\Phi(B)X_t = \varepsilon_t, \;\; \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{4.4}$$

where $\Phi(z) = 1 - \Phi_1 z - \Phi_2 z^2 - \ldots - \Phi_p z^p$, $\Phi_p \neq 0$ and $\Phi(z) \neq 0$ for $|z| \leq 1$, so that the process $\{X_t\}$ is causal. Multiplying each side of (4.4) by X_{t-j} for $j = 0, \ldots, p$, and taking expectations, we obtain the $p+1$ Yule-Walker equations, given by,

$$\begin{aligned}
\Gamma_p \Phi^{'} &= \gamma_p^{'}, \\
c_X(0) - \gamma_p \Phi^{'} &= \sigma^2,
\end{aligned}$$

where Γ_p is the $p \times p$ covariance matrix with elements $c_X(i-j)$, $i, j = 1, \ldots, p$, $\Phi = (\Phi_1, \ldots, \Phi_p)$, and $\gamma_p = (c_X(1), \ldots, c_X(p))$. Note that the covariance between X_t and ε_t is found to be σ^2 since, writing the causal representation of the process

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

we obtain $E\{X_t \varepsilon_t\} = \psi_0 \sigma^2$, by continuity of the linear product and using the fact that the solution to (4.4) can be shown to be mean-square convergent (Brockwell and Davis, 1991). It is also easy to verify that $\psi_0 = 1$ given the fact that for $|z| \leq 1$, $\psi(z) = 1/\phi(z)$.

The method of moments here consists in writing the Yule-Walker equations where the theoretical autocovariances $c_X(\tau; \Phi)$ are replaced with the estimated au-

tocovariances $\hat{c}_X(\tau)$, i.e. we solve,

$$\hat{\Gamma}_p\hat{\Phi}' = \hat{\gamma}_p'.$$

The existence and uniqueness of the solution $\hat{\Phi}$ is ensured as long as $\hat{c}_X(0)$ is positive, since in that case it can be shown that the matrix $\hat{\Gamma}_p$ is non-singular. Therefore, $\hat{\Phi}'$ is uniquely given by,

$$\hat{\Phi}' = \hat{\Gamma}_p^{-1}\hat{\gamma}_p'.$$

The method of moments can be achieved in linear complexity as it only requires the estimation of the variance and autocovariance for the p first lags, followed by the inversion of a $p \times p$ matrix. The method of moments is often less efficient than other estimators such as Maximum Likelihood Estimators (MLE), which are discussed in the next section. However, the method of moments is consistent under some very mild conditions and can serve as a preliminary estimation procedure. Also note that for the class of AR processes, it has the same asymptotic efficiency as MLE estimators (Brockwell and Davis, 1991, page 240).

4.2 Maximum likelihood for time series

Maximum Likelihood Estimators are obtained by maximizing the likelihood function. The likelihood function is defined over the set of parameter vectors Θ, and maps each parameter vector $\theta \in \Theta$ to the probability (or probability density for continuous distributions) of the observed sample under the probability distribution specified by θ. More formally, if a finite sample $X_0,\ldots,X_{N-1}$ follows a joint distribution given by $p_X(x_1,\ldots,x_n;\theta)$, then the likelihood function is defined by,

$$L(\theta) = p_X(X_0,\ldots,X_{N-1};\theta).$$

The log-likelihood function is simply defined as $l(\theta) = \log L(\theta)$, and the MLE is obtained by maximizing the likelihood, or equivalently the log-likelihood,

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta}\{l(\theta)\}.$$

Finding the global maximum of the log-likelihood can be achieved via using optimization procedures or analytically in special cases. However, it is often essential that such procedures are initialized with good starting values, to avoid convergence to local maxima. Another reason for requiring good initial starting values of the optimization procedure is the heavy computational cost of exact likelihood. Computation of the exact likelihood for a stationary Gaussian process can be achieved with order $\mathcal{O}(N^2)$ operations using the innovations algorithm (Brockwell and Davis, 1991). More generally, the Gaussian likelihood may be computed for a stationary linear time series, with a $\mathcal{O}(N^2)$ complexity if the sampling is regular, making use of the Toeplitz structure of the covariance matrix (Jain, 1979). To further reduce computational times, it is common to make use of likelihood approximations, commonly termed as quasi-likelihoods. This is the topic of the next section of this chapter.

4.3 Quasi-likelihood for stationary time series

Large data sets have become ubiquitous due to the rise of automatic high-frequency measurements and storage capacities. Despite the growth of computational power and the development of parallel-computing, it is common that computing the exact Gaussian likelihood becomes too expensive in terms of computational time. One way to reduce the computational time is to use likelihood approximations, commonly termed as *quasi-likelihoods* or *pseudo-likelihoods*. In the theory of stationary stochastic processes, one well-studied such approximation to the Gaussian likelihood is the Whittle likelihood (Whittle, 1953), which can be computed in $\mathcal{O}(N \log N)$ elementary operations.

We now describe the Whittle likelihood for stationary stochastic processes. Let $\{X_t\}$ be a stationary stochastic process with mean zero and with autocovariance sequence $c_X(\tau)$ which we assume to be absolutely summable. The corresponding spectral density is denoted $S_X(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} c_X(\tau) e^{-i\omega\tau}$. The finite symmetric Toeplitz autocovariance matrix $\Gamma_N = \{c_X(i-j)\}_{i,j=0,\dots,N-1}$, which appears in the Gaussian likelihood, can be approximated by the circulant $N \times N$ symmetric matrix

with first row given by (Grenander and Szegö, 2001),

$$\frac{2\pi}{N} \sum_{k=0}^{N-1} S_X \left(\frac{2k\pi}{N} \right) e^{i\frac{2k\pi}{N}\tau}, \ \tau = 0, \ldots, N-1.$$

The corresponding quasi-likelihood is then given by,

$$l_W(\theta) = \sum_{\omega \in \Omega_N} \left\{ \log S_X(\omega; \theta) + \frac{\hat{S}_X(\omega)}{S_X(\omega; \theta)} \right\}, \tag{4.5}$$

where $\hat{S}_X(\omega)$ is the periodogram of the time series, which can be computed in $\mathcal{O}(N \log N)$ elementary operations using the Fast Fourier Transform algorithm, making the computation of the Whittle likelihood $\mathcal{O}(N \log N)$ itself, an improvement compared to the $\mathcal{O}(N^2)$ necessary for the computation of the exact likelihood in general. The Whittle likelihood estimator is the parameter vector that minimizes the quantity $l_W(\theta)$ over the parameter set Θ. Its distribution converges in law to that of the Gaussian likelihood estimator for stationary linear stochastic processes (Hannan, 1973) under increasing sample size.

The counter-part to the computational ease from using the Whittle likelihood as compared to the Gaussian likelihood is a finite-sample bias and possibly variance inflation. The bias can be explained by the fact that the periodogram itself is a biased estimator of the spectral density, except in the case of white noise processes. Truncated time series result in *blurring* or *leakage*, in the sense that we are not able to estimate the spectral density at one specific frequency but rather a local average of the spectral density (Thomson, 1982). More precisely, the expectation of the periodogram is (Bloomfield, 2000),

$$\mathrm{E}\left\{ \hat{S}_X(\omega) \right\} = \int_{-\pi}^{\pi} S_X(\omega - \lambda) \mathscr{F}^{(N)}(\lambda) d\lambda, \tag{4.6}$$

where the Féjer kernel $\mathscr{F}^{(N)}(\lambda)$ is the Fourier transform of the triangle kernel $1 - |\tau|/N$, $\tau = -(N-1), \ldots, N-1$, which satisfies for $\lambda \in [-\pi, \pi]$,

$$\mathscr{F}^{(N)}(\lambda) = \begin{cases} \dfrac{\sin^2\left(\frac{N\lambda}{2}\right)}{N\sin^2\left(\frac{\lambda}{2}\right)} & \forall \lambda \in \mathbb{R} \setminus \Omega_N \\ N & \text{if } \lambda = 0, \\ 0 & \forall \lambda \in \Omega_N \setminus \{0\} \end{cases} \quad . \tag{4.7}$$

The impact of leakage increases with high dynamic range and curvature of the spectral density (Prieto et al., 2007), and disappears for white noise, in which case the spectral density is constant. For white noise the spectrum has no dynamic range or curvature. Additionally, when sampling from a continuous-time process, a common issue is the aliasing effect, where any frequency above the Nyquist frequency cannot be estimated, and folds back into the estimated spectrum. The bias of the periodogram can be corrected partly by tapering (Thomson, 1982, Percival and Walden, 1993), which consists in replacing the periodogram by,

$$\hat{S}_X^{(t)}(\omega) = \frac{1}{N}\left| \sum_{t=0}^{N-1} h_t X_t e^{-i\omega t} \right|^2, \ \omega \in \mathbb{R}, \tag{4.8}$$

for some appropriate choice of $\{h_t\}$ admitting a Finite Fourier transform concentrated around zero (Dahlhaus, 1988). An alternative method to reduce the bias of the Whittle estimate is to forward model the expectation of the periodogram and therefore account for both aliasing and leakage. In Section 5.2 we review a quasi-likelihood estimate for stationary time series based on this idea (Sykulski et al., 2016a), and where the spectral density $S_X(\omega)$ in (4.5) is replaced by the finite-sample parametric expected periodogram.

4.4 Inference for locally-stationary time series

In this section we review the problem of inference for locally-stationary stochastic processes, which were reviewed in Section 2.6. Keeping with the notation of that section for locally-stationary processes, we recall that the time-varying spectral density in the rescaled time $u \in [0,1]$ is defined as $f(u, \lambda) = |A(u, \lambda)|^2$. We consider the problem of inferring a parametric form of the time-varying spectral density, i.e. we consider a family of time-varying spectral densities $f_\theta(u, \lambda)$ indexed by some

parameter $\theta \in \Theta$. Accounting for model misspecification, we do not assume that there always exists a parameter value $\theta \in \Theta$ such that $f(u, \lambda) = f_\theta(u, \lambda)$. However, we have seen in Section 2.6 that the asymptotic Kullback-Leibler distance between the true probability density function and the probability density function associated with the spectral density $f_\theta(u, \lambda)$ is given by, under the assumption of Gaussianity,

$$\mathscr{L}(\theta) = \int_0^1 \int_{-\pi}^\pi \left\{ f_\theta(u, \lambda) + \frac{f(u, \lambda)}{f_\theta(u, \lambda)} \right\} d\lambda du.$$

Therefore estimation in that situation can be viewed as the problem of finding the value θ_0 that minimizes the asymptotic Kullback-Leibler information divergence $\mathscr{L}(\theta)$. It is shown by Dahlhaus (1997) that we can obtain a consistent estimator of θ_0 by minimizing the function,

$$\mathscr{L}_T(\theta) = \sum_{j=0}^{M-1} \int_{-\pi}^\pi \left\{ \log f_\theta(u_j, \lambda) + \frac{I_N(u_j, \lambda)}{f_\theta(u_j, \lambda)} d\lambda \right\}, \tag{4.9}$$

where $I_N(u, \lambda)$ is a local tapered periodogram, defined by,

$$I_N(u, \lambda) = \frac{1}{H_N} \left| \sum_{k=0}^{N-1} h\left(\frac{k}{N}\right) X_{\lfloor Tu \rfloor - N/2 + k, T} \exp(-i\lambda k) \right|^2,$$

and the points u_j (or equivalently $t_j = Tu_j$) in (4.9) are obtained by translation, according to,

$$t_j = N/2 + Sj, \quad j = 0, \dots, M-1.$$

The taper $h(\cdot)$ is a continuous function on $[0, 1]$, and the quantity H_N is defined by,

$$H_N = \sum_{k=0}^{N-1} h\left(\frac{k}{N}\right)^2.$$

The asymptotic consistency is established upon conditions on N, T and S. As one would expect, the size of the local periodograms, given by N, must increase, but at a slower rate than T, so that asymptotically the periodogram behaves as for stationary processes with increasing sample size. More precisely, Dahlhaus (1997) estab-

lishes consistency under the condition $T^{1/4} << N << T^{1/2}/\log T$ while $T \to \infty$, and $S = N$ or $S/N \to 0$. The framework of infill asymptotics, which corresponds to T going to infinity, also allows us to understand the consequences of fitting a stationary model to a non-stationary time series. In our real-world data application of Chapter 8, we consider a situation where the non-stationarity is so strong within a time window that stationary model estimates are badly biased. Choosing smaller time windows may then reduce the bias due to the non-stationarity, at the expense of a large variance due to the small sample size. We will show how the non-stationary process considered in Chapter 8 can be modelled as a modulated process with a significant correlation contribution, and how the inference methods developed in Chapter 5 allows us to bypass this bias-variance trade-off.

4.5 Another approach to Gaussian likelihood approximation

In this section we briefly review another approach to approximating the Gaussian likelihood. Let $\{X_t\}$ be a stationary Gaussian process with autocovariance sequence $c_X(\tau; \theta)$, where $\theta \in \Theta$ is a vector of $\mathbb{R}^p$. Let $\mathbf{X}$ be a length-N sample from this process, and denote $C_X(\theta)$ the corresponding autocovariance matrix. The minimization of the likelihood function is (almost) equivalent to solving the score equations, which take the following form in the Gaussian case,

$$0 = \text{Tr} \left\{ \left(C_X(\theta)^{-1} \mathbf{X} (C_X(\theta)^{-1} \mathbf{X})^T - C_X(\theta)^{-1} \right) \frac{\partial C_X}{\partial \theta_j}(\theta) \right\} \quad j = 1, \ldots, p. \quad (4.10)$$

For large N, the right hand side of the above equation is intractable. Anitescu et al. (2012) propose a matrix-free approach to (4.10), where only matrix-vector computations are necessary. Key to their methodology is the use of the Hutchison estimator to estimate the trace, where for any matrix A,

$$\text{Tr}\, A = E\left\{ U^T A U \right\}, \quad (4.11)$$

with U a random vector with independent components taking values 1 and -1 with equal probability. The reason for choosing this distribution rather than the Gaussian distribution for the random vector U is due to the smaller variance of the obtained estimator. The expectation in (4.11) can be approximated by a sample average approximation, which simply consists in generating a finite number m of random vectors U and averaging,

$$\operatorname{Tr} A \approx \sum_{j=1}^{m} U^{T} A U.$$

We then replace (4.10) by the following stochastic system of equations,

$$0 = \sum_{j=1}^{m} \left\{ U_{j}^{T} \left(C_{X}(\theta)^{-1} \mathbf{X} (C_{X}(\theta)^{-1} \mathbf{X})^{T} - C_{X}(\theta)^{-1} \right) \frac{\partial C_{X}}{\partial \theta_{j}}(\theta) U_{j} \right\} \quad j = 1, \ldots, p.$$

The stochastic solution θ^{*} to these equations can be shown to converge normally in probability to the true parameter vector.

Chapter 5

Inference for modulated processes

In this chapter we treat the problem of inference for non-stationary modulated processes. We focus on the study of scalar modulated processes, but extending the theory to our class of bivariate modulated processes from Section 3.5 is straightforward using complex-valued representations. We consider $\{X_t\}$ a stationary stochastic process, modulated by a sequence $\{g_t\}$. The modulated process is obtained through pointwise multiplication, $\widetilde{X}_t = g_t X_t$, as described in Definition 8. Inference for modulated processes means inference about the model posited for the latent stationary process. Dividing the observed modulated process by the observed modulation sequence is not considered as a viable solution in the most general case, as g_t may take the value zero for some time points, and generally we may observe the aggregation of the modulated process with another process. We seek to derive an estimation method that is consistent in the framework of increasing sample size. This is made possible despite the non-stationarity of the observed processes, as we model the non-stationarity via modulation which ensures the growth of information in time if the modulation sequence satisfies certain sufficient conditions which we studied in Sections 3.2 and 3.3.

The problem of inference for modulated processes has been treated mostly under the assumption of asymptotically stationary modulated processes. This constrains the modulation sequence $\{g_t\}$ to be such that its sample autocovariance sequence converges for all lags to a non-zero value. The autocovariance sequence of the latent process X_t can then be estimated by dividing the sample autocovariance

sequence of the modulated processes by this non-zero limit at all lags, as proposed by Parzen (1963), see (3.6). These second-order moment estimates can then be used in method of moments estimators or in Gaussian likelihood methods. Spectral estimates can be constructed for the latent process simply by Fourier transforming their estimated autocovariance sequence. However the non-negative definiteness of the estimated autocovariance sequence of the latent process is no longer ensured, due to the division operation in their definition. Thus the resulting spectral estimates may be negative at some frequencies. Parametric inference in contrast ensures the validity of the fitted model. We present a parametric estimation method based on an adaptation of the Whittle likelihood. In Chapter 6 we will prove the consistency of this method for the class of modulated processes with a significant correlation contribution, which extends that of asymptotically stationary modulated processes. This chapter starts with a short review of sampling properties of modulated processes in Section 5.1, with a focus on the properties of the periodogram. We then review an adaptation of the Whittle likelihood for stationary time series in Section 5.2, based on the finite sample expectation of the periodogram. In Section 5.3 we present an extension of this idea for the estimation of modulated processes. In Section 5.4 we consider the situation where the modulation sequence is not observed, as opposed to the general case considered in this thesis, but admits a parametric form. Assuming the parameters of the modulation sequence can be inferred from observing the modulated process, we adapt the inference procedure presented in Section 5.3 to this situation. In this chapter, Section 5.1 constitutes reviewed material, except for Proposition 8. Section 5.2 is novel and is published in Sykulski et al. (2016a). Sections 5.3 and 5.4 are novel and published in Guillaumin et al. (2017), and also relate to developments of Dunsmuir and Robinson (1981b), who treated the problem of parametric spectral estimation of asymptotically stationary modulated processes.

5.1 Sampling properties of modulated processes

In this section we shall review and study some distributional properties of the periodogram of a modulated stochastic process. Dunsmuir and Robinson (1981a) used the periodogram as the basis for designing quasi-likelihood methods for asymptotically stationary modulated time series, with an emphasis on treating the problem of missing data. Similarly, in Section 5.3 we will use the results of this section to formulate a quasi-likelihood using the periodogram, for our class of modulated processes with significant correlation contribution.

We denote $\widetilde{\mathbf{X}} = \{\widetilde{X}_t : t = 0, \cdots, N-1\}$ as a single realization of a length-N sample of a modulated process $\{\widetilde{X}_t\}$ defined in Definition 8. The unobserved sample of the latent stationary process is denoted $\mathbf{X} = \{X_t : t = 0, \cdots, N-1\}$ accordingly. The stationary latent process is modelled by a parameter vector $\theta \in \Theta$. For the modulated process $\{\widetilde{X}_t\}$, the latent time series $\{X_t\}$ is not observed, so we instead compute the periodogram of the modulated (and observed) process itself, $\hat{S}_{\widetilde{X}}^{(N)}(\omega)$, and we define the expected periodogram to be

$$\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) = \mathrm{E}\left\{\hat{S}_{\widetilde{X}}^{(N)}(\omega); \theta\right\}, \quad \omega \in \mathbb{R}.$$

Note that this quantity is also 2π-periodic. It is necessary to understand how modulation in the time domain will affect the expected periodogram. Proposition 6 gives more insight on how $\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta)$ relates to the modulating sequence g_t and the spectral density $S_X(\omega; \theta)$ of the latent stationary process $\{X_t\}$.

Proposition 6 (Expectation of the periodogram of a modulated time series)**.** *The expectation of the periodogram of the modulated time series takes the form*

$$\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) = 2\pi \int_{-\pi}^{\pi} S_X(\omega - \lambda; \theta) S_g^{(N)}(\lambda) d\lambda, \ \forall \omega \in \mathbb{R}, \tag{5.1}$$

which is a periodic convolution. Here $S_g^{(N)}(\lambda)$ is the squared value of the Fourier

Transform of the finite sequence $\{g_t\}_{t=0,\cdots,N-1}$ *i.e.*

$$S_g^{(N)}(\lambda) = \frac{1}{2\pi N}\left|\sum_{t=0}^{N-1} g_t e^{-i\lambda t}\right|^2,$$

defined for $\lambda \in \mathbb{R}$ *and which is* 2π-*periodic.*

Proof. The proof for this proposition, which is a well-known result, can be found in Dunsmuir and Robinson (1981a, p. 562). □

When g_t is unity everywhere, which corresponds to observing the stationary latent process directly, the quantity $2\pi S_g^{(N)}(\lambda)$ is the usual Féjer kernel as defined by (4.7), and which behaves asymptotically (as N tends to infinity) as a Dirac delta-function centred at zero. This explains why the periodogram is, asymptotically, an unbiased estimator of the spectral density of a stationary process up to a multiplicative factor of 2π (Brockwell and Davis, 1991).

When g_t is such that the modulated process is asymptotically stationary, Dunsmuir and Robinson (1981b) approximate $\frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty}\gamma(\tau)e^{i\omega\tau}$, where $\gamma(\tau) = \gamma_g(\tau)c_X(\tau)$ using the notation of (3.5), for ω at Fourier frequencies by,

$$\widetilde{S}_{\widetilde{X}}^{(D)}(\omega;\theta) = \frac{(2\pi)^2}{N}\sum_{\lambda\in\Omega_N} S_X(\omega-\lambda;\theta)S_g^{(N)}(\lambda). \tag{5.2}$$

When $\{g_t\}$ is such that the modulated process $\widetilde{X}_t$ has a significant correlation contribution, we derive the exact value of $\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta)$ by using the theoretical auto-covariances of the latent model, in a similar fashion as in Sykulski et al. (2016a) for stationary processes. This is the result of Proposition 7, which follows.

Proposition 7 (Computation of the expected periodogram)**.** *Let* $\omega \in \mathbb{R}$. *We have*

$$\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta) = 2\mathscr{R}\left\{\sum_{\tau=0}^{N-1}\overline{c}_{\widetilde{X}}^{(N)}(\tau;\theta)e^{-i\omega\tau}\right\} - \overline{c}_{\widetilde{X}}^{(N)}(0;\theta), \tag{5.3}$$

where $\overline{c}_{\widetilde{X}}^{(N)}(\tau;\theta)$ *is defined in (3.3). By defining* $\overline{c}_{\widetilde{X}}^{(N)}(-\tau;\theta) = \overline{c}_{\widetilde{X}}^{(N)}(\tau;\theta)$ *for* $\tau =$

$1, \cdots, N-1$ *we can (equivalently) express this relationship as*

$$\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) = \sum_{\tau=-(N-1)}^{N-1} \overline{c}_{\widetilde{X}}^{(N)}(\tau; \theta) e^{-i\omega\tau},$$

Proof. See Appendix A.5. □

Therefore the expectation of the periodogram of $\widetilde{\mathbf{X}}$ is the discrete Fourier transform of the expected sample autocovariance sequence. This is true even though we have not assumed stationarity; it is simply a consequence of the relation between the formal definitions of (3.3) and (4.1). Note that calculating the Fourier transform of the sequence $\overline{c}_{\widetilde{X}}^{(N)}(\tau; \theta)$ will always give a real-valued positive $\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta)$ for $\theta \in \Theta$, as the latter is defined as the expectation of the squared modulus of the Fourier transform of the process.

Proposition 7 can be used to compute the expected periodogram of an asymptotically stationary modulated process. In such cases, the difference between (5.2) and (5.3) is that (5.2) is a finite approximation of (5.1), whereas (5.3) is exact. To justify the use of the expected periodogram in the setting of modulated processes, we investigate the dependence of the expectation of the periodogram on the parameter of the latent process under the assumption of significant correlation contribution.

Proposition 8 (Identifiability of the expected periodogram)**.** *If the modulated process has a significant correlation contribution, the expected periodogram is a one-to-one (i.e. injective) mapping from the parameter set Θ to the set of non-negative continuous functions on $[-\pi, \pi]$, for a large enough sample size. More specifically, for two distinct parameter vectors θ and θ', the expected periodograms $\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta)$ and $\overline{S}_{\widetilde{X}}^{(N)}(\omega; \widetilde{\theta})$ cannot be equal for all Fourier frequencies $\frac{2\pi}{N}\left(-\lceil \frac{N}{2} \rceil + 1, \cdots, -1, 0, 1, \cdots, \lfloor \frac{N}{2} \rfloor\right).$*

Proof. Let $\theta, \widetilde{\theta} \in \Theta$ be distinct parameter vectors and let N be a positive integer. Let Γ be as given by Definition 9. By the assumption of significant correlation contribution, the finite sequences $\{c_X(\tau; \theta) : \tau \in \Gamma\}$ and $\{c_X(\tau; \widetilde{\theta}) : \tau \in \Gamma\}$ are not equal. Since $\overline{c}_{\widetilde{X}}^{(N)}(\tau; \theta) = c_g^{(N)}(\tau)c_X(\tau; \theta)$ for $\tau \in \Gamma$, and according to (3.8), for N large enough (and independently from the chosen parameter vectors θ and $\widetilde{\theta}$)

the sequences $\{\overline{c}_{\widetilde{X}}^{(N)}(\tau;\theta) : \tau \in \Gamma\}$ and $\{\overline{c}_{\widetilde{X}}^{(N)}(\tau;\widetilde{\theta}) : \tau \in \Gamma\}$ are not equal. Hence for N large enough the sequences $\{\overline{c}_{\widetilde{X}}^{(N)}(\tau;\theta) : \tau = -(N-1), \cdots, N-1)\}$ and $\{\overline{c}_{\widetilde{X}}^{(N)}(\tau;\widetilde{\theta}) : \tau = -(N-1), \cdots, N-1)\}$ are not equal. Their finite Fourier transforms $\{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta) : \omega \in \Omega_N\}$ and $\{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\widetilde{\theta}) : \omega \in \Omega_N\}$, are by the bijective nature of the Fourier transform, not equal either. $\qquad\square$

This means that for two distinct parameters vectors $\theta, \widetilde{\theta} \in \Theta$, we will have two distinct expected periodograms. This is a necessary condition for an estimation procedure based on the expected periodogram. We will propose such an estimation procedure in Section 5.3, and derive its consistency and convergence rate in Chapter 3.

5.2 Finite sample bias elimination for stationary stochastic processes

In this section we review a modification of the Whittle likelihood proposed by Sykulski et al. (2016a) intended to reduce its bias for finite samples. We discussed a first approach in Section 4.3 which consists in tapering the time series sample, to reduce the effect of the convolution (4.6). Another approach is to model the outcome of the convolution itself. We consider a length-N realization $X_0, \ldots, X_{N-1}$ from a stochastic stationary Gaussian process modelled by a parameter vector $\theta \in \Theta$. The idea is to replace in (4.5) the spectral density by the finite sample expectation of the periodogram. The expected periodogram from a stationary time series is known to be the convolution of the spectral density with the Féjer kernel. It can be computed using Proposition 7 by setting $g_t = 1$ for $t = 0, \ldots, N-1$, which leads to,

$$\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta) = \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) c_X(\tau;\theta) e^{-i\omega\tau}. \tag{5.4}$$

The modified version of the Whittle likelihood is called *de-biased Whittle likelihood*, taking the form of:

$$\ell_M(\theta) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) + \frac{\hat{S}_{\widetilde{X}}^{(N)}(\omega)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta)} \right\}. \tag{5.5}$$

In Chapter 6 we prove the consistency of the estimator obtained by minimizing $\ell_M(\theta)$ for a class of stationary Gaussian processes that satisfy some mild assumptions. We will also review the asymptotic $\mathcal{O}(N^{-\frac{1}{2}})$ convergence rate of this method. Sykulski et al. (2016a) also noted that tapering can be used to further remove bias from this form of estimates.

As a remark, a third option exists to reduce the bias of the Whittle estimate, namely, we can divide $\hat{c}_X(\tau)$ by $\frac{N-|\tau|}{N}$ and Fourier transform to obtain an unbiased estimate of the spectral density (Guyon, 1982). However we lose the non-negative semi-definiteness of the estimated autocovariance sequence and increase its variance and mean square error for large lags, as noted by Dahlhaus and Künsch (1987).

5.3 Estimation of modulated processes

In Chapter 3 we have explored a class of univariate and bivariate modulated processes. The next stage is to describe their efficient inference. In this section we describe how the parameters of the latent model for $\{X_t\}$ can be inferred from observing a single realization of the modulated process $\{\widetilde{X}_t\}$. Most authors have focused on the problem of estimating modulated processes under the assumption of asymptotic stationarity as defined in Definition 5 (Parzen, 1963, Dunsmuir and Robinson, 1981a,c, Toloi and Morettin, 1989). Although non-parametric estimates have been the key concern in most of the relevant literature, there have been instances where parametric estimation has been considered, see for instance Dunsmuir and Robinson (1981c). Parametric estimation ensures that the estimated autocovariance sequence is non-negative definite, as opposed to using non-parametric estimates of the form given in (3.6). Parametric estimation is also preferable when the true model is known, as it uses the observed degrees of freedom more efficiently. Herein we

consider the problem of parametric estimation for our class of modulated processes with a significant correlation contribution, which, we recall, is a generalization of asymptotically stationary modulated processes. We propose an adaptation of the Whittle likelihood (Whittle, 1953), based on the expected periodogram.

We wish to infer the parameter vector θ of the latent univariate stationary process $\{X_t\}$ within the parameter set Θ, based on the sample $\widetilde{\mathbf{X}} = \widetilde{X}_0, \cdots, \widetilde{X}_{N-1}$ and the known modulating sequence $\{g_t : t = 0, \cdots, N-1\}$. Because it has been assumed that the latent process is a zero-mean Gaussian process, the same is true for the modulated process. The vector $\widetilde{\mathbf{X}}$ is multivariate Gaussian with an $N \times N$ autocovariance matrix $C_{\widetilde{X}}(\theta) = \left\{ c_{\widetilde{X}}(t_1, t_2; \theta) \right\}_{t_1, t_2 = 0, \cdots, N-1}$, where the components of this matrix are given by $c_{\widetilde{X}}(t_1, t_2; \theta) = g_{t_1} g_{t_2} c_X(t_2 - t_1; \theta)$. However, the parameter vector θ of the latent process $\{X_t\}$ can be uniquely determined from the modulated process $\{\widetilde{X}_t\}$ only if $\theta \to \left\{ c_{\widetilde{X}}(t_1, t_2; \theta) : t_1, t_2 \in \mathbb{N} \right\}$ is injective, i.e. there is no $\theta' \in \Theta$ such that $\theta \neq \theta'$ and $c_{\widetilde{X}}(t_1, t_2; \theta) = c_{\widetilde{X}}(t_1, t_2; \theta') \ \forall t_1, t_2 \in \mathbb{N}$. This necessary condition is clearly achieved under the assumption of a modulated process with significant correlation contribution. The negative of the exact time-domain Gaussian log-likelihood is proportional to

$$\ell_G(\theta) = \frac{1}{N} \log \left| C_{\widetilde{X}}(\theta) \right| + \frac{1}{N} \widetilde{\mathbf{X}}^T C_{\widetilde{X}}(\theta)^{-1} \widetilde{\mathbf{X}}, \tag{5.6}$$

where $\left| C_{\widetilde{X}}(\theta) \right|$ denotes the determinant of $C_{\widetilde{X}}(\theta)$. Note that one may need to remove from $\widetilde{\mathbf{X}}$ points where g_t is zero, to ensure that the determinant of the covariance matrix is non-zero, and since those observations carry no information about θ. We minimize ℓ_G to obtain the time-domain MLE, i.e.

$$\hat{\theta}_G^{(N)} = \arg \min_{\theta \in \Theta} \ell_G(\theta).$$

Due to the drawbacks of exact likelihood which we described in Chapter 4, we propose a computationally efficient estimation method for the parameters of the latent model based on the periodogram of the modulated time series. In Section 4.3 we reviewed the Whittle likelihood for stationary time series. In Section 5.2 we

reviewed how the bias from this method can be partly suppressed by replacing the spectral density by the expectation of the periodogram. We now further use this idea, and present a similar quasi-likelihood for the class of modulated processes with significant correlation contribution in the following definition.

Definition 13 (Spectral maximum quasi-likelihood estimator for univariate modulated processes). *Let $\{\widetilde{X}_t\}$ be a modulated process with significant correlation contribution and let $\widetilde{\mathbf{X}}$ be its length-N sample. We define the following quasi-likelihood:*

$$\ell_M(\theta) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) + \frac{\hat{S}_{\widetilde{X}}^{(N)}(\omega)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta)} \right\}, \qquad (5.7)$$

where $\overline{S}_{\widetilde{X}}^{(N)}(\omega)$ is defined in Section 5.1 as the expectation of the periodogram of the modulated time series, and is computed using Proposition 7. The corresponding estimator of the parameter vector θ is obtained by a minimization procedure over the parameter set,

$$\hat{\theta}_M^{(N)} = \arg\min_{\theta \in \Theta} \ell_M(\theta). \qquad (5.8)$$

The sequence $\{c_g^{(N)}(\tau) : \tau = 0, \cdots, N-1\}$ defined in (3.4) is necessary in the computation of the expected periodogram. It requires $\mathcal{O}(N \log N)$ computations as it can be computed as the biased sample autocovariance sequence of $g_0, \cdots, g_{N-1}$ via a Fast Fourier Transform. This initial step is carried out independently of inferring the parameter of interest θ. Then any computation of $\{\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) : \omega \in \Omega_N\}$ for any value of the parameter vector θ will require $\mathcal{O}(N \log N)$ computations. Indeed we can compute $\{\overline{c}_{\widetilde{X}}^{(N)}(\tau; \theta) : \tau = 0, \cdots, N-1\}$ in $\mathcal{O}(N)$ computations using (3.3) and the precomputed $\{c_g^{(N)}(\tau) : \tau = 0, \cdots, N-1\}$. Then the quantity $\{\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) : \omega \in \Omega_N\}$ is then computed via a fast Fourier transform (see Proposition 7). The reason for separating the initial step of computing $\{c_g^{(N)}(\tau) : \tau = 0, \cdots, N-1\}$ from the rest of the computation is that it is carried out independently of the parameter value and stored, and therefore outside any call to a minimization procedure over the parameter set Θ involving the expected periodogram.

In the trivial case of a modulation sequence equal to unity everywhere, then

the quasi-likelihood of Definition 13 does not exactly equal the Whittle likelihood of (4.5). This is because the spectral density $S_X(\omega)$ would be replaced by the expected periodogram $\overline{S}_{\widetilde{X}}^{(N)}(\omega)$, which is the convolution of the true spectral density with the Fejér kernel (see (5.2)). For stationary time series, this type of estimator was investigated in Sykulski et al. (2016a), and was found to significantly reduce bias in parameter estimation as compared with standard Whittle estimation, as discussed in Section 5.2. For modulated processes that are asymptotically stationary, this signifies the difference between using (5.2) and the quantity defined by Proposition 7 to fit the periodogram.

The same estimator to (13) can be used for the complex-valued proper time series $\widetilde{Z}_t$ considered in Section 3.5, i.e. we define our estimator,

$$\hat{\theta}_M^{(N)} = \arg\min_{\theta \in \Theta} \ell_M(\theta), \tag{5.9}$$

with the objective function given by,

$$\ell_M(\theta) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \overline{S}_{\widetilde{Z}}^{(N)}(\omega; \theta) + \frac{\hat{S}_{\widetilde{Z}}^{(N)}(\omega)}{\overline{S}_{\widetilde{Z}}^{(N)}(\omega; \theta)} \right\}, \tag{5.10}$$

The comments on computational aspects hold for the complex-valued case as well. In Chapter 6, we will prove consistency of the frequency-domain estimator (5.7) and its optimal $\mathcal{O}(N^{-1/2})$ convergence rate.

5.4 Unobserved modulation sequence

Most contributions to the literature on modulated processes have focused on the situation where the modulation sequence is observed (Jones, 1962, Parzen, 1963, Scheinok, 1965, Toloi and Morettin, 1989, Jiang and Hui, 2004), and in particular to applications to missing observations. However in practice it may happen that the modulation function is known only up to a parametric form. This situation has been considered for instance by Dunsmuir and Robinson (1981b) where it is assumed that the modulation sequence admits a parametric spectral representation.

Similarly, in this section we consider the situation where the modulation admits

a parametric representation determined by a parameter $\theta_g \in \Theta_g$. We denote $\{g_t(\theta_g)\}$ the parametric modulation sequence. We assume we only observe the modulated process. If a consistent estimator of the parameter vector θ_g is available, say $\hat{\theta}_g$, we estimate the latent process model parameters by computing $c_{\hat{g}}^{(N)}(\tau)$ according to,

$$c_{\hat{g}}^{(N)}(\tau) = \frac{1}{N} \sum_{t=0}^{N-\tau-1} g_t(\hat{\theta}_g) g_{t+\tau}(\hat{\theta}_g),$$

and we then minimize the quasi-likelihood (5.7) where the deterministic $c_g(\tau)$ is replaced by the estimated $c_{\hat{g}}^{(N)}(\tau)$ in the computation of $\overline{S}_X^{(N)}(\omega; \theta)$.

When estimating the parameters of the modulation sequence directly from the modulated time series is a difficult problem, we may incorporate these parameters in the quasi-likelihood minimization procedure. More precisely, a natural estimator for θ and θ_g would be,

$$(\hat{\theta}, \hat{\theta}_g) = \arg\min \ell_M(\theta, \theta_g), \tag{5.11}$$

where,

$$\ell_M(\theta, \theta_g) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta, \theta_g) + \frac{\hat{S}_{\widetilde{X}}^{(N)}(\omega)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta, \theta_g)} \right\},$$

and

$$\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta, \theta_g) = \sum_{\tau=-(N-1)}^{N-1} \overline{c}_{\widetilde{X}}^{(N)}(\tau; \theta, \theta_g) e^{-i\omega\tau},$$

with,

$$\overline{c}_{\widetilde{X}}^{(N)}(\tau; \theta, \theta_g) = c_g^{(N)}(\tau; \theta_g) c_X(\tau; \theta),$$

and with

$$c_g^{(N)}(\tau; \theta_g) = \frac{1}{N} \sum_{t=0}^{N-\tau-1} g_t(\theta_g) g_{t+|\tau|}(\theta_g).$$

Note that this estimation procedure requires that there exists Γ a subset of $\mathbb{N}$ such that $[\theta_g, \theta] \mapsto [\overline{c}_{\widetilde{X}}^{(N)}(\tau; \theta_g, \theta) : \tau \in \Gamma]$ be asymptotically injective, see Definition 10. We now consider two examples of parametric forms of the modulation sequence

where each estimation methods described above are applicable, respectively.

1. *Periodic modulation.* Let T be a positive integer strictly greater than one, and define

$$g_t = \alpha_{t \bmod T} + A\eta_t,$$

where $\{\eta_t\}$ is a standard normal white noise process independent of the latent stationary process, $A \geq 0$ is a fixed unknown constant, and $\alpha_0, \ldots, \alpha_{T-1}$ are unknown constants. The modulation sequence $\{g_t\}$ is periodic with period T up to the white noise process $\{\eta_t\}$ but is assumed completely unknown. This can be seen as an extension of the periodically missing data scheme considered earlier in Section 3.4, Example 1. Note that this model is identifiable only if we impose some normalization constraint on the $\alpha_0, \ldots, \alpha_{T-1}$, such as,

$$\sum_{k=0}^{T-1} \alpha_k^2 = 1, \tag{5.12}$$

so that the expected sample variance of the modulated process equals the variance of the latent process. A natural estimator of the parameters $\alpha_0, \ldots, \alpha_{T-1}$ of the modulation sequence, assuming the period T is know, is given by,

$$\hat{\alpha}_k^{(N)} = \sqrt{\frac{\sum_{i=0}^{\lfloor N/T \rfloor} \widetilde{X}_{k+iT}^2}{\sum_{l=0}^{T} \sum_{i=0}^{\lfloor N/T \rfloor} \widetilde{X}_{l+iT}^2}}, \quad k = 0, \ldots, T-1. \tag{5.13}$$

Indeed the numerator and denominator will respectively converge in probability to $\alpha_k \sqrt{\mathrm{var} X_t}$ and $\sqrt{\mathrm{var} X_t}$, because of the constraint (5.12), because of the stationarity of the T sub-samples and the properties of the sample variance. We do not estimate the value of A as it does not impact the quantity $\bar{c}_{\widetilde{X}}^{(N)}(\tau; \theta, \theta_g)$.

2. *Linearly increasing frequency.* Let $\gamma \in [-\pi, \pi]$ and $\Delta > 0$. Define,

$$\beta_t = \gamma + \Delta \frac{2t - (N-1)}{2(N-1)}, \tag{5.14}$$

and

$$g_t = e^{i \sum_{u=1}^{t} \beta_u}.$$

It is a matter of calculus to verify that we the modulation sequence is then given by,

$$c_g^{(N)}(\tau) = \frac{\sin\left[\frac{\Delta\tau}{2(N-1)}(N-\tau)\right]}{N \sin\left[\frac{\Delta\tau}{2(N-1)}\right]} e^{\left\{i\left(\gamma\tau + \frac{\Delta\tau}{2(N-1)}\right)\right\}}, \ \tau \in \mathbb{N},$$

which can directly be used (with computational complexity $\mathcal{O}(N)$) to compute the expected periodogram, although a Fast Fourier Transform will still be required.

In Section 7.3.1 we present simulations based on the models of both examples described above.

Chapter 6

Asymptotic theory

Previous work by Dunsmuir and Robinson (1981b) established the consistency of a quasi-likelihood estimator based on (5.2) under the assumption of asymptotic stationarity as reviewed in Definition 3.2. In this chapter we study the asymptotic behaviour of the quasi-likelihood estimator for modulated processes that we introduced in Chapter 5. The class of processes for which we prove consistency is that of modulated processes with a significant correlation contribution, which extends the results of Dunsmuir and Robinson (1981b). We demonstrate that the frequency domain estimator $\hat{\theta}_M^{(N)}$ from (5.8), which for simplicity we denote $\hat{\theta}^{(N)}$ in this chapter, is a consistent estimator of the latent process parameter vector in the case of a univariate modulated process with a significant correlation contribution. This result is stated in Theorem 4. Extension to the class of bivariate modulated processes introduced in Section 3.5 follows directly. We then show that the estimator $\hat{\theta}^{(N)}$ converges with the classical rate of $\mathscr{O}(N^{-\frac{1}{2}})$, where N is the sample size of the time series. This is the result of Theorem 5. The work is this chapter is novel and published in Guillaumin et al. (2017).

6.1 Assumptions and lemmas

To establish consistency, we shall view $l_M(\theta)$, the quasi-likelihood introduced in Section 5.3, as a random function defined over the parameter set Θ. We shall go through the usual following steps (Taniguchi, 1979) to determine the properties of $\hat{\theta}^{(N)}$,

1. Prove that the expectation of the quasi-likelihood $l_M(\theta)$, as a function, admits a unique minimum which is the true parameter vector, for large enough sample sizes. This is the result of Lemma 5.

2. Prove that the variance of the quasi-likelihood function decreases uniformly towards zero over the parameter set Θ. This is a consequence of Lemma 7.

3. Use the two above results to prove that the minimum of the quasi-likelihood behaves "similarly" to the minimum of the expectation of the likelihood function and therefore converges in probability to the true parameter vector. This is the result of Theorem 4.

These steps are standard, however result 1 is not sufficient to guarantee our result, as the expectation of the quasi-likelihood function does not converges to a fixed function. For that reason we will need an additional result, given by Lemma 6.

To guarantee consistency we require the standard Whittle assumptions to be satisfied:

1. The parameter set $\Theta \subset \mathbb{R}^d$ is compact with a non-null interior, and the true parameter θ lies in the interior of Θ.

2. For all $\theta \in \Theta$, we have $\sum_{\tau \in \mathbb{N}} |c_X(\tau; \theta)| < \infty$ (short memory). The spectral densities are continuous with respect to both variables ω, θ, and bounded below and above by two positive constants.

3. The spectral densities are continuously differentiable with respect to both variables ω and θ. By continuity on a compact set the derivatives with respect to ω are bounded above, independently of θ.

4. The process $\widetilde{X}_t$ is a modulated process with a significant correlation contribution. We recall that this implies the existence of a finite subset $\Gamma \subset \mathbb{N}$ such that the mapping $\theta \mapsto \{c_X(\tau) : \tau \in \Gamma\}$ is one-to-one. We also assume that the modulating sequence $\{g_t\}$ is bounded above in absolute value by some finite constant $g_{\max} > 0$.

We start with the following two lemmas which yield uniform bounds of the expected periodogram and its derivative.

Lemma 3 (Boundedness of the expected periodogram). *For all $\theta \in \Theta$ and $N \in \mathbb{N}$, the expected periodogram $\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta)$ is bounded below (by a positive real number) and above independently of N and θ. We denote these bounds $\overline{S}_{\widetilde{X},\min}$ and $\overline{S}_{\widetilde{X},\max}$ respectively.*

Proof. We denote $S_{X,\max} = \max_{\theta,\omega} S_X(\omega; \theta)$ and $S_{X,\min} = \min_{\theta,\omega} S_X(\omega; \theta)$.

1. We first show the existence of the upper bound. According to Proposition 6 the expected periodogram can be expressed, for $\omega \in [-\pi, \pi]$, $\theta \in \Theta$ and $N \in \mathbb{N}$, by

$$\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) = 2\pi \int_{-\pi}^{\pi} S_X(\omega - \lambda; \theta) S_g^{(N)}(\lambda) d\lambda.$$

Therefore,

$$\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) \leq 2\pi S_{X,\max} \int_{-\pi}^{\pi} S_g^{(N)}(\lambda) d\lambda = S_{X,\max} \frac{1}{N} \sum_{t=0}^{N-1} |g_t|^2,$$

by Parseval equality, and finally,

$$\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) \leq g_{\max}^2 S_{X,\max},$$

and this by assumption is finite.

2. Similarly, we show the existence of a lower bound. According to the assumption of a modulated process with a significant correlation contribution, there exists a non-negative integer $\tau \in \Gamma$ and a positive real number α_τ such that for N large enough, $c_g^{(N)}(\tau) \geq \alpha_\tau$. Then,

$$\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) \geq 2\pi S_{X,\min} \int_{-\pi}^{\pi} S_g^{(N)}(\lambda) d\lambda = S_{X,\min} \frac{1}{N} \sum_{t=0}^{N-1} |g_t|^2$$

$$\geq S_{X,\min} \frac{1}{N} \sqrt{\sum_{t=0}^{N-\tau-1} |g_t|^2} \sqrt{\sum_{t=\tau}^{N-1} |g_t|^2}$$

$$\geq \quad S_{X,\min} \frac{1}{N} \left| \sum_{t=0}^{N-\tau-1} g_t^* g_{t+\tau} \right|,$$

by Cauchy-Schwartz inequality. Hence we get for N large enough $\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta) \geq \alpha_\tau S_{X,\min} > 0$. This proves the stated result.

$\square$

Lemma 4 (Boundedness of the derivative of the expected periodogram). *The derivative of the expected periodogram with respect to ω exists and is bounded in absolute value independently of θ and N.*

Proof. We have for all $\omega \in [-\pi, \pi]$, $\theta \in \Theta$ and $N \in \mathbb{N}$ that the form of $\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta)$ is given by

$$\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta) = 2\pi \int_{-\pi}^{\pi} S_X(\omega - \lambda; \theta) S_g^{(N)}(\lambda) d\lambda.$$

We obtain (where the change of order of differentiation and integration is a consequence of the differentiability of the functions $\omega \to S_X(\omega;\theta)$ and the fact that the spectral densities are bounded above),

$$\left| \frac{\partial \overline{S}_{\widetilde{X}}^{(N)}}{\partial \omega}(\omega;\theta) \right| = 2\pi \left| \int_{-\pi}^{\pi} \frac{\partial S_X}{\partial \omega}(\omega - \lambda; \theta) S_g^{(N)}(\lambda) d\lambda \right|$$

$$\leq 2\pi \max_{\omega,\theta} \left\{ \left| \frac{\partial S_X}{\partial \omega}(\omega;\theta) \right| \right\} \int_{-\pi}^{\pi} S_g^{(N)}(\lambda) d\lambda$$

$$\leq g_{\max}^2 \max_{\omega,\theta} \left\{ \left| \frac{\partial S_X}{\partial \omega}(\omega;\theta) \right| \right\},$$

which concludes the proof. $\square$

In analogue to Taniguchi (1979) for stationary processes, we introduce the following quantity,

$$D^{(N)}(\gamma, f) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma) + \frac{f(\omega)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma)} \right\},$$

for all positive integer N, $\gamma \in \Theta$ and non-negative real-valued function f defined on

Ω_N. We also define

$$T^{(N)}(f) = \arg\min_{\gamma \in \Theta} D^{(N)}(\gamma, f).$$

This minimum for fixed f and N is well defined since the set Θ is compact and since the function $\gamma \mapsto D^{(N)}(\gamma, f)$ is continuous. However in cases where the minimum is reached not uniquely but at multiple parameter values, $T^{(N)}(f)$ will denote any of these values, chosen arbitrarily. Note that, by the definition of our frequency domain estimator, we have $\hat{\theta}^{(N)} = T^{(N)}\left(\hat{S}_{\widetilde{X}}^{(N)}(\cdot)\right)$. We derive three lemmas that will be required in proving Theorem 4 which establishes consistency.

Lemma 5. *We have, for N large enough, $T^{(N)}(\overline{S}_{\widetilde{X}}^{(N)}(\cdot; \theta)) = \theta$, uniquely.*

Proof. We will use repeatedly the fact that the function $x \to x - \log x$, defined on the set of positive real numbers, admits a global minimum for $x = 1$ where it takes value 1. It is an increasing function on $(1, \infty)$ and decreasing on $(0, 1)$. This is easily seen by studying the derivative. Now let N be a natural integer. We have for all $\gamma \in \Theta$

$$
\begin{aligned}
D(\gamma, \overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta)) &= \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \overline{S}_{\widetilde{X}}^{(N)}(\omega; \gamma) + \frac{\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega; \gamma)} \right\} \\
&= \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) + \overbrace{\frac{\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega; \gamma)} - \log \left[\frac{\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega; \gamma)} \right]}^{\geq 1} \right\} \\
&\geq \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) + 1 \right\},
\end{aligned}
$$

where we have an equality if and only if $\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) = \overline{S}_{\widetilde{X}}^{(N)}(\omega; \gamma)$ for all $\omega \in \Omega_N$, which for N large enough is equivalent to $\gamma = \theta$ according to Proposition 8. $\square$

This shows that for all N large enough, the function $\gamma \to D\left(\gamma, \overline{S}_{\widetilde{X}}^{(N)}(\cdot; \theta)\right)$ reaches a global minimum at the true parameter vector θ. However because $\overline{S}_{\widetilde{X}}^{(N)}(\cdot; \theta)$ is changing with N and is not expected to converge to a given function, we need the following stronger result.

Lemma 6. *If $\{\gamma_N\}_{N \in \mathbb{N}} \in \Theta^{\mathbb{N}}$ is a sequence of parameter vectors such that $D\left(\gamma_N, \overline{S}_{\widetilde{X}}^{(N)}(\cdot;\theta)\right) - D\left(\theta, \overline{S}_{\widetilde{X}}^{(N)}(\cdot;\theta)\right)$ converges to zero when N goes to infinity, then γ_N converges to θ.*

Proof. We prove this in three steps.

1. We have for a positive integer N,

$$
\begin{aligned}
D\left(\gamma_N, \overline{S}_{\widetilde{X}}^{(N)}(\cdot;\theta)\right) \quad &- \quad D\left(\theta, \overline{S}_{\widetilde{X}}^{(N)}(\cdot;\theta)\right) \qquad\qquad (6.1)\\
&= \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \frac{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma_N)} - \log \frac{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma_N)} - 1 \right\}.
\end{aligned}
$$

   Assume this converges to zero as N goes to infinity. For any non-negative integer τ smaller than N we can write,

$$
\overline{c}_{\widetilde{X}}^{(N)}(\tau;\gamma_N) - \overline{c}_{\widetilde{X}}^{(N)}(\tau;\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma_N) - \overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta) \right) e^{i\omega\tau} d\omega,
$$

   so we have the following bound,

$$
\left| \overline{c}_{\widetilde{X}}^{(N)}(\tau;\gamma_N) - \overline{c}_{\widetilde{X}}^{(N)}(\tau;\theta) \right| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma_N) - \overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta) \right| d\omega. \quad (6.2)
$$

2. Now we assume, with the intent to arrive at a contradiction, that this quantity does not converge to zero. Then there exists an increasing function $\phi(N)$, defined on the set of non-negative integers and taking values in the set of non-negative integers and $\varepsilon > 0$ such that

$$
\left| \overline{c}_{\widetilde{X}}^{(\phi(N))}(\tau;\gamma_{\phi(N)}) - \overline{c}_{\widetilde{X}}^{(\phi(N))}(\tau;\theta) \right| \geq \varepsilon, \ \forall N \in \mathbb{N}. \qquad (6.3)
$$

   Fix $N \in \mathbb{N}$. Denote M the upper bound (independent of N) of the integrand in (6.2) using lemma 3. Let $B_{\phi(N)} \subset [-\pi, \pi]$ be the inverse image of $[\varepsilon/2, \infty)$ by the function $\omega \mapsto \left| \overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega;\gamma_{\phi(N)}) - \overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega;\theta) \right|$. Let $\lambda_{\phi(N)}$ be the

Lebesgue measure of the Borel set $B_{\phi(N)}$. We have,

$$\varepsilon \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)}) - \overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \theta) \right| d\omega \leq \frac{\lambda_{\phi(N)}}{2\pi} M + \frac{2\pi - \lambda_{\phi(N)}}{2\pi} \frac{\varepsilon}{2},$$

and therefore

$$\lambda_{\phi(N)} \geq \frac{\pi\varepsilon}{M - \frac{\varepsilon}{2}}. \tag{6.4}$$

Since $B_{\phi(N)}$ is defined such that,

$$\left| \overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)}) - \overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \theta) \right| \geq \frac{\varepsilon}{2}, \ \forall \omega \in B_{\phi(N)}, \tag{6.5}$$

it follows that, dividing each side of (6.5) by $\overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)})$,

$$\left| \frac{\overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \theta)}{\overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)})} - 1 \right| \geq \frac{\varepsilon}{2\overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)})} \geq \frac{\varepsilon}{2\overline{S}_{\widetilde{X},\max}}, \ \forall \omega \in B_{\phi(N)}.$$

We therefore have that for all $\omega \in B_{\phi(N)}$, $\left| \overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \theta) / \overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)}) - 1 \right|$ is bounded below by $c\varepsilon$, where $c = 1/(2\overline{S}_{\widetilde{X},\max})$ is a positive constant independent of N. Denote

$$
\begin{aligned}
b &: \quad x \to x - \log x - 1, x > 0, \\
\overline{b}^{(\phi(N))} &: \quad \omega \mapsto b\left(\frac{\overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \theta)}{\overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)})} \right), \quad \omega \in [-\pi, \pi].
\end{aligned}
$$

For all $\omega \in B_{\phi(N)}$, $\overline{b}^{(\phi(N))}(\omega)$ is bounded below by $d = \min(b(1 + c\varepsilon), b(1 - c\varepsilon))$ (where $d > 0$ is a constant that depends on ε but not on N) because of the properties of the function $b(x)$ which we recalled at the beginning of the proof of lemma 5. The function $b(x)$ has a bounded derivative on any interval of the form $[a_1, a_2]$ where $0 < a_1 < a_2 < \infty$. Since

$$\frac{\overline{S}_{\widetilde{X},\max}}{\overline{S}_{\widetilde{X},\min}} \geq \frac{\overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \theta)}{\overline{S}_{\widetilde{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)})} \geq \frac{\overline{S}_{\widetilde{X},\min}}{\overline{S}_{\widetilde{X},\max}} > 0,$$

and using Lemma 3 and Lemma 4, the function $\bar{b}^{(\phi(N))}(\omega)$ has a bounded derivative. We denote the corresponding bound $l_{\max}$, which is independent of N.

Recalling that $\lambda_{\phi(N)}$ is the measure of $B_{\phi(N)}$, there exist $T = \lfloor \frac{N\lambda_{\phi(N)}}{4\pi} \rfloor$ increasing elements $v_1, \cdots, v_T \in B_{\phi(N)}$ such that $v_{i+1} - v_i \geq \frac{4\pi}{N}, i = 1, \cdots, T-1$. Then there exist $T-1$ Fourier frequencies $v'_1, \cdots, v'_{T-1}$, such that $v_i < v'_i < v_{i+1}$. Then we have,

$$\left| \sum_{i=1}^{T-1} \bar{b}^{(\phi(N))}(v'_i) - \bar{b}^{(\phi(N))}(v_i) \right| \leq \sum_{i=1}^{T-1} \left| \bar{b}^{(\phi(N))}(v'_i) - \bar{b}^{(\phi(N))}(v_i) \right|$$

$$\leq \sum_{i=1}^{T-1} (v'_i - v_i) l_{\max} \leq 2\pi l_{\max},$$

which implies

$$\sum_{i=1}^{T-1} \bar{b}^{(\phi(N))}(v'_i) \geq \sum_{i=1}^{T-1} \bar{b}^{(\phi(N))}(v_i) - 2\pi l_{\max}$$

$$\geq (T-1)d - 2\pi l_{\max}.$$

Because T is of order N, we conclude that (6.1) cannot converge to zero. We arrive at a contradiction.※

By this contradiction we obtain that for all integer τ, $\bar{c}_{\tilde{X}}^{(N)}(\tau; \gamma_N) - \bar{c}_{\tilde{X}}^{(N)}(\tau; \theta)$ converges to zero when N goes to infinity.

3. In particular for $\tau \in \Gamma$, if we denote $\alpha_\tau = \liminf_{N\to\infty} \left| c_g^{(N)}(\tau) \right| > 0$, we have for N large enough,

$$\left| \bar{c}_{\tilde{X}}^{(N)}(\tau; \gamma_N) - \bar{c}_{\tilde{X}}^{(N)}(\tau; \theta) \right| = \left| c_g^{(N)}(\tau) \left(c_X(\tau; \gamma_N) - c_X(\tau; \theta) \right) \right|$$

$$\geq \alpha_\tau \left| c_X(\tau; \gamma_N) - c_X(\tau; \theta) \right|,$$

so that (with the assumption of significant correlation contribution) $|c_X(\tau; \gamma_N) - c_X(\tau; \theta)|$ converges to zero as N tends to infinity. Because of the compacity of Θ, and using the fact that the function $\theta \mapsto \{c_X(\tau) : \tau \in \Gamma\}$ is

one-to-one and continuous, this yields the stated lemma.

This concludes the proof. □

We now show that the functions $D\left(\gamma, \overline{S}_{\widetilde{X}}^{(N)}(\cdot; \theta)\right)$ and $D\left(\gamma, \hat{S}_{\widetilde{X}}^{(N)}(\cdot)\right)$, defined on Θ, behave asymptotically *in the same way*. For this, we first need the following lemma where we bound the asymptotic variance of some linear functionals of the periodogram.

Lemma 7. *Let* $\left\{a^{(N)}(\omega) : \omega \in [-\pi, \pi)\right\}_{N \in \mathbb{N}}$ *be a family of real-valued functions, uniformly bounded by a positive real number. Then it follows,*

$$\mathrm{var}\left\{\frac{1}{N} \sum_{\omega \in \Omega_N} a^{(N)}(\omega)\hat{S}_{\widetilde{X}}^{(N)}(\omega)\right\} = \mathcal{O}\left(\frac{1}{N}\right). \tag{6.6}$$

Proof. Let $a_{\max}$ be a finite positive constant such that $|a^{(N)}(\omega)| \leq a_{\max}, \forall \omega \in [-\pi, \pi), \forall N \in \mathbb{N}$. We start by looking at the covariance matrix of the Fourier transform. We shall denote the Fourier transform, for a fixed positive integer N,

$$J_{\widetilde{X}}^{(N)}(\omega) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \widetilde{X}_t e^{-i\omega t} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} g_t X_t e^{-i\omega t}, \quad \omega \in \Omega_N.$$

Since the expectation of the latent process is assumed to be zero, the same holds for the Fourier transform by the linearity of the Fourier transform. As g_t is deterministic and from the linear equation above we see that the covariance matrix elements can be expressed in the following way:

$$\mathrm{cov}\left\{J_{\widetilde{X}}^{(N)}(\omega), J_{\widetilde{X}}^{(N)}(\omega')\right\} = \frac{1}{N}\left(G_{\omega}^{(N)}\right)^H C_X^{(N)}(\theta) G_{\omega'}^{(N)}, \quad \omega, \omega' \in \Omega_N, \tag{6.7}$$

where subscript H denotes the Hermitian transpose, $C_X^{(N)}(\theta)$ denotes the finite theoretical autocovariance matrix of the latent process (i.e. with elements $c_X(i-j; \theta)$, $i, j = 0, \cdots, N-1$), and $G_{\omega}^{(N)}$ is the vector $[g_t e^{i\omega t} : t = 0, \cdots, N-1]^T$. Using Isserlis' theorem (Isserlis, 1918) and the assumption of Gaussianity of the latent process (which in turns implies the Gaussianity of the Fourier transform of the modulated process), the covariances of the periodogram are related to the covariances of the

Fourier transform according to the simple following relation

$$\text{cov}\left\{\hat{S}_{\tilde{X}}^{(N)}(\omega),\hat{S}_{\tilde{X}}^{(N)}(\omega')\right\} = \left|\text{cov}\left\{J_{\tilde{X}}^{(N)}(\omega),J_{\tilde{X}}^{(N)}(\omega')\right\}\right|^{2}. \tag{6.8}$$

This can be written as

$$\begin{aligned}
\text{cov}\left\{\hat{S}_{\tilde{X}}^{(N)}(\omega),\hat{S}_{\tilde{X}}^{(N)}(\omega')\right\} &= \frac{1}{N^2}\left(G_{\omega}^{(N)}\right)^{H}C_{X}^{(N)}(\theta)G_{\omega'}^{(N)}\left[\left(G_{\omega}^{(N)}\right)^{H}C_{X}^{(N)}(\theta)G_{\omega'}^{(N)}\right]^{H} \\
&= \frac{1}{N^2}\left(G_{\omega}^{(N)}\right)^{H}C_{X}^{(N)}(\theta)G_{\omega'}^{(N)}G_{\omega'}^{(N)H}C_{X}^{(N)}(\theta)^{H}G_{\omega}^{(N)}.
\end{aligned}$$

We then have

$$\begin{aligned}
\text{var}&\left\{\frac{1}{N}\sum_{\omega\in\Omega_N}a^{(N)}(\omega)\hat{S}_{\tilde{X}}^{(N)}(\omega)\right\} = \\
&\frac{1}{N^2}\sum_{\omega\in\Omega_N}\sum_{\omega'\in\Omega_N}a^{(N)}(\omega)a^{(N)}(\omega')\frac{1}{N^2}\left(G_{\omega}^{(N)}\right)^{H}C_{X}^{(N)}(\theta)G_{\omega'}^{(N)}G_{\omega'}^{(N)H}C_{X}^{(N)}(\theta)^{H}G_{\omega}^{(N)} \\
\leq\ &\frac{a_{\max}^2}{N^4}\sum_{\omega\in\Omega_N}\sum_{\omega'\in\Omega_N}\left(G_{\omega}^{(N)}\right)^{H}C_{X}^{(N)}(\theta)G_{\omega'}^{(N)}G_{\omega'}^{(N)H}C_{X}^{(N)}(\theta)^{H}G_{\omega}^{(N)} \\
=\ &\frac{a_{\max}^2}{N^4}\sum_{\omega\in\Omega_N}\left(G_{\omega}^{(N)}\right)^{H}C_{X}^{(N)}(\theta)\sum_{\omega'\in\Omega_N}\left\{G_{\omega'}^{(N)}G_{\omega'}^{(N)H}\right\}C_{X}^{(N)}(\theta)^{H}G_{\omega}^{(N)},
\end{aligned}$$

where the first inequality is legitimate as the covariances of the periodogram are positive real-valued numbers (see (6.8)), and where the last equality is obtained after factorizing. Now we use the fact that

$$\sum_{\omega'\in\Omega_N}\left\{G_{\omega'}^{(N)}G_{\omega'}^{(N)H}\right\} = N\text{diag}(g_0^2,\cdots,g_{N-1}^2). \tag{6.9}$$

Indeed, the (t_1,t_2)-th term of the left hand side of (6.9) is given by

$$\sum_{\omega'\in\Omega_N}g_{t_1}g_{t_2}e^{i\omega'(t_1-t_2)} = \sum_{k=0}^{N-1}g_{t_1}g_{t_2}e^{\frac{i2k\pi(t_1-t_2)}{N}} = g_{t_1}g_{t_2}\sum_{k=0}^{N-1}e^{\frac{i2k\pi(t_1-t_2)}{N}}, \tag{6.10}$$

where we recognize the finite sum of the geometric sequence of term $e^{\frac{i2\pi(t_1-t_2)}{N}}$,

which is N if $t_1 = t_2$, and otherwise,

$$\sum_{k=0}^{N-1} e^{\frac{i2k\pi(t_1-t_2)}{N}} = \sum_{k=0}^{N-1} \left(e^{\frac{i2\pi(t_1-t_2)}{N}} \right)^k = \frac{1 - \left(e^{\frac{i2\pi(t_1-t_2)}{N}} \right)^N}{1 - e^{\frac{i2\pi(t_1-t_2)}{N}}} = 0. \qquad (6.11)$$

Therefore

$$\text{var}\left\{ \frac{1}{N} \sum_{\omega \in \Omega_N} a(\omega)\hat{S}_{\tilde{X}}^{(N)}(\omega) \right\} \leq \frac{a_{\max}^2}{N^3} \sum_{\omega \in \Omega_N} \left(G_\omega^{(N)} \right)^H C_X^{(N)}(\theta)\Gamma_g C_X^{(N)}(\theta)^H G_\omega^{(N)},$$

where $\Gamma_g = \text{diag}(g_0^2, \cdots, g_{N-1}^2)$ is the diagonal matrix with elements $g_0^2, \ldots, g_{N-1}^2$. Thus,

$$\text{var}\left\{ \frac{1}{N} \sum_{\omega \in \Omega_N} a(\omega)\hat{S}_{\tilde{X}}^{(N)}(\omega) \right\} \leq \frac{a_{\max}^2 g_{\max}^2}{N^3} \sum_{\omega \in \Omega_N} \left(G_\omega^{(N)} \right)^H C_X^{(N)}(\theta) C_X^{(N)}(\theta)^H G_\omega^{(N)}.$$

Therefore we now have

$$\text{var}\left\{ \frac{1}{N} \sum_{\omega \in \Omega_N} a(\omega)\hat{S}_{\tilde{X}}^{(N)}(\omega) \right\} \leq \frac{a_{\max}^2 g_{\max}^2}{N^3} \sum_{\omega \in \Omega_N} \left\| C_X^{(N)}(\theta)^H G_\omega^{(N)} \right\|_2^2, \qquad (6.12)$$

where $\| \cdot \|_2$ denotes the Euclidean norm on $\mathbb{C}^N$. For all $\mathbf{U} \in \mathbb{C}^N$, the matrix $C_X^{(N)}(\theta)$ is Hermitian, so it can be written PDP^H where D is a diagonal matrix and where P is unitary, so that,

$$\left\| C_X^{(N)}(\theta)\mathbf{U} \right\|_2 \leq \|\mathbf{U}\|_2 \max_{\eta \in \text{sp}\left(C_X^{(N)}(\theta) \right)} |\eta|, \qquad (6.13)$$

where $\text{sp}\left(C_X^{(N)}(\theta) \right)$ is the set of eigenvalues of $C_X^{(N)}(\theta)$. Furthermore we have from Horn and Johnson (1985, p. 394) that, recalling that the spectral density $S_X(\omega; \theta)$ is assumed to be continuous in ω,

$$\max_{\eta \in \text{sp}\left(C_X^{(N)}(\theta) \right)} |\eta| = \max_{\mathbf{U} \in \mathbb{C}^n} \left\{ \frac{\mathbf{U}^H C_X^{(N)}(\theta)\mathbf{U}}{\mathbf{U}^H \mathbf{U}} \right\} \leq S_{X,\max}. \qquad (6.14)$$

Combining (6.12)-(6.14) and replacing $\mathbf{U}$ by $G_\omega^{(N)}$,

$$\text{var}\left\{\frac{1}{N}\sum_{\omega\in\Omega_N}a(\omega)\hat{S}_{\widetilde{X}}^{(N)}(\omega)\right\} \leq \frac{\left(S_{X,\max}\,a_{\max}\,g_{\max}^2\right)^2}{N}, \tag{6.15}$$

as $\left\|G_\omega^{(N)}\right\|_2 \leq g_{\max}\sqrt{N}$. Remembering that $\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta) = \text{E}\left\{\hat{S}_{\widetilde{X}}^{(N)}(\omega);\theta\right\}$, we thus have

$$\sum_{\omega\in\Omega_N}a^{(N)}(\omega)\hat{S}_{\widetilde{X}}^{(N)}(\omega) = \sum_{\omega\in\Omega_N}a^{(N)}(\omega)\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta) + \mathscr{O}_P\left(\frac{1}{\sqrt{N}}\right).$$

$\square$

6.2 Consistency and convergence rate

We are now able to state a consistency theorem for our estimator $\hat{\theta}^{(N)}$.

Theorem 4 (Consistency of the frequency domain estimator). *We have $\hat{\theta}^{(N)} \xrightarrow{P} \theta$ in probability.*

Proof. The proof is based on Taniguchi (1979). Denote, for any $\gamma\in\Theta$, $\overline{h}^{(N)}(\gamma;\theta) = D\left(\gamma,\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta)\right)$ and $\hat{h}^{(N)}(\gamma) = D\left(\gamma,\hat{S}_{\widetilde{X}}^{(N)}(\omega)\right)$. We have,

$$\begin{aligned}
\overline{h}^{(N)}(\gamma;\theta) - \hat{h}^{(N)}(\gamma) &= \frac{1}{N}\sum_{\omega\in\Omega_N}\left\{\log\overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma) + \frac{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma)}\right.\\
&\quad - \left.\left(\log\overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma) + \frac{\hat{S}_{\widetilde{X}}^{(N)}(\omega)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma)}\right)\right\}\\
&= \frac{1}{N}\sum_{\omega\in\Omega_N}\frac{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\theta) - \hat{S}_{\widetilde{X}}^{(N)}(\omega)}{\overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma)}.
\end{aligned}$$

We have shown in Lemma 3 that $\overline{S}_{\widetilde{X}}^{(N)}(\omega;\gamma)$ is bounded below in both variables ω and γ by a positive real number, independently of N. Therefore, making use of Lemma 7 we have

$$\sup_{\gamma\in\Theta}\left|\overline{h}^{(N)}(\gamma;\theta) - \hat{h}^{(N)}(\gamma)\right| \xrightarrow{P} 0, \quad (N\to\infty), \tag{6.16}$$

where the letter P indicates that the convergence is in probability, as the difference is of stochastic order $\mathscr{O}\left(N^{-\frac{1}{2}}\right)$. In particular (6.16) implies that

$$\left|\min_{\gamma}\overline{h}^{(N)}(\gamma;\theta) - \min_{\gamma}\hat{h}^{(N)}(\gamma)\right| \leq \sup_{\gamma\in\Theta}\left|\overline{h}^{(N)}(\gamma;\theta) - \hat{h}^{(N)}(\gamma)\right| \xrightarrow{P} 0$$

i.e.

$$\left|\overline{h}^{(N)}\left(T^{(N)}(\overline{S}^{(N)}_{\widetilde{X}}(\omega;\theta));\theta\right) - \hat{h}^{(N)}\left(T^{(N)}(\hat{S}^{(N)}_{\widetilde{X}}(\omega))\right)\right| \xrightarrow{P} 0. \qquad (6.17)$$

Relation (6.16) also implies that

$$\left|\overline{h}^{(N)}\left(T^{(N)}(\hat{S}^{(N)}_{\widetilde{X}}(\omega));\theta\right) - \hat{h}^{(N)}\left(T^{(N)}(\hat{S}^{(N)}_{\widetilde{X}}(\omega))\right)\right| \xrightarrow{P} 0, \qquad (6.18)$$

so that using the triangle inequality with (6.17) and (6.18) we get,

$$\left|\overline{h}^{(N)}\left(T^{(N)}(\hat{S}^{(N)}_{\widetilde{X}}(\omega));\theta\right) - \overline{h}^{(N)}\left(T^{(N)}(\overline{S}^{(N)}_{\widetilde{X}}(\omega;\theta));\theta\right)\right| \xrightarrow{P} 0.$$

We then obtain the stated theorem making use of Lemma 6. $\qquad\square$

We now study the convergence rate of our frequency domain estimator. For this we first need the following two lemmas. Although the Hessian matrix of the likelihood is not expected to converge for modulated processes with a significant correlation contribution, we can show that its norm is bounded below by a positive real number. For this we need to strengthen the assumption of significant correlation contribution. Assuming that the spectral densities of the latent process are twice continuously differentiable with respect to θ, we assume that the Jacobian determinant of the mapping $\theta \mapsto [c_X(\tau;\theta) : \tau \in \Gamma]^T$ taken at the true parameter value θ, i.e. the determinant of the matrix with elements $\frac{\partial c_X(\tau_i;\theta)}{\partial \theta_j}$ (with $\Gamma = \{\tau_1, \tau_2, \cdots, \tau_d\}$ here), exists and is non-zero.

Lemma 8. *Let* $\mathbf{U}_1, \cdots, \mathbf{U}_d$ *a family of vectors of* $\mathbb{R}^d$ *with rank* d. *Let* $\alpha_1, \cdots, \alpha_d$ *be positive real numbers. There exists a positive constant* $C > 0$ *such that for all*

$\mathbf{V} \in \mathbb{R}^d$,

$$\sum_{i=1}^d \alpha_i^2 \left(\mathbf{U}_i^T \mathbf{V}\right)^2 \geq C \|\mathbf{V}\|_2^2,$$

where $\|\cdot\|_2$ denotes the Euclidean norm on $\mathbb{R}^N$.

Proof. We first show that the proposition is true for all vectors in $\mathscr{C} = \left\{\mathbf{V} \in \mathbb{R}^d : \|\mathbf{V}\|_2 = 1\right\}$, which is compact. The function $\mathscr{S} : \mathbf{V} \mapsto \sum_{i=1}^d \alpha_i^2 \left(\mathbf{U}_i^T \mathbf{V}\right)^2$ is continuous, so the image of $\mathscr{C}$ by $\mathscr{S}$ is compact. Since $\mathscr{S}$ takes non-negative values, the image of $\mathscr{C}$ by $\mathscr{S}$ either contains zero or there exists a constant $C > 0$ such that it is bounded below by C. The image of $\mathscr{C}$ by $\mathscr{S}$ cannot contain zero, as otherwise there would be a vector of $\mathbb{R}^d$ with norm 1 whose scalar product with vectors $U_i, i = 1, \cdots, d$ is zero, which is impossible as we have assumed that the family $U_1, \cdots, U_d$ has rank d. Therefore there exists a constant $C > 0$ such that,

$$\sum_{i=1}^d \alpha_i^2 \left(\mathbf{U}_i^T \mathbf{V}\right)^2 \geq C, \ \forall \mathbf{V} \in \mathscr{C}.$$

Now in general, if $\mathbf{V}$ is any non-zero vector in $\mathbb{R}^d$, we have, using the result we have derived for vectors of $\mathscr{C}$,

$$\begin{aligned}
\sum_{i=1}^d \alpha_i^2 \left(\mathbf{U}_i^T \mathbf{V}\right)^2 &= \|\mathbf{V}\|_2^2 \sum_{i=1}^d \alpha_i^2 \left(\mathbf{U}_i^T \frac{\mathbf{V}}{\|\mathbf{V}\|_2}\right)^2 \\
&\geq \|\mathbf{V}\|_2^2 C.
\end{aligned}$$

If $\mathbf{V} = 0$ the result is obvious. This concludes the proof in the general case. $\square$

Lemma 9. *We have, for $i = 1, \ldots, d$,*

$$\frac{\partial l_M}{\partial \theta_i}(\theta) = \mathcal{O}_P\left(\frac{1}{\sqrt{N}}\right).$$

The Hessian matrix of the function $l_M(\theta)$ satisfies

$$H(\theta) = \mathscr{I}(\theta) + \mathcal{O}_P\left(\frac{1}{\sqrt{N}}\right),$$

where the matrix norm of $\mathscr{I}(\theta)$ is bounded below by a positive value, independently

of N.

Proof. 1. Direct calculations give that the score function takes the form, for $i = 1, \cdots, d$,

$$\frac{\partial l_M}{\partial \theta_i}(\theta) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \frac{\frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\omega; \theta)}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) \right)^2} \left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) - \hat{S}_{\tilde{X}}^{(N)}(\omega) \right) \right\}.$$

Since by definition $\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) = \mathrm{E}\left\{ \hat{S}_{\tilde{X}}^{(N)}(\omega); \theta \right\}$, the expectation of the score function at the true parameter vector is zero. Applying Lemma 7, and the fact that $\frac{\frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\omega; \theta)}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) \right)^2}$ is bounded above in absolute value independently of N (as a direct consequence of Lemma 3 and Lemma 4), we get the stated result.

2. Again by direct calculation we obtain the following Hessian matrix:

$$\frac{\partial^2 l_M^{(N)}}{\partial \theta_i \partial \theta_j}(\theta) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \left[\frac{\frac{\partial^2 \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i \partial \theta_j}(\omega; \theta) \left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) \right)^2}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) \right)^4} \right. \right.$$
$$\left. - 2 \frac{\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\omega; \theta) \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_j}(\omega; \theta)}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) \right)^4} \right]$$
$$\left. \times \left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) - \hat{S}_{\tilde{X}}^{(N)}(\omega) \right) + \frac{\frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\omega; \theta) \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_j}(\omega; \theta)}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) \right)^2} \right\}.$$

The expectation of the Hessian matrix is therefore

$$\mathscr{I}^{(N)}(\theta) = \mathrm{E}\left\{ \frac{\partial^2 l_M^{(N)}}{\partial \theta \partial \theta^T}(\omega; \theta) \right\} = \frac{1}{N} \sum_{\omega \in \Omega_N} \frac{\frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta}(\omega; \theta) \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta^T}(\omega; \theta)}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) \right)^2},$$

where we use the notation $\frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta^T}(\omega; \theta)$ to denote the transpose of the gradient vector $\frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta}(\omega; \theta)$. For any of the $\omega \in \Omega_N$ (to which corresponds a term in

the above sum), and for any vector $\mathbf{U} \in \mathbb{R}^d$,

$$\mathbf{U}^T \frac{\partial \overline{S}_{\widetilde{X}}^{(N)}}{\partial \theta}(\omega; \theta) \frac{\partial \overline{S}_{\widetilde{X}}^{(N)}}{\partial \theta^T}(\omega; \theta)\mathbf{U} = \left| \frac{\partial \overline{S}_{\widetilde{X}}^{(N)}}{\partial \theta^T}(\omega; \theta)\mathbf{U} \right|^2 \geq 0,$$

so that the matrix $\mathscr{I}^{(N)}(\theta)$ is non-negative definite as a sum of non-negative definite matrices. Now to show that the matrix $\mathscr{I}(\theta)$ is positive definite, let $\mathbf{U} = [u_1, \cdots, u_d]^T \in \mathbb{R}^d$ non-zero. We have

$$
\begin{aligned}
\mathbf{U}^T \mathscr{I}^{(N)}(\theta) \, \mathbf{U} &= \frac{1}{N} \sum_{\omega \in \Omega_N} \frac{1}{\left(\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) \right)^2} \mathbf{U}^T \frac{\partial \overline{S}_{\widetilde{X}}^{(N)}}{\partial \theta}(\omega; \theta) \frac{\partial \overline{S}_{\widetilde{X}}^{(N)}}{\partial \theta^T}(\omega; \theta)\mathbf{U} \\
&= \frac{1}{N} \sum_{\omega \in \Omega_N} \frac{1}{\left(\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) \right)^2} \left| \frac{\partial \overline{S}_{\widetilde{X}}^{(N)}}{\partial \theta^T}(\omega; \theta)\mathbf{U} \right|^2 \\
&= \frac{1}{N} \sum_{\omega \in \Omega_N} \frac{1}{\left(\overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) \right)^2} \left(\sum_{i=1}^d \frac{\partial \overline{S}_{\widetilde{X}}^{(N)}}{\partial \theta_i}(\omega; \theta)u_i \right)^2 \\
&\geq \frac{1}{N \left(\sup_{\omega \in \Omega_N} \overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) \right)^2} \sum_{\omega \in \Omega_N} \left(\sum_{i=1}^d \frac{\partial \overline{S}_{\widetilde{X}}^{(N)}}{\partial \theta_i}(\omega; \theta)u_i \right)^2 .
\end{aligned}
$$

Seeing $\left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^d u_i \frac{\partial \overline{S}_{\widetilde{X}}^{(N)}}{\partial \theta_i}(\omega; \theta) \right\}_{\omega \in \Omega_N} \Leftrightarrow \left\{ \sum_{i=1}^d u_i \frac{\partial \overline{c}_{\widetilde{X}}^{(N)}}{\partial \theta_i}(\tau; \theta) \right\}_{\tau=-(N-1), \cdots, N-1}$ as a finite Fourier pair and applying Parseval's equality we obtain that,

$$
\begin{aligned}
\mathbf{U}^T \mathscr{I}^{(N)}(\theta) \, \mathbf{U} &\geq \frac{\sum_{\tau=-(N-1)}^{N-1} \left(\sum_{i=1}^d u_i \frac{\partial \overline{c}_{\widetilde{X}}^{(N)}}{\partial \theta_i}(\tau; \theta) \right)^2}{\left(\sup_{\omega \in \Omega_N} \overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) \right)^2} \\
&= \frac{\sum_{\tau=-(N-1)}^{N-1} \left(\sum_{i=1}^d u_i c_g^{(N)}(\tau) \frac{\partial c_X}{\partial \theta_i}(\tau; \theta) \right)^2}{\left(\sup_{\omega \in \Omega_N} \overline{S}_{\widetilde{X}}^{(N)}(\omega; \theta) \right)^2},
\end{aligned}
$$

by definition of $\overline{c}_{\widetilde{X}}^{(N)}(\tau; \theta)$ and noting that $c_g^{(N)}(\tau)$ does not depend on θ.

Therefore,

$$\mathbf{U}^T \mathscr{I}^{(N)}(\boldsymbol{\theta}) \, \mathbf{U} \geq \frac{1}{\left(\sup_{\omega \in \Omega_N} \overline{S}_{\widetilde{X}}^{(N)}(\omega; \boldsymbol{\theta})\right)^2} \sum_{\tau \in \Gamma} \left(\sum_{i=1}^{d} u_i c_g^{(N)}(\tau) \frac{\partial c_X}{\partial \theta_i}(\tau; \boldsymbol{\theta})\right)^2,$$

as long as N is larger than the greater integer value in Γ. Denote $\alpha_\tau = \liminf_{N \to \infty} \left|c_g^{(N)}(\tau)\right| > 0$ for $\tau \in \Gamma$, we obtain that for N large enough (see (3.8)),

$$\begin{aligned}
\mathbf{U}^T \mathscr{I}^{(N)}(\boldsymbol{\theta}) \, \mathbf{U} \quad &\geq \quad \frac{\sum_{\tau \in \Gamma} c_g^{(N)}(\tau)^2 \left(\sum_{i=1}^{d} u_i \frac{\partial c_X}{\partial \theta_i}(\tau; \boldsymbol{\theta})\right)^2}{\left(\sup_{\omega \in \Omega_N} \overline{S}_{\widetilde{X}}^{(N)}(\omega; \boldsymbol{\theta})\right)^2} \\
&\geq \quad \frac{\sum_{\tau \in \Gamma} \alpha_\tau^2 \left(\sum_{i=1}^{d} u_i \frac{\partial c_X}{\partial \theta_i}(\tau; \boldsymbol{\theta})\right)^2}{\left(\sup_{\omega \in \Omega_N} \overline{S}_{\widetilde{X}}^{(N)}(\omega; \boldsymbol{\theta})\right)^2}.
\end{aligned}$$

Now according to the assumption of significant correlation contribution, the mapping $\boldsymbol{\theta} \mapsto [c_X(\tau) : \tau \in \Gamma]^T$ is one-to-one, so its Jacobian taken at the true parameter vector $\boldsymbol{\theta}$ is non-zero. Therefore the family $\frac{\partial c_X}{\partial \theta}(\tau; \boldsymbol{\theta}) : \tau \in \Gamma$ has rank d and we can apply Lemma 8, i.e. we can conclude that there exists a positive constant $C > 0$ such that,

$$\mathbf{U}^T \mathscr{I}^{(N)}(\boldsymbol{\theta}) \, \mathbf{U} \geq C \|U\|_2^2.$$

This implies that the norm of the expected Hessian matrix is bounded below by a positive real-number. Similarly to the gradient, using Lemma 7, we obtain the stated result for the Hessian.

This concludes the proof. $\qquad\square$

Theorem 5 (Convergence rate). *We have* $\hat{\boldsymbol{\theta}}^{(N)} = \boldsymbol{\theta} + \mathscr{O}_P\left(\frac{1}{\sqrt{N}}\right).$

Proof. We have, by Taylor expansion with Lagrange form of the remainder term,

$$\nabla l_M(\hat{\boldsymbol{\theta}}^{(N)}) = \mathbf{0} = \nabla l_M(\boldsymbol{\theta}) + H(\widetilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}),$$

where $\widetilde{\theta}$ lies between $\hat{\theta}^{(N)}$ and θ. Therefore,

$$\hat{\theta}^{(N)} - \theta = -H(\widetilde{\theta})^{-1}\nabla l_M(\theta).$$

We have shown that $\hat{\theta}^{(N)}$ converges in probability to θ. By continuity of the Hessian, and using the results of Lemma 9, we obtain

$$\hat{\theta}^{(N)} - \theta = -\left[\mathscr{I}(\theta) + \mathscr{O}_P\left(\frac{1}{\sqrt{N}}\right) + o_P(1)\right]^{-1}\mathscr{O}_P\left(\frac{1}{\sqrt{N}}\right) = \mathscr{O}_P\left(\frac{1}{\sqrt{N}}\right).$$

This concludes the proof. □

Chapter 7

Simulation studies

In this chapter we present simulation results that show the usefulness of our class of modulated processes with a significant correlation contribution, and that of the inference procedure described in Chapter 5. In Section 7.1 we first consider some applications to the study of missing observations in stationary stochastic processes. We then consider another application in Section 7.2 where the observed process is a sum of two modulated processes, and the modulation sequences are known. In Section 7.3 we consider two situations where the modulation sequence is known up to a parametric form, but unobserved, according to the models described in Section 5.4. The material in this chapter is novel.

7.1 Missing observations

Authors who have treated the topic of modulated stochastic processes have mainly focused on applying this framework to the analysis of time series with missing observations. We reviewed in Section 3.4 how missing observations in a time series $\{X_t\}$ can be taken into account by viewing the observed time series as the realization of a modulated process where the latent process is the stationary process of interest, and the modulation sequence is taking values one and zero, depending on whether a point is observed or missed, respectively. In Section 3.4, we showed how the class of modulated processes with a significant correlation contribution applies to missing observations and extends the class of asymptotically stationary modulated stochastic processes, allowing for more sophisticated schemes of missing observa-

tions. In this section we present results from simulations in that context, for different schemes of missing data. In Section 7.1.1, we consider an application to the classical problem of periodically missing observations, studied by Parzen (1963). In Section 7.1.2 we consider an application to a randomly missing data scheme. The latter is designed such that the resulting observed process does not fit in the class of asymptotically stationary modulated stochastic processes but is however a modulated stochastic process with a significant correlation contribution.

7.1.1 Periodically missing observations

Let $\{X_t\}$ be a stationary process, modelled by a parameter vector θ belonging to a compact parameter set Θ. Define the modulating sequence g_t to take value one if the value of the process is observed at time t, and 0 if it is missed, as in (3.9). In this section we focus on the case of periodically missing observations studied by Parzen (1963), where g_t is defined to take value one for k consecutive times and value zero for l consecutive times, where k and l are two positive integers. Parzen (1963) showed that the modulated process defined by $\widetilde{X}_t = g_t X_t$ is asymptotically stationary as long as $k > l$, and gave an explicit formulation to the limiting sequence $\gamma(\tau)$ defined in (2.6) that depends on k and l. We consider the situation where the stationary process X_t is a causal AR(2) process defined by the difference equation

$$X_t - 0.9X_{t-1} + 0.1X_{t-2} = \varepsilon_t, \ t \in \mathbb{Z}. \tag{7.1}$$

where $\{\varepsilon_t\}$ is a white noise process with variance unity. The autocovariance function $c_X(\tau)$ of the process defined by equation (7.1) can be obtained numerically for lags $0, 1, \ldots, N-1$ in a recursive way by solving the Yule-Walker equations for lags zero, one and two, and then using the following recursive relation,

$$c_X(\tau+2) = 0.9c_X(\tau+1) - 0.1c_X(\tau), \ \ \tau \geq 1.$$

We simulate a periodically missing pattern as described above with $k = 5$ and with various values for l, ranging from zero to six. For all (k, l) combinations the modulated process satisfies the conditions of having a significant correlation contribution,

with $\Gamma = \{0,1,2\}$. Additionally, we observe that we are able to achieve estimation in the situation where $l > k$ due to the parametric model for the stationary process and our definition of significant correlation contribution. For each (k,l) ($l = 0$ corresponding to observing the whole time series) we simulate 5,000 independent time series, with sizes ranging from 256 to 2,048. We aggregate the obtained estimates and provide RMSEs in Figure 7.1. We note that irrespective of the value of l, the RMSEs seem to decrease to zero as the sample size increases, even for $l > k$. Naturally, it seems that the RMSEs increase with the value of l, with the notable exception of $l = 1$ which performs worse than $l = 2$ and $l = 3$ for all three parameters. This is surprising, but may be a consequence of the shape of the expected periodogram.
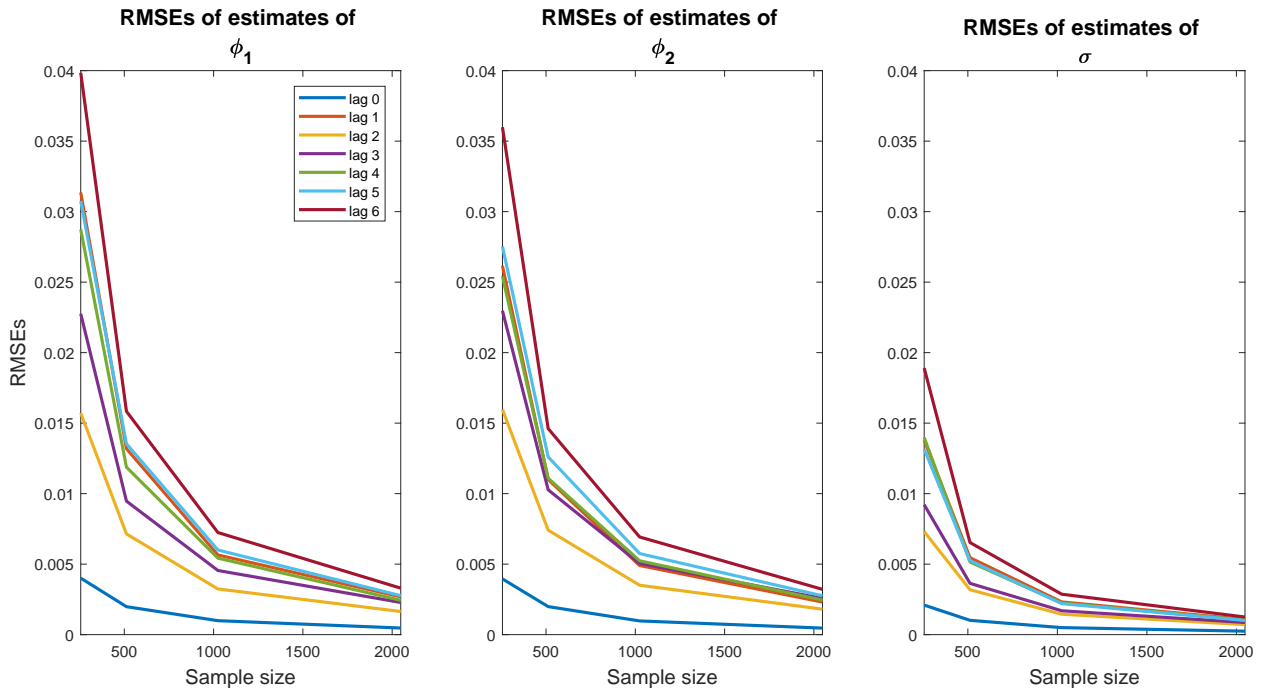
Figure 7.1: Root Mean Square Errors of estimates of the stationary AR(2) process defined by $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = \varepsilon_t$, where $\phi_1 = 0.9$, $\phi_2 = -0.1$, and ε_t is i.i.d. $\mathcal{N}(0,\sigma^2)$, where $\sigma = 1$. The observed process is subject to a (k,l)-periodic missing data pattern, which corresponds to observing k points, missing l, observing k, and so on. We fix $k = 5$ and let l varies from 0 to 6. To observe the improvement with the sample size, we consider sample sizes 256, 512, 1,024 and 2,048. For each combination of the simulation parameters, the RMSE is obtained by aggregating the estimates from 5,000 independently generated time series.

7.1.2 Randomly missing observations

In this section we show that the estimator defined in Definition 13, Section 5.3, can be used for the random missing data scheme 3 of Section 3.4. We simulate a real-valued first-order autoregressive process with parameters $0 \leq a < 1$ and σ according to

$$X_t = aX_{t-1} + \varepsilon_t, \ \ t \geq 1, \tag{7.2}$$

where $X_0 \sim \mathcal{N}\left[0, \sigma^2/(1-a^2)\right]$, and $\{\varepsilon_t\}$ is a Gaussian white noise process with mean zero and variance σ^2. The process $\{X_t\}$ is the latent process of interest. To account for the missing data, we generate a modulated time series $\widetilde{X}_t = g_t X_t$ and assume we only observe the time series $\{\widetilde{X}_t\}$, from which we estimate the parameters of the process $\{X_t\}$. The sequence $\{g_t\}$ takes its values in the set $\{0, 1\}$ and is generated according to

$$g_t \sim \mathscr{B}(p_t), \tag{7.3}$$

where $\mathscr{B}(p)$ represents the Bernoulli distribution with parameter p, and where we set

$$p_t = \frac{1}{2} + \frac{1}{4}\cos\left(\frac{2\pi}{10}t\right). \tag{7.4}$$

The observed modulating sequence $\{g_t\}$, made of zeros and ones, is clearly non-stationary as it does not admit a constant expectation in t. Therefore a spectral representation of the second-order structure of the random modulating sequence $\{g_t\}$, as required in Dunsmuir and Robinson (1981b), does not exist, and the modulated process is not asymptotically stationary. We simulate and estimate such a model for different sample sizes ranging from $N = 128$ to $N = 16384$. For each value of N, we independently simulate 2,000 time series and for each time series we estimate $\{a, \sigma\}$. The outcomes of our simulation study are reported in Table 7.1. The bias, variance and mean square error rapidly decrease with increasing N, while the computational time only increases gradually with N such that the methods are still computationally efficient for long time series.

Table 7.1: Performance of the estimator defined by (5.7) for the missing data problem defined in (7.2)–(7.4). The unknown parameters are set as $a = 0.8$, and $\sigma = 1$. The results are averaged over 2,000 independently generated time series for each sample size N. The average CPU times for the optimization are given in seconds, as performed on a 3.60Ghz Intel i7-4790 processor (4 cores).

Sample size	128	512	1024	2048	4096	8192	16384
Estimate of parameter a							
Bias	-2.08e-02	-4.81e-03	-3.09e-04	-4.17e-04	-2.73e-04	-4.94e-04	-1.52e-04
Variance	1.07e-02	2.71e-03	1.28e-03	6.31e-04	3.09e-04	1.44e-04	7.10e-05
MSE	1.12e-02	2.73e-03	1.28e-03	6.31e-04	3.09e-04	1.44e-04	7.10e-05
Estimate of parameter σ							
Bias	-1.71e-02	-5.49e-03	-6.89e-03	-2.79e-03	-1.31e-03	2.29e-04	-1.74e-04
Variance	3.37e-02	8.76e-03	4.17e-03	1.94e-03	9.80e-04	4.15e-04	2.27e-04
MSE	3.40e-02	8.79e-03	4.22e-03	1.95e-03	9.81e-04	4.15e-04	2.27e-04
Computational time							
Average CPU time (s)	7.82e-03	1.63e-02	2.77e-02	2.40e-02	3.99e-02	6.81e-02	1.22e-01

7.2 Linear combination of modulated processes

In this section we present some simulations of the process described by the example model of Section 3.6, and estimate its parameters by fitting the parametric expected periodogram to the observed periodogram. More precisely, the process $\{X_{1,t}\}$ is a Gaussian white noise process with variance σ_1^2, and the process $\{X_{2,t}\}$ is an $AR(1)$ process with regression parameter r and innovation variance σ_2^2. Both processes are modulated, with respective modulation sequences chosen as follows,

$$\begin{cases} g_{1,t} &= \mathscr{D}_{\gamma_1,\Delta_1}(g_{1,t-1}+A_1\varepsilon_{1,t}) \\ g_{2,t} &= \mathscr{D}_{\gamma_2,\Delta_2}(g_{2,t-1}+A_2\varepsilon_{2,t}) \end{cases} \quad t \geq 1,$$

where the constants $A_1 = A_2 = 0.2, \gamma_1 = 10, \Delta_1 = 5, \gamma_2 = 30, \Delta_2 = 15$ are fixed and where $\mathscr{D}$ is the function defined by,

$$\mathscr{D}_{\gamma,\Delta}(x) = \min\left[\max(x, \gamma - \Delta), \gamma + \Delta\right], \tag{7.5}$$

so that the modulation sequences are both bounded random walks. They are initialized according to,

$$\begin{cases} g_{1,0} &= \mathscr{D}_{\gamma_1,\Delta_1}(\gamma_1 + A_1\varepsilon_{1,t}) \\ g_{2,0} &= \mathscr{D}_{\gamma_1,\Delta_2}(\gamma_2 + A_2\varepsilon_{2,t}) \end{cases}. \tag{7.6}$$

The observed process is given by,

$$\widetilde{Z}_t = a_1\widetilde{X}_{1,t} + a_2\widetilde{X}_{2,t},$$

with $a_1 = 2$ and $a_2 = 1$ fixed and known. We also assume that we observe the modulation sequences $g_{1,t}$ and $g_{2,t}$. Note that we cannot however retrieve $X_{1,t}$ or $X_{2,t}$ from observing $\widetilde{Z}_t$ in parallel with $g_{1,t}$ and $g_{2,t}$.

We generated 10,000 time series, each of length 8,192, based on this stochastic process model, and estimated the parameters σ_1, r and σ_2, which were fixed to 1, 0.9 and 2 respectively. The quasi-likelihood used here is that from (5.7), where the expected periodogram is obtained by applying (5.3), where the expected modulation sequence is obtained from (3.16).

The initializations were started with the parameter vector $[2\ 0.5\ 1]^T$. We present histograms of estimates of σ_1, r and σ_2 in Figure 7.2. We note that estimates seem somewhat normally distributed, and among the 10,000 repeats the minimization of the likelihood always converged to a point *near* the true parameter vector except for two samples. We also plot a $3D$ color map of one sample likelihood function with parameter σ_1 fixed to its true value around the true parameter vector in Figure 7.3. We note that the likelihood function for that time series does not seem to have any local minima other than the global minimum, at least within the range of values considered in the plot, despite the observed process being an aggregation of two modulated processes.

7.3 Unobserved modulation sequence

In this section we present estimations in the situation where the modulation sequence is only known up to a parametric form, as described in Section 5.4.

7.3.1 Unobserved periodic modulation with white noise

In this section we present a simulation based on the model described in Example 1 of Section 5.4. The latent process considered is an $AR(1)$ process with regression parameter r and innovation variance σ^2. The modulation sequence is given by,

$$g_t = \alpha_{t \bmod T} + A\eta_t, \tag{7.7}$$

where we fix $T = 4, \alpha_0 = 3, \alpha_1 = 4, \alpha_2 = 12, \alpha_3 = 8$, and where $\{\eta_t\}$ is a standard normal white noise process independent of the latent process. We consider a range of values for the parameters r and A, while σ is fixed to 1. For each combination of (A, r), we simulate 2000 time series, and aggregate estimates. The inference method uses (5.13) to estimate $\alpha_0, \ldots, \alpha_{T-1}$ and (5.7) subsequently, as described in Section 5.4. We present the bias, variance, and RMSE of $\hat{r}$ in Figure 7.4. We observe that the RMSE increases with small values of r (corresponding to a smaller peak of the spectral density) and large values of A (corresponding to more noisy estimates of the periodic pattern of the modulation sequence). We note that we
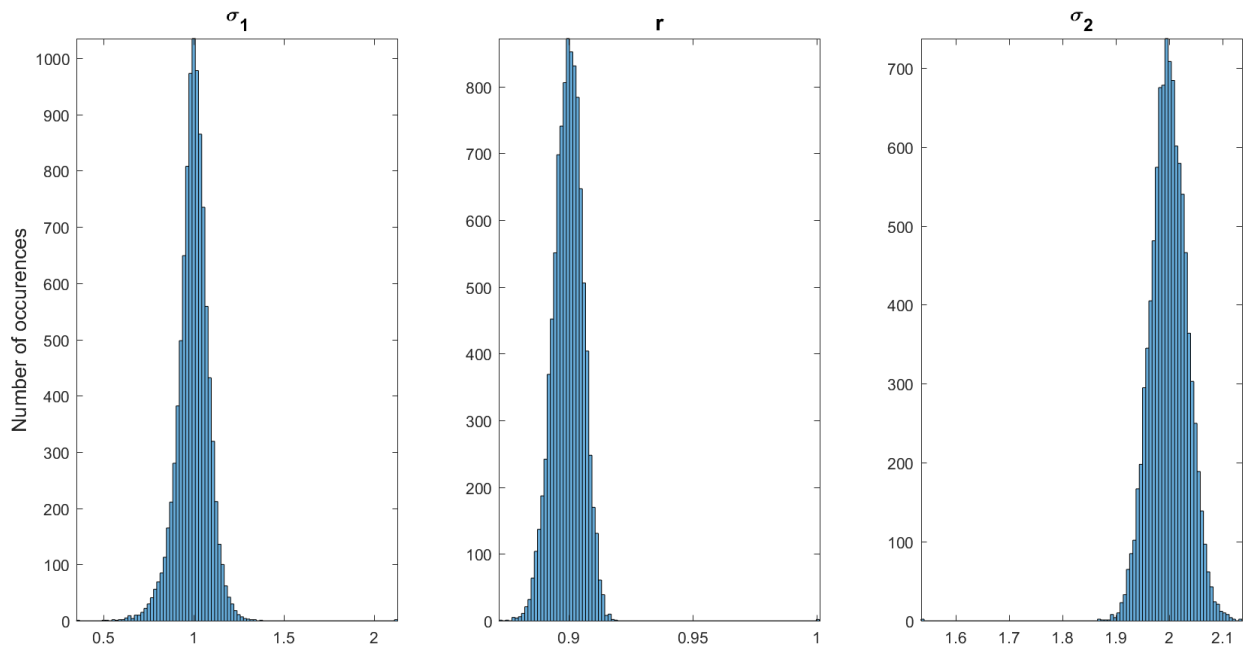
Figure 7.2: Histogram of estimates of the parameters $\sigma_1 = 1$, $r = 0.9$ and $\sigma_2 = 2$ of the processes $\{X_{1,t}\}$ and $\{X_{2,t}\}$, from observing $Z_t = 2g_{1,t}X_{1,t} + g_{2,t}X_{2,t}$ and the modulation sequences $\{g_{1,t}\}$ and $\{g_{2,t}\}$, which are generated according to the random walks defined in (7.6).
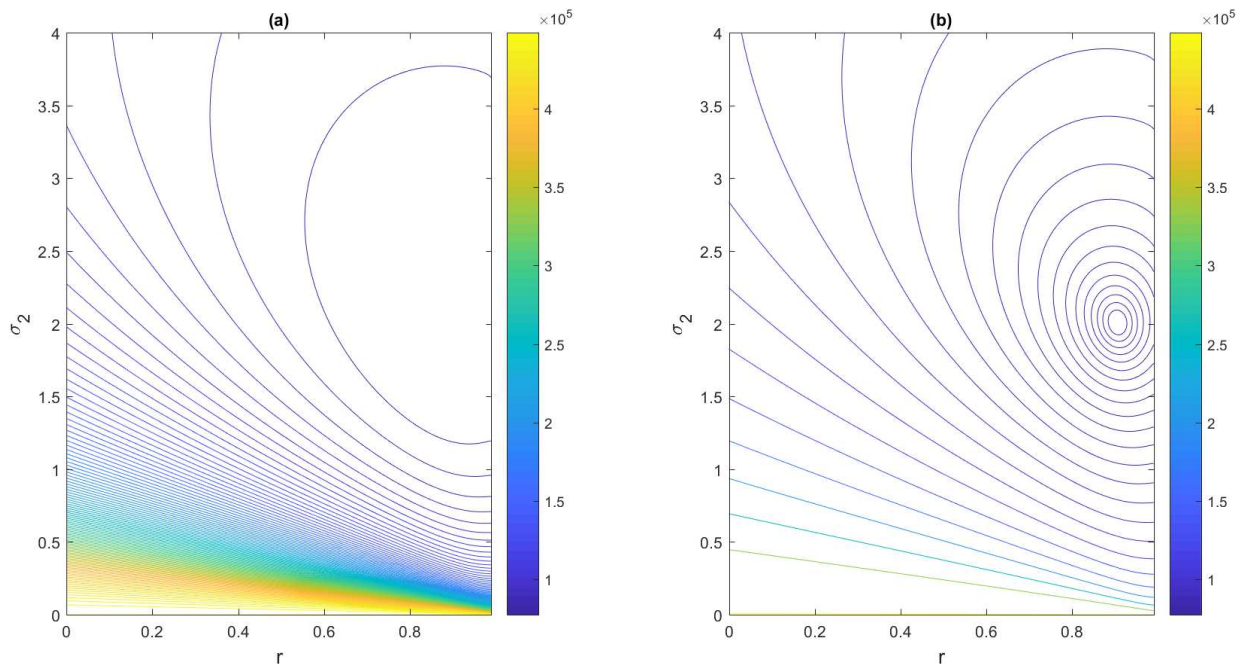

Figure 7.3: Contour plots of the likelihood function for one simulated sample from the model of Section 7.2, with the parameter σ_1 fixed to its true value, 1. (a) 100 levels of the likelihood function, with a constant step from the minimum value to the maximal value; (b) 100 levels of the likelihood function with unequal steps.

.

may get more accurate estimates by implementing an EM procedure, with the two following steps repeated until convergence:

1. *E-step*. Given the current estimates of r, σ, $\alpha_0, \ldots, \alpha_{T-1}$, derive the conditional variances at each time point. Update the estimates $\alpha_0, \ldots, \alpha_{T-1}$ using the ratio of the sample variances by the mean conditional variances (subsampling according to the value of T as in (5.13).

2. *M-step*. Maximize the quasi-likelihood function using the new estimates of $\alpha_0, \ldots, \alpha_{T-1}$, and update r, σ.
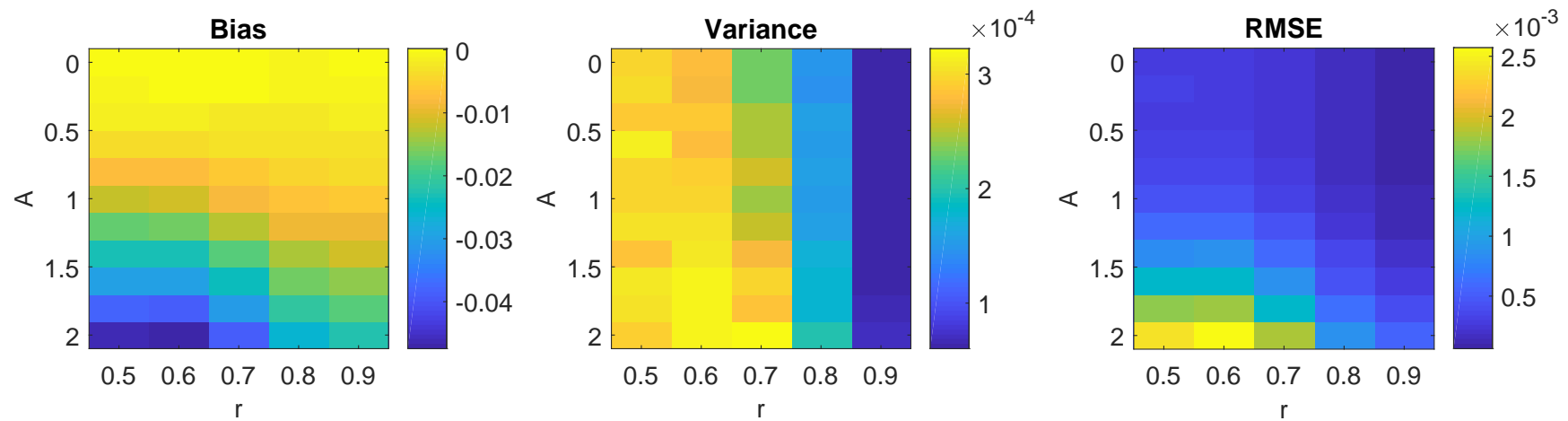
Figure 7.4: Biases, variances and RMSEs of the estimate of the parameter *r*, for values of *r* ranging from 0.5 to 0.9 with a step of 0.1, and values of *A* ranging from 0 to 2 with a step of 0.2.

7.3.2 Linearly increasing frequency

In this section we simulate the complex-valued process defined by (3.14) with the frequency parameter β_t having the parametric form described in Example 2 of Section 5.4. In this problem we have to estimate $\{\gamma, \Delta\}$ from β_t as well as $\{r, \sigma\}$ from Z_t. We perform a Monte Carlo simulation with a fixed sample size of $N = 512$, where we simulate 5,000 independent time series each with parameters set to $r = 0.9$, $\sigma = 10$, $\gamma = 0.8$, and $\Delta = \frac{2\pi}{3}$. We report the biases, variances and MSEs with the stationary and non-stationary methods in Table 7.2. The non-stationary estimator used is (5.11). As the simulated stochastic process (3.14) is Markovian, it is also possible to implement exact maximum likelihood in $\mathcal{O}(N)$ elementary operations for this specific problem, and we report these values in the table also for comparison. Our non-stationary inference method performs relatively close to that of exact maximum likelihood, despite the challenge of having to estimate parameters of the modulating sequence, as well as the latent process. The stationary method performs poorly, as with previous examples, as stationary modelling is not appropriate for such rapidly-varying oscillatory structure.

Table 7.2: Performance of estimators with the stationary and non-stationary methods for the model of (3.14) with β_t evolving according to (5.14). The parameters are set as $r = 0.9$, $\sigma = 10$, $\gamma = 0.8$, and $\Delta = 2\pi/3$. The results are averaged over 5,000 independently generated time series for each sample size N. N/A stands for Not Applicable.

Estimated parameter	r	σ	γ	Δ
Exact likelihood				
Bias	-1.32e-03	2.11e-02	2.12e-03	-3.61e-03
Variance	1.87e-04	5.21e-02	2.37e-04	2.83e-03
MSE	1.88e-04	5.26e-02	2.42e-04	2.85e-03
Stationary frequency domain likelihood				
Bias	-1.54e-01	5.21e+00	2.59e-03	N/A
Variance	5.51e-04	8.69e-01	9.46e-03	N/A
MSE	2.42e-02	2.80e+01	9.47e-03	N/A
Non-stationary frequency domain likelihood				
Bias	-1.71e-03	6.82e-03	1.14e-03	-3.71e-02
Variance	2.40e-04	1.53e-01	2.08e-03	1.64e-02
MSE	2.43e-04	1.53e-01	2.08e-03	1.78e-02

Chapter 8

Inference for oceanographic time series

In this chapter we analyse real-world bivariate velocity data from instruments navigating the oceans in order to model ocean surface currents around the Earth. In particular we focus on the equatorial regions, a region less well understood in the literature due to modelling difficulties (Early, 2012). In Section 8.1 we describe succinctly the importance of modelling ocean surface currents and the challenge we address with our models and inference method. In Section 8.2 we describe the Global Drifter Program (GDP), which maintains and develops a massive database of ocean surface measurements from instruments called drifters. Then we present our modelling of velocities obtained from the drifters. Specifically we model jointly the latitudinal and longitudinal velocities as the aggregation of two independent complex-valued processes as discussed by Sykulski et al. (2016c). Here we improve the model of this reference by allowing one of the processes in the model to become strongly non-stationary. Specifically we use the complex-valued AR(1) process defined by (3.14) and do not approximate β_t by a constant, as in Sykulski et al. (2016c). This motivates us to use the parametric inference method introduced in Section 5. To investigate this estimator and compare with alternative approaches, in particular stationary modelling and estimation methods, we present two simulation studies in Section 8.3, the first one being based on a dynamical model of the ocean surface currents with white noise input accounting for the wind forcing, and the sec-

ond one being a simulation of the model from Section 3.5.3. Finally, we infer phys-
ical quantities that describe the ocean surface currents for selected drifters from the
GDP subject to a strong modulation. All data and code used here are available for
download at `www.ucl.ac.uk/statistics/research/spg/software`
and all results in this chapter are exactly reproducible by using this code. All the
material in this chapter, except for the stationary model described in Section 8.2.1
which was introduced in Sykulski et al. (2016c), is novel and was published in Guil-
laumin et al. (2017).

8.1 Oceanography background

Studying the ocean currents is essential to understanding their impact on climate.
For instance, the relatively warm temperature in the UK, given its latitude, are in
part the result of warm waters being brought to the UK by the Gulf Stream from
more southern latitudes. Understanding ocean currents is also essential to be able
to model oil spills or to recover objects lost at sea (Biastoch et al., 2017). One
main pattern of the ocean currents, in terms of relative contribution to the kinetic
energy of a particle of water at the surface of the oceans (Ferrari and Wunsch, 2009),
is the *inertial effect* (Early, 2012). This effect is caused by what is commonly
known among physicists as the Coriolis force. More precisely, as the observer is in
a rotating frame, the studied particle is subject to a *force* whose amplitude depends
on the rate of rotation, and whose direction is orthogonal to the velocity vector in
the rotating frame. In oceanography the Coriolis effect results in oscillations called
inertial oscillations, whose frequencies f directly depend on the rotating rate of the
Earth Ω and the latitude ζ through the relation (Early, 2012),

$$f = 2\Omega \sin \zeta. \tag{8.1}$$

The rotation rate of the Earth Ω is computed as $2\pi/T$ where $T = 86,164.1$s is one
sidereal day in seconds. A sidereal day corresponds to the time it takes for the Earth
to complete one rotation around its own axis. One of the main interests in modelling
the velocities of the ocean currents is to understand the impact of the wind forcing

and the corresponding damping time scale (Elipot et al., 2010). This requires to correctly distinguish the contribution to the kinetic energy coming from the general background, which can be modelled by a bi-dimensional Matérn process (Sykulski et al., 2016c), and that coming from the inertial effect.

In most areas of the oceans, the Lagrangian velocities of a surface particle can reasonably be modelled as locally-stationary, allowing us to apply stationary estimation methods over rolling and overlapping temporal windows. However, the study of areas with strong latitudinal mean flows is complicated by fast-varying Coriolis frequencies, causing the inertial oscillations to display highly non-stationary structures. The usual trade-off in this situation usually consists in a choice between small temporal windows to limit the variations in the generating mechanism, at the expense of increased variances of estimates, and large temporal windows, with the risk of highly-biased estimates due to the stationary approximation. We shall address this challenge using the class of bivariate modulated processes introduced in Section 3.5.3, and the estimators presented in Section 5.3.

8.2 The Global Drifter Program

The GDP database (www.aoml.noaa.gov/phod/dac) is a collection of measurements obtained from buoys known as surface drifters, which drift freely with ocean currents and regularly communicate measurements to passing satellites at unequally spaced time intervals averaging 1.4hrs. The data is then interpolated onto a regular temporal grid using for example the approach of Elipot et al. (2016). The measurements include position, and often sea surface temperature. Since 2005, over 11,000 drifters have been deployed, with approximately 70 million position recordings obtained. The surface positions are differentiated to obtain bivariate velocity time series, represented as complex-valued time series $Z_t = X_t + iY_t$ where X_t is the longitudinal velocity and Y_t is the latitudinal velocity (Sykulski et al., 2016c). The analysis of this data is crucial to our understanding of ocean circulation (Lumpkin and Pazos, 2007), which is known to play a primary role in determining the global climate system, see e.g. Andrews et al. (2012). Furthermore, GDP data are used to
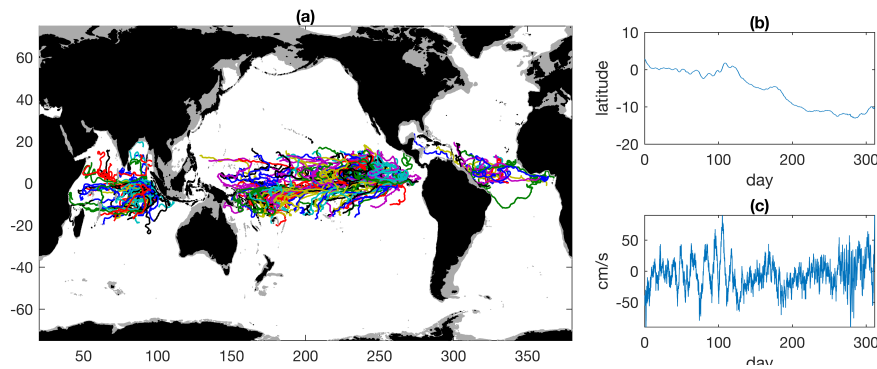
Figure 8.1: (a) The trajectories of the 200 drifters from the Global Drifter Program, analysed in Section 8.4, that exhibit the greatest change in Coriolis frequency (f) across 60 inertial cycles, as described in that section; (b) a segment of data of the latitudinal positions over time from Drifter ID#43594; and (c) a segment of data of the latitudinal velocities from this drifter in cm/s. This figure is produced using the jLab toolbox (Lilly, 2016).

understand the dispersion characteristics of the ocean, which are critical in correctly modelling oil spills (Abascal et al., 2010) and more generally assist in developing theoretical understanding of ocean fluid dynamics (Griffa et al., 2007), which is necessary for global climate modelling.

In Fig. 8.1(a), we display the trajectories of 200 drifters which either traverse or are near the equator (latitudes between ± 20 degrees), interpolated onto a 2 hour grid from raw position observations available at the GDP web site. We focus on a single drifter trajectory, drifter ID#43594, in panels (b) and (c), displaying both its latitudinal position and velocity respectively, the latter of which is obtained by differencing the positions. It appears that the velocity time series is strongly non-stationary, as it has oscillations which appear to change in frequency over time. This occurs because the drifters are changing latitude—and the *Coriolis frequency*, denoted f, is equal to twice the rotation rate of the Earth (Ω) multiplied by the sine of the latitude (ζ), as per (8.1). Note that the Coriolis frequency f is a signed quantity, implying that oscillations occur in opposite rotational clockwise/anti-clockwise directions from one hemisphere to the other. The Coriolis frequency is positive in the Northern hemisphere whereas the oscillations occur in the mathematical negative

sense. Therefore we define the inertial frequency,

$$\omega^{\{f\}} = -\frac{f}{2\pi K}, \tag{8.2}$$

as the negative of the Coriolis frequency f divided by $2\pi K$, where $K = 86,400.002$s is one solar day in seconds, so that $\omega^{\{f\}}$ is in cycles per day.

8.2.1   Stochastic modelling

The stochastic modelling of Lagrangian trajectories was investigated in Sykulski et al. (2016c), where the term "Lagrangian" is used because the moving object making the observations (i.e. the drifter) is the frame of reference, as opposed to fixed-point measurements known as *Eulerian* observations. In that paper, the Lagrangian velocity time series was modelled as a stationary Gaussian complex-valued time series, with the following 6-parameter power spectral density

$$S(\omega) \;=\; \frac{A^2}{(\omega - \omega^{\{f\}})^2 + \lambda^2} + \frac{B^2}{(\omega^2 + h^2)^\alpha}, \quad \omega \in \mathbb{R}, \tag{8.3}$$
$$A > 0, \;\; \lambda > 0, \;\; h > 0, \;\; B > 0, \;\; \alpha > \frac{1}{2},$$

where ω is given in cycles per day.

The first component of (8.3) is the spectral density of a complex-valued Ornstein-Uhlenbeck (OU) process (Arató et al., 1962), and is used to describe the effect of inertial oscillations at frequency $\omega^{\{f\}}$. Denoting $\widetilde{Z}_{\text{OU}}(t)$ the OU component, where $\widetilde{Z}_{\text{OU}}(t)$ is complex-valued, these oscillations are described by the following stochastic differential equation (SDE),

$$d\widetilde{Z}_{\text{OU}}(t) = (-\lambda + i2\pi\omega^{\{f\}})\widetilde{Z}_{\text{OU}}(t)dt + AdW(t), \tag{8.4}$$

where t is expressed in days and $W(t)$ is a complex-valued Brownian process with independent real and imaginary parts. The damping parameter $\lambda > 0$ ensures that the OU process is mean-reverting. The corresponding continuous complex-valued

autocovariance is given by

$$s(\tau) = \frac{A^2}{2\lambda} \exp\left\{-\lambda|\tau| + i2\pi\omega^{\{f\}}\tau\right\},$$

and the sampled process (Arató et al., 1999) $\widetilde{Z}_{\text{OU},t} = \widetilde{Z}_{\text{OU}}(t\Delta)$, where we recall that $\Delta = 1/12$ day is the sampling rate corresponding to the 2hr grid, is a complex-valued AR(1),

$$\widetilde{Z}_{\text{OU},t} = re^{i2\pi\omega^{\{f\}}\Delta}\widetilde{Z}_{\text{OU},t-1} + \varepsilon_t. \tag{8.5}$$

Here $\{\varepsilon_t\}$ is a Gaussian complex-valued white noise process with variance σ^2 and independent real and imaginary parts. The autocovariance sequence of the stationary sampled process is given by,

$$c_{\widetilde{Z}_{\text{OU}}}(\tau) = \frac{\sigma^2}{1-r^2}r^\tau e^{i2\pi\tau\Delta\omega^{\{f\}}}. \tag{8.6}$$

The transformation between the parameters of the complex-valued OU and the complex-valued AR(1) are given by,

$$\sigma^2 = \frac{A^2(1-e^{-\lambda\Delta})}{2\lambda\Delta}, \quad r = e^{-\lambda\Delta}. \tag{8.7}$$

The second component of (8.3) is the spectral density of a stationary proper Matérn process (Gneiting et al., 2010), denoted $Z_M(t)$, and is used to describe two-dimensional background turbulence, see Lilly et al. (2017). Although the parameter $\omega^{\{f\}}$ is varying as the drifter changes latitude, this parameter is fixed to its mean value in each trajectory segment in Sykulski et al. (2016c). This leaves five remaining parameters to estimate, $\{A, \lambda, B, h, \alpha\}$, in different regions of the ocean.

The model of (8.3) is stationary—slowly-varying non-stationarity in the data is accounted for by windowing the data into segments of approximately 60 inertial periods, and treating the process as locally-stationary within each window. The estimated parameters can then be aggregated spatially to quantify the heterogeneity of ocean dynamics. This method works well on relatively quiescent and stationary regions of the ocean; however this method cannot account for the rapidly-varying

non-stationarity evident in Fig. 8.1, and leads to model misfit and biased parameter estimates, as we shall now investigate in detail.

8.2.2 Modulated time series modelling and estimation

We now apply the methodological contributions of this thesis to improve the model of (8.3) for highly non-stationary time series, such as those observed in Fig. 8.1(a). We do this by accounting for changes in the inertial frequency, $\omega^{\{f\}}$, within each window of observation, by using the framework of modulated processes. We denote $\omega^{\{f\}}(t)$ the continuous time-varying inertial frequency and $\omega_t^{\{f\}} = \omega^{\{f\}}(t\Delta)$ the inertial frequency value at each observed time step, $t = 0, \cdots, N-1$. The adapted version of the SDE (8.4) is then given by,

$$d\widetilde{Z}_{\text{OU}}(t) = \left(-\lambda + i2\pi\omega^{\{f\}}(t)\right)\widetilde{Z}_{\text{OU}}(t)dt + AdW(t). \tag{8.8}$$

In analogue to the proof of Proposition 4 it is shown that the sampled process $\widetilde{Z}_{\text{OU},t} = \widetilde{Z}_{\text{OU}}(t\Delta)$ satisfies,

$$\widetilde{Z}_{\text{OU},t} = re^{i2\pi\int_{\Delta(t-1)}^{\Delta t}\omega^{\{f\}}(u)du}\widetilde{Z}_{\text{OU},t-1} + \varepsilon_t.$$

As the inertial frequency is only observed at sampled points, we approximate the term $\int_{\Delta(t-1)}^{\Delta t}\omega^{\{f\}}(u)du$ by $\Delta\omega_t^{\{f\}}$. Specifically, we use the model of (3.14) for complex-valued time series, i.e.

$$\widetilde{Z}_{\text{OU},t} = re^{i2\pi\Delta\omega_t^{\{f\}}}\widetilde{Z}_{\text{OU},t-1} + \varepsilon_t, \ \ t \geq 1, \tag{8.9}$$

where ε_t has the same properties as in (8.5) and the transformation between the parameters $\{A, \lambda, \omega_t^{\{f\}}\}$ of the non-stationary complex-valued OU process (8.8) and the parameters $\{\sigma, r, \omega_t^{\{f\}}\}$ of the non-stationary complex-valued AR(1) process (8.9) are given by (8.7).

The required methodology has been developed in Section 3.5 for bivariate (or complex-valued) time series. We only perform the modulation on the complex OU component in (8.3); the Matérn component for the turbulent background is

unchanged and is considered to be stationary in the window, as it is not in general affected by changes in $\omega^{\{f\}}$. However the two components are observed in aggregation, and for this reason we cannot simply demodulate the observed non-stationary signal to recover a stationary signal. Instead, to jointly estimate the parameters $\{A, \lambda, B, h, \alpha\}$, we first compute the modulating sequence associated with (8.9), g_t, using (3.15) from Proposition 4 and accounting for the temporal sample rate Δ,

$$g_t = e^{i\sum_{u=1}^{t} 2\pi\Delta\omega_u^{\{f\}}}, \tag{8.10}$$

for $t = 0, \cdots, N-1$. Then we obtain the expected periodogram of the OU component, by computing $c_g(\tau)$ according to (3.13), then $\overline{c}_{Z_{OU}}^{(N)}(\tau)$, where we use the autocovariance of a stationary complex-valued OU process

$$c_{Z_{OU}}(\tau; r, \sigma) = \frac{\sigma^2}{1-r^2} r^\tau,$$

and finally Fourier transforming $\overline{c}_{Z_{OU}}^{(N)}(\tau)$ according to (5.3). Next, we compute the expected periodogram of the stationary Matérn as outlined in Sykulski et al. (2016c). Note that this can also be computed from the autocovariance of a Matérn, which we denote $c_{Z_M}(\tau)$, using (5.3), and by setting $g_t = 1$ for all t, from which the modulation kernel (3.4) is simply the triangle kernel $1 - |\tau|/N$. Finally, we additively combine the expected periodograms, i.e.

$$\overline{S}^{(N)}(\omega; \theta) = \sum_{\tau=-(N-1)}^{N-1} \left[c_g(\tau)c_{Z_{OU}}(\tau) + \left(1 - \frac{|\tau|}{N}\right) c_{Z_M}(\tau) \right] e^{-i2\pi\omega\tau\Delta},$$

for $\omega \in [-\frac{1}{2\Delta}, \frac{1}{2\Delta}]$, and then minimize the objective function, given in (5.10), to obtain parameter estimates for $\{A, \lambda, B, h, \alpha\}$.

Note that the modulation of a complex-valued AR(1) process by (8.10) will not lead to an asymptotically stationary process, as in general we cannot expect the quantities $c_g^{(N)}(\tau)$ to converge. However, we can see from Fig. 8.1(a) that the drifters of our dataset have latitudes comprised between ± 20 degrees. Therefore the terms $2\pi\Delta\omega_t^{\{f\}}$ in (8.9) are comprised between ± 0.3591 radians (recalling (8.1)

and (8.2)), so that the conditions of Proposition 5 are verified. Hence the sampled inertial component is a modulated process with a significant correlation contribution, which motivates the use of our estimator (5.9). Note that significant correlation contribution is achieved here because of both the latitudes of the drifters and the sampling rate used.

The assumption of Gaussianity is reasonable for modelling the velocity of instruments from the GDP as is discussed in Section 2.4 of LaCasce (2008) and references therein. To further inspect this, we tested the hypothesis of Gaussianity for the specific equatorial drifter velocity dataset that is the subject of our study. The code for this test is available online at `www.ucl.ac.uk/statistics/research/spg/software/modulated`. Because the time series values are correlated in time, it would be incorrect to apply a Gaussianity test to all the velocities, as such tests typically require the samples to be independent (Paparoditis and Politis, 2012). Therefore we perform our tests on the differenced process (i.e. the accelerations, noting that if the accelerations are Gaussian then the velocities are Gaussian too), where correlation will decay more quickly in time. Furthermore, to account for any remaining correlations we sub-sample the differenced time series. To choose a relevant sampling step we selected a decorrelation length common to all our velocity time series based on an analysis inspired by Paparoditis and Politis (2012). We averaged the biased sample autocorrelation sequences and selected the first lag from which the averaged autocorrelation sequence is within the 95% confidence interval obtained under the hypothesis of zero correlation. The sub-samples are tested via a Kolmogorov-Smirnov test with size 5%. We obtain a rejection rate of 6.5% for the overall dataset (including both longitudinal and latitudinal velocities), which is broadly consistent with a type I error rate of 5%. This shows that the assumption of Gaussianity is reasonable.

8.3 Simulation studies

In this section, before we analyse real data from the GDP, we use simulated time series to test our proposed modelling and inference method. In Section 8.3.1 we sim-

ulate velocity time series according to a dynamical model for inertial oscillations, with simulated white noise input accounting for the wind forcing. We generate time series for a range of realistic latitudinal velocities, and show the necessity of accounting for the changing latitudes and Coriolis frequencies to correctly estimate the damping time scale of the inertial oscillations, as compared to a model assuming a constant Coriolis frequency. In Section 8.3.2 we simulate time series according to the bivariate modulated process described in Section 3.5.3 and where we take the time-varying frequency β_t of (3.14) to be generated according to a bounded random walk. We generate time series for a range of sample sizes, showing the decrease in the root mean squares errors as the time series become larger.

8.3.1 Testing with a dynamical model

We first test the accuracy of the non-stationary modelling and parameter estimation for drifters by analysing output in a controlled setting using a dynamical model for inertial oscillations. The model propagates particles on an ocean surface forced by winds—simulated white noise in our simulations—with a fixed damping parameter, similarly to the damped-slab model of Pollard and Millard (1970), but uses the correct spherical dynamics for the Earth from Early (2012), so that the oscillations occur at the correct Coriolis frequency given the particle's latitude and so that the model remains valid at the equator. The damping timescale parameter is fixed globally *a priori* in the model and the goal is to see if it can be accurately estimated using parametric time series models.

The numerical model is constructed such that the particle can also be given a linear mean flow, denoted $U + iV$. If this mean flow has a significant vertical component V, then the particle will cross different latitudes and the frequency of inertial oscillations will significantly change over a single analysis window. We display particle trajectories from the dynamical model in Fig. 8.2, with various realistic mean flow values, where the spherical dynamics can clearly be seen for larger latitudinal mean flow values. We observe that the particles subject to small mean flows display stationary oscillation patterns, whereas for the particles with a large latitudinal mean flow, the oscillation frequency appears to diminish as the particle approaches
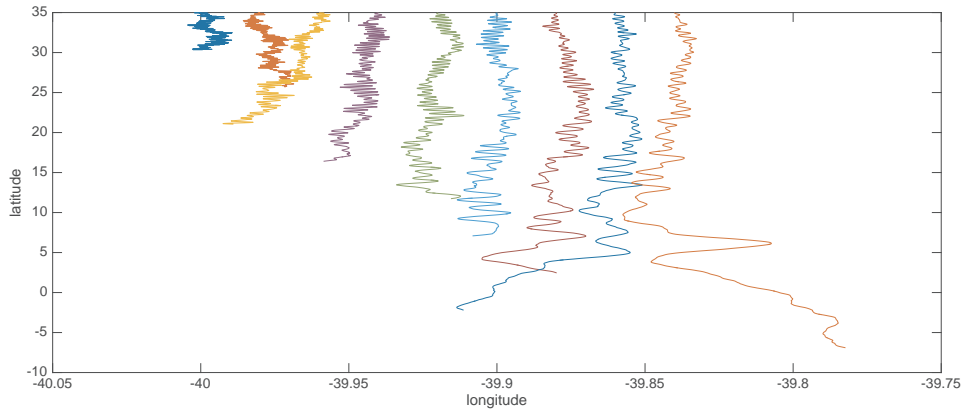
Figure 8.2: Trajectories of 9 particles from the dynamical model of Pollard and Millard (1970) with the correct spherical dynamics of the Earth as described in Early (2012) and with the damping timescale set to 4 days. All particle trajectories are started at 35° N and 40° W with increasing latitudinal mean flow from $V = 0.1$ to $V = 0.9$ m/s going from left to right (U is set to zero for this example). The drifters are offset in longitude by 0.02 degrees for representation.

latitude zero. A more complete description of the numerical model is available in the online code.

To explore the performance of the estimation of damping timescales, we assess the performance of the parameter estimates of our non-stationary model, by performing a Monte Carlo study based on the dynamical model described in the previous paragraph. We generate 100 trajectories, each of length 60 days and sampled every 2 hours, for a given damping timescale ($1/\lambda$) and latitudinal mean flow (V). We estimate the damping parameter using the stationary and non-stationary methods, and average the estimated damping timescales $1/\lambda$ over the 100 time series. We note that this model has no background turbulence, and so we set $B = 0$ in (8.3) such that there is no Matérn component present. We then repeat this analysis over a range of realistic values for $1/\lambda$ and V. The average estimates of $1/\lambda$ are reported in Fig. 8.3. The stationary method breaks down for large mean flows and long damping timescales, with large overestimates of λ, confirming our observations from the real-data analysis. The non-stationary method performs well across the entire range of values. We note that long damping timescales are generally harder to estimate, as λ becomes close to zero and is estimated over relatively fewer frequencies. We have not reported mean square errors, but we found the parameter biases to be the
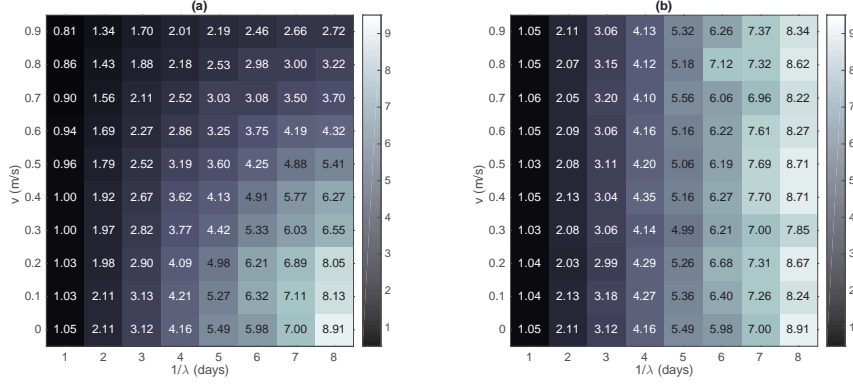
Figure 8.3: Mean estimates of the damping timsescale $1/\lambda$ with (a) the stationary model (8.3) and (b) the non-stationary model of Section 8.2.2, applied to 100 realizations of the dynamical model described in Section 8.3.1. The experiment is performed over a grid of latitudinal mean flow values V from 0 to 0.9 cm/s, and over a range of true damping timescales $1/\lambda$ from 1 to 8 days. The estimated damping timescale values, averaged over 100 repeat experiments, is written in each cell and shaded according to the colorbar.

main contribution to the errors, so it follows that the non-stationary method remains strongly preferable.

8.3.2 Testing with Stochastic Model Output

In this section we test with purely stochastic output, which allows us to extensively compare biases, errors and computational times of the stationary and non-stationary methods in a much larger Monte Carlo study. We continue using the bivariate model of (3.14) which is suitable for inertial oscillations, except this time we change β_t according to a stochastic process. Specifically, we set as our generative mechanism for the frequencies β_t,

$$\beta_0 = \mathscr{D}_{\gamma,\Delta}(\gamma + A\varepsilon_t) \tag{8.11}$$

$$\beta_t = \mathscr{D}_{\gamma,\Delta}(\beta_{t-1} + A\varepsilon_t), \tag{8.12}$$

where $\gamma \in [-\pi, \pi)$, $\Delta > 0$, $A > 0$ and $\{\varepsilon_t\}$ is a standard normal white noise, and $\mathscr{D}(\cdot)$ is the function defined by (7.5). We recall that this choice of $\mathscr{D}_{\gamma,\Delta}(x)$ constrains β_t in the interval $[\gamma - \Delta, \gamma + \Delta]$. This way the frequencies β_t are generated according to a bounded random walk, i.e. a random walk which is constrained to stay within a fixed bounded interval. According to Proposition 5, if Δ is smaller than $\pi/2$, then

this ensures that the modulated process belongs to the class of modulated processes with a significant correlation contribution, and our estimator (5.10) is consistent.

In our simulations we have set $\gamma = \pi/2$, $\Delta = 1$, $A = 1/20$. We simulate for a range of sample sizes ranging from $N = 128$ to $N = 4096$. For each sample size N, we independently simulate 2000 time series and estimate $\{r, \sigma\}$ for each series to report ensemble-averaged biases, errors, and computational times. The results are reported in Table 8.1. The bias and Mean Square Error (MSE) of the estimated parameters with the stationary method (with β_t replaced by its average value) are seen to increase with increasing sample size. This is because the random walk of β_t increases the range of β_t with larger N, such that the non-stationarity of the time series is increasing. Conversely, the non-stationary method accounts for these rapidly changing modulating frequencies, and the bias and MSE of parameter estimates rapidly decrease with increasing N. The average CPU time of the non-stationary model fitting is of the same order as that of the CPU time of the stationary model fitting, as the method is still $\mathscr{O}(N \log N)$ in computational efficiency.

Table 8.1: Performance of estimators with the stationary and non-stationary methods for the model of (3.14) with β_t evolving according to the bounded random walk described by (8.11). The parameters are set as $r = 0.8$, $\sigma = 1$, $\gamma = \pi/2$, $\Delta = 1$, and $A = 1/20$. The results are averaged over 2,000 independently generated time series for each sample size N. The average CPU times for the optimization are given in seconds, as performed on a 3.60Ghz Intel i7-4790MQ processor (4 cores).

Sample size (N)	128	256	512	1024	2048	4096
Stationary frequency domain likelihood						
Bias (r)	-2.33e-02	-3.27e-02	-4.81e-02	-6.97e-02	-9.32e-02	-1.12e-01
Variance (r)	1.80e-03	1.04e-03	1.13e-03	1.54e-03	1.40e-03	8.27e-04
MSE (r)	2.34e-03	2.11e-03	3.44e-03	6.40e-03	1.01e-02	1.33e-02
Bias (σ)	3.73e-02	6.16e-02	9.28e-02	1.34e-01	1.75e-01	2.07e-01
Variance (σ)	3.72e-03	2.87e-03	3.33e-03	4.44e-03	3.98e-03	2.16e-03
MSE (σ)	5.11e-03	6.67e-03	1.19e-02	2.24e-02	3.46e-02	4.49e-02
Average CPU time (sec)	8.26e-03	1.21e-02	1.46e-02	2.19e-02	3.68e-02	6.44e-02
non-stationary frequency domain likelihood						
Bias (r)	-4.08e-03	-1.95e-03	-1.37e-03	-2.51e-04	-2.28e-04	1.10e-04
Variance (r)	1.64e-03	7.52e-04	3.98e-04	2.07e-04	1.07e-04	5.30e-05
MSE (r)	1.65e-03	7.56e-04	4.00e-04	2.07e-04	1.07e-04	5.30e-05
Bias (σ)	-4.00e-03	-2.87e-03	-1.48e-03	-1.12e-03	-7.77e-04	-6.15e-04
Variance (σ)	2.56e-03	1.29e-03	6.57e-04	3.55-e04	2.03e-04	1.08e-04
MSE (σ)	2.58e-03	1.29e-03	6.59e-04	3.57e-04	2.03e-04	1.08e-04
Average CPU time(sec)	9.03e-03	1.14e-02	1.36e-02	2.01e-02	3.33e-02	6.21e-02

8.4 Inference for equatorial drifters

In this section we apply the methods that we have developed in this thesis to the analysis of drifters trajectories from the Global Drifter Program. We focus on drifter trajectories with strong relative changes in the Coriolis frequency. To do so, the entire drifter dataset is split into segments of 60 inertial periods in length, accounting for the variation of the inertial period along drifter trajectories, and with 50% overlap between segments. The standard deviation of the inertial frequency along each data segment is taken, and the 200 segments exhibiting the largest ratio of the standard deviation of the inertial frequency, to the magnitude of its mean value along the segment, are identified for use in this study. These trajectories exhibit the largest fractional changes in the inertial frequency, and as shown in Fig. 8.1(a), are all located in the vicinity of the equatorial region where the inertial frequency vanishes. We compare the likelihood estimates and parametric fits for the stationary model (8.3), with those for the non-stationary version of this model described in the previous subsection. In particular, the damping timescale $1/\lambda$ is of primary interest in oceanography (Elipot et al., 2010). In Fig. 8.5, we display the Whittle likelihood fits of each model to segments of data from drifter IDs #79243, #54656 and #71845, all of which are among the trajectory segments displayed in the left-hand panel of Fig. 8.1. We also include model fits to a 60-inertial period window of drifter ID#44312, which is investigated in detail in Sykulski et al. (2016c), as this South Pacific drifter is from a more quiescent region of the ocean, and does not exhibit significant changes in $\omega^{\{f\}}$. For the South Pacific drifter in Fig. 8.5(d), both fits are almost equivalent (and hence are overlaid), capturing the sharp peak in inertial oscillations at approx 1.2 cycles per day. For the three equatorial drifters, the stationary model (8.3) has been fit using (5.10) with the inertial frequency set to the average of $\omega_t^{\{f\}}$ across the window. The stationary model is a relatively poor fit to the observed time series spectra. The non-stationary modulated model, which incorporates changes in $\omega^{\{f\}}$, is a better fit, capturing the spreading of inertial energy between the maximum and minimum values of $\omega_t^{\{f\}}$.

In this analysis, we have excluded frequencies higher than 0.8 cycles per day

from all the likelihood fits to the equatorial drifters (the Nyquist is 6 cycles per day for this 2-hourly data), to ignore contamination from tidal energy occurring at 1 cycle per day or higher, which is not part of our stochastic model. Furthermore, we also only fit to the side of the spectrum dominated by inertial oscillations, as the model is not always seen to be a good fit on the other side of the spectrum. The modelling and inference approach is therefore semi-parametric (Robinson, 1995).

The significance of the misfit of the stationary model is that parameters of the model may be under- or over-estimated as the model attempts to compensate for the misfit. For example, the damping parameter of the inertial oscillations, λ, will likely be overestimated in the stationary model, as it is used to try to capture the spread of energy around $\omega^{\{f\}}$, which is in fact mostly caused by the changing value of $\omega^{\{f\}}$, rather than a true high value of λ.

To investigate this further, we perform the analysis with all 200 drifters shown in Fig. 8.1. In Fig. 8.4(a), we show a scatter plot of the estimates of $1/\lambda$, known as the damping timescale, as estimated by both models. In general, the damping timescales are larger with the non-stationary model (consistent with a smaller λ), where the median value is 3.42 days, rather than 1.3 days with the stationary model. Previous estimates of the damping timescale in the literature have not included data from the equatorial region, so while direct comparisons are not possible, the former estimates are found to be more consistent with previous estimates at higher latitudes where values of around 3 days are reported in Elipot et al. (2010), and values ranging from 2 to 10 days are reported in Watanabe and Hibiya (2002).

The non-stationary model does not require more parameters to be fitted than the stationary model; both have 5 unknown parameters. Therefore there is no need to penalize the non-stationary model using model choice or likelihood ratio tests. Even though the models are not nested, comparing the likelihood of the two approaches can be informative. We can directly compare the negative quasi-likelihood value of each model using (5.7) with the stationary model (denoted $\ell_S(\hat{\theta}_S)$) and the non-stationary model (denoted $\ell_{NS}(\hat{\theta}_{NS})$). A histogram of the difference between the likelihoods for the 200 drifters is shown in Fig. 8.4(b), where positive values
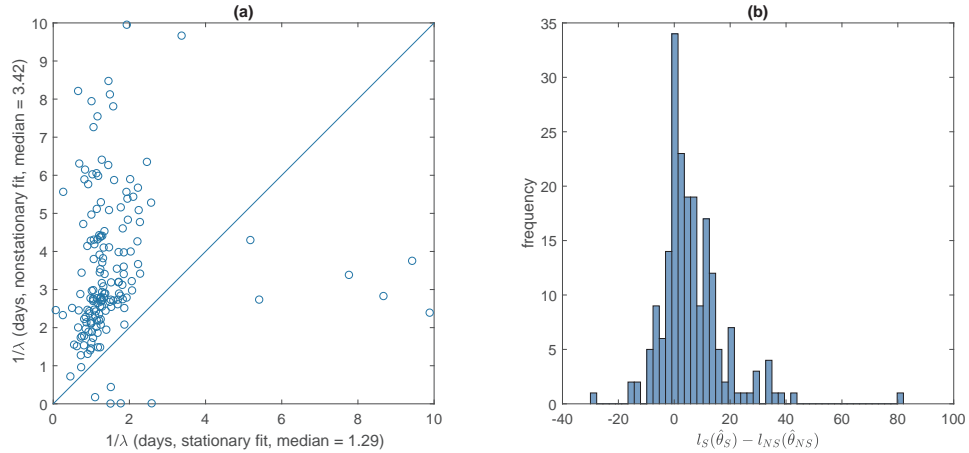
Figure 8.4: Subplot (a) is a scatter plot of the damping timescale $1/\lambda$ as estimated by the stationary and non-stationary models, for each of the 200 trajectories displayed in Fig. 8.1; Subplot (b) is a histogram of the difference between the log-likelihoods of the non-stationary and stationary models for the same 200 trajectories.

indicate that the likelihood of the non-stationary model is higher. Overall, the non-stationary model has a smaller negative likelihood in 146 out of the 200 trajectories and is therefore seen to be the better model in general.

There are other regions of the global oceans, in addition to the equator, where the non-stationary methods of this thesis may significantly improve parameter estimates of drifter time series. These include drifters which follow currents that traverse across different latitudes, such as the Gulf Stream or the Kuroshio. Analysis of such data is an important avenue of future investigation.
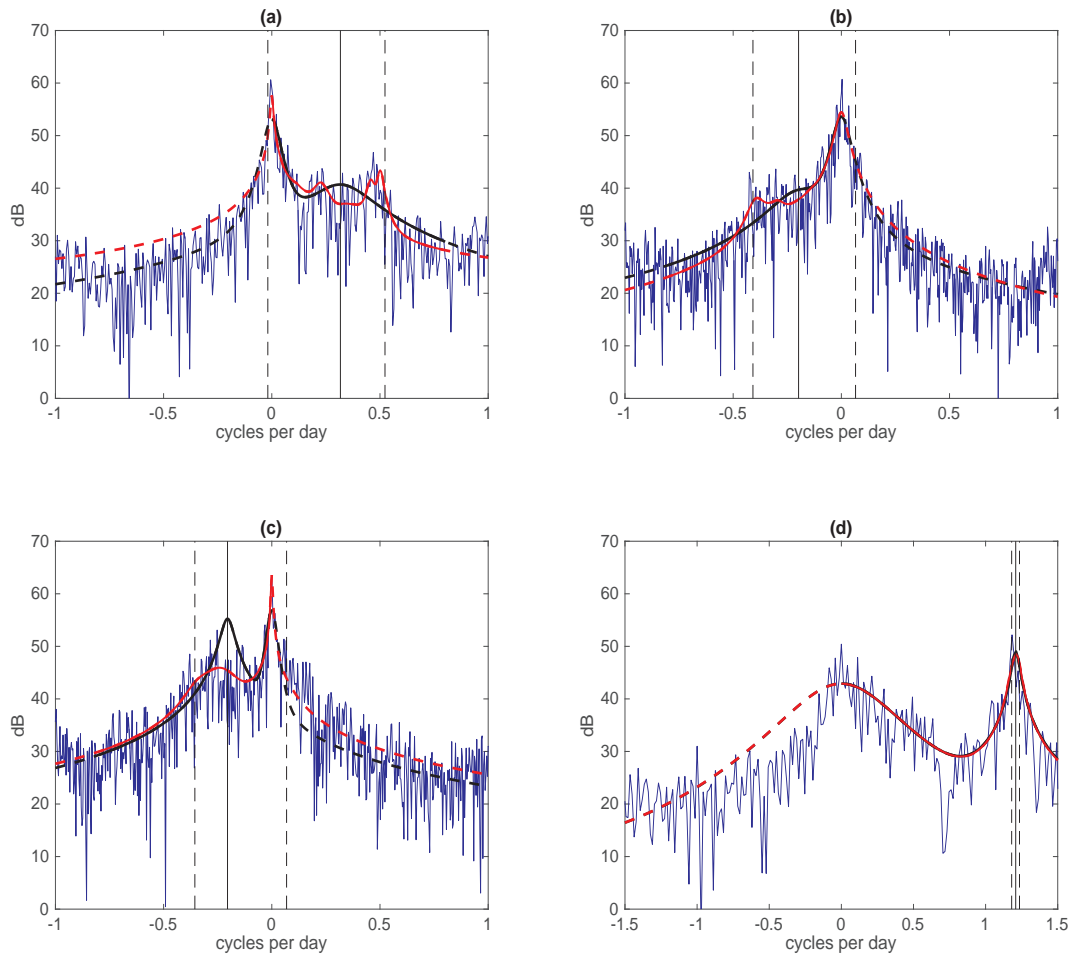
Figure 8.5: Fitted expected periodogram using either the stationary model (black) or the non-stationary model (red) to the periodogram (blue) for segments of data from drifter IDs (a) #79243, (b) #54656, (c) #71845 and (d) #44312. The solid black vertical line is the average inertial frequency, and the dashed vertical black lines are the minimum and maximum observed inertial frequency over the observed time window. The models are fit in the frequency range of 0 to 0.8 cycles per day in (a)–(c), and from 0 to 1.5 cycles per day in (d) as this drifter is at a higher latitude of 37° S where inertial oscillations occur at a frequency of about 1.2 cycles per day. The fitted models are shown in solid lines within the frequency range, and in dashed lines outside the frequency range.

Chapter 9

Future work and conclusions

Non-stationary time series occur in many real-data applications, and hence have attracted the attention of a large amount of the time series literature in recent decades. The first difficulty in defining non-stationary models is to ensure their validity in terms of the resulting covariance matrices. This issue is exacerbated in multi-dimensional time series. The second difficulty is that, when defining a non-stationary stochastic processes model, one has to choose a way in which the information will repeat itself despite the non-stationarity, so that averaging makes sense when trying to estimate parameters or obtain summary statistics. Finally, the computational burden of the maximum likelihood is well-known for the analysis of many stationary time series, and is obviously not expected to disappear in the case of non-stationary time series. One approach is the use of likelihood approximations called quasi-likelihoods. Deriving such approximations and establishing their consistency is usually an additional difficulty for non-stationary time series.

In this thesis we have proposed a class of non-stationary processes based on the framework of modulated processes (Parzen, 1963). The framework of modulation makes a lot of the problems described in the above paragraph easier. Firstly, modulation, and more generally defining a process pointwise as $f_t(X_t)$ where $f_t(\cdot)$ are functions indexed by time and $\{X_t\}$ is a stochastic process, ensures the validity of the so-defined process, and in the simple case of modulation the covariance of the modulated process can easily be derived from the covariance of the latent process. In this thesis we also introduced a class of non-stationary bivariate modulated

processes, where the modulation can efficiently account for frequency modulation. Secondly, the way the information repeats itself in the observed modulated process can be expressed in terms of the relation between the modulation sequence and the latent stationary process. A simple possible assumption is that the sample autocovariances of the modulation sequence converge to a non-zero limit, leading to an asymptotically stationary modulated process (Parzen, 1963, Dunsmuir and Robinson, 1981c, Toloi and Morettin, 1989). The central contribution of this thesis is to extend this idea to allow for more general forms of the modulation sequence, by introducing the class of modulated processes with a *significant correlation contribution*, where we require that the sample autocovariance sequence of the modulation sequence be asymptotically bounded below, at least for a set of lags that entirely determine the parameter within its range of possible values. Finally, we introduced a quasi-likelihood in the frequency domain, which is based on the Whittle likelihood for stationary processes and we conserve the $\mathcal{O}(N \log N)$ computational order. We proved its consistency as well as its rate of convergence under mild assumptions on the latent process and the modulation sequence.

From a broader point of view, a key contribution of this thesis is to show how parametric modelling may permit forward modelling of statistics from time series samples. While most methods in time series spectral analysis intend to limit the bias of estimators, we have directly modelled the expectation of such an estimator, and made use of the parametric modelling to transform the estimation problem in an inverse problem. We note that Fuentes (2007) proposed a similar method to estimate the parameters of irregularly sampled random fields.

As a direct extension of the work presented in this thesis, in future work we may consider establishing the asymptotic distribution of the quasi-likelihood estimator introduced, in order to be able to provide confidence intervals. Another challenging possible extension is to consider the estimation of a latent parametric model when the modulation sequence is not observed and does not admit a parametric form, in contrast to the problem we considered in Section 5.4. This cannot be achieved for any modulation sequence and will require some assumptions, such as

smoothness assumptions. We note the difference with locally-stationary processes, as we may still be able to achieve consistent estimates of the latent model parameter under increasing sample size.

Concerning our real-data application, future work will consist in establishing a global map of the world oceans showing the variations of key physical quantities such as the damping timescale which we analysed in Section 8.4. This will pose several challenges, such as extending our methods to replicated modulated time series, and choosing an optimal window length that may depend on the variations of the Coriolis frequencies and possibly on the estimated parameter (in which case the estimation would go through several estimations).

The Whittle likelihood can be applied more generally to the study of random fields (Whittle, 1954), that is to say stochastic processes on a multidimensional space, for instance $\mathbb{R}^2$, rather than on the time line. Although one may think that results from time series analysis can be extended to random fields directly, there are two issues which require more consideration. Firstly, the main difficulty in going from dimension 1 to dimension $d > 1$ is referred to as *edge effects* (Dahlhaus and Künsch, 1987). This refers to the fact that the bias of the periodogram decreases with rate $N^{-1/d}$ where N is the total sample size, making small sample effects of time series even more cumbersome in the analysis of random fields. Secondly, the restrictions on covariance functions are stronger. More specifically, if $\{X_s : s \in D\}$ is a homogeneous (or stationary) random field on $D \in \mathbb{R}^d$, and if $H \subset D$ is a subset with dimension $d - 1$, the autocovariance function of the process $\{X_s : s \in H\}$ clearly belongs to the set of stationary autocovariance functions of processes on $\mathbb{R}^{d-1}$, but the reciprocal is not true (Yaglom, 1987). In particular, the set of autocovariance functions from lines of random fields on $\mathbb{R}^2$ is a strict subset of non-negative definite functions. Similarly to this thesis work, we can easily define a non-stationary random field via modulation. Eliminating the bias of the Whittle likelihood by replacing the spectral densities by the finite sample expectations of the periodogram may be even more valuable than in dimension 1, due to edge effects. Moreover, calculation techniques for the expected periodogram can be extended to

random fields, including applications to non-rectangular regular grids, which is of great interest to practitioners as spatial data very often come from non-rectangular shapes. This provides with an enticing possible extension of the work presented in this thesis.

Appendix A

Proofs

The proofs in this appendix are standard. All the proofs in the main body of this thesis are by the author.

A.1   Proof of Lemma 1

Proof. Let $\{X_t : t \in \mathbb{Z}\}$ be a stationary process, with autocovariance function $c_X(\tau)$. Let $n \geq 1$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. We have that,

$$\text{var}\left\{\sum_{i=1}^{n} \lambda_i X_i\right\} = \sum_{i,j=1}^{n} \lambda_i \lambda_j c_X(i-j),$$

which is always non-negative from basic properties of the variance. Moreover, for any given $\tau \in \mathbb{Z}$, $c_X(-\tau) = \text{E}\{X_0 X_{-\tau}\} = \text{E}\{X_{-\tau} X_0\} = c_X(\tau)$, so that the autocovariance function of the process X_t is symmetric. Respectively, it can be shown that for any symmetric non-negative definite sequence, there exists a stochastic process whose autocovariance function is identical to that given sequence. This part of the proof can be found for example in Brockwell and Davis (1991). □

A.2   Proof of Proposition 1

Proof. Since $|c_X(\tau)e^{-i\omega\tau}| \leq c_X(\tau)$, the trigonometric series $\sum_{\tau \in \mathbb{Z}} c_X(\tau)e^{-i\omega\tau}$ converges normally. Therefore we can invert integration and summation, and in partic-

ular, for any natural integer k,

$$\int_{-\pi}^{\pi} \sum_{\tau=-\infty}^{\infty} c_X(\tau) e^{-i\omega\tau} e^{i\omega k} d\omega = \sum_{\tau=-\infty}^{\infty} \int_{-\pi}^{\pi} c_X(\tau) e^{-i\omega(\tau-k)} = c_X(k).$$

This concludes the proof, as it shows that $c_X(\tau) = \int_{-\pi}^{\pi} f(\omega) e^{-i\omega\tau}, \forall \tau \in \mathbb{Z}$. $\quad\square$

A.3 Proof of Lemma 2

Proof. The result is a direct consequence of the fact that the autocovariance function is non-negative definite (see Lemma 1), and from Herglotz's theorem on spectral representations of this class of functions. Equation (2.2) is a known result from Fourier theory. $\quad\square$

A.4 Proof of Theorem 3

Proof.   1. Assume $lim_{\tau\to\infty} c_X(\tau) = 0$. Then we have,

$$\begin{aligned}
\text{var}\left\{\frac{1}{N}\sum_{t=0}^{N-1} X_t\right\} &= \frac{1}{N^2} \sum_{i,j=1,\dots,N} \text{cov}\{X_i, X_j\} \\
&= \frac{1}{N} \sum_{|\tau|=0,\cdots,N-1} \left(1 - \frac{\tau}{N}\right) c_X(\tau) \\
&\leq \frac{1}{N} \sum_{|\tau|=0,\cdots,N-1} |c_X(\tau)|.
\end{aligned}$$

The last quantity converges to zero, since the average of a convergent sequence converges to the limit of that sequence. Therefore, and since it is unbiased, the sample mean converges in mean square, and thus in probability, to μ.

2. Assume that $\sum_{\tau\in\mathbb{Z}} |c_X(\tau)|$ converges. We then have,

$$\text{var}\left\{\frac{1}{N}\sum_{t=0}^{N-1} X_t\right\} = \frac{1}{N} \sum_{|\tau|=0,\cdots,N-1} \left(1 - \frac{\tau}{N}\right) c_X(\tau).$$

Let $f_n(x) = \left(1 - \frac{\lfloor |x| \rfloor}{n}\right) c_X(\lfloor |x| \rfloor) 1_{[0,n]}(|x|)$ and $f(x) = c_X(\lfloor |x| \rfloor)$, we have,

$$\frac{1}{N} \sum_{|\tau|=0,\cdots,N-1} \left(1 - \frac{\tau}{N}\right) c_X(\tau) = \int_{-\infty}^{\infty} f_n(x) dx,$$

and $f_n(x) \to f(x)$ and $|f_n(x)| < |f(x)|$ with $f(x)$ integrable. By the dominated convergence theorem we obtain,

$$\sum_{|\tau|=0,\cdots,N-1} \left(1 - \frac{\tau}{N}\right) c_X(\tau) \to \sum_{\tau \in \mathbb{Z}} c_X(\tau),$$

leading to the stated result.

$\square$

A.5   Proof of Proposition 7

Proof. The proof, which is standard (Brockwell and Davis, 1991, page 334), follows directly from (4.1) and (3.3) in a few lines of algebra after aggregating along the diagonal of the covariance matrix. $\square$

Bibliography

A. J. Abascal, S. Castanedo, R. Medina, and M. Liste. Analysis of the reliability of a statistical oil spill response model. *Marine Pollution Bulletin*, 60(11):2099–2110, 2010.

C. Allefeld, H. Atmanspacher, and J. Wackermann. Mental states as macrostates emerging from brain electrical dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19:015102, 2009.

M. R. Allen and A. W. Robertson. Distinguishing modulated oscillations from coloured noise in multivariate datasets. *Climate Dynamics*, 12(11):775–784, 1996.

T. Andrews, J. M. Gregory, M. J. Webb, and K. E. Taylor. Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, 39(9), 2012.

M. Anitescu, J. Chen, and L. Wang. A matrix-free approach for solving the parametric Gaussian process maximum likelihood problem. *Society for industrial and applied mathematics*, 34(3):A1460–A1487, 2012.

M. Arató, A. N. Kolmogorov, and Ya. G. Sinai. Evaluation of the parameters of a complex stationary Gauss-Markov process. *Doklady Akademii Nauk SSSR*, 146: 747–750, 1962.

M. Arató, S. Baran, and M. Ispány. Functionals of complex Ornstein-Uhlenbeck processes. *Computers & Mathematics with Applications*, 37(1):1 – 13, 1999.

A. Biastoch, J. V. Durgadoo, and S. Rühs. Tracing the drift of MH370 debris throughout the Indian Ocean. In *EGU General Assembly Conference Abstracts*, volume 19 of *Geophysical Research Abstracts*, page 7553, April 2017.

P. Bloomfield. Spectral analysis with randomly missing observations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(3):369–380, 1970.

P. Bloomfield. *Fourier analysis of time series: an introduction*. Wiley Series in Probability and Statistics, 2000.

P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer, 2nd edition, 1991.

P. M. T. Broersen, S. De Waele, and R. Bos. Autoregressive spectral analysis when observations are missing. *Automatica*, 40(9):1495–1504, 2004.

W. Clinger and J. W. V. Ness. On unequally spaced time points in time series. *The Annals of Statistics*, 4(4):736–745, 07 1976.

R. Dahlhaus. Small Sample Effects in Time Series Analysis: A New Asymptotic Theory and a New Estimate. *The Annals of Statistics*, 16(2):808–841, 1988.

R. Dahlhaus. On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications*, 62:139–168, 1996.

R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25(1):1–37, 1997.

R. Dahlhaus and H. Künsch. Edge effects and efficient parameter estimation for stationary random fields. *Biometrika Trust*, 74(4):877–882, 1987.

W. Dunsmuir and P. M. Robinson. Estimation of time series models in the presence of missing data. *Journal of the American Statistical Association*, 76(375):560–568, 1981a.

W. Dunsmuir and P. M. Robinson. Parametric estimators for stationary time series with missing observations. *Advances in Applied Probability*, 13(1):129–146, 1981b.

W. Dunsmuir and P. M. Robinson. Asymptotic theory for time series containing missing and amplitude modulated observations. *Sankhyā: The Indian Journal of Statistics, Series A*, 43(3):260–281, 1981c.

K. O. Dzhaparidze and A. M. Yaglom. Spectrum parameter estimation in time series analysis. In P.R. Krishnaiah, editor, *Developments in statistics*, volume 4, chapter 1. New York: Academic Press, 1983.

J. J. Early. The forces of inertial oscillations. *Quarterly Journal of the Royal Meteorological Society*, 138(668):1914–1922, 2012.

S. Elipot, R. Lumpkin, and G. Prieto. Modification of inertial oscillations by the mesoscale eddy field. *Journal of Geophysical Research: Oceans*, 115(C09010), 2010.

S. Elipot, R. Lumpkin, R. C. Perez, J. M. Lilly, J. J. Early, and A. M. Sykulski. A global surface drifter data set at hourly resolution. *Journal of Geophysical Research: Oceans*, 121(5):2937–2966, 2016.

R. Ferrari and C. Wunsch. Ocean circulation kinetic energy: reservoirs, sources, and sinks. *Annual Review of Fluid Mechanics*, 41(1):253–282, 2009.

M. Fuentes. Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*, 102(477):321–331, 2007.

W. A. Gardner, A. Napolitano, and L. Paura. Cyclostationarity: half a century of research. *Signal Processing*, 86(4):639–697, 2006.

E. G. Gladyshev. Periodically and almost-periodically correlated random processes with a continuous time parameter. *Theory of Probability and its Applications*, 8 (2):173–177, 1963.

T. Gneiting, W. Kleiber, and M. Schlather. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105 (491):1167–1177, 2010.

U. Grenander and G. Szegö. *Toeplitz forms and their applications*. University of California Press, 2001.

Y. Grenier. Time-dependent ARMA modeling of nonstationary signals. *IEEE Transactions on Acoustics, speech, and signal processing*, 31(4):899–911, 1983.

A. Griffa, A. D. Kirwan, A. J. Mariano, T. Özgökmen, and T. Rossby. *Lagrangian analysis and prediction of coastal and ocean dynamics*. Cambridge University Press, 2007.

A. P. Guillaumin, A. M. Sykulski, S. C. Olhede, J. J. Early, and J. M. Lilly. Analysis of non-stationary modulated time series with applications to oceanographic surface flow measurements. *Journal of Time Series Analysis*, 38(5):668–710, 2017.

X. Guyon. Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika Trust*, 69(1):95–105, 1982.

E. J. Hannan. The asymptotic theory of linear time-series models. *Journal of Applied probability*, 10(1):130–145, 1973.

R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, 1985.

H. L. Hurd and A. Miamee. *Periodically Correlated Random sequences: Spectral Theory and Practice*. Series in Probability and Statistics. Wiley, 2007.

L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2): 134–139, 1918.

J. R. Jain. An efficient algorithm for a large toeplitz set of linear equations. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(6):612–615, 1979.

J. Jiang and Y. V. Hui. Spectral density estimation with amplitude modulation and outlier detection. *Annals of the Institute of Statistical Mathematics*, 56(4):611–630, 2004.

R. H. Jones. Spectral analysis with regularly missed observations. *The Annals of Mathematical Statistics*, 33(2):455–461, 1962.

R. H. Jones. Maximum likelihood fitting of ARMA Models to time series with missing observations. *Technometrics*, 22(3):389–395, 1980.

J. H. LaCasce. Statistics from lagrangian observations. *Progress in Oceanography*, 77(1):1–29, 2008.

A. Le Breton. A note on maximum likelihood estimation for the complex-valued first-order autoregressive process. *Statistics & Probability Letters*, 7(2):171–173, 1988.

J. M. Lilly. jLab: A data analysis package for matlab, v. 1.6.2, http://www.jmlilly .net/jmlsoft.html, 2016.

J. M. Lilly and S. C. Olhede. Analysis of modulated multivariate oscillations. *IEEE Transactions on Signal Processing*, 60(2):600–612, 2012.

J. M. Lilly, A. M. Sykulski, J. J. Early, and S. C. Olhede. Fractional brownian motion, the matérn process, and stochastic modeling of turbulent dispersion. *Nonlinear Processes in Geophysics*, 24(3):481–514, 2017.

M. Loève. Fonctions aléatoires du second ordre. *Comptes Rendus de l'Académie des Sciences*, page 220, 1945.

R. Lumpkin and M. Pazos. Measuring surface currents with Surface Velocity Program drifters: the instrument, its data, and some results. In A. Griffa, A. D. Kirwan, A. J. Mariano, T. Özgökmen, and T. Rossby, editors, *Lagrangian analysis and prediction of coastal and ocean dynamics*, chapter 2, pages 39–67. Cambridge University Press, 2007.

P. McCullagh. Quasi-likelihood functions. *The Annals of Statistics*, 11(1):59–67, 1983.

G. Mélard and A. Herteleer-De Schutter. Contributions to evolutionary spectral theory. *Journal of Time Series Analysis*, 10(1):41–63, 1989.

K. S. Miller. Complex Gaussian processes. *Siam Review*, 11(4):544–567, 1969.

E. Paparoditis and D. N. Politis. Nonlinear spectral density estimation : thresholding the correlogram. *Journal of Time Series Analysis*, 33(March):386–397, 2012.

E. Parzen. Spectral analysis of asymptotically stationary time series. Technical report, DTIC Document, 1961.

E. Parzen. On spectral analysis with missing observations and amplitude modulation. *Sankhyā: The Indian Journal of Statistics. Series A*, 25(4):383–392, 1963.

D. B. Percival and A. T. Walden. *Spectral analysis for physical applications*. Cambridge University Press, 1993.

R. T. Pollard and R. C. Millard. Comparison between observed and simulated wind-generated inertial oscillations. *Deep Sea Research and Oceanographic Abstracts*, 17(4):813–821, 1970.

M. B. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(2):204–237, 1965.

M. B. Priestley. *Non-linear and non-stationary time series analysis*. Academic Press, 1988.

G. A. Prieto, R. L. Parker, D. J. Thomson, F. L. Vernon, and R. L. Graham. Reducing the bias of multitaper spectrum estimates. *Geophysical Journal International*, 171(3):1269–1281, 2007.

M. M. Rao. Harmonizable, Cramèr, and Karhunen classes of processes. *Handbook of Statistics*, 5:279–310, 1985.

P. M Robinson. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23(5):1630–1661, 1995.

G. Rünstler. Modelling phase shifts among stochastic cycles. *The Econometrics Journal*, 7:232–248, 2004.

P. A. Scheinok. Spectral analysis with randomly missed observations: the binomial case. *The Annals of Mathematical Statistics*, 36(3):971–977, 1965.

P. J. Schreier and L. L. Scharf. *Statistical signal processing of complex-valued data*. Cambridge University Press, 2010.

A. M. Sykulski, S. C. Olhede, A. P. Guillaumin, J. M. Lilly, and J. J. Early. The de-biased Whittle likelihood for second-order stationary stochastic processes. *arXiv preprint: 1605.06718v1*, 2016a.

A. M. Sykulski, S. C. Olhede, and J. M. Lilly. A widely linear complex autoregressive process of order one. *IEEE Transactions on Signal Processing*, 64(23): 6200–6210, 2016b.

A. M. Sykulski, S. C. Olhede, J. M. Lilly, and E. Danioux. Lagrangian time series models for ocean surface drifter trajectories. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(1):29–50, 2016c.

M. Taniguchi. On estimation of parameters of Gaussian stationary processes. *Journal of Applied Probability*, 16(3):575–591, 1979.

D. J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.

C. M. C. Toloi and P. A. Morettin. Spectral analysis for amplitude-modulated time series. *Journal of Time Series Analysis*, 14(4):409–432, 1989.

H. Tong. Some comments on spectral representations of non-stationary stochastic processes. *Journal of Applied Probability*, 10(4):881–885, 1973.

H. Tong. On time-dependent linear transformations of non-stationary stochastic processes. *Journal of Applied Probability*, 11(1):53–62, 1974.

A. W. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

A. M. Walker. Periodogram analysis for complex-valued time series. In T. Subba Rao, editor, *Developments in Time Series Analysis*, pages 149–163. Chapman and Hall, 1993.

M. Watanabe and T. Hibiya. Global estimates of the wind-induced energy flux to inertial motions in the surface mixed layer. *Geophysical Research Letters*, 29(8): 64–1–64–3, 2002.

R. W. M. Wedderburn. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61(3):439–447, 1974.

P. Whittle. Estimation and information in stationary time series. *Arkiv för Matematik*, 2(5):423–434, 1953.

P. Whittle. On stationary processes in the plane. *Biometrika*, 41(3):434–449, 1954.

A. M. Yaglom. *Correlation Theory of Stationary and Related Random Functions*. Springer, 1987.