Pontificia Universidad Católica de Chile
Faculty of Mathematics
Department of Statistics

Fitting time-varying parameters to astronomical time series

Darlin Macarena Soto Vásquez

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR IN STATISTICS

Ph.D. Advisor : Giovanni Motta, Texas A&M University, USA
Academic UC Advisor : Manuel Galea, Pontificia Universidad Católica de Chile, Chile

July 2022

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
DEPARTMENT OF STATISTICS

The undersigned hereby certify that they have read and recommend to the Faculty of Mathematics for acceptance a thesis entitled **Fitting time-varying parameters to astronomical time series** by **Darlin Macarena Soto Vásquez** in partial fulfillment of the requirements for the degree of **Doctor in Statistics**.

Dated: July the 4th, 2022

Ph.D. Advisor	Giovanni Motta Texas A&M University, USA
Academic UC Advisor	Manuel Galea Pontificia Universidad Católica de Chile, Chile
Examining Committee	Jonathan Acosta Pontificia Universidad Católica de Chile, Chile
	Ronny Vallejos Universidad Técnica Federico Santa María, Chile
	Mauricio Zevallos Universidade Estadual de Campinas, Brasil

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Date: July 2022

Author	:	Darlin Macarena Soto Vásquez
Title	:	Fitting time-varying parameters to astronomical time series
Department	:	Statistics
Degree	:	Doctor in Statistics
Convocation	:	July
Year	:	2022

TO LILIANA VÁSQUEZ, DESIDERIO SOTO
AND CRISTÓBAL SOTO

Acknowledgements

I would first like to thank my PhD advisor, professor Giovanni Motta, for his guidance, help, and support during this process. His expertise was invaluable to developing this research and his feedback pushed me to continue working and bringing my work to a higher level.

I want to thank Professor Márcio Catelan and Dr. Malgorzata Sobolewska for introducing me to the world of astrophysics. Their scientific knowledge helped me understanding how to model astronomical light curves by means of time series analysis.

I also want to thank the professors in the Statistics department, for giving me the opportunity to be part of the Ph.D. program where I learned a lot. I would also like to thank my academic advisor professor Manuel Galea, as well as the members in the Faculty of Mathematics, for all their administrative support.

Finally, I gratefully acknowledge the financial support awarded by the National Agency for Research and Development (ANID) / Scholarship Program / Doctorado Nacional /2017-21171100.

Contents

List of Tables

List of Figures

List of Symbols

$\boldsymbol{y}$ Data vector

$\boldsymbol{z}$ White noise error

$m(t)$ Trend at time t

$g_{\ell,k}(t)$ Amplitude at time t

$\mathbb{E}[y]$ Expectation of the random variable y

$\mathbb{V}\mathrm{ar}[y]$ Variance of the random variable y

$\mathbb{C}\mathrm{ov}[x, y]$ Covariance between the two random variables x and y

$r_x(h)$ Autocovariance at lag h of the process $\{x_i\}$

$\rho_x(h)$ Autocorrelation at lag h of the process $\{x_i\}$

$P_x(\lambda)$ Spectral density at frequency λ of the process $\{x_i\}$

$W_x(\lambda)$ Spectral window at frequency λ of the process $\{x_i\}$

$\phi_1, \ldots, \phi_p$ Autoregressive coefficients of ARMA processes

$\theta_1, \ldots, \theta_q$ Moving-average coefficients of ARMA processes

$\phi_1(u), \ldots, \phi_p(u)$ Time-varying autoregressive coefficients of LSARMA processes

$\theta_1(u), \ldots, \theta_q(u)$ Time-varying moving-average coefficients of LSARMA processes

$\boldsymbol{B}(t)$ Base of B-splines at time t

d Order of the B-splines

$\tau_1, \ldots, \tau_{2K+1}$ Regularization parameters associated with the trend and amplitudes

$\tau_1^{\phi}, \ldots, \tau_p^{\phi}$ Regularization parameters associated with the time-varying autoregressive coefficients

$\tau_1^{\theta}, \ldots, \tau_q^{\theta}$ Regularization parameters associated with the time-varying moving-average coefficients

r Order of the penalization

$\mathcal{F}\{\cdot\}$ Discrete Fourier Transform

$\mathcal{F}^{-1}\{\cdot\}$ Inverse Discrete Fourier Transform

Chapter 1

Introduction

The night sky seems immutable to the human eye, except for some astronomical objects. In fact, the brightness of stars seems to be unchanged from our childhood until now. However, the interest to understand the mysterious luminous objects in the sky is not new. Many ancient civilizations, such as Egyptian and Greek, carried out quantitative measurements and suggested that a few stars brightness changed over time. Many years later, as the telescopic studies increased more variable stars were found with a wide range of features.

Nowadays, our ability to see deeper into the universe has improved and more samples of cosmic populations are available to be studied. Celestial objects exhibit a wide range of variability in brightness at different wavebands. The study of the brightness of a celestial object belongs to a discipline called "time domain astronomy". Time domain astronomy is the study of how the brightness of astronomical objects changes with time, including the time-dependent behaviour. Examples of studies in the time domain are the pulsar variability, and the variability of accreting black holes, variable stars, and the Sun. To carry out these studies enormous investments in telescopes for repeated measurements over time are made, as for example, the Vera C. Rubin Observatory constructed in the high Atacama Desert of Chile (Feigelson et al., 2018, 2021).

Gaining knowledge from astronomical datasets requires a wide range of sophisticated statistical methods. In this thesis we provide two novel approaches to analyze light curves

of variable astronomical objects. First, we introduce a model to describe periodic variable stars. Second, we propose a model to describe the variability of X-ray binary systems.

Our model for periodic variable stars allows to describe their behavior from a more flexible point of view compared to existing approaches. More specifically, we introduce a model for the description of the so-called modulated light curves that accounts for a smoothly time-varying trend and amplitudes without assuming a closed-form. We estimate the curves of this model in both cases of uncorrelated and autocorrelated errors. Moreover, in the case of unequally spaced time series, we introduce a procedure to detect serial correlation in the residuals of our fitted model.

In the case of X-ray binary systems, the objective is to propose a model to describe the nonstationarity observed in the frequency domain using the spectral density. For this reason, we adopt locally stationary autoregressive moving-average (ARMA) processes. Under some assumptions, we show that the spectral density of an autoregressive moving-average process is suitable to describe the spectral density of X-ray binary systems. Then, assuming a nonparametric form for the time-varying coefficients of locally stationary processes, we propose a method to estimate these coefficients and calculate their confidence intervals.

This introductory chapter is organized as follows. In Section 1.1 we summarize the keys ingredients of the most popular time series models used in astronomy. In Section 1.2 we review the main concepts in time series such as mean function, covariance function, stationarity, autocovariance function, spectral density, among others. We also review the definitions of some processes as autoregressive moving-average processes, continuous-time autoregressive moving-average processes, and the models with trend and seasonality, and the spectral density for continuous unequally spaced time series. All these models involve time-invariant parameters. In Section 1.3 we review examples of nonstationary astronomical objects and, in more detail, two classes of models used in astronomy to describe their nonstationary. The first class is the modulation models used to fit periodic variable stars (the nonstationary is observed in the mean), and the second one is represented by the X-ray binary systems where the periodogram is used to estimate their spectral density (the non-

stationary is observed in the errors). For both models we also illustrate their limitations, and clarify why making the use of time-varying parameters is more appropriate. In Section 1.4 we introduce the definition and the spectral density of locally stationary processes, and in Section 1.5 we review the literature on some methods to estimate their time-varying coefficients. Finally in Section 1.6 we present the outline of the thesis, and in Section 1.7 the current status of our published work. In this thesis we denote vectors with bold letters, and matrices with bold capital letters.

1.1 Time series in astronomy

As we mentioned in the introduction, time domain astronomy is the study of how astronomical objects change with time. Models to study these changes involve ARMA, ARIMA, ARFIMA, ARCH, GARCH processes for equally spaced data, CARMA and IAR processes for unequally spaced data, and models that include the seasonality. ARMA, ARIMA, and ARFIMA models are attractive for astronomical time series analysis for various reasons. Some of them are 1) models are flexible and can fit several of irregular, quasi-period, or smooth light curves, 2) the number of parameters to be estimated is low, and therefore the computational time is moderate, and 3) they can be extended to the multivariate case, combinations of stochastic and deterministic parts, and moderate irregular observation spacing. A mathematical limitation of ARMA, ARIMA, and ARFIMA models is the restriction to evenly spaced datasets. In astronomy, observations are limited by daily and annual celestial cycles, as well as telescope allocation constraints. As a consequence, the observations are unequally spaced. To deal with this, astronomers proposed to use CARMA models. The main reasons to use CARMA models in astronomy are 1) models give a treatment of irregular time sampling and 2) models have favorable mathematical properties for astronomical use (Feigelson et al., 2018). The favorable property of CARMA processes is that the spectral density can be expressed as a sum of Lorentzian functions (Kelly et al., 2014). Many spectral densities of accretion disk systems with quasi-periodic oscillations and red noise can be described by Lorentzian functions.

Examples of fitting ARMA processes to astronomical data are variate. For example, Feigelson et al. (2018) fitted ARMA, ARIMA, and ARFIMA processes to the light curve of variable stars KIC 005880320, KIC 004276716, HATS-2b, and RR Hyi. Related to CARMA processes, Graham et al. (2015) used these processes to describe the light curve of the quasar PG 1302-102, Guo et al. (2017) applied the models to observations of SDSS Stripe 82 quasars, and Kelly et al. (2014) fitted these models to the light curve of the X-ray binary XTE 1550-564, two Active Galactic Nuclei, and two variable stars.

1.2 Time series models with time-invariant parameters

In this section, we review the concepts of stationarity, nonstationarity, covariance and autocovariance functions, and spectral density defined for time series. We also introduce important parametric families of stationary and nonstationary time series with time-invariant parameters, as the autoregressive moving-average, or ARMA, processes (see Section 1.2.1), the continuous-time autoregressive moving-average, or CARMA, processes (see Section 1.2.2), and time series with harmonic components (see Section 1.2.3). Finally, we introduce the concept of unequally spaced data and its spectral density (see Section 1.2.4).

Let $\{x_t,\, t = 1, \ldots, T\}$ be a time series with $\mathbb{E}\left[x_t^2\right]$. The *mean function* of $\{x_t\}$ is

$$\mu_x(t) = \mathbb{E}\left[x_t\right],$$

and *covariance function* of $\{x_t\}$ is

$$\gamma_x(r, s) = \mathbb{C}\text{ov}\left[x_r, x_s\right] = \mathbb{E}\left[\{x_r - \mu(r)\}\{x_s - \mu(s)\}\right].$$

$\{x_t\}$ is a *strictly stationary* time series if, for all $n > 0$ and $h \in \mathbb{N}$

$$\mathbb{P}(x_1 < x_1, \ldots, x_n < x_n) = \mathbb{P}(x_{1+h} < x_1, \ldots, x_{n+h} < x_n).$$

$\{x_t\}$ is a *weakly stationary* time series if

i) $\mu_x(t)$ is independent of t,

ii) $\gamma_x(t+h,t)$ is independent of t for each h.

Whereas the concept of strict stationarity is difficult to satisfy for any time series, the concept of weakly stationary is a relaxer requirement because we focus only on the first and second moments. We use the term *stationary* to refer to weakly stationary. When $\{x_t\}$ is neither strictly nor weakly stationary, $\{x_t\}$ is *nonstationary*. It means that the statistical properties are changing over time.

The *autocovariance function* of $\{x_t\}$ at lag h is

$$r_x(h) = \mathbb{C}\text{ov}\left[x_{t+h}, x_t\right], \quad h \in \mathbb{Z},$$

and the *autocorrelation function* at lag h (ACF), is defined as

$$\rho_x(h) = \frac{r_x(h)}{r_x(0)}.$$

$\{x_t\}$ is a *linear process* if for all t it has the representation

$$x_t = \psi(B)z_t = \sum_{j=-\infty}^{\infty} \psi_j \, z_{t-j},$$

where B is the backward shift operator, $B^k z_t = z_{t-k}$, z_t is a white noise process with zero-mean and variance σ_z^2, and the coefficients $\{\psi_j\}$ are absolutely summable:

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

$\{z_t\}$ is a *white noise*, or WN, process with mean zero and variance σ_z^2 if it satisfies

$$\mathbb{E}\left[z_t\right] = 0 \quad \text{and} \quad \mathbb{E}\left[z_i \, z_j\right] = \delta_{\{i=j\}}\sigma_z^2, \quad \text{for all } i, j = 1, \ldots, T,$$

where $\delta_{\{i=j\}} = 1$ if $i = j$ and zero otherwise.

A linear process $\{x_t\}$ can be also written in terms of the spectral representation, that is,

$$x_t = \sum_{-\pi}^{\pi} \exp(i\omega t) A(\omega) d\xi(\omega),$$

where $A(\omega)$ is a transfer function and $\xi(\omega)$ is a orthogonal increments process on $[-\pi, \pi]$ such that

$$\mathbb{C}\text{ov}\left[\xi(\omega), \xi(\lambda)\right] = \frac{\sigma_z^2}{2\pi}\delta(\omega - \lambda)d\lambda d\omega.$$

The *spectral density* (PSD) of a linear process $\{x_t\}$ in function of the coefficient $\{\psi_j\}$ is given by

$$P_x(\omega) = \frac{\sigma_z^2}{2\pi}|\psi(\exp(i\omega))|^2.$$

The PSD in terms of the autocovariance function is

$$P_x(\omega) = \frac{1}{2\pi}\sum_{-\infty}^{\infty} r_x(h)\exp(i\omega h),$$

and in terms of the transfer function is

$$P_x(\omega) = \frac{\sigma_z^2}{2\pi}|A(\omega)|^2.$$

1.2.1 Autoregressive moving-average processes

Let $\{x_t,\, t = 1, \ldots, T\}$ be a set observations occurring at certain discrete time $1, \ldots, T$. An ARMA(p, q) process $\{x_t,\, t = 1, \ldots, T\}$ with autoregressive coefficients $\phi_1, \ldots, \phi_p$, and moving-average coefficients $\theta_1, \ldots, \theta_q$, is defined to be a solution of the equation

$$x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} = z_t + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q}, \quad \{z_t\} \sim WN(0, \sigma_z^2), \tag{1.1}$$

where $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ are constants, the polynomials $1 - \phi_1 v - \cdots - \phi_p v^p$ and $1 + \theta_1 v + \cdots + \theta_q v^q$ have no common factors, and $\{z_t\}$ is a WN process with zero-mean and variance σ_z^2. An autoregressive AR(p) process corresponds to an ARMA$(p, 0)$ process given by

$$x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + z_t, \quad \{z_t\} \sim WN(0, \sigma_z^2).$$

An ARMA(p, q) process is *invertible* if all the roots v of the polynomial $\theta(v) = 1 + \theta_1 v + \cdots + \theta_q v^q$ satisfy $|v| > 1$. This means that the noise z_t can be expresses as

$$z_t = \sum_{j=0}^{\infty} \pi_j x_{t-j},$$

with constants π_j such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$.

An ARMA(p, q) process is *causal* if all the roots v of the polynomial $\phi(v) = 1 - \phi_1 v - \cdots - \phi_p v^p$ satisfy $|v| > 1$. This means that x_t can be expresses as

$$x_t = \sum_{j=0}^{\infty} \psi_j z_{t-j},$$

with constants ψ_j such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

The ARMA(p, q) process in equation (1.1) has the PSD

$$P_x(\omega) = \frac{\sigma_z^2}{2\pi} \frac{\left| 1 + \sum_{j=1}^{q} \theta_j \exp(-i\omega j) \right|^2}{\left| 1 - \sum_{j=1}^{p} \phi_j \exp(-i\omega j) \right|^2}, \quad \omega = 2\pi f, \quad -\infty < f < \infty,$$

and the autocovariance function at lag h in terms of the PSD is given by

$$r_x(h) = \int_{-\pi}^{\pi} \exp(i\omega h) P_x(\omega)\, d\omega, \quad h \in \mathbb{Z}.$$

1.2.2 Continuous-time autoregressive moving-average processes

A Gaussian CARMA(p, q) process $\{x(t),\, t \geq 0\}$ with $0 \leq q < p$, autoregressive coefficients $a_1, \ldots, a_p$, and moving-average coefficients $b_1, \ldots, b_q$, is defined as a strictly stationary solution of the pth-order linear differential equation,

$$D^p x(t) + a_1 D^{p-1} x(t) + \cdots + a_p x(t) = b_0 D w(t) + b_1 D^2 w(t) + \cdots + b_q D^{q+1} w(t), \quad (1.2)$$

where D^j denotes j-fold differentiation with respect to t, $\{w(t)\}$ is standard Brownian motion, $a_1, \ldots, a_p$, and $b_1, \ldots, b_q$, are constants, the polynomials $v^p + a_1 v^{p-1} + \cdots + a_p$ and $b_0 + b_1 v + \cdots + b_q v^q$ have no common zeroes, and the coefficients satisfy $b_q \neq 0$, $b_j = 0$ for $q < j \leq p$.

A CARMA(p, q) process is *causal* if the roots v of the polynomial $v^p + a_1 v^{p-1} + \cdots + a_p$ have negative real parts.

The PSD of the CARMA(p, q) process defined in equation (1.2) is

$$P_{x}(\omega) = \frac{1}{2\pi} \frac{\left| \sum_{k=0}^{q} b_{k} (i\omega)^{k} \right|^{2}}{\left| \sum_{j=0}^{p} a_{p-j} (i\omega)^{j} \right|^{2}}, \quad \omega = 2\pi f, \quad -\infty < f < \infty,$$

with $a_{0} = 1$, and the autocovariance function at lag h in terms of the PSD is given by

$$r_{x}(h) = \mathbb{C}\text{ov}\left[x(t+h), x(t)\right] = \int_{-\infty}^{\infty} \exp(-i\omega h) P_{x}(\omega) d\omega, \quad h \in \mathbb{R}.$$

1.2.3 Time series models with trend and seasonality

Many time series are influenced by a trend and seasonally varying factors such as the weather. The effect of which can be modeled by a periodic component with a fixed known period. In order to represent the trend and seasonal effect, we can define the model as a sum of a deterministic signal and a random noise. The deterministic part consists of the *trend component* $m(t)$ and the *periodic function* $s(t)$. Thus, the model is represented by

$$x_{t} = m(t) + s(t) + z_{t}, \tag{1.3}$$

where the trend $m(t)$ is a slowly changing aperiodic function, $s(t)$ is function of t with period d ($s(t-d) = s(t)$), and z_{t} has zero mean. A form for $s(t)$ is a sum of K harmonic components (or sine waves) given by

$$s(t) = a_{0} + \sum_{k=1}^{K} \left\{ a_{k} \cos(2\pi f_{k} t) + b_{k} \sin(2\pi f_{k} t) \right\},$$

where $a_{0}, a_{1}, \ldots, a_{K}$, and $b_{1}, \ldots, b_{K}$, are unknown parameters and $f_{1}, \ldots, f_{K}$, are fixed frequencies. The period associated with the frequency f_{k} is $1/f_{k}$.

1.2.4 Spectral density of unequally spaced data

In time series, data can be equally or unequally spaced. We defined an *equally spaced time series* if time is evenly sampled, that is, time satisfies the condition $t_{i} = t_{0} + i\Delta$

with $i \in \mathbb{N}$ and $\Delta > 0$ the constant data spacing. If time does not satisfy the condition $t_i = t_0 + i\Delta$, then the time series is called *unequally spaced*. Notice that, in some cases, unequally spaced time series can be seen as an equally spaced time series with gaps. Let x_i the observation at time t_i. An example of an equally spaced time series (first example) is

$$\{(x_1, t_1 = 3), (x_2, t_2 = 5), (x_3, t_3 = 7), (x_4, t_4 = 9), (x_5, t_5 = 11), (x_6, t_6 = 13), (x_7, t_7 = 15)\}$$

where $t_i = t_0 + i\Delta$ with $t_0 = 1$ and $\Delta = 2$, whereas an example of an unequally spaced time series (second example) is

$$\{(x_1, t_1 = 1), (x_2, t_2 = 3), (x_3, t_3 = 4), (x_4, t_4 = 10), (x_5, t_5 = 11), (x_6, t_6 = 13), (x_7, t_7 = 14)\}.$$

Notice that, it is not possible to identify t_0 and Δ in the second example. A third example of an unequally spaced time series is

$$\{(x_1, t_1 = 3), (x_2, t_2 = 5), (x_3, t_3 = 11), (x_4, t_4 = 13), (x_5, t_5 = 15)\}$$

which corresponds to the time series in the first example removing the observations at time $t = 3$ and 4. Therefore, for unequally spaced time series we distinguish two cases. In the first case, we can identify t_0 and Δ, but we do not observe x_i for all i (third example), and a the second case, we do not identify neither t_0 nor Δ for all i (second example).

Consider the case where the observed process $\{x_t\}$ has a purely continuous spectrum so that its PSD function exists for all f. The most standard models including the ARMA process in equation (1.1) and the CARMA process in equation (1.2) which have purely continuous spectral. The harmonic process in equation (1.3) has a purely discrete spectrum. Given a sample of T equally spaced observations $x_1, x_2, \ldots, x_t$, occurring at certain discrete time $1, 2, \ldots, T$. We can estimate the PSD of an ARMA or CARMA processes using the periodogram through the relationship (see Priestley, 1981, page 418)

$$\mathbb{E}\left[\frac{I_x(\omega)}{T}\right] \to 2\pi P_x(\omega), \tag{1.4}$$

where $I_x(\omega)$ is the periodogram defined by

$$I_x(\omega) = \left|\sum_{t=1}^{T} x_t \exp(-i\omega t)\right|^2.$$

The result in equation (1.4) shows that, for equally spaced time series, $I_x(\omega)$ is an asymptotically unbiased estimate of $P_x(\omega)$. However, when the time series are unequally spaced, the relationship in equation (1.4) does not hold, and the periodogram is not an asymptotically unbiased estimator of the PSD.

Let $\{x(t)\}$ be a *continuous* zero-mean stationary times series with spectral density

$$P_x(\omega) = \int_{-\infty}^{\infty} r_x(h) \exp(i\omega h) dh, \quad -\infty < \omega < \infty,$$

and autocovariance function given by

$$r_x(h) = \int_{-\infty}^{\infty} P_x(\omega) \exp(-i\omega h) d\omega, \quad h \in \mathbb{R},$$

and consider the observations of $x(t)$ at unequally spaced times $t_1, t_2, \ldots, t_N$, given by $x(t_1), x(t_2), \ldots, x(t_N)$. Deeming (1975) proved that the expectation of the periodogram of the observations $x(t_1), x(t_2), \ldots, x(t_N)$ is equal to the *continuous convolution* of the spectral density of $x(t)$ with a spectral window (see equation (36) in Deeming, 1975), that is,

$$\mathbb{E}\left[I_x(\omega)\right] = \mathbb{E}\left[\sum_{k=1}^{N}\sum_{j=1}^{N} x(t_j)x(t_k)\exp(i\lambda[t_k - t_j])\right] = P_x(\omega) \star W_x(\omega),$$

where $W_x(\lambda)$ is the *power spectral window* given by

$$W_x(\omega) = \sum_{j=1}^{N}\sum_{k=1}^{N} \exp(i\omega[t_k - t_j]),$$

and $P_x(\omega) \star W_x(\omega)$ is the continuous convolution of $P_x(\omega)$ with $W_x(\omega)$ defined as

$$P_x(\omega) \star W_x(\omega) = \int_{-\infty}^{\infty} P_x(\lambda) W_x(\omega - \lambda) d\lambda.$$

1.3 Why time-varying parameters?

In this section, we review examples of nonstationary astronomical time series and the models used by astronomers to describe them. In Section 1.3.1 the nonstationarity is given

by the time-varying mean, whereas in Section 1.3.2 the nonstationarity is observed in the time-varying PSD. In Section 1.3.1 we describe modulation models and the relationship with Blazhko stars, and in Section 1.3.2 we review the PSD of X-ray binaries and the models based on Lorentzian and power law functions used to fit their PSDs.

1.3.1 Modulation models

The Blazhko effect is a periodic amplitude and phase variation in the light curves of RR Lyrae variable stars. In astronomy, the Blazhko effect is usually interpreted as a modulation phenomenon. Modulation is the process of transmitting a low-frequency signal into a high-frequency wave, called the *carrier wave*, by changing its amplitude, frequency, and/or phase angle through the *modulating signal*. The function of the carrier wave is to carry the message or modulating signal from the transmitter to the receiver. The superposition of the signal and the carrier wave results in the so-called *modulated signal*.

In the following sections, we review three types of modulation given in Benkő et al. (2011): amplitude modulation, frequency modulation, and amplitude-frequency modulation. Finally, we review how the amplitude and frequency modulations are combined to represent the Blazhko effect.

Amplitude modulation

Amplitude modulation (AM) changes the amplitude of the carrier signal. Let the carrier wave $c(t)$ be a sinusoidal signal of the form

$$c(t) = U_c \sin(2\pi f_c t + \phi_c),$$

where the constant parameters U_c, f_c, and ϕ_c are the amplitude, frequency, and phase of the carrier wave, respectively.

Let $U_m(t)$ represent a waveform that is the message to be transmitted, or *modulating* signal. The transmitter uses the information signal $U_m(t)$ to vary the amplitude of the

carrier U_c to produce the *amplitude modulated* signal U_{AM}:

$$U_{\text{AM}}(t) = [U_c + U_m(t)]\sin(2\pi f_c t + \phi_c) = [U_c + U_m(t)]\frac{c(t)}{U_c} = \left[1 + \frac{U_m}{U_c}\right] c(t). \tag{1.5}$$

In the simplest case, when the modulating signal is sinusoidal, that is,

$$U_m(t) = U_m^A \sin(2\pi f_m t + \phi_m^A), \tag{1.6}$$

the amplitude-modulated signal in equation (1.5) is

$$U_{\text{AM}}(t) = \left[U_c + U_m^A \sin(2\pi f_m + \phi_m^A)\right] \sin(2\pi f_c t + \phi_c). \tag{1.7}$$

Clearly, a more complex example of amplitude modulation arises when $K \geq 1$, where K denotes the number of harmonic components. Suppose the carrier wave $c(t)$ is a linear combination of sine harmonics:

$$c(t) = a_0 + \sum_{k=1}^{K} a_k \sin(2\pi k f_0 t + \phi_k),$$

and the modulating signal is sinusoidal and given again by equation (1.6). Following the same idea as in equation (1.5), the amplitude-modulated signal in equation (1.7) is

$$\begin{aligned} U_{\text{AM}}(t) &= \left[1 + \frac{U_m(t)}{U_c}\right] c(t) \\ &= \left[1 + \frac{U_m^A \sin(2\pi f_m t + \phi_m^A)}{U_c}\right] \left[a_0 + \sum_{k=1}^{K} a_k \sin(2\pi k f_0 t + \phi_k)\right]. \end{aligned} \tag{1.8}$$

If we call $h = U_m^A/U_c$, and use the basic trigonometrical identities $\sin(a)\sin(b) = \frac{1}{2}[\cos(a-b) - \cos(a+b)]$ and $\sin(a) = \cos(a - \frac{\pi}{2})$, equation (1.8) can be written as

$$\begin{aligned} U_{\text{AM}}(t) = a_0 &+ \sum_{k=1}^{K} a_k \sin(2\pi k f_0 t + \phi_k) + a_0 h \sin(2\pi f_m t + \phi_m^A) \\ &+ \sum_{k=1}^{K} \frac{a_k h}{2} \sin(2\pi(k f_0 - f_m)t + (\phi_k - \phi_m) + \pi/2) \\ &- \sum_{k=1}^{K} \frac{a_k h}{2} \sin(2\pi(k f_0 + f_m)t + (\phi_k + \phi_m) + \pi/2). \end{aligned} \tag{1.9}$$

This example shows that when the time-varying amplitude $U_m(t)$ in equation (1.5) takes a sinusoidal form, the amplitude modulated model with *time-varying* amplitude in equation (1.8) can be written as a model with *time-invariant* parameters as in equation (1.9). This implies that, when frequencies and phases are known, the parameters $\{a_k,\ 0 \leq k \leq K\}$ in equation (1.9) can be estimated by ordinary least squares.

Frequency modulation

Frequency modulation (FM) changes the frequency of the carrier signal. We assume the sinusoidal carrier wave to be

$$c(t) = U_c \sin(\Theta(t)),$$

where $\Theta(t) = 2\pi f_c t + \phi_c$ is the angular part of the function. Suppose that the modulating signal is $U_m(t)$, then the modulated signal is given by

$$U_{\text{FM}}(t) = U_c \sin\left(2\pi f_c t + 2\pi k_{\text{FM}} \int_0^t U_m(\tau)\mathrm{d}\tau + \phi_c\right). \tag{1.10}$$

where k_{FM} is the frequency deviation.

In the simplest case, when the modulating signal is represented by a sinusoidal wave with amplitude U_m^F and frequency f_m, the integral of such a signal is

$$\int_0^t U_m(\tau)\mathrm{d}\tau = \frac{U_m^F}{2\pi f_m} \sin(2\pi f_m t + \phi_m),$$

and the frequency-modulated signal in equation (1.10) is

$$U_{\text{FM}}(t) = U_c \sin\left[2\pi f_c t + \frac{k_{\text{FM}} U_m^F}{f_m} \sin(2\pi f_m t + \phi_m) + \phi_c\right]. \tag{1.11}$$

Amplitude-Frequency modulation

In practice, modulated signals can be a mixture of amplitude and frequency modulations, which can be used to described Blazhko RR Lyrae stars (Benkő et al., 2011). We review

the simplest case when both AM and FM are sinusoidal. Combining the amplitude modulated signal in equation (1.7) and the frequency modulated signal in equation (1.11), the *amplitude and frequency modulated* signal is thus

$$U_{\text{Comb}}(t) = [U_c + U_m^A \sin(2\pi f_m t + \phi_m)] \sin\left[2\pi f_c t + \frac{k_{\text{FM}} U_m^F}{f_m} \sin(2\pi f_m t + \phi_m) + \phi_c\right].$$

Blazhko modulation

Benkő et al. (2011) and Benkő (2018) proposed models to describe Blazhko effect. Benkő et al. (2011) proposed to following model:

$$\mu^*(t) = a_0^A a_0 + a_0 g^A(t) + \sum_{k=1}^{K} \left[a_0^A a_k + a_k g^A(t)\right] \sin[2\pi k f_0 t + \varphi_k + k g^F(t)], \qquad (1.12)$$

where a_k and f_0 denote amplitude and frequency, respectively, and

$$g^M(t) = \sum_{j=1}^{\ell^M} a_j^M \sin(2\pi j f_m t + \varphi_j^M), \quad M = A \text{ or } F. \qquad (1.13)$$

More recently, Benkő (2018) introduced a similar model:

$$\mu^*(t) = m_0 + \sum_{r=1}^{\ell} b_r \sin(2\pi r f_m t + \varphi_r^b) + \sum_{k=1}^{K} \left[a_k + g_k^A(t)\right] \sin[2\pi k f_0 t + \varphi_k + g_k^F(t)], \qquad (1.14)$$

where $\sum_{r=1}^{\ell} b_r \sin(2\pi r f_m t + \varphi_r^b)$ corresponds to the mean light-curve variation during the Blazhko cycle, and

$$g_k^M(t) = \sum_{j=1}^{\ell_k^M} a_{kj}^M \sin(2\pi j f_m t + \varphi_{kj}^M), \quad M = A \text{ or } F. \qquad (1.15)$$

The functions $g^M(t)$ and $g_k^M(t)$ in equations (1.13) and (1.15) are the modulating functions with subscripts $M = A$ and $M = F$ denoting amplitude and frequency modulation, respectively. The main pulsation frequency is denoted by f_0, whereas f_m is the modulating frequency. The models in equations (1.12)-(1.13) and (1.14)-(1.15) have two limitations.

Firstly, a parametric model is assumed for the amplitude and frequency modulations in equations (1.13) and (1.15), not allowing to capture more complex shapes, and secondly, a nonlinear least squares algorithms, such as the Levenberg-Marquardt algorithm, is used to fit the parametric nonlinear models in equations (1.12) and (1.14). These methods require initial values close to the solution, which in some applications are difficult to find.

Examples of Blazhko stars

Examples of Blazhko RR Lyrae stars studied by Benkő et al. (2014) are shown in Figure 1.1. The first plot corresponds to the light curve called V808 Cyg. The envelope shape suggests a highly nonsinusoidal amplitude modulation and the Blazhko cycle is close to the length of the observing quarters. The second plot corresponds to the V445 Lyr. The light curve of this star shows strong and complicated amplitudes changes. Finally, the third plot corresponds to the V783 Cyg. The Blazhko effect of V873 Cyg is described by sinusoidal amplitude and frequency modulations. For more details about the stars see Benkő et al. (2014).

1.3.2 Astronomical objects with time-varying PSD

X-ray binaries are the brightest X-ray sources in our galaxy and contain two stars that rotate around each other. One is a normal star and the other is a collapsed star, such as a white dwarf, a neutron star, or a black hole. These two stars produce X-rays if the stars are close enough together that material is drawn from the normal star and spirals in via an accretion disk onto the compact star. X-rays come from the inner region of the accretion disk.

Many X-ray binary systems exhibit distinct spectral states in the frequency domain. The classification of these states is not rigorously defined and is an active topic of debate. Five states have been frequently quoted: the off state, the low-hard state (LS), the intermediate state (IS), the high-soft state (HS), and the very high state (VHS). The HS is dominated by a soft spectrum with little or no timing nose. On the other hand, the LS is dominated

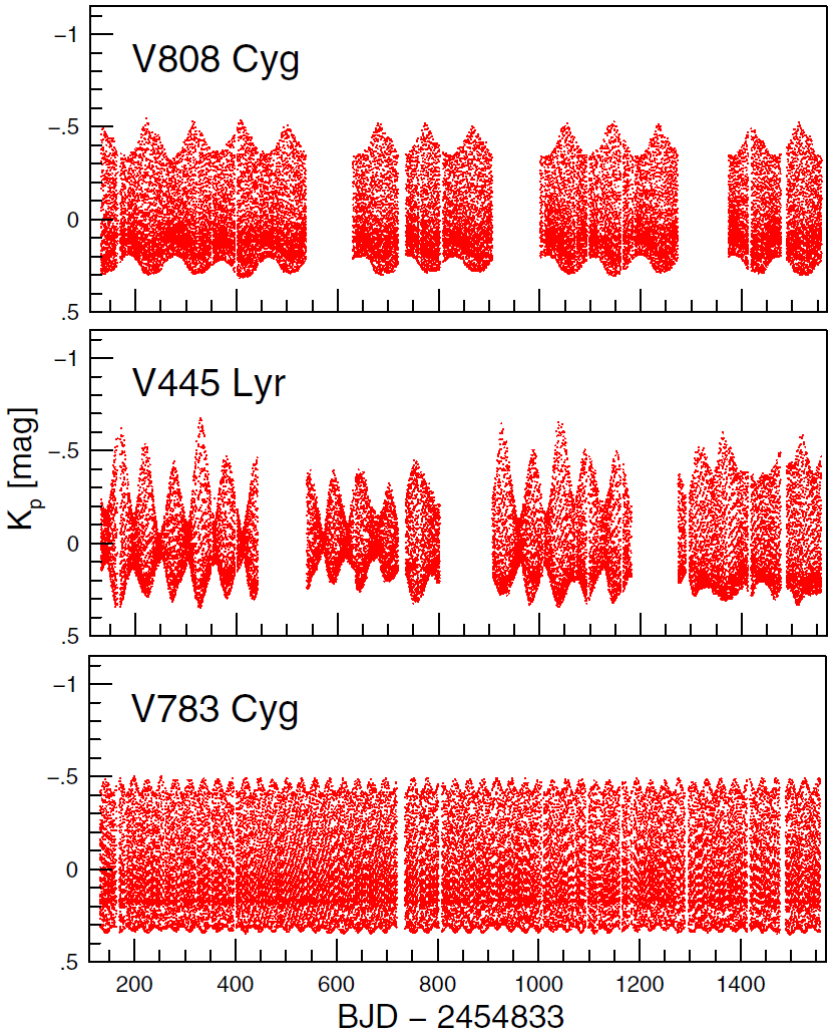

Figure 1.1 This figure has been extracted from Benkő et al. (2014) and it presents three examples of Blazhko stars mentioned in Section 1.3.1.

by a power-law component in the X-ray spectrum, which is often interpreted as being associated with a hot electro corona, and shows quasid-periodic oscillations (QPO). The IS/VHS show both an ultrasoft spectral component and a power-law tail (for more details about these states see Esin et al., 1997). In the next sections, we review the methods used by astronomers to study the PSD of X-ray light curves. For more details see Uttley et al. (2014).

Computation of the PSD

X-ray time series analysis is focused on Fourier analysis techniques in the frequency domain. Uttley et al. (2014) mentioned two reasons for this. Firstly, the PSD is an easy way to describe the underlying structure of a stochastic variable process, and secondly, Fourier techniques give good results when they are used to analysis of the very large, high-time-resolution light curves.

The PSD can be estimated from the periodogram, which is the modulus-squared of the discrete Fourier transform of the light curve. The *discrete Fourier transform* (DFT) $X_1, \ldots, X_{N/2}$ of a light curve $x_1, \ldots, x_{N-1}$ consisting of fluxes measured in N contiguous time bins of width Δ is given by:

$$X_n = \sum_{k=0}^{N-1} x_k \exp(i2\pi nk/N),$$

where x_k is the kth value of the light curve and X_n is the discrete Fourier transform at each Fourier frequency $f_n = n/(N\Delta)$, where $n = 1, 2, \ldots, N/2$. Thus the minimum frequency is the inverse of the duration of the observation, $T_{obs} = N\Delta$, and the maximum is the Nyquist frequency, $f_N = 1/(2\Delta)$.

The *periodogram* is simply given by

$$|X_n|^2 = \left| \sum_{k=0}^{N-1} x_k \exp(i2\pi nk/N) \right|^2 .$$

In practice the periodogram is further normalised to give the same units as the PSD:

$$P_n = \frac{2\Delta}{\overline{x}^2 N} |X_n|^2,$$

where $\overline{x}$ is the mean flux of the light curve and the normalized periodogram P_n is thus expressed in units of fractional variance per Hz, so that this normalization is often called the "rms-squared" normalization.

For a process with noise, the periodogram is a random realization of the underlying PSD. Since the underlying PSD of the process is the physically interesting quantity, one bins up

the periodogram to obtain an estimate of the PSD, so that the PSD in a frequency bin ν_j averaged over M segments and K frequencies per segment is given by:

$$\overline{P}(\nu_j) = \frac{1}{KM} \sum_{n=i,i+K-1} \sum_{m=1,M} P_{n,m}, \tag{1.16}$$

where $\overline{P}(\nu_j)$ is the estimate of the PSD obtained from the average of the periodogram in the bin ν_j and $P_{n,m}$ is the value of a single sample of the peridogram measured from the mth segment with a frequency f_n that is contained within the frequency bin ν_j (which contains frequencies in the range f_i to f_{i+K-1}).

Modeling the PSD

The PSD of several black holes and neutron stars during the low state can be modeled by a broken power law function, whereas during the hard state can be modeled using Lorentzian functions (see Pottschmidt et al., 2003; Kalemci et al., 2003). The fit is done in the frequency domain using the periodogram in equation (1.16) and power law and Lorentzian functions. Lorentzian functions have the form

$$L_i(f) = \frac{1}{\pi} \frac{2R_i^2 Q_i f_i}{f_i^2 + 4Q_i^2(f - f_i)^2},$$

where the subscript i denotes each Lorentzian component, f_i is the resonance frequency, Q_i is a quality factor, R_i is a normalization constant. A useful quantity of the Lorentzian is the frequency at which its contribution to the total rms variability is maximum (hereafter peak frequency)

$$\nu_i = f_i \left(\frac{1}{4Q_i^2} + 1 \right)^{1/2}.$$

The power law function has the form

$$Pl(f) = Af^{-\alpha} \exp(-f/f_c),$$

where A is the normalization constant, α is the power-law index, and f_c is the turnover frequency.

Examples of X-ray binary systems with time-varying PSD

In this section, we review two examples of time-varying PSD. The first example corresponds to the black hole XTE J1650-500, studied by Kalemci et al. (2003). Kalemci et al. (2003) studied the XTE J1650-500 through the PSD for 7 observations (7 segments of time). For each observation, the PSD was modeled using Lorentzian functions. Figure 1.2 shows the PSD for the 7 observations (see the number in the top-right position on each plot), the fitted curves (solid lines), and the Lorentzian functions used in the fit (dashed lines). As we can see, the PSD is changing over time. In fact, for observation 2, a weak narrow peak at 8.71 Hz begins to be observed.

In addition to being interested in studying the evolution of PSD, astronomers study the evolution of some peaks observed in the PSD, called quasi-periodic oscillations (QPO). In Figure 1.2, Kalemci et al. (2003) found that most of the observations need three Lorentzian functions to describe the PSD. Narrow Lorentzian functions are associated with QPOs, and in this case, the resonance frequency f_i is very close to the peak frequency ν_i. Figure 1.3 shows the evolution of the peak frequencies ν_1 and ν_2 associated with the two wide Lorentzians functions and the resonance frequency of the QPO. There is an overall shift to lower frequencies with time for both the wide Lorentzian functions and the QPO.

The second example of time-varying PSD is the black hole Cygnus X-1 studied by Axelsson, M. et al. (2005). Axelsson, M. et al. (2005) described the variability properties of Cyg X-1 in terms of the PSD. The light curve is studied from 1996 to 2003 and the PSD during this period is fitted using one or two Lorentzian functions and/or an exponential power law. Figure 1.4 shows the gradual change from normal hard state to canonical soft state PDS of Cyg X-1.

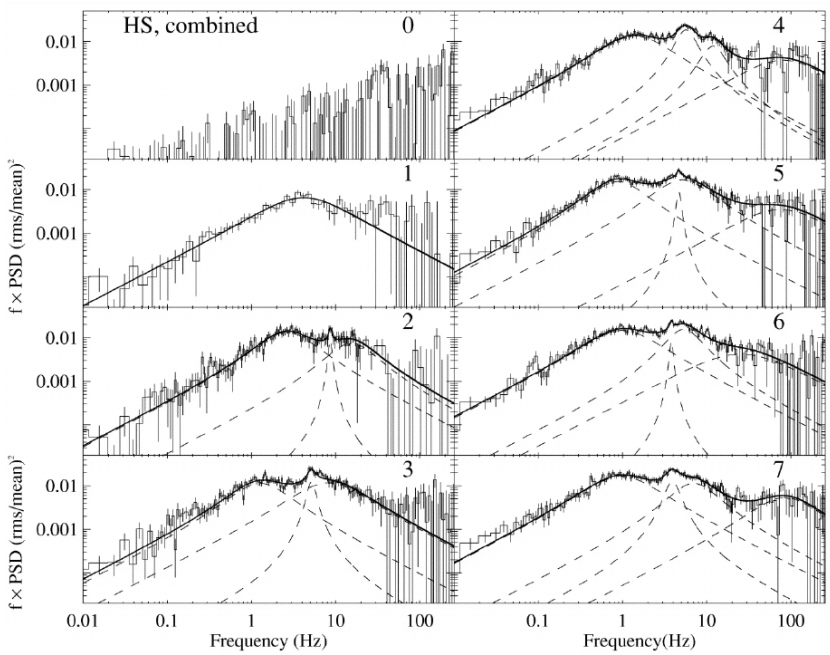

Figure 1.2 This figure has been extracted from Kalemci et al. (2003) and it presents the evolution of the PSD and fitted curves using Lorentzian functions for the XTE J1650-500 mentioned in Section 1.3.2. The panel numbered by "0" is the PSD from the combined light curves in the HS. The rest of the panels numbered by 1,. . . ,7, represent the 7 observations. Except for the HS PSD, the solid line represents the overall fit, and the dashed lines represent each component.

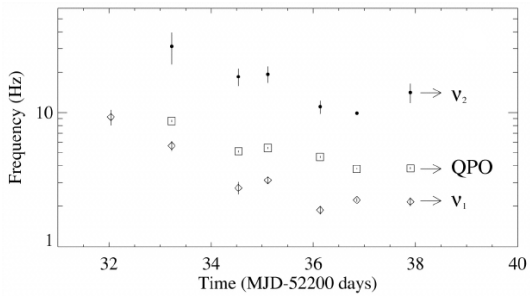

Figure 1.3 This figure has been extracted from Kalemci et al. (2003) and it presents the evolution of the peaks of the Lorentzian functions used to fit the PSD of the XTE J1650-500 mentioned in Section 1.3.2. Filled circles represent the peak frequency ν_2 associated to the Lorentzian L_2, squares represent QPO frequency, and diamonds represent the peak frequency ν_1 associated to the Lorentzian L_1.

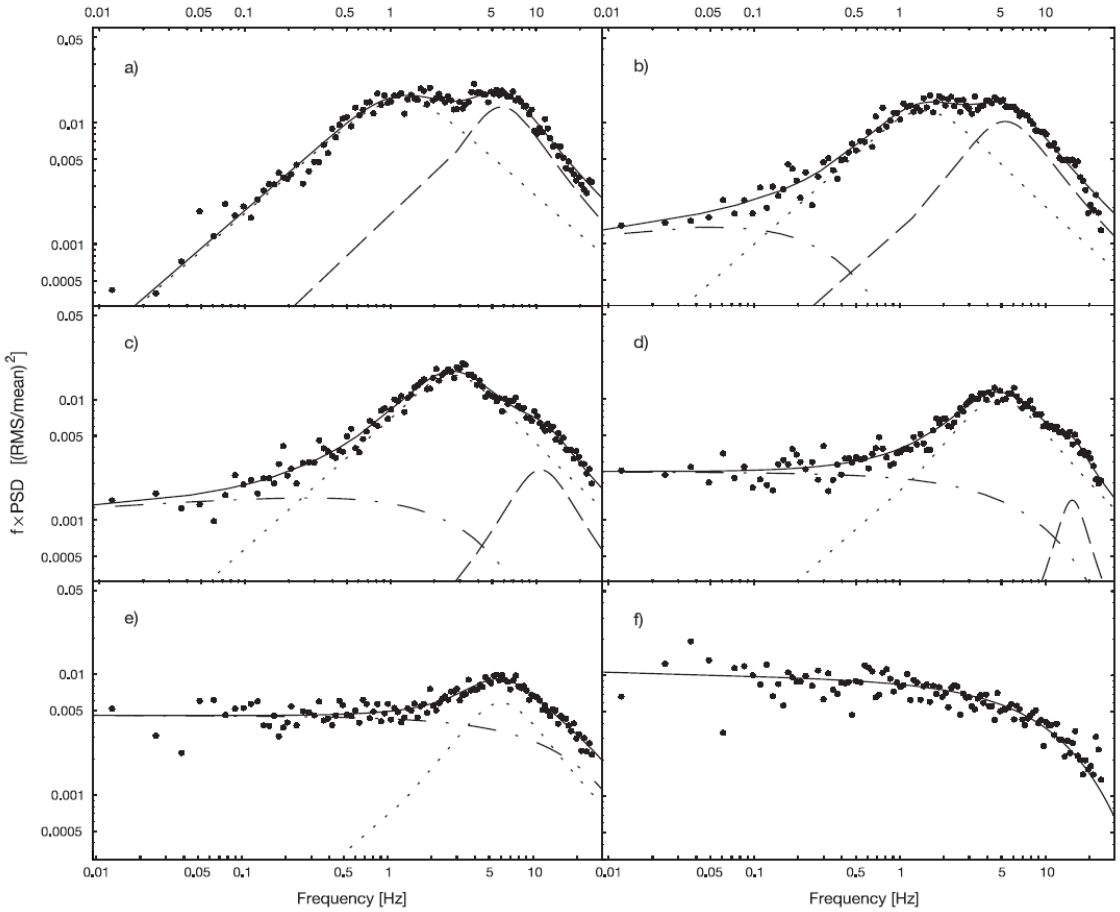

Figure 1.4 This figure has been extracted from Axelsson, M. et al. (2005) and it presents the evolution of the PSD and fitted curves using Lorentzian and power law functions for the Cyg X-1 mentioned in Section 1.3.2. Evolution from the hard state (panel a)) to soft state (panel f)). The solid line represents the fitted curves, the dashed and dotted lines represent the Lorentzian functions and the dash-dotted line represents the power law function used to fit the PSD.

The are mainly two limitations in studying the time-variation of the spectral density in the way astronomers has been doing so far. First, dividing the time-span into segments (and then computing the Fourier transform on each segment) requires the selection of the break points, which is often arbitrary. Second, for each segment it is necessary to fit a model based on Lorentzian and power law functions. As a consequence of these two

limitations, it is not possible to observe *smooth* changes over time of the PSD and the Lorentzian functions.

The are mainly two approaches to model time-varying parameters: *structural breaks* and *local stationarity*. The structural breaks models assume that the coefficients change abruptly right after the so-called *change points*. Aue and Horváth (2013) give an account of some of the recent work on structural breaks in time series models. The estimation is usually based on the popular cumulative sum, CUSUM. Both structural breaks in the mean as well as in the variance and covariance/correlation structure belong to this approach. CUSUM procedures are nonparametric by design.

To overcome the limitations discussed above, we work within the framework of local stationarity, which assumes instead that the parameters (mean, variance and autocovariance) change *slowly* over time. This type of nonstationarity is the one we adopt in this thesis.

1.4 Locally stationary processes and time-varying spectral density

In this section, we introduce the definition of locally stationary processes proposed by Dahlhaus (1996), which is based on the concept of evolutionary spectrum. A sequence of stochastic processes $\{x_{t,T},\, t = 1, \ldots, T\}$ is called *locally stationary* with transfer function A^0 and trend μ if there exists a representation

$$x_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\omega t) A^0_{t,T}(\omega) d\xi(\omega),$$

where the following holds.

i) $\xi(\omega)$ is a stochastic process on $[-\pi, \pi]$ with $\overline{\xi(\omega)} = \xi(-\omega)$ and

$$\text{cum}\left\{d\xi(\omega_1), \ldots, d\xi(\omega_k)\right\} = \eta\left(\sum_{j=1}^{k} \omega_j\right) g_k\left(\omega_1, \ldots, \omega_{k-1}\right) d\omega_1 \ldots d\omega_k,$$

where cum $\{\cdots\}$ denotes the cumulant of the kth order,

$$g_1 = 0, \quad g_2(\omega) = 1, \quad |g_k(\omega_1, \ldots, \omega_{k-1})| \leq \text{const}_k, \quad \text{for all } k,$$

and $\eta(\omega) = \sum_{j=-\infty}^{\infty} \delta(\omega + 2\pi j)$ is the period 2π extension of the Dirac delta function.

ii) There exists a constant K and a $2\pi-$periodic function $A : [0,1] \times \mathbb{R} \to \mathbb{C}$ with $A(u, -\omega) = \overline{A(u,\omega)}$ and

$$\sup_{t,\omega} \left| A^0_{t,T}(\omega) - A\left(\frac{t}{T}, \omega\right) \right| \leq \frac{K}{T}, \quad \text{for all } T,$$

and $A(u,\omega)$ and $\mu(u)$ are assumed to be continuous in u, where t and $u = t/T$ denotes time points in the interval $[1,T]$ and the rescaled interval $[0,1]$, respectively.

The smoothness of A in u guarantees that the process has locally a stationary behavior. If μ and A^0 do not depend on t and T then X does not depend on T as well and we obtain the spectral representation of a stationary process. Thus, the classical asymptotic theory for stationary processes is a special case of Dahlhaus's approach.

The function $P(u,\omega) = |A(u,\omega)|^2$ is called the *time-varying spectral density* at time $u \in [0,1]$ and frequency $\omega \in [-\pi, \pi]$ of a locally stationary process.

Let $\{x_{t,T}\}$ be an autoregressive moving-average process with time-varying coefficients, i.e. the solution of

$$x_{t,T} = \sum_{j=1}^{p} \phi_j\left(\frac{t}{T}\right) x_{t-j,T} + z_t + \sum_{j=1}^{q} \theta_j\left(\frac{t}{T}\right) z_{t-j}, \quad t = 1, \ldots, T, \quad \{z_t\} \sim WN(0, \sigma_z^2), \tag{1.17}$$

where $\{z_t\}$ is a WN process with mean zero and variance σ_z^2. If $1 - \sum_{j=1}^{p} \phi_j(u) x^j \neq 0$ for all $|x| \leq 1 + c$ with $c > 0$ uniformly in u and the coefficients functions $\phi_j(u)$ are continuous in u, then the process in equation (1.17) is a locally stationary process in the sense that equation (1.17) has a solution of the form

$$x_{t,T} = \int_{-\pi}^{\pi} \exp(i\omega t) A^0_{t,T}(\omega) d\xi(\omega)$$

with

$$A(u,\omega) = \frac{\sigma_z}{\sqrt{2\pi}} \frac{1 + \sum_{j=1}^{q} \theta_j(u) \exp(-i\omega j)}{1 - \sum_{j=1}^{p} \phi_j(u) \exp(-i\omega j)},$$

and time-varying PSD

$$P(u,\omega) = \frac{\sigma_z^2}{2\pi} \frac{\left| 1 + \sum_{j=1}^{q} \theta_j(u) \exp(-i\omega j) \right|^2}{\left| 1 - \sum_{j=1}^{p} \phi_j(u) \exp(-i\omega j) \right|^2}.$$

The locally stationary process in equation (1.17) is called *locally stationary autoregressive moving-average process* and it is denoted by LSARMA(p, q). A *locally stationary autoregressive process*, denoted by LSAR(p), corresponds to a LSARMA$(p, 0)$ process given by

$$x_{t,T} = \sum_{j=1}^{p} \phi_j \left(\frac{t}{T} \right) x_{t-j,T} + z_t, \quad t = 1, \ldots, T, \quad \{z_t\} \sim WN(0, \sigma_z^2),$$

with time-varying PSD

$$P(u,\omega) = \frac{\sigma_z^2}{2\pi} \frac{1}{\left| 1 - \sum_{j=1}^{p} \phi_j(u) \exp(-i\omega j) \right|^2}.$$

1.5 Estimation of time-varying parameters

The variation over time can be observed in the time-varying mean $\mu(u)$ or in the time-varying coefficients $\phi_1(u), \ldots, \phi_p(u)$, and $\theta_1(u), \ldots, \theta_q(u)$, with $u \in [0, 1]$ of a LSARMA(p, q) process. The model choice for $\mu(t)$, $\phi_j(u)$ and $\theta_k(u)$ depends on the prior knowledge of their functional form. Thus, depending on the form, we distinguish three models, parametric, semiparametric, and nonparametric model. When the statistical model is a parameterized family of distributions with finite-dimensional parameters, such a model is referred to as a *parametric model*. When the parameters lie in a subset of an infinite dimensional space or the form of distribution is not completely specified, such a model is often called a *nonparametric model* (Fan and Yao, 2003). Parametric models differ from nonparametric models in that the shape of the coefficients is not predetermined but can be adjusted to

capture unusual or unexpected features of the coefficients. A combination of parametric and nonparametric models is known as *semiparametric models*. Therefore, if the forms of $\mu(u)$, $\phi_j(u)$ and $\theta_k(u)$ are know and they are correct, a parametric model can be used to describe $\mu(u)$, $\phi_j(u)$ and $\theta_k(u)$ well. However, if a wrong functional form is chosen, this can lead to misunderstanding of the process, wrong conclusions, and erroneous forecasting (Fan and Yao, 2003).

A usual assumption about the form of the time-varying coefficients is that they may be approximated by a linear combination of a small number of known functions (parametric model). For instance, Rao (1970) proposed a procedure to fit time-varying AR(p) process based on the weighted least squares method assuming that the first three terms of the Taylor series expansion give a good approximation for the time-varying parameters $\phi_j(u) = \phi_{j,0} + \phi_{j,1}u + \phi_{j,2}u^2/2$. Similar ideas with various approximations in a finite dimensional linear space of approximation may be found, for example, Dahlhaus (1997) assumed polynomial form for the time-varying coefficients of a LSAR(p) process to study the relationship between the estimation given by minimizing the local version of the Whittle function and the least squares estimation, Grenier (1983) assumed a polynomial form of order m for the coefficients of a time-varying ARMA process, this allowed the extension of several well-known techniques of stationary spectral estimation to the nonstationary case, and Dahlhaus (2000) gave a new approximation to the Gaussian likelihood of a multivariate locally stationary process assuming that the time-varying coefficients are described by a finite-dimensional set, that is, the parameter curves themselves are parameterized.

A more flexible assumption about the form of the time-varying coefficient can be made using basis functions. For example, Dahlhaus et al. (1999) used an orthonormal wavelet basis representation for the coefficients. The nonparametric estimation of the coefficients is made in two steps. In the first step, the empirical wavelet coefficients are obtained from the solution of a least-squares minimization problem, and in the second step, a soft or hard thresholding is applied. A similar idea was adopted by Dahlhaus and Neumann (2001), who proposed a semiparametric version of LSARMA processes, where the time-varying mean is estimated using kernel estimator and coefficients is approximated by an

orthonormal wavelet basis. The estimation of the empirical wavelet coefficients is obtained by minimizing an empirical version of the Kullback-Leibler distance.

1.6 Outline of the dissertation

The organization of the dissertation is presented in Figure 1.5.

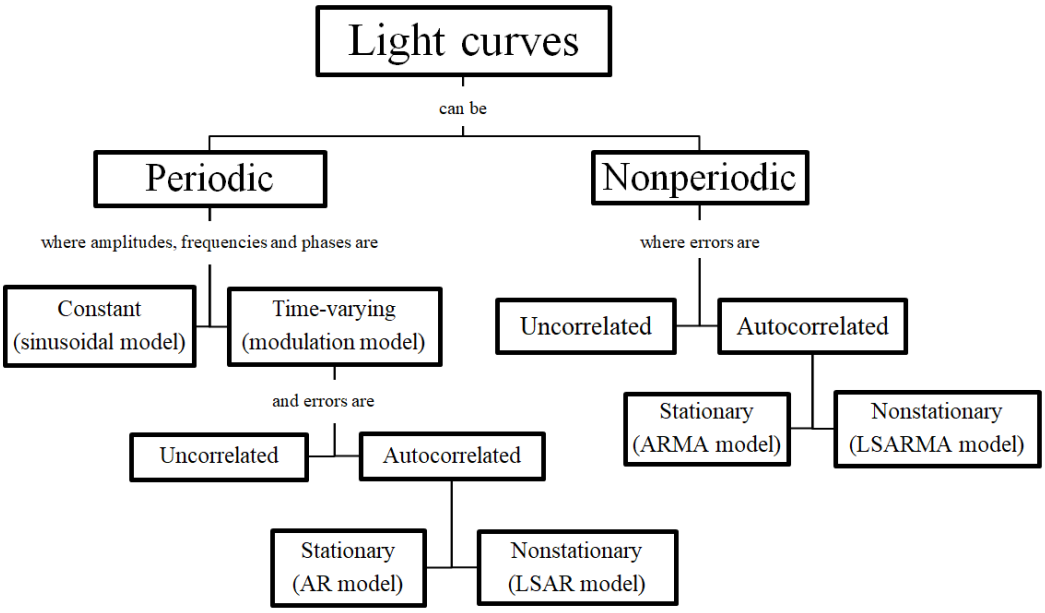

Figure 1.5 Conceptual map about light curves from the point of view of time series analysis.

In Chapter 2, we introduce a new mathematical model for the description of the so-called modulated light curves, as found in periodic variable stars that exhibit smoothly time-varying parameters such as amplitude, frequency, and/or phase, with uncorrelated errors. The model is given by

$$y_i = \mu(t_i) + z_i, \qquad i = 1, \ldots, N, \qquad \{z_i\} \sim WN(0, \sigma_z^2),$$

$$\mu(t_i) = m(t_i) + \sum_{k=1}^{K} \{g_{1,k}(t_i)\cos(w_k t_i) + g_{2,k}(t_i)\sin(w_k t_i)\}.$$

In Chapter 3, we extend the model proposed in Chapter 2 considering autocorrelated errors in two ways. First, we assume that the errors follow an autoregressive (AR) process:

$$\begin{aligned} y_i &= \mu(t_i) + \varepsilon_i, \qquad i = 1, \ldots, N, \\ \mu(t_i) &= m(t_i) + \sum_{k=1}^{K} \{g_{1,k}(t_i)\cos(w_k t_i) + g_{2,k}(t_i)\sin(w_k t_i)\}, \\ \varepsilon_i &= \sum_{j=1}^{p} \phi_j \varepsilon_{i-j} + z_i, \quad \{z_i\} \sim WN(0, \sigma_z^2). \end{aligned}$$

Second, we allow for non-stationarity and model the errors as being locally stationary AR processes, that is,

$$\begin{aligned} y_{i,N} &= \mu\left(\tfrac{i}{N}\right) + \varepsilon_{i,N}, \qquad i = 1, \ldots, N, \\ \mu\left(\tfrac{i}{N}\right) &= m\left(\tfrac{i}{N}\right) + \sum_{k=1}^{K} \left\{g_{1,k}\left(\tfrac{i}{N}\right)\cos(N w_k \tfrac{i}{N}) + g_{2,k}\left(\tfrac{i}{N}\right)\sin(N w_k \tfrac{i}{N})\right\}, \\ \varepsilon_{i,N} &= \sum_{j=1}^{p} \phi_j\left(\tfrac{i}{N}\right) \varepsilon_{i-j,N} + z_i, \quad \{z_i\} \sim WN(0, \sigma_z^2), \end{aligned} \tag{1.18}$$

where the time-varying AR coefficients $\{\phi_j(u),\, j \leq p\}$ are smooth functions of rescaled time $u \in [0, 1]$.

In Chapter 4, we adopt locally stationary ARMA processes (LSARMA) to describe the time variability observed in X-ray binaries. An LSARMA(p, q) process is defined as solution of

$$x_{i,N} = \sum_{j=1}^{p} \phi_j\left(\frac{i}{N}\right) x_{i-j,N} + z_i + \sum_{j=1}^{q} \theta_j\left(\frac{i}{N}\right) z_{i-j}, \quad \{z_i\} \sim WN(0, \sigma_z^2), \quad i = 1, \ldots, N,$$

with corresponding spectral density

$$P_x(u, f) = \frac{\sigma_z^2}{2\pi} \frac{\left| 1 + \sum_{j=1}^{q} \theta_j(u)\exp(-i2\pi f\Delta)^j \right|^2}{\left| 1 - \sum_{j=1}^{p} \phi_j(u)\exp(-i2\pi f\Delta)^j \right|^2}, \quad -\frac{1}{2\Delta} < f < \frac{1}{2\Delta}, \quad u \in [0, 1].$$

We prove that the spectral density of an LSARMA(p, q) processes can be expresses as the sum of p components

$$P_x(u, f) = \sum_{j=1}^{p} P_{x,j}(u, f),$$

where the components $P_{x,j}(u, f)$, $j = 1, \ldots, p$, have local extremums at the frequencies 0 and $1/(2\Delta)$, and/or at the frequency

$$f_{\max,\ell}(u) = \frac{1}{2\pi\Delta} \arccos\left(\frac{-K_{2,\ell}(u) \pm \sqrt{K_{2,\ell}^2(u) - 4K_{1,\ell}(u)K_{3,\ell}(u)}}{2K_{1,\ell}(u)} \right), \quad \ell = j, k,$$

where the coefficients $K_{2,\ell}(u)$ depend on the LSARMA parameters $\{\phi_j(u),\, j \leq p\}$ and $\{\theta_k(u),\, \leq q\}$, $u \in [0, 1]$. The mathematical expression of the argmax $f_{\max,\ell}(u)$ allows to estimate the peak-frequency (in frequency domain) by replacing the coefficients $K_{j,\ell}$, $j = 1, 2, 3$, with the corresponding estimates (obtained in time domain).

1.7 Our contribution and Published work

In this Ph.D. thesis we deliver several important results for the analysis of astronomical time series. Our contribution can be summarized as follows.

For periodic variable stars observed unequally over time, in Chapter 2 we introduce a novel semi-parametric model with smoothly time-varying trend and amplitudes. The time-varying parameters are modeled with B-splines and estimated with penalized least squares. The estimation of our time-varying curves translates into the estimation of time-invariant parameters that can be performed by ordinary least squares, with the following two advantages: modeling and forecasting can be implemented in a parametric fashion, and we are able to cope with missing observations. We then derive the mathematical definition of the spectral density for unequally spaced time series. We also present a new method to estimate the spectral density of unequally spaced times series, which is needed for the analysis of the residuals.

In Chapter 3 we extend the results of Chapter 2 to the case where the errors are autocorrelated, stationary and nonstationary. First, we introduce a novel version of the Cochrane-Orcutt algorithm (Cochrane and Orcutt, 1949) that is suitable for penalized least square (PLS) regression. Then, we adapt the Cochrane-Orcutt algorithm procedure for PLS to the case where the errors follow a locally stationary autoregressive process.

For nonperiodic variable stars we introduce in Chapter 4, for both stationary and locally stationary ARMA models, an important tool for the study of time series observations coming from X-ray binaries. First, we prove that the PSD of a stationary ARMA(p, q) process can be expressed as a sum of p functions. We provide a mathematical description of these p functions and find a closed form for the frequency at which these functions have a local maximum or minimum. This result makes ARMA models suitable for fitting X-rays binary systems and, in some cases, it avoids having to calculate the frequency at which the PSD has a peak. Then we extend this result to LSARMA processes and, assuming that the time-varying coefficients are smooth, we establish an estimation method (for the time-varying coefficients) that combines B-splines with the Hannan-Rissanen algorithm (Hannan and Rissanen, 1982).

The dissertation is a collection of three papers. The first paper has been recently published, whereas the other two manuscripts are drafts of work in progress to be submitted. Chapter 2 is fully based on our paper Motta et al. (2022), joint with Dr. Márcio Catelan. Chapter 3 corresponds to a work in progress co-authored with my Ph.D. advisor Dr. Giovanni Motta, and Chapter 4 is also a work in progress co-authored with Dr. Giovanni Motta and my collaborator Dr. Malgorzata Sobolewska.

Chapter 2

A nonparametric approach for periodic time series with uncorrelated errors

2.1 Introduction

RR Lyrae stars are important astrophysical tools for the measurement of distances and studies of the astrophysical properties of old stellar populations. They are moderately bright, evolved low-mass stars, currently in the core helium-burning phase, also known as the horizontal branch. Their periods are typically in the range of between about 0.2 and 1.0 d, which together with their characteristic light-curve shapes, allow them to be relatively easily identified in time series photometric surveys. An overview of their properties can be found in the monographs by Smith (2004) and Catelan and Smith (2015).

In spite of their astrophysical importance, RR Lyrae stars are still not fully understood. Indeed, one of the longest-standing problems in stellar astrophysics is also one that specifically affects RR Lyrae stars: the so-called *Blazhko effect* (Blažko, 1907). It is a long-term modulation of an RR Lyrae's light curve, over timescales ranging from a few to hundreds of days (for recent reviews, see Catelan and Smith, 2015; Gillet et al., 2019). The Blazhko effect is particularly common amongst fundamental-mode (ab-type) pulsators (e.g., Plachy

et al., 2019), but is also present, to a lesser extent, in first-overtone (c-type) RR Lyrae stars (e.g., Netzel et al., 2018).

Over the decades since it was first described, the Blazhko effect has persistently defied theoretical explanations as to its cause (e.g., Gillet et al., 2019). Gradual strengthening and weakening of turbulent convection in the stellar envelope (Stothers, 2006), a 9:2 resonance between the fundamental and ninth-overtone radial modes (Buchler and Kolláth, 2011), and interaction between fundamental and first-overtone modes in the "either-or" region of the instability strip (Gillet, 2013) are the most recent candidates, but no consensus has yet been reached as to the root cause of the Blazhko effect, due in large part to the difficulties involved in the nonlinear hydrodynamical modeling of the phenomenon.

In this chapter, we introduce a model for time series observations of variable stars having a smoothly time-varying trend and amplitudes. More precisely, we develop a semi-parametric method for unequally spaced time series measuring the brightness of a modulated variable star. Our approach is flexible because it avoids assumptions about the functional form of the trend and amplitudes. The estimation of our time-varying curves translates into the estimation of time-invariant parameters that can be performed by ordinary least squares, with the following two advantages: modeling and forecasting can be implemented in a parametric fashion, and we are able to cope with missing observations. We also study the spectral density of the residuals obtained from the fit of our novel model.

In order to detect serial correlation in the residuals, in this chapter we derive the definition of the spectral density for unequally spaced time series. There are many reasons why astronomical time series are not sampled equidistantly, and the gaps can be either regular or random. From the Earth, stars cannot be observed during the day, which introduces regular gaps in the time series. Also, for about half a year, most objects become unobservable, as they are up on the sky at the same time as the Sun, which introduces yearly gaps. There could be clouds or high winds, forcing the closure of telescopes, producing random gaps. There could be high-priority alerts overriding the observations, or the telescope could be available only on certain nights.

In some cases, observations are unevenly spaced due to missing values. Astronomical

data sets often contain missing values, and this limitation is sometimes due to incomplete observations or varying survey depths. Even if telescopes are recording and storing information systematically, that is, at a regular cadence, there are a few things that can alter the regular sampling. For example, an astronomer might decide to increase the exposure time if there are clouds obscuring the target, to try to increase the signal-to-noise ratio. Conversely, if the observing conditions are excellent, the astronomer might decide to decrease the exposure times (and hence the cadence) to avoid saturating the detector, for example. Missing values are usually handled via imputation, that is, the gap generated by the missing value is "filled in" by an estimated value. If observations are missing because of the survey, imputation can be performed using statistical models.

However, in astrostatistics, missing value problems are sometimes inherently brought about by the manner in which physical processes are recorded. In particular, telescopes are not located in the center of the solar system. Since the speed of light is finite, this results in a time delay between the arrival times of signals at our position and at the center of the solar system. This is typically corrected for by referring the times of observations to either Heliocentric Julian Dates (HJD) or Barycentric Julian Dates (BJD), which refer to the center of the Sun or the entire solar system, respectively. Thus, even if telescopes record data strictly evenly according to the local time at the observatory (e.g., one observation performed every night at local midnight), this correction will slowly change between observations, modifying what was initially a regular grid to an irregular one. Also, this correction is different for every source on the sky, even though sources close to each other may have very similar corrections. Therefore, for some astronomical data sets where missing values may arise from the manner in which observations of a physical process are collected, or even the nature of the physical process itself (e.g., sudden, extreme dimming events that may occasionally render an object impossible to detect for a certain amount of time), the imputation method may not be applicable (see Chattopadhyay, 2017). Our novel approach, which involves the classical periodogram, has the advantage of not relying on any imputation method.

We divide the chapter into seven main sections. In Section 2.2 we introduce our novel

model and clarify analogies and differences as compared with previous approaches. In Section 2.3, the present status of important ingredients of amplitude and frequency modulations is critically discussed. In Section 2.4 we present the method we adopt to estimate the time-varying parameters. In Section 2.5 we present a new method to estimate the spectral density of unequally spaced times series, which is needed for the analysis of the residuals. Section 2.6 provides simulation results, whereas Section 2.7 illustrates the advantages of using our novel method by means of an application to an RR Lyrae variable star. Finally, our main conclusions are summarized in Section 2.8.

2.2 Model definition

Light curves of variable stars are typically fitted using harmonic models with a linear (or constant) trend and time-invariant amplitudes (see equations (1) and (5) in Richards et al., 2011). This type of model would be inappropriate when the underlying trend and amplitudes change over time in a more complex way. Eilers et al. (2008) proposed a model with one harmonic component ($K = 1$) where the trend and amplitudes vary smoothly over time. In this chapter, we extend the model by Eilers et al. (2008) to the case of $K \geq 1$ harmonic components, where the amplitudes associated with each harmonic component vary smoothly over time. We estimate our model by means of P-splines (Eilers and Marx, 1996), which are a combination of B-splines and penalties. The estimation of the time-varying curves translates into the estimation of time-invariant parameters that can be performed by the least squares method, with the following three advantages: it is computationally fast, forecasting can be implemented in a parametric fashion, and we can cope with missing observations.

Compared to local smoothers (such as kernel smoothers), the main advantage of regression spline in the context of time series is that the unknown parameters are time-invariant and thus they can be estimated globally rather than locally. As a consequence, forecasting only requires good estimates of the global unknown parameters. Finally, the use of B-splines in regression allows us to rewrite the estimation problem as a least squares fit,

avoiding the use of numerical methods – such as Newton Raphson – which can be time consuming.

Let $\{y_i \equiv y_{t_i},\, i = 1, \ldots, N\}$ be a set of observations occurring at certain discrete times $t_1, \ldots, t_N$. In the case of equally spaced observations, $t_i = t_0 + i\Delta$ where i is an integer, and $\Delta > 0$ is the constant data spacing. Then $|t_i - t_k| = \Delta|i - k|$, and typically $\Delta = 1$. Astronomical light curves are often observed unequally in time, that is, the data spacing of observation times is not constant.

We decompose the observed light curve into the sum of a deterministic signal μ and a random noise z. The deterministic part $\mu(t)$ consists of a trend $m(t)$ and a *modulated* periodic signal. The modulated periodic signal is a linear combination of K cosines and sines, with weights given by the *modulating* functions $g(t)$:

$$\begin{aligned} y_i &= \mu(t_i) + z_i, \qquad i = 1, \ldots, N, \qquad \{z_i\} \sim WN(0, \sigma_z^2), \\ \mu(t_i) &= m(t_i) + \sum_{k=1}^{K} \{g_{1,k}(t_i)\cos(w_k t_i) + g_{2,k}(t_i)\sin(w_k t_i)\}, \end{aligned} \tag{2.1}$$

or in matrix notation $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{z}$, where $\boldsymbol{y} = (y_1, \ldots, y_N)^\top$ is the vector of observations at time $\boldsymbol{t} = (t_1, \ldots, t_N)^\top$, $\boldsymbol{\mu} = [\mu(t_1), \ldots, \mu(t_N)]^\top$ is the expectation of $\boldsymbol{y}$, $m(t_i)$ is the smooth time-varying trend at time t_i, the $g_{\ell,k}(t_i)$'s are smooth time-varying amplitudes of the cosine and sine waves at time t_i, respectively, $w_k = 2\pi f_k$ is the angular frequency, and f_k is the ordinary frequency. Since the errors are zero-mean, the expectation of the observed brightness at time t_i is equal to the deterministic part of the signal at time t_i, that is, $\mathbb{E}[y_i] = \mu(t_i)$.

We refer to $m(\cdot)$ as the "trend", that is, the (typically) aperiodic change in the mean of the light curve. By contrast, we call "amplitudes" the functions $g(\cdot)$'s that weigh the periodic variation (of this average brightness) of cosine and sine waves. Both trend and amplitudes are typically restricted to be sinusoidal, whereas in this chapter our trend $m(t)$ and our amplitude functions $g_{\ell,k}(t)$'s are general smooth functions and not necessarily sinusoidal. In Section 2.3 we clarify the mathematical connection between the standard modulation models and our novel modulation model in equation (2.1). The error vector

$\boldsymbol{z} = (z_1, \ldots, z_N)^\top$ is a WN process with mean zero and variance σ_z^2.

The approaches based on time-varying parameters are receiving growing interest in various areas of astrophysics. For instance, Kelly et al. (2014) simulated a light curve that switches from one CARMA process to another. More precisely, they constructed a nonstationary light curve by generating two CARMA processes of the same order ($p = 5, q = 3$), but with different parameters (see Kelly et al., 2014, Section 4.3):

$$\boldsymbol{\theta}(t) = \begin{cases} \boldsymbol{\theta}_1 & t_1 \leq t < t_0 \\ \boldsymbol{\theta}_2 & t_0 \leq t \leq t_N, \end{cases}$$

where $\boldsymbol{\theta}(t) = [\alpha_1(t), \ldots, \alpha_p(t), \beta_1(t), \ldots, \beta_q(t), \sigma^2(t)]^\top$. The vector $\boldsymbol{\theta}(t)$ is a step-wise function that is constant before and after t_0. Wong et al. (2016) adopt a Poisson model for the photon counts. They define $\lambda(t_j, w_i)$ as the expected count per unit time and per unit wavelength averaged over the bin centered at (t_j, w_i), and detect change points π such that $\{\lambda(t_j, w_i)|t_j \leq \pi\} \neq \{\lambda(t_j, w_i)|t_j > \pi\}$. Wong et al. (2016) estimate the number of change points and their values. Xu et al. (2021) develop a method for modeling a time series of images, and assume that the arrival times of the photons follow a Poisson process. They assume that all image stacks between any two adjacent change points (in the time domain) share the same unknown piecewise constant function. Xu et al. (2021) estimate the number and the locations of all of the change points (in the time domain), as well as all of the unknown piecewise constant functions between any pairs of the change points.

In the next section, we review the models proposed by Benkő et al. (2011) and Benkő (2018) for Blazhko light curves. Interestingly, our model in equation (2.1) generalizes the models by Benkő et al. (2011) and Benkő (2018) in the sense that the modulating functions $g_{\ell,k}(\cdot)$ are not confined to the class of parametric (sinusoidal or nonsinusoidal) functions.

2.3 Modeling Blazhko light curves

In this section, we review briefly the models proposed by Benkő et al. (2011) and Benkő (2018), and we compare them with our novel model in equation (2.1). To describe Blazhko

light curves, Benkő et al. (2011) proposed to fit the following model:

$$\mu^*(t) = a_0^A a_0 + a_0 g^A(t) + \sum_{k=1}^{K} \left[a_0^A a_k + a_k g^A(t)\right] \sin[2\pi k f_0 t + \varphi_k + k g^F(t)], \tag{2.2}$$

where a_k and f_0 denote amplitude and frequency, respectively, and

$$g^M(t) = \sum_{j=1}^{\ell^M} a_j^M \sin(2\pi j f_m t + \varphi_j^M), \quad M = A \text{ or } F. \tag{2.3}$$

More recently, Benkő (2018) introduced a similar model:

$$\mu^*(t) = m_0 + \sum_{r=1}^{\ell} b_r \sin(2\pi r f_m t + \varphi_r^b) + \sum_{k=1}^{K} \left[a_k + g_k^A(t)\right] \sin[2\pi k f_0 t + \varphi_k + g_k^F(t)], \tag{2.4}$$

where $\sum_{r=1}^{\ell} b_r \sin(2\pi r f_m t + \varphi_r^b)$ corresponds to the mean light-curve variation during the Blazhko cycle, and

$$g_k^M(t) = \sum_{j=1}^{\ell_k^M} a_{kj}^M \sin(2\pi j f_m t + \varphi_{kj}^M), \quad M = A \text{ or } F. \tag{2.5}$$

The functions $g^M(t)$ and $g_k^M(t)$ in equations (2.3) and (2.5) are the modulating functions with subscripts $M = A$ and $M = F$ denoting amplitude and frequency modulation, respectively. The main pulsation frequency is denoted by f_0, whereas f_m is the modulating frequency. In this chapter we improve the models in equations (2.2)-(2.3) and (2.4)-(2.5) from two different viewpoints. From the modeling viewpoint, we relax the assumption of parametric amplitude and frequency modulations. From the estimation viewpoint, we do not rely on the nonlinear least squares algorithms, such as the Levenberg-Marquardt algorithm, that are typically used to fit parametric nonlinear models. These methods require initial values close to the solution, which in some applications are difficult to find.

Both models proposed by Benkő et al. (2011) and Benkő (2018) and given by equations (2.2) and (2.4), respectively, are a special case of our model defined by equation (2.1). To

see this, let us define

$$\begin{aligned} v(t) &= a_0^A a_0 + a_0 g^A(t), \\ w_{1,k}(t) &= \left[a_0^A a_k + a_k g^A(t)\right] \sin[\varphi_k + k g^F(t)], \quad k = 1, \ldots, K, \\ w_{2,k}(t) &= \left[a_0^A a_k + a_k g^A(t)\right] \cos[\varphi_k + k g^F(t)], \quad k = 1, \ldots, K, \end{aligned} \tag{2.6}$$

and

$$\begin{aligned} u(t) &= m_0 + \sum_{r=1}^{\ell} b_r \sin(2\pi r f_m t + \varphi_r^b), \\ h_{1,k}(t) &= \left[a_k + g_k^A(t)\right] \sin[\varphi_k + g_k^F(t)], \quad k = 1, \ldots, K, \\ h_{2,k}(t) &= \left[a_k + g_k^A(t)\right] \cos[\varphi_k + g_k^F(t)], \quad k = 1, \ldots, K. \end{aligned} \tag{2.7}$$

We now show how equations (2.6) and (2.7) allow to us compare our model in equation (2.1) with the models proposed by Benkő et al. (2011) and Benkő (2018), respectively. Comparing the models in equations (2.1) and (2.2), time-varying trend and amplitudes of the model in equation (2.1) are expressed as

$$\begin{aligned} m(t) &= v(t), \\ g_{\ell,k}(t) &= w_{\ell,k}(t), \quad \ell = 1, 2, \quad k = 1, \ldots, K. \end{aligned} \tag{2.8}$$

At the same time, comparing the model in equation (2.1) with the model in equation (2.4), the time-varying trend and amplitudes of the model in equation (2.1) are

$$\begin{aligned} m(t) &= u(t), \\ g_{\ell,k}(t) &= h_{\ell,k}(t), \quad \ell = 1, 2, \quad k = 1, \ldots, K, \end{aligned} \tag{2.9}$$

the ordinary frequency being $f_k = k f_0$.

As we can see in equations (2.8) and (2.9), the functions $m(t)$ and $g_{\ell,k}(t)$ incorporate the amplitude and frequency modulation functions $g^M(t)$ and $g_k^M(t)$ in equations (2.3) and (2.5). In this sense, the limitation of our approach is that it does not aim at identifying the amplitude and frequency modulating functions $g^M(t)$ and $g_k^M(t)$ in equations (2.3) and (2.5). That said, the benefit of our approach from the estimation viewpoint is twofold. An

important advantage of our model in equation (2.1) over the models in equations (2.2) and (2.4) is that the modulating frequency f_m does not need to be estimated. In other words, in order to describe statistically a Blazhko light curve using our model in equation (2.1), we only need to estimate f_0. If the observed time series is indeed a Blazhko light curve, the modulating frequency f_m is included in the nonparametric trend $m(t)$ and amplitude $g_{\ell,k}$ of our model in equation (2.1). Moreover, assuming that the frequencies are known, for our model in equation (2.1) we only need to estimate the functions $m(\cdot)$ and $g_{\ell,k}(\cdot)$, whereas for the model in equations (2.2) and (2.4) the estimated parameters are the amplitudes a_0^A, a_0, m_0, a_k's, b_r's, and a_{kj}^M's, and the phases φ_r^b's, φ_k's, φ_j^M's, and φ_{kj}^M's.

2.4 Estimation

In Section 2.4.1 we define estimators of the unknown trend $m(\cdot)$, amplitudes $\{g_{\ell,k}(\cdot),\ \ell = 1, 2,\ k = 1, \ldots, K\}$, and variance σ_z^2 of the model in equation (2.1), and in Section 2.4.2 we explain how to select the tuning parameters associated with the B-splines and the penalization used in the estimation method. We denote by N the sample size, $T = t_N - t_1$ the time span, J the number of B-splines that form the basis, d the degree of the B-splines, K the number of harmonics components, r the order of the penalty, and M the number of replications in Monte Carlo simulations.

We performed our calculations using the R Language for Statistical Computing (R Core Team, 2021). Our codes combine existing functions (available as part of `R` packages) with our own development. The computations implemented in this chapter are available as a GitHub public code repository[1].

2.4.1 Penalized Least squares

As mentioned in Section 2.2, we use B-splines to estimate the trend and amplitudes of the model given by equation (2.1). The smooth trend function $m(t_i)$ is modeled as a linear

[1] https://github.com/DarlinSoto/Modulation-models.

combination of B-splines basis

$$m(t_i) = \sum_{j=1}^{J} \alpha_j B_j(t_i), \quad i = 1, \ldots, N,$$

which can be written in matrix notation as

$$\boldsymbol{m} = \mathbf{B}\boldsymbol{\alpha},$$

where $\boldsymbol{m} = [m(t_1), \ldots, m(t_N)]^\top$, $\mathbf{B} = [B_{ij}] = [B_j(t_i)]$ is the $N \times J$ basis matrix ($i = 1, \ldots, N$, $j = 1, \ldots, J$) and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)^\top$. The exact definition of B-splines is given in Appendix A.

The smooth amplitude functions, $g_{\ell,k}(t_i)$, $\ell = 1, 2$, are modeled in the same way:

$$g_{1,k}(t_i) = \sum_{j=1}^{J} \beta_{k,j} B_j(t_i), \qquad g_{2,k}(t_i) = \sum_{j=1}^{J} \gamma_{k,j} B_j(t_i), \qquad k = 1, \ldots, K.$$

In matrix notation

$$\boldsymbol{g}_{1,k} = \mathbf{B}\boldsymbol{\beta}_k \text{ and } \boldsymbol{g}_{2,k} = \mathbf{B}\boldsymbol{\gamma}_k, \qquad k = 1, \ldots, K,$$

where $\boldsymbol{\beta}_k = (\beta_{k,1}, \ldots, \beta_{k,J})^\top$, $\boldsymbol{\gamma}_k = (\gamma_{k,1}, \ldots, \gamma_{k,J})^\top$, and $\boldsymbol{g}_{\ell,k} = [g_{\ell,k}(t_1), \ldots, g_{\ell,k}(t_N)]^\top$, $\ell = 1, 2$, $k = 1, \ldots, K$. Thus, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_k$, and $\boldsymbol{\gamma}_k$, $k = 1, \ldots, K$, are vectors associated with the trend and amplitudes, respectively. We define the $N \times N$ matrices $\mathbf{C}_k$ and $\mathbf{S}_k$, $k = 1, \ldots, K$, as

$$\mathbf{C}_k = \text{diag}\{\cos(w_k t_1), \ldots, \cos(w_k t_N)\} \text{ and } \mathbf{S}_k = \text{diag}\{\sin(w_k t_1), \ldots, \sin(w_k t_N)\}.$$

Thus, the model can be expressed as

$$\boldsymbol{\mu} = \mathcal{B}\boldsymbol{\theta},$$

where $\mathcal{B}$ is the $N \times c$ design matrix given by

$$\mathcal{B} = [\mathbf{B}|\mathbf{C}_1\mathbf{B}| \ldots |\mathbf{C}_K\mathbf{B}|\mathbf{S}_1\mathbf{B}| \ldots |\mathbf{S}_K\mathbf{B}],$$

with $c = J(2K + 1)$, and

$$\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_K^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_K^\top)^\top$$

is the vector of regression coefficients of length c.

The *ordinary least squares* (OLS) estimator of $\boldsymbol{\theta}$ is the vector $\widehat{\boldsymbol{\theta}}_{\text{OLS}}$ which minimizes the sum of squares

$$M_{\boldsymbol{\theta}} = ||\boldsymbol{y} - \mathcal{B}\boldsymbol{\theta}||^2.$$

Equating to zero the partial derivatives with respect to each component of $\boldsymbol{\theta}$ and assuming (as we shall) that $\mathcal{B}^\top\mathcal{B}$ is nonsingular, the estimator of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}}_{\text{OLS}} = (\mathcal{B}^\top\mathcal{B})^{-1}\mathcal{B}^\top\boldsymbol{y}.$$

The OLS estimate also maximizes the likelihood of the observations when the errors $z_1, \ldots, z_N$ are independent and identically distributed (iid) and Gaussian.

The size of the basis determines the amount of smoothing of the fitted curves. The larger the value of J, the bumpier the fitting will be. To avoid overfitting, Eilers and Marx (1996) proposed a penalty on the (high-order) finite differences of the coefficients

$$M^*_{\boldsymbol{\theta}} = ||\boldsymbol{y} - \mathcal{B}\boldsymbol{\theta}||^2 + \tau_1||\mathbf{D}_r\boldsymbol{\alpha}||^2 + \sum_{k=1}^{K}\left\{\tau_{2k}||\mathbf{D}_r\boldsymbol{\beta}_k||^2 + \tau_{2k+1}||\mathbf{D}_r\boldsymbol{\gamma}_k||^2\right\},$$

where $\{\tau_k,\ k = 1, \ldots, 2K + 1\}$ are positive regularization parameters that control the smoothness of the curve, penalizing the coefficients that are far apart from one another. If $\tau_k = 0$, $k = 1, \ldots, 2K + 1$, we have the standard normal equations of linear regression with a B-splines basis. The larger the value of τ_k, the closer the coefficient $\boldsymbol{\theta}$ is to zero. When $\tau_k \to \infty$ we obtain a polynomial fit of degree $r - 1$ with rth order differences. The matrix $\mathbf{D}_r$ constructs rth-order differences of a vector $\boldsymbol{\eta}$ as

$$\mathbf{D}_r\boldsymbol{\eta} = \Delta^r\boldsymbol{\eta}.$$

The first difference of $\boldsymbol{\eta}$, $\Delta^1\boldsymbol{\eta}$, is the vector with elements $\boldsymbol{\eta}_{l+1} - \boldsymbol{\eta}_l$. Repeated differencing applied to $\Delta\boldsymbol{\eta}$ results in higher differences, such as $\Delta^2\boldsymbol{\eta}$ and $\Delta^3\boldsymbol{\eta}$. For $r = 1$ and $r = 2$,

the matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ are given by

$$\mathbf{D}_1 = \begin{pmatrix} -1 & 1 & 0 & 0 & \dots \\ 0 & -1 & 1 & 0 & \dots \\ 0 & 0 & -1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots \\ 0 & 1 & -2 & 1 & \dots \\ 0 & 0 & 1 & -2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The penalties can be represented as $\boldsymbol{\theta}^\top \mathbf{P} \boldsymbol{\theta}$ with the block-diagonal matrix $\mathbf{P} = \mathbf{T} \otimes \mathbf{D}_r^\top \mathbf{D}_r$ and $\mathbf{T} = \text{diag}\{\tau_1, \tau_2, \tau_3, \dots, \tau_{2K+1}\}$. Then, minimizing

$$M^*_{\boldsymbol{\theta}} = ||\boldsymbol{y} - \mathcal{B}\boldsymbol{\theta}||^2 + \boldsymbol{\theta}^\top \mathbf{P} \boldsymbol{\theta}$$

with respect to $\boldsymbol{\theta}$, the *penalized ordinary least squares estimator* (POLS) of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}}_{\text{POLS}} = (\mathcal{B}^\top \mathcal{B} + \mathbf{P})^{-1} \mathcal{B}^\top \boldsymbol{y}. \tag{2.10}$$

The prediction of y at time t_i is given by

$$\widehat{y}_i = \widehat{\mu}(t_i) = \boldsymbol{\mathcal{B}}(t_i)^\top \widehat{\boldsymbol{\theta}}_{\text{POLS}}, \tag{2.11}$$

where $\boldsymbol{\mathcal{B}}(t_i)$ is the ith row of $\mathcal{B}$, the residuals are $\widehat{\boldsymbol{z}} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$, with $\widehat{\boldsymbol{y}} = (\widehat{y}_1, \dots, \widehat{y}_N)^\top$, and the mean square error (MSE) is $\text{MSE} = N^{-1} \sum_{i=1}^N (y_i - \widehat{y}_i)^2$.

The estimators of the trend $\boldsymbol{m}$ and amplitudes $\{\boldsymbol{g}_{\ell,k},\ \ell = 1, 2,\ k = 1, \dots, K\}$, are

$$\widehat{\boldsymbol{m}} = \mathbf{B}\widehat{\boldsymbol{\alpha}}, \quad \widehat{\boldsymbol{g}}_{1,k} = \mathbf{B}\widehat{\boldsymbol{\beta}}_k, \quad \widehat{\boldsymbol{g}}_{2,k} = \mathbf{B}\widehat{\boldsymbol{\gamma}}_k. \tag{2.12}$$

Another parameter of interest is the variance of the errors, σ_z^2, which can be estimated by

$$\widehat{\sigma}_z^2 = [N - \text{tr}(\widehat{\mathbf{S}})]^{-1} \sum_{i=1}^N \left\{ y_i - \boldsymbol{\mathcal{B}}(t_i)^\top \widehat{\boldsymbol{\theta}}_{\text{POLS}} \right\}^2$$

where $\widehat{\mathbf{S}} = \mathcal{B}(\mathcal{B}^\top \mathcal{B} + \mathbf{P})^{-1} \mathcal{B}^\top$.

In addition to the point estimate, interval estimation for $\widehat{y}_i$ is often of interest and is easy to construct. Assuming that the error terms $\{z_i,\ i = 1, \dots, N\}$ follow a Gaussian

distribution with zero mean and variance σ_z^2, the $(1-\alpha)\times 100\%$ prediction interval for $\mu(t_i)$, with $i=1,\ldots,N$, is

$$\boldsymbol{\mathcal{B}}^\top(t_i)\widehat{\boldsymbol{\theta}}_{\text{POLS}} \pm z(1-\alpha/2)\sqrt{\boldsymbol{\mathcal{B}}^\top(t_i)\mathbb{V}\text{ar}\left[\widehat{\boldsymbol{\theta}}_{\text{POLS}}\right]\boldsymbol{\mathcal{B}}(t_i)}, \tag{2.13}$$

where $z(1-\alpha/2)$ denotes the $(1-\alpha/2)$ quantile of the standard Gaussian distribution, and

$$\mathbb{V}\text{ar}\left[\widehat{\boldsymbol{\theta}}_{\text{POLS}}\right] = \sigma_z^2(\mathcal{B}^\top\mathcal{B}+\mathbf{P})^{-1}\mathcal{B}^\top\mathcal{B}(\mathcal{B}^\top\mathcal{B}+\mathbf{P})^{-1}.$$

The $(1-\alpha)\times 100\%$ confidence interval for the trend $m(t_i)$ is

$$\mathscr{B}(t_i)^\top\widehat{\mathbf{Q}}_m\widehat{\boldsymbol{\theta}}_{\text{POLS}} \pm z(1-\alpha/2)\sqrt{\mathscr{B}(t_i)^\top\widehat{\mathbf{Q}}_m\mathbb{V}\text{ar}\left[\widehat{\boldsymbol{\theta}}_{\text{POLS}}\right]\widehat{\mathbf{Q}}_m^\top\mathscr{B}(t_i)}, \tag{2.14}$$

and the $(1-\alpha)\times 100\%$ confidence intervals for the amplitudes $g_{\ell,k}(t_i)$, $\ell=1,2$, $k=1,\ldots,K$, are

$$\mathscr{B}(t_i)^\top\widehat{\mathbf{Q}}_g(\ell,k)\widehat{\boldsymbol{\theta}}_{\text{POLS}} \pm z(1-\alpha/2)\sqrt{\mathscr{B}(t_i)^\top\widehat{\mathbf{Q}}_g(\ell,k)\mathbb{V}\text{ar}\left[\widehat{\boldsymbol{\theta}}_{\text{POLS}}\right]\widehat{\mathbf{Q}}_g(\ell,k)^\top\mathscr{B}(t_i)}, \tag{2.15}$$

where $\mathscr{B}(t_i)$ is the ith row of the matrix $\mathscr{B}$, and $\mathscr{B}=[\mathbf{B}|\mathbf{B}|\ldots|\mathbf{B}]$ is a matrix of dimension $N\times c$. The $c\times c$ matrices $\widehat{\mathbf{Q}}_m$, $\{\widehat{\mathbf{Q}}_g(\ell,k),\ \ell=1,2,\ k=1,\ldots,K\}$ satisfy $\widehat{\mathbf{Q}}_m\widehat{\boldsymbol{\theta}}_{\text{POLS}}=(\widehat{\boldsymbol{\alpha}}^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top$ and

$$\begin{aligned}
&\widehat{\mathbf{Q}}_g(1,1)\widehat{\boldsymbol{\theta}}_{\text{POLS}}=(\mathbf{0}_J^\top,\widehat{\boldsymbol{\beta}}_1^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top, && \widehat{\mathbf{Q}}_g(2,1)\widehat{\boldsymbol{\theta}}_{\text{POLS}}=(\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top,\widehat{\boldsymbol{\gamma}}_1^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top,\\
&\widehat{\mathbf{Q}}_g(1,2)\widehat{\boldsymbol{\theta}}_{\text{POLS}}=(\mathbf{0}_J^\top,\mathbf{0}_J^\top,\widehat{\boldsymbol{\beta}}_2^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top, && \widehat{\mathbf{Q}}_g(2,2)\widehat{\boldsymbol{\theta}}_{\text{POLS}}=(\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top,\mathbf{0}_J^\top,\widehat{\boldsymbol{\gamma}}_2^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top,\\
&\qquad\vdots && \qquad\vdots\\
&\widehat{\mathbf{Q}}_g(1,K)\widehat{\boldsymbol{\theta}}_{\text{POLS}}=(\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top,\widehat{\boldsymbol{\beta}}_K^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top, && \widehat{\mathbf{Q}}_g(2,K)\widehat{\boldsymbol{\theta}}_{\text{POLS}}=(\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top,\widehat{\boldsymbol{\gamma}}_K^\top)^\top.
\end{aligned}$$

2.4.2 Automatic selection of the tunable parameters

Before calculating the estimator in equation (2.10), it is necessary to select the tuning parameters $\boldsymbol{\tau}=(\tau_1,\tau_2,\ldots,\tau_{2K+1})^\top$. To choose the tuning parameters, we propose to use the Akaike information criterion (AIC).

The AIC penalizes the log-likelihood of a fitted model by considering the effective number of parameters. The definition of AIC given by Hastie et al. (2004) is

$$\text{AIC}(\boldsymbol{\tau}) = \overline{\text{err}}(\boldsymbol{\tau}) + 2\frac{\text{df}}{N}\widehat{\sigma}_0^2,$$

where $\overline{\text{err}}(\boldsymbol{\tau})$ corresponds to the mean square error in the case of Gaussian errors, df is the effective number of parameters, N is the number of observations used to fit the model, and $\widehat{\sigma}_0^2$ is given by the variance of the residuals from the $\widehat{y}_i$ that are computed when $\boldsymbol{\tau} = \mathbf{0}_{2K+1}$.

The value for $\boldsymbol{\tau}$ is chosen by minimizing the AIC, which is computed as

$$\text{AIC}(\boldsymbol{\tau}) = \frac{1}{N}\sum_{i=1}^{N}\left\{y_i - \boldsymbol{\mathcal{B}}(t_i)^\top \widehat{\boldsymbol{\theta}}_{\text{POLS}}\right\}^2 + 2\frac{\text{tr}(\widehat{\mathbf{S}})}{N}\widehat{\sigma}_0^2, \tag{2.16}$$

The AIC given by equation (2.16) can also be used to select the number of B-splines J, the degree d of the B-spline, the order of penalty r, and the number of harmonic components K.

In Figure 2.1, we have generated $N = 500$ observations from the model described in equation (2.1), with the Gaussian errors $\{z_i,\ 1 \leq i \leq 500\}$ being simulated using the `R` function `rnorm`. We consider the following artificial signal:

$$\mu(t_i) = -0.05t_i - (-0.0002t_i + 0.0003t_i^2)\cos(0.2\pi t_i) + (1 - 0.0005t_i)\sin(0.2\pi t_i),$$

with the errors following a Gaussian distribution with zero mean and variance $\sigma_z^2 = 1$. Time t is unequally spaced and was obtained from a uniform distribution $U(\theta_1, \theta_2)$ with $\theta_1 = 0$ and $\theta_2 = 55$ using the `R` function `runif`. In the first plot of Figure 2.1 the observations $\boldsymbol{y}$ are represented by the gray points, and the mean $\boldsymbol{\mu}$ by the black curve. The orange, blue, and green curves illustrate three possible estimates for $\boldsymbol{y}$ obtained using the method described in Section 2.4.1 with increasing smoothing parameters. The orange line is the fit obtained with $\tau_j = 0$, $j = 1, 2, 3$: the corresponding $\widehat{\boldsymbol{y}}$ matches the data well, but fits the true $\boldsymbol{\mu}$ poorly because it is wiggly. The blue curve is obtained using the smoothing parameters $\tau_j = 30$, $j = 1, 2, 3$, and the green curve is obtained using $\tau_j = 200$, $j = 1, 2, 3$. In the second plot of Figure 2.1, we observe that the optimal tuning parameters are $\tau_j = 30$, $j = 1, 2, 3$, and as the values of $\boldsymbol{\tau}$ increase the obtained curve fits the observed data less closely.

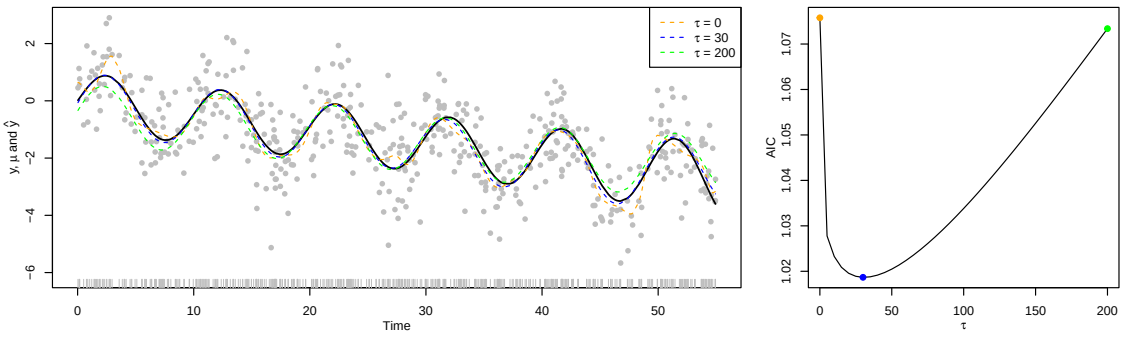

Figure 2.1 Automatic selection of the tunable parameters presented in Section 2.4.2. Left: data (gray dots) simulated according to the model defined by equation (2.1), with $N = 500$, $\mu(t_i) = -0.05t_i - (-0.0002t_i + 0.0003t_i^2)\cos(0.2\pi t_i) + (1 - 0.0005t_i)\sin(0.2\pi t_i)$ (black curve), and where the errors follow a Gaussian distribution with zero mean and variance $\sigma_z^2 = 1$. Time is unequally spaced, obtained from a uniform distribution $U(0, 55)$ (gray ticks on the horizontal axis). We illustrate three estimates of $\boldsymbol{y}$ corresponding to three different specifications of τ_j, with $j = 1, 2, 3$: $\tau_j = 0$ (orange curve), $\tau_j = 30$ (blue curve), and $\tau_j = 200$ (green curve). Right: values of the AIC in equation (2.16), obtained from the simulated and estimated light curve, corresponding to forty-one equally spaced values of τ ranging from 0 to 200. The three points (orange, blue, and green) on the AIC curve correspond to the three fits presented in the left-hand plot of the figure.

2.5 Detecting serial correlation

A statistical model is an approximation to the true process that generates the observed data. After fitting the model given by equation (2.1), it is necessary to check whether the residuals obtained from the fit behave like a white noise process. A significant departure from this assumption suggests the inadequacy of the assumed form of the model. Thus, it is important to assess whether the residuals follow a white noise process.

Detecting serial correlation becomes more challenging when the available observations are unequally spaced in time. If the observations are unequally spaced, so are the errors. In order to study the spectral density of the residuals obtained from the fitted model, in this section we derive the mathematical definition of the spectrum for unequally spaced time series.

Before presenting our approach, we briefly review the results given by Deeming (1975) about the relationships between the periodogram, the spectral density (PSD), and the autocorrelation function for *continuous* time series. Then, we extend the results given by Deeming (1975) to the case of *discrete* time series.

Let $\{\varepsilon_i\}$ be a *continuous*, zero-mean stationary times series with spectral density

$$P_\varepsilon(\lambda) = \int_{-\infty}^{\infty} r_\varepsilon(h) \exp(i\lambda h) dh, \quad -\infty < \lambda < \infty,$$

and autocovariance function

$$r_\varepsilon(h) = \int_{-\infty}^{\infty} P_\varepsilon(\lambda) \exp(-i\lambda h) d\lambda, \quad h \in \mathbb{R}.$$

Consider a time series $\varepsilon_1, \ldots, \varepsilon_N$ with spectrum $P_\varepsilon(\cdot)$ and autocovariance function $r_\varepsilon(\cdot)$, and assume that the observations $\varepsilon_1, \ldots, \varepsilon_N$ are obtained at *unequally spaced* times $t_1, \ldots, t_N$, respectively. The periodogram of $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)^\top$ at frequency λ is defined as

$$I_\varepsilon(\lambda) = \sum_{k=1}^{N} \sum_{j=1}^{N} \varepsilon_k \varepsilon_j \exp(i\lambda[t_k - t_j]), \quad \lambda = 2\pi f. \tag{2.17}$$

Deeming (1975) proved that the expectation of the periodogram of $\{\varepsilon_i\}$ in equation (2.17) is

$$\mathbb{E}\left[I_\varepsilon(\lambda)\right] = P_\varepsilon(\lambda) \star W_\varepsilon(\lambda), \tag{2.18}$$

where $W_\varepsilon(\lambda)$ is the power spectral window given by

$$W_\varepsilon(\lambda) = \sum_{j=1}^{N} \sum_{k=1}^{N} \exp(i\lambda[t_k - t_j]),$$

and $P_\varepsilon(\lambda) \star W_\varepsilon(\lambda)$ is the continuous convolution of $P_\varepsilon(\lambda)$ with $W_\varepsilon(\lambda)$ defined as

$$P_\varepsilon(\lambda) \star W_\varepsilon(\lambda) = \int_{-\infty}^{\infty} P_\varepsilon(\omega) W_\varepsilon(\lambda - \omega) d\omega.$$

The following lemma states that it is possible to extend the result in equation (2.18) to the case of a *discrete* zero-mean stationary times series that is generated according to *equally*

spaced times but observed at *unequally spaced* times. The lemma applies to unequally spaced time points t_i with index i belonging to a subset $\mathcal{I}$ of the set $\mathbb{N} = \{1, 2, \dots\}$ of positive integers.

Lemma 1 *Let* $\{\varepsilon_i,\ t_i = t_0 + i\Delta,\ \Delta > 0,\ i \in \mathcal{I} \subseteq \mathbb{N}\}$ *be a zero-mean, stationary, discrete time series with spectral density*

$$P_\varepsilon(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \exp(i\lambda h\Delta) r_\varepsilon(h), \quad -\infty < \lambda < \infty, \tag{2.19}$$

with autocovariance function defined as $r_\varepsilon(h) = \mathbb{E}\left[\varepsilon_k \varepsilon_j\right]$, *with* $k = j + |h|$, $h \in \mathbb{Z}$, *that can be expressed in term of the spectral density in equation* (2.19) *as*

$$r_\varepsilon(h) = \frac{2\pi}{N_\mathcal{I}} \sum_{j=1}^{N_\mathcal{I}} \exp(-i\lambda_j h\Delta) P_\varepsilon(\lambda_j), \quad h \in \mathbb{Z}, \tag{2.20}$$

where $\lambda_j = 2\pi f_j$, *with* $f_j = j/(N_\mathcal{I}\Delta)$ *and* $N_\mathcal{I} = \max\{\mathcal{I}\}$. *Then, the expectation of the periodogram in equation* (2.17) *obtained from* $\{\varepsilon_i, i \in \mathcal{I}\}$ *is*

$$\mathbb{E}\left[I_\varepsilon(\lambda)\right] = \frac{2\pi}{N_\mathcal{I}} P_\varepsilon(\lambda) * W_\varepsilon(\lambda), \tag{2.21}$$

with power spectral window given by

$$W_\varepsilon(\lambda) = \sum_{k\in\mathcal{I}} \sum_{j\in\mathcal{I}} \exp(i\lambda[t_k - t_j]) \tag{2.22}$$

and $P_\varepsilon(\lambda) * W_\varepsilon(\lambda)$ *is the discrete convolution of* $P_\varepsilon(\lambda)$ *with* $W_\varepsilon(\lambda)$ *defined as*

$$P_\varepsilon(\lambda) * W_\varepsilon(\lambda) = \sum_{j=1}^{N_\mathcal{I}} P_\varepsilon(\omega_j) * W_\varepsilon(\lambda - \omega_j), \quad \omega_j = 2\pi f_j, \quad f_j = \frac{j}{N_\mathcal{I}\Delta}.$$

The PSD defined in equation (2.19) have period 2π and is even. Therefore it suffices to confine attention to the value of f on the interval $(-\pi, \pi]$.

Our result in equation (2.21) differs from the result by Deeming (1975) in equation (2.18). Deeming (1975) proved that the expectation of both discrete and continuous Fourier transforms of a *continuous stochastic process* $f(t)$; (in the sense of equations (31) and (32) in

Deeming, 1975) is equal to the *continuous convolution* of the spectral density of $f(t)$ with a spectral window (see equations (33) and (36) in Deeming, 1975). In Lemma 1, instead, we prove that the expectation of the discrete Fourier transform of the *discrete stochastic process* ε_i is equal to the *discrete convolution* of the spectral density of ε_i with a spectral window (up to the constant $2\pi/N_{\mathcal{I}}$).

When the time series is generated according to an equally spaced stochastic process and the observations are equally spaced, the periodogram is an (asymptotically) unbiased estimator of the spectral density (see Priestley, 1981, page 418). However, when the observations are unequally spaced it does not make sense to estimate the spectral density in the same way. This is due to the *power spectral window* $W_\varepsilon(\lambda)$ in equations (2.21)-(2.22). Nevertheless, as we show in the following proposition, it is possible to disentangle the spectral density $P_\varepsilon(\lambda)$ from the spectral window $W_\varepsilon(\lambda)$.

Proposition 1 *Let $\mathcal{F}\{g_j\}[k]$ denote the Discrete Fourier Transform of the sequence of m numbers $g_1, \ldots, g_m$ into another sequence $h_1, \ldots, h_m$, that is,*

$$h_k = \mathcal{F}\{g_j\}[k] = \sum_{j=1}^{m} g_j \exp(-ikj2\pi/m), \quad k = 1, \ldots, m. \tag{2.23}$$

Accordingly, define $\mathcal{F}^{-1}\{h_k\}[j]$ as the Inverse Discrete Fourier Transform of the sequence $h_1, \ldots, h_m$ into another sequence $g_1, \ldots, g_m$, that is,

$$g_j = \mathcal{F}^{-1}\{h_k\}[j] = \frac{1}{m}\sum_{k=1}^{m} h_k \exp(ikj2\pi/m), \quad j = 1, \ldots, m. \tag{2.24}$$

Assume that $\{\varepsilon_i, i \in \mathcal{I}\}$ satisfy the same conditions as in Lemma 1. Then we can write the spectral density $P_\varepsilon(\lambda_j)$ in equation (2.19) *at frequency $\lambda_j = 2\pi f_j$, with $f_j = j/(N_{\mathcal{I}}\Delta)$, as*

$$P_\varepsilon(\lambda_j) = \frac{N_{\mathcal{I}}}{2\pi}\mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{\mathbb{E}\left[I_\varepsilon(\lambda_j)\right]\}[k]}{\mathcal{F}\{W_\varepsilon(\lambda_j)\}[k]}\right\}[j], \qquad j = 1, \ldots, N_{\mathcal{I}}. \tag{2.25}$$

The proofs of Lemma 1 and Proposition 1 are given in Appendix C. Equation (2.25) suggests that in order to estimate $P_\varepsilon(\lambda)$, we need the value of Δ and $W_\varepsilon(\lambda)$, as well as the estimate of $\mathbb{E}\left[I_\varepsilon(\lambda)\right]$. Notice that, the errors $\boldsymbol{z} = (z_1, \ldots, z_N)^\top$ of our model in equation

(2.1) are assumed to be serially uncorrelated. The following algorithm is proposed to establish whether the unequally spaced residuals obtained when fitting our model in equation (2.1) are uncorrelated.

(i) Obtain the residuals $\widehat{\boldsymbol{z}} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$.

(ii) For each $j = 1, \ldots, N_{\mathcal{I}}$, define $I_{\widehat{z}}(\lambda_j)$ as the periodogram in equation (2.17) computed upon the residuals $\widehat{\boldsymbol{z}}$ obtained in step i), with $\lambda_j = \frac{2\pi j}{N_{\mathcal{I}}\Delta}$.

(iii) For each $j = 1, \ldots, N_{\mathcal{I}}$, define $W_{\widehat{z}}(\lambda_j)$ as the power spectral window in equation (2.22) computed upon the residuals $\widehat{\boldsymbol{z}}$ obtained in step i), with $\lambda_j = \frac{2\pi j}{N_{\mathcal{I}}\Delta}$.

(iv) Smooth the periodogram obtained in step ii) over frequencies, and denote the smoothed periodogram by $\tilde{I}_{\widehat{z}}(\lambda_j)$.

(v) Calculate the Discrete Fourier Transform in equation (2.23) of the power spectral window and the periodogram obtained in steps (iii) and (iv), respectively.

(vi) For each frequency $\lambda_j = \frac{2\pi j}{N_{\mathcal{I}}\Delta}$, define the estimated spectral density of the errors $\boldsymbol{z}$ as

$$\widehat{P}_z(\lambda_j) = \frac{N_{\mathcal{I}}}{2\pi}\mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{\tilde{I}_{\widehat{z}}(\lambda_j)\}[k]}{\mathcal{F}\{W_{\widehat{z}}(\lambda_j)\}[k]}\right\}[j], \qquad j = 1, \ldots, N_{\mathcal{I}}, \tag{2.26}$$

where the inverse Fourier transform $\mathcal{F}^{-1}$ is given by equation (2.24).

(vii) If the estimated spectral density obtained in step (vi) does not vary significantly over frequencies, conclude that the errors are uncorrelated over time.

2.6 Simulation results

In this section we provide Monte Carlo simulations to illustrate the performance of the estimators $\widehat{\mu}(t)$, $\widehat{m}(t)$, $\widehat{g}_{\ell,k}(t)$, $\ell = 1, 2$, $k = 1, \ldots, K$, defined by equations (2.11) and (2.12), and the estimator of the spectral density in equation (2.25).

In Section 2.6.1 we simulate unequally spaced observations from the model in equation (2.1), under two scenarios. In the first scenario both the trend and amplitudes are sinusoidal, whereas in the second scenario the trend and amplitudes are polynomial. In Section 2.6.2 we simulate a Blazhko light curve and fit the model in equation (2.1). Finally in Section 2.6.3 we evaluate the performance of the estimator of the spectral density defined in equation (2.25) of a discrete unequally spaced time series.

2.6.1 Simulating the time-varying model

In this section we generate the data according to the model described by equation (2.1) with $K = 2$, $N = 500$, and time t is unequally spaced, obtained form a uniform distribution $U(\theta_1, \theta_2)$ with $\theta_1 = 0$ and $\theta_2 = 1$. In order to illustrate the flexibility of our novel method, we consider two different scenarios for the trend and amplitudes. In the first scenario, we simulate a sinusoidal trend and amplitudes as $m(t) = \sin(2\pi t)$, $g_{1,1}(t) = \cos(9\pi t)$, $g_{2,1}(t) = \sin(6\pi t)$, $g_{1,2}(t) = \cos(4\pi t)$, $g_{2,2}(t) = \sin(7\pi t)$, with frequencies $w_1 = 40\pi$, and $w_2 = 100\pi$. In the second scenario, we simulate (global) polynomial trend and amplitudes as $m(t) = 0.2t - 5t^2 + 5.5t^3$, $g_{1,1}(t) = 4t^3 - 5t^2$, $g_{2,1}(t) = -0.5 - 0.5t + 2.5t^2 - 0.5t^3$, $g_{1,2}(t) = -t + t^2 + 1.3t^3$, $g_{2,2}(t) = 0.5 + 2t^2 - 3t^3$, with frequencies $w_1 = 30\pi$, and $w_2 = 40\pi$. In both scenarios, we assume that the error terms $\{z_i,\ i = 1, \ldots, N\}$ is a Gaussian white noise sequence with zero mean and variance $\sigma_z^2 = 2$.

In both scenarios, we simulate $M = 200$ realizations of the model in equation (2.1). For each $j = 1, \ldots, M$, we compute the estimate $\widehat{\boldsymbol{\theta}}^{(j)}$ defined by equation (2.10). In the first scenario, we select the smoothing parameter $\boldsymbol{\tau} = (50, 1, 2, 10, 1)^\top$, a total number of B-splines $J = 33$ of order $d = 3$, and an order penalty $r = 2$. In the second scenario, we choose the smoothing parameter $\boldsymbol{\tau} = (3, 3, 3, 3, 3)^\top$, a total number of B-splines $J = 6$ of order $d = 3$, and an order penalty $r = 4$. Figure 2.2 shows our estimates of $\boldsymbol{\mu}$, $\boldsymbol{m}$, $\{\boldsymbol{g}_{\ell,k},\ \ell = 1, 2,\ k = 1, 2\}$, and their 95% confidence intervals. Figure 2.2 shows that our model in equation (2.1) fits well the simulated data in both the sinusoidal and polynomial scenarios. That is, the trend and amplitudes are well fitted in both scenarios.

The 95% confidence intervals are constructed in a nonparametric fashion using quantiles; see Appendix B for more details.

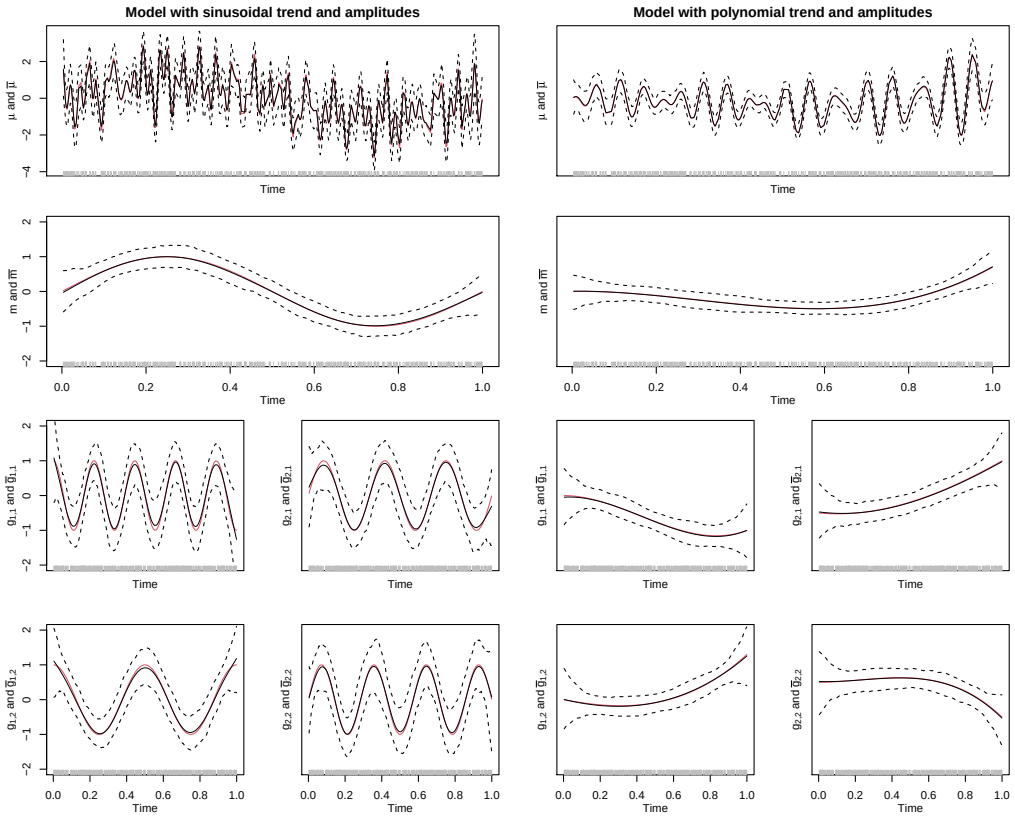

Figure 2.2 Simulation scenarios of Section 2.6.1: data generated from the model in equation (2.1) with sinusoidal and polynomial time-varying trends and amplitudes. Time t is unequally spaced. The first column shows the fit of the model with a sinusoidal trend and amplitudes, whereas the second column shows the fit of the model with a polynomial trend and amplitudes. The first row shows the true $\mu(t)$ (red solid line), together with the average $\overline{\mu}(t) = \frac{1}{M}\sum_{j=1}^{M}\widehat{\mu}^{(j)}(t)$ of the estimates $\widehat{\mu}^{(j)}(t)$ (black solid line). The second row shows the true trend $m(t)$ (red solid line) together with the average $\overline{m}(t) = \frac{1}{M}\sum_{j=1}^{M}\widehat{m}^{(j)}(t)$ of the estimates $\widehat{m}^{(j)}(t)$ (black solid line). The third and fourth rows show the true amplitudes $g_{\ell,k}(t)$ (red solid line), $\ell = 1, 2$, $k = 1, 2$, together with the average $\overline{g}_{\ell,k}(t) = \frac{1}{M}\sum_{j=1}^{M}\widehat{g}_{\ell,k}^{(j)}(t)$ of the estimates $\widehat{g}_{\ell,k}^{(j)}(t)$ (black solid line), with $M = 200$. The nonparametric quantiles (black dashed lines) are the confidence intervals corresponding to the 2.5th and 97.5th order statistics, respectively, see Appendix B.

2.6.2 Simulating a Blazhko RR Lyrae light curve characterized by amplitude modulation

We simulate a Blazhko RR Lyrae light curve with amplitude modulation according to Benkő et al. (2011) as

$$\begin{aligned} y_i &= \mu(t_i) + z_i, \quad i = 1, \ldots, N = 1000, \\ \mu(t_i) &= \left[1 + \frac{U_m(t_i)}{U_c}\right] c(t), \\ c(t) &= a_0 + \sum_{k=1}^{4} a_k \sin(2\pi k f_0 t_i + \varphi_k), \\ U_m(t) &= a_m \sin(2\pi f_m t + \varphi_m), \end{aligned} \tag{2.27}$$

where $c(t)$ is the carrier wave with four harmonic components, $U_m(t)$ is the modulating signal, $U_c = a_m/h$ is the amplitude of the unmodulated light curve, and $\{z_i, i = 1, \ldots, N\}$ are the error terms. The values of the parameters used in equation (2.27) and the time-design are obtained from Benkő et al. (2011). In particular, $a_m = 0.1$ mag, $h = 1.2$, $a_0 = 0.01$ mag, $f_m = 0.05$ days^{-1}, $\varphi_m = 270°$, and the values $\{a_k, \varphi_k, 1 \leq k \leq 4\}$ are presented in Table 2.1.

Parameters of Simulated RR Lyrae Star

k	kf_0 (days^{-1})	a_k (mag)	φ_k (°)
1	2	0.401	5.490
2	4	0.171	144.040
3	6	0.133	285.25
4	8	0.097	81.290

Table 2.1 Parameters (frequencies, amplitudes, and phases) obtained from Benkő et al. (2011), as explained in Section 2.6.2.

We convert the Blazhko phase φ_m and the main phases $\{\varphi_k, 1 \leq k \leq 4\}$ in equation (2.27) from degrees to radians using the `R` function `NISTdegTOradian` (avail-

able in the `R` package `NISTunits2016` by Gama, 2016). The original time design $\{t_j,\, j = 1, \ldots, 28799\}$ in Benkő et al. (2011) is equally spaced. However, variable stars are often observed at irregular intervals. For this reason, in our simulation exercise we sample a subset of the original time points and use this subset to evaluate the performance of our method. We obtain the time design $\{t_i,\, i = 1, \ldots, 1000\}$ in equation (2.27) by sampling the original, equally spaced time design $\{t_j,\, j = 1, \ldots, 28799\}$. We end up with $N = 1000$ unequally spaced observations ranging from $t = 0.03819$ days to $t = 69.37847$ days. The error terms $\{z_i,\ i = 1, \ldots, N\}$ is a Gaussian white noise sequence with zero mean and variance $\sigma_z^2 = 0.005$.

If we consider our novel model in equation (2.1) with time-varying trend and amplitudes specified as

$$\begin{aligned} m(t_i) &= a_0 \left[1 + U_m(t_i)/U_c\right], \\ g_{1,k}(t_i) &= a_k \cos(\varphi_k)[1 + U_m(t_i)/U_c], \quad k = 1, \ldots, 4, \\ g_{2,k}(t_i) &= a_k \sin(\varphi_k)[1 + U_m(t_i)/U_c], \quad k = 1, \ldots, 4, \end{aligned} \tag{2.28}$$

we can rewrite the model in equation (2.27) as a special case of our model given by equation (2.1). The main advantage of fitting the model in equation (2.1) instead of the model in equation (2.27), is that one does not need to estimate the parameters a_m, h, a_0, φ_m, a_k, φ_k, $k = 1, \ldots, 4$, f_m. Moreover, we do not need to adopt any specific functional form for $m(\cdot)$ and $g(\cdot)$, such as those given by equation (2.28), because they are well approximated by B-splines.

We fit the model in equation (2.1) with $K = 4$ to the data generated according to the model in equation (2.27). We assume that the frequencies f_k, $k = 1, \ldots, 4$, of each harmonic component are known; see Table 1. Also, we use a total of $J = 18$ B-splines of degree $d = 3$, an order penalty $r = 1$, and the smoothing parameters $\boldsymbol{\tau} = (5, 1, 0.1, 0.1, 0.1, 0.1, 1, 0.1, 4)^\top$.

We fit the model in equation (2.1) to the simulated data obtained from the model in equation (2.27), and present the results in Figure 2.3. The first row shows the simulation of the amplitude modulated RR Lyrae light curve given by equation (2.27) (gray points),

together with the true and fitted curve (solid red and solid black lines, respectively). The second row shows the residuals, and the third and last fourth rows show the true trend and amplitudes (red lines) given by equation (2.28) and their fits (black lines). We observe from Figure 2.3 that the model in equation (2.1) fits well the simulated data. That is, the trend and amplitudes are well fitted, and the residuals satisfy the assumption of zero mean and constant variance. The 95% confidence intervals are constructed in a parametric fashion; see equations (2.13) - (2.15).

2.6.3 Estimating the spectral density of unequally spaced time series

In this section we estimate the spectral density of unequally spaced time series by means of our novel estimator in equation (2.25). To this end, we simulate unequally spaced observations generated from the following AR(2) process:

$$\begin{aligned} \varepsilon_i &= \phi_1 \varepsilon_{i-1} + \phi_2 \varepsilon_{i-2} + z_i \\ \{z_i\} &\sim \mathcal{N}(0, \sigma_z^2) \\ t_i &= t_0 + i\Delta, \end{aligned} \tag{2.29}$$

where $i = 1, \ldots, N$, with $N = 500$ equally spaced observations, starting time $t_0 = 0.67$, and $\Delta = 0.33$, and $\{z_i,\, i = 1, \ldots, N\}$ a Gaussian white noise sequence with zero mean and variance σ_z^2. In order to simulate a realistic AR(2) process, we use the coefficients of the sunspot numbers in Example 3.2.9 of Brockwell and Davis (2016), where $\phi_1 = 1.318$, $\phi_2 = -0.634$, and $\sigma_z^2 = 289.2$. These ϕ coefficients ensure the existence of a causal solution

$$\varepsilon_i = \sum_{j=0}^{\infty} \psi_j z_{i-j} \tag{2.30}$$

of equation (2.29). The time series in equation (2.30) is *causal* in the sense that ε depends upon current and past (rather than future) values of the error term z. We simulate $M = 500$ times the AR(2) model given by equation (2.29) obtaining the observations $\varepsilon_1^{(m)}, \ldots, \varepsilon_N^{(m)}$, $m = 1, \ldots, M$. Then, in order to obtain unequally spaced observations we use the following three steps.

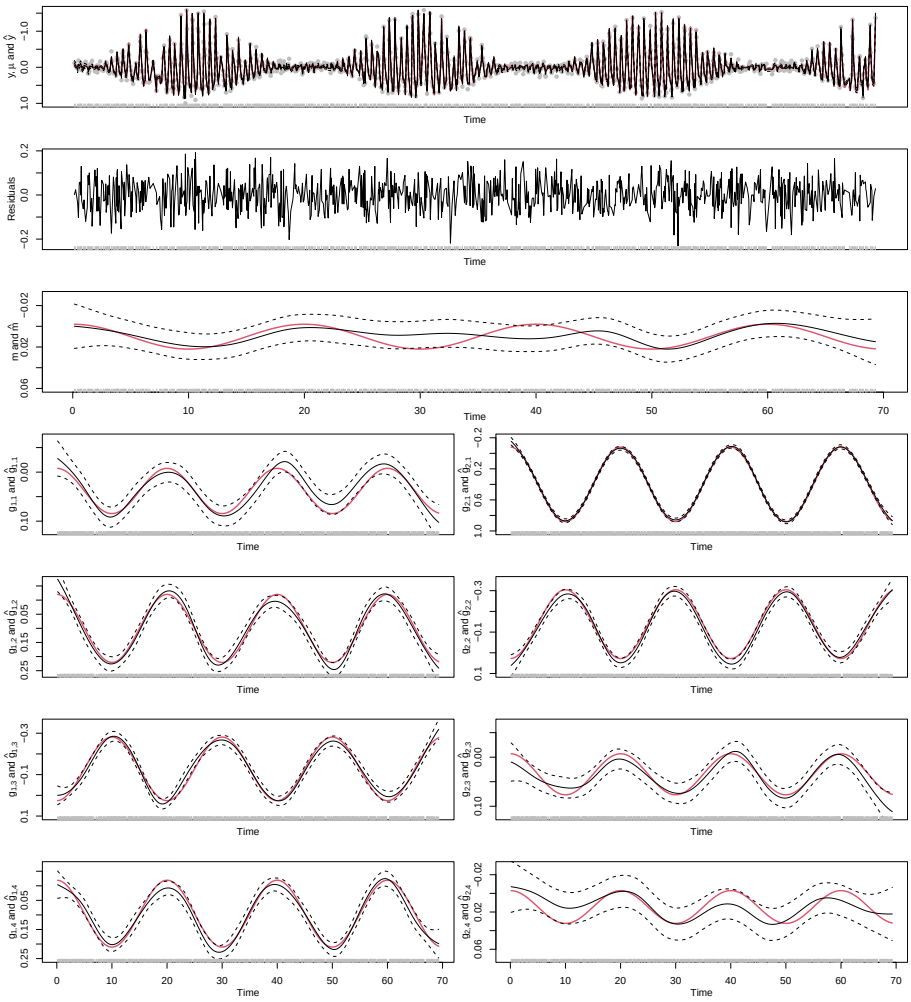

Figure 2.3 Simulation scenario of Section 2.6.2: simulated Blazhko RR Lyrae light curve characterized by amplitude modulation. The first row shows the light curve data of an RR Lyrae star simulated according to the model given by equation (2.27) (gray points), the curve $\mu(t)$ (solid red line), and the prediction $\widehat{y}_i$ (solid black line). The second row shows the residuals. The third and last four rows show the time-varying trend and amplitudes (red solid lines) together with the estimated trend and amplitudes (solid black lines). Trend and amplitudes are simulated according to equation (2.28). The parametric 95% confidence intervals (dashed black lines) are obtained according to equations (2.13) - (2.15).

1. We divide time into 50 blocks, where each block has 10 observations, in a way to preserve the original time series structure.

2. In order to preserve the autocorrelation between the observations, we select randomly 30 blocks and collect the time points corresponding to these blocks, obtaining a new set of time points $\{t_i^*, i = 1, \ldots, n\}$, with $n = 300$ observations. In contrast to the simulation schemes of Sections 2.6.1 and 2.6.2 where time was sampled randomly, here data sets with uniformly sampled subsets are produced. While the former sampling is close to the data distribution of large ground-based surveys, the latter is the typical sampling of photometric space telescopes that are dedicated to high-cadence time-series observations, such as Kepler (Koch et al., 2010).

3. Finally, we collect the observations $\varepsilon_i^{(m)}$ corresponding to the new set of time points $\{t_i^*, i = 1, \ldots, n\}$ and rename them as $e_i^{(m)} = \varepsilon_i^{(m)}$, with $e_i^{(m)}$ being observed at time t_i^*, $i = 1, \ldots, n$.

Thus, we obtain the unequally spaced observations $e_1^{(m)}, \ldots, e_n^{(m)}$, which represent a subset of the equally spaced time series $\varepsilon_1^{(m)}, \ldots, \varepsilon_N^{(m)}$. For each $m = 1, \ldots, M$, and each fixed frequency $\lambda_j = 2\pi f_j$, $f = j/(N\Delta)$, $j = 1, \ldots, N$, we compute the periodogram of $e_1^{(m)}, \ldots, e_n^{(m)}$ as

$$I_e^{(m)}(\lambda_j) = \sum_{k=1}^{n} \sum_{d=1}^{n} e_k^{(m)} e_d^{(m)} \exp(i\lambda_j [t_k^* - t_d^*]),$$

the average of the periodograms $I_e^{(m)}(\lambda_j)$ as

$$\overline{I}_e(\lambda_j) = \frac{1}{M} \sum_{m=1}^{M} I_e^{(m)}(\lambda_j),$$

and the power spectral window of $e_1^{(m)}, \ldots, e_n^{(m)}$ as

$$W_e(\lambda_j) = \sum_{d=1}^{n} \sum_{k=1}^{n} \exp(i\lambda_j [t_k^* - t_d^*]).$$

For each frequency λ_j, $j = 1, \ldots, N$, replacing $\mathbb{E}\left[I_e(\lambda_j)\right]$ with $\overline{I}_e(\lambda_j)$ and substituting $W_e(\lambda_j)$ in equation (2.25), the estimated spectral density of the unequally spaced time series $e_1^{(m)}, \ldots, e_n^{(m)}$ is given by

$$\widehat{P}_e(\lambda_j) = \frac{N}{2\pi}\mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{\overline{I}_e(\lambda_j)\}[k]}{\mathcal{F}\{W_e(\lambda_j)\}[k]}\right\}[j], \tag{2.31}$$

A smooth version of the estimated spectral density in equation (2.31) is

$$\widetilde{P}_e(\lambda_j) = \tfrac{1}{N_\lambda}\sum_{i=1}^{N_\lambda} K_h(\lambda_j - \lambda_i)\widehat{P}_e(\lambda_j). \tag{2.32}$$

The rescaled kernel function is defined as $K_h(x) = \frac{1}{h}K(x/h)$, where K is a second-order kernel and h is the bandwidth. For this application, we used the Gaussian kernel $K(y) = \frac{1}{\sqrt{2\pi}}\exp(-y^2/2)$ and a bandwidth $h = 0.3$.

Figure 2.4 compares the underlying spectral density $P_\varepsilon(\lambda_j)$ of the equally spaced time series $\{\varepsilon_i\}$, with the estimated spectral densities $\widehat{P}_e(\lambda_j)$ and $\widetilde{P}_e(\lambda_j)$ of the unequally spaced time series $\{e_i\}$. The underlying spectral density of the equally spaced time series $\{\varepsilon_i\}$, generated by the AR(2) process in equation (2.29), is given by

$$P_\varepsilon(\lambda_j) = \frac{\sigma_z^2}{2\pi}\left[1 + \phi_1^2 + \phi_2^2 + 2\phi_2 + 2(\phi_1\phi_2 - \phi_1)\cos(\lambda_j\Delta) - 4\phi_2\cos^2(\lambda_j\Delta)\right]^{-1}. \tag{2.33}$$

The estimated spectral density $\widehat{P}_e(\lambda_j)$ of the unequally time series $\{e_i\}$ is given in equation (2.31), and its smooth version $\tilde{P}_e(\lambda_j)$ in equation (2.32). Figure 2.4 shows that the estimated spectral density of the unequally time series, $\widetilde{P}_e(\lambda_j)$, fits very well the true spectral density $P_\varepsilon(\lambda_j)$.

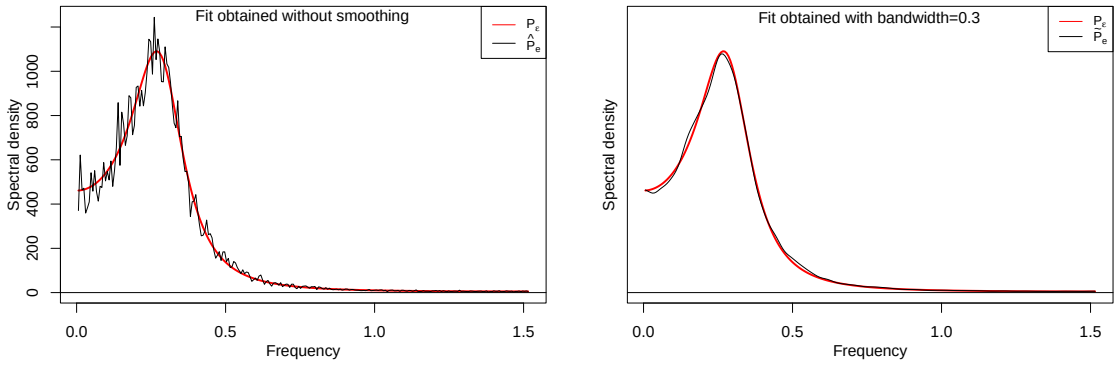

Figure 2.4 Simulation scenario of Section 2.6.3: estimated spectral density of the unequally spaced time series sampled by blocks. Left: comparison between the true spectral density in equation (2.33) (red line) and the estimated spectral density $\widehat{P}_e(\lambda_j)$ in equation (2.31) (black line) for $j = 1 \ldots, N/2$. Right: comparison between the true spectral density in equation (2.33) (red line) and the smooth estimated spectral density $\widetilde{P}_e(\lambda_j)$ in equation (2.32) (black line) for $j = 1 \ldots, N/2$. The true spectral density corresponds to the equally spaced time series ε_i which follows the AR(2) process given by equation (2.29), whereas the estimated spectral density is computed from the unequally spaced observations $e_1^{(m)}, \ldots, e_n^{(m)}$, $m = 1, \ldots, M$. The unequally spaced time series $e_1^{(m)}, \ldots, e_n^{(m)}$ is obtained as a subset of the time series $\varepsilon_1^{(m)}, \ldots, \varepsilon_N^{(m)}$. In this example $N = 500$, $n = 300$, and $M = 500$.

2.7 Application to real data

In this section, we fit our model in equation (2.1) and the model proposed by Benkő (2018) in equation (2.4) to the same light curve: the V783 Cyg, KIC 5559631. This time series has 61,351 unequally spaced observations and is available online from the Konkoly Observatory of the Hungarian Academy of Sciences webpage.[2] We choose this particular light curve for two reasons. Firstly, the Blazhko effect of the V783 Cyg time series is known to be characterized by a sinusoidal amplitude and frequency modulation (Benkő et al., 2014). The light curve of V783 Cyg can be described by $K = 15$ significant harmonics with a sinusoidal amplitude and frequency modulations (Benkő et al., 2014), which makes V783

[2] https://konkoly.hu/KIK/data_en.html

Cyg an ideal target for comparing the fits obtained with the models in equations (2.1) and (2.4). Secondly, these two modulations are well captured and fitted by our novel model in equation (2.1).

In order to reduce the computational time and to satisfy the condition $t_i = t_0 + i\Delta$ with $\Delta > 0$ and $i \in \mathcal{I} \subseteq \mathbb{N}$ (which is required by Proposition 1), we analyze a ≈ 78 days segment of this light curve from $t = 827.44$ days to $t = 904.9$ days. For this segment, the time-origin and the time-spacing take the values $t_0 = 827.42$ days and $\Delta = 0.0204345$ days, respectively, with a total of $N = 2101$ unequally spaced observations.

When fitting the models in equations (2.1) and (2.4), the main pulsation and modulation frequencies are not estimated: they take the values $f_0 = 1.611084$ days^{-1} and $f_m = 0.036058$ days^{-1} (see Benkő et al., 2014), respectively. In addition, to the $K = 15$ significant harmonics fitted by Benkő et al. (2014), we found, after pre-whitening and fitting our model in equation (2.1), four significant frequencies taking the values $f'_{11} = 18.3254$ days^{-1}, $f'_{12} = 19.9365$ days^{-1}, $f'_{13} = 21.5476$ days^{-1}, and $f'_{14} = 23.1587$ days^{-1}. The values we obtain for $\{f'_j,\, 11 \leq j \leq 14\}$ demonstrate that these frequencies are not harmonics of the form kf_0, which might suggest that these four are independent frequencies. Interestingly, however, we find that the latter belong to a set of fourteen "reflection frequencies" of the form $\{f'_j = 2f_N - (30-j)f_0,\, 1 \leq j \leq 14\}$, where $f_N = 24.46$ days^{-1} is the Nyquist frequency.

Among these fourteen frequencies, only the last six $\{f'_j,\, 9 \leq j \leq 14\}$ exhibit significant peaks in the Lomb-Scargle periodogram (computed according to Lomb, 1976). However, to avoid overfitting, we only consider the four frequencies $\{f'_j,\, 11 \leq j \leq 14\}$ corresponding to last four peaks of the estimated power spectrum (see the last row of Figure 2.5, bottom right panel). In summary, the only truly independent frequencies are f_0, f_m, and f_N, the other frequencies $\{f_k = kf_0,\, k = 1, \ldots, 15\}$ and $\{f'_j = 2f_N - (30-j)f_0,\, 1 \leq j \leq 14\}$ being linear combinations (or harmonics) of those.

The frequencies f'_j do not depend on the Blazkho frequency f_m, as the information about the Blazhko effect is captured by the time-varying trend $m(\cdot)$ and amplitudes $\{g_{\ell,k}(\cdot),\, \ell = 1, 2,\, k = 1, \ldots, K\}$, see equations (2.8)-(2.9) and Figure 2.6. We use a different notation

(f' rather than f) to avoid confusion, since in this case $f_{_N} < 16 f_0 = 25.77$ days^{-1}, and the four frequencies we are considering take value < 24 days^{-1}.

After fitting the models in equations (2.1) and (2.4), we compute their residuals and estimate their spectral densities using equation (2.26). To estimate the spectral densities according to the procedure in Section 2.5, we adopt the Gaussian kernel $K(y) = \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)$ with a bandwidth $h = 7.2$. We fitted both models with a PC having a 2.7 GHz 12-core Intel Xeon E5 processor and 64 GB of 1866 MHz DDR3 memory. Fitting our novel model in equation (2.1) required seventeen minutes and thirteen seconds, whereas fitting the model by Benkő (2018) in equation (2.4) required twelve minutes and twenty-eight seconds.

The description provided so far applies to both fits of models in equations (2.1) and (2.4). We now provide, separately, computational details about the estimation of these two models. Then in Sections 2.7.1 and 2.7.2 we compare and interpret the fits.

To fit our novel model in equation (2.1), we apply the methodology described in Section 2.4. When fitting our model in equation (2.1) we consider two sets of harmonic components. The first set is given by the harmonic components with frequencies $\{f_k = k f_0,\, k = 1, \ldots, 15\}$ provided by Benkő et al. (2014), weighted by our amplitudes $\{g_{\ell,k}(t_i),\, \ell = 1, 2,\, k = 1, \ldots, 15\}$. For the second set, the harmonic components are characterized by the four amplitudes $\{g'_{\ell,j}(t_i),\, \ell = 1, 2,\, j = 11, \ldots, 14\}$ weighting the corresponding four frequencies $\{f'_j,\, j = 11, \ldots, 14\}$. That is, we fit the following extended version

$$\begin{aligned} \mu(t_i) = m(t_i) \quad &+ \quad \sum_{k=1}^{15} \{g_{1,k}(t_i) \cos(w_k t_i) + g_{2,k}(t_i) \sin(w_k t_i)\} \\ &+ \quad \sum_{j=11}^{14} \{g'_{1,j}(t_i) \cos(w'_j t_i) + g'_{2,j}(t_i) \sin(w'_j t_i)\} \end{aligned}$$

of the model in equation (2.1), with $\omega_k = 2\pi f_k$ and $\omega'_j = 2\pi f'_j$. The resulting fitted model involves a total of $39J$ parameters. Before fitting our model, we select the smoothing parameters $\boldsymbol{\tau}$ and the number of B-splines J. These parameters are selected by the AIC criterion described in Section 2.4.2. To simplify the selection of the smoothing parameters

$\boldsymbol{\tau}$, we consider the case $\tau_2 = \cdots = \tau_{11}$, $\tau_{12} = \cdots = \tau_{21}$, $\tau_{22} = \cdots = \tau_{31}$, and $\tau_{32} = \cdots = \tau_{39}$. We pick the smoothing parameter τ_1 over the grid $\{0, 0.1, 10\}$, the parameters $\{\tau_k,\ k = 2, \ldots, 39\}$ over the grid $\{0, 0.1, 10, 100\}$, and the total number J of B-splines (of degree $d = 3$) over the grid $\{8, 13, 23, 33\}$. We apply the AIC formula in equation (2.16). The lowest AIC value occurs for $J = 33$ B-splines, $\tau_1 = 0$, $\tau_k = 0.1$, $k = 2, \ldots, 21$, $\tau_k = 0$, $k = 22, \ldots, 31$, and $\tau_k = 10$, $k = 32, \ldots, 39$.

To fit the model in equation (2.4), we implement the Levenberg–Marquardt algorithm using the `R` function `nls.lm` (available in the `R` package `minpack.lm` by Elzhov et al., 2016), with $K = 15$ and $\ell = \ell_k^A = \ell_k^F = 1$, $k = 1, \ldots, K$, for a total of 93 parameters.

2.7.1 Comparing the accuracy of the fits

The MSE corresponding to the fit of our model in equation (2.1) is 0.000001, whereas the MSE of the model in equation (2.4) is 0.000008. That is, the MSE of the model in equation (2.1) is about 12.5% of the MSE of the model in equation (2.4). Fitting the model in equation (2.1) involves 1287 parameters, whereas the number of parameters estimated with the model in equation (2.4) is 93. The larger number of parameters needed to fit the model in equation (2.1) is due to the semi-parametric form of the trend and amplitudes, which does not impose any particular shape to the underlying functions we estimate.

Figure 2.5 compares the fits of the models in equations (2.1) and (2.4). The first row shows the fitted curves, the second row shows the residuals, and the third and fourth rows show the estimated spectral density of the residuals. Although the fitted curves (first row) look very similar, the residuals are significantly different. Indeed, the residuals obtained with the model in equation (2.1) are compatible with the assumption of stationary and uncorrelated errors. By contrast, the residuals obtained with the model in equation (2.4) exhibit a time-dependent trend. Moreover, the estimated spectral densities in the last two rows of Figure 2.5 show that the model in equation (2.1) delivers residuals with a flat estimated spectral density, mimicking the behavior of the spectral density of white noise errors, whereas for the model in equation (2.4) shows that some harmonic components

should be added to the model (see the peaks between the frequencies 17 days^{-1} and 24 days^{-1} in the last row).

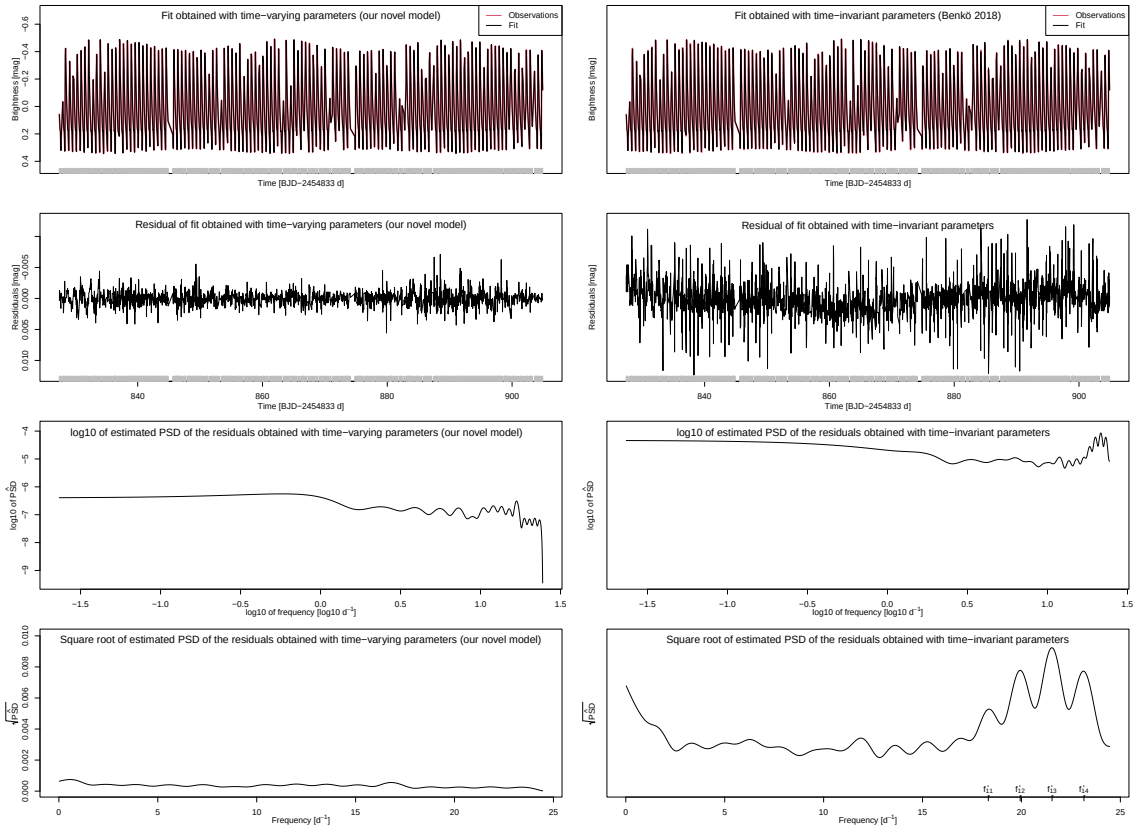

Figure 2.5 Comparison of fitted models to the light curve of V783 Cyg in Section 2.7.1. The first column corresponds to the fit of the model in equation (2.1), whereas the second column corresponds to the fit of the model in equation (2.4). From top to bottom: the first row shows the Brightness mag (red solid lines) together with the fits (black solid lines). The second row shows the residuals resulting from the fits. The last two rows show the spectral density of the residual obtained with equation (2.26) under different transformations (log-10 scale and square root).

2.7.2 Comparing the estimated time-varying parameters

In Section 2.3 we have shown that the model in equation (2.4) is a special case of our novel model in equation (2.1). To establish whether the fitted model in equation (2.1) matches (or differs from) the fitted model in equation (2.4), we now compare the estimates of $m(t)$ and $g_{\ell,k}(t)$ obtained by fitting the model in equation (2.1) with the estimates of $u(t)$ and $h_{\ell,k}(t)$ defined in equation (2.7) obtained by fitting the model in equation (2.4).

Figure 2.6 shows the estimates $\widehat{m}(t)$, $\{\widehat{g}_{\ell,k}(t),\ \ell = 1, 2,\ k = 1, \ldots, 15\}$, and $\{\widehat{g}'_{\ell,j}(t),\ \ell = 1, 2,\ j = 11, \ldots, 14\}$ (black lines), together with the estimates $\widehat{u}(t)$ and $\{\widehat{h}_{\ell,k}(t),\ \ell = 1, 2,\ k = 1, \ldots, 15\}$ (red lines). The estimated trend $\widehat{m}(t)$ is similar to the sinusoidal $\widehat{u}(t)$. Similarly, the first eight estimated harmonic components $\{\widehat{g}_{\ell,k}(t),\ \ell = 1, 2,\ k = 1, \ldots, 8\}$ and $\{\widehat{h}_{\ell,k}(t),\ \ell = 1, 2,\ k = 1, \ldots, 8\}$ are very close to each other. The next seven estimated harmonic components $\{\widehat{g}_{\ell,k}(t),\ \ell = 1, 2,\ k = 9, \ldots, 15\}$ and $\{\widehat{h}_{\ell,k}(t),\ \ell = 1, 2,\ k = 9, \ldots, 15\}$ are still similar but in some cases are slightly different. Nevertheless, these small differences do not have a significant impact on the fitted curves, because the last harmonic components have less of a contribution to the fit than the first ones. For the four estimated harmonic components associated with the frequencies $\{f'_j,\ j = 11, \ldots, 14\}$, which were fitted only for the model in equation (2.1) – and were not fitted for the model in equation (2.4) – we observe that the four corresponding time-varying amplitudes $\{\widehat{g}'_{\ell,j}(t),\ \ell = 1, 2,\ j = 11, \ldots, 14\}$ are allowed to have either a sinusoidal or a nonsinusoidal form. This finding is in accordance with the form of Amplitude Modulation and Frequency Modulation of Blazhko stars described by Benkő (2018). Finally, in Figure 2.6, we see that most of the confidence intervals of $\widehat{g}_{\ell,k}(t)$ contain $\widehat{h}_{\ell,k}(t)$. Therefore we conclude the following. Although the modulation frequency f_m is not a parameter of our model in equation (2.1), we are able to describe, through the estimated time-varying trend $\widehat{m}(t)$ and amplitudes $\widehat{g}_{\ell,k}(t)$, the Blazhko effect resulting from the amplitude and frequency modulation considered by the model in equation (2.4).

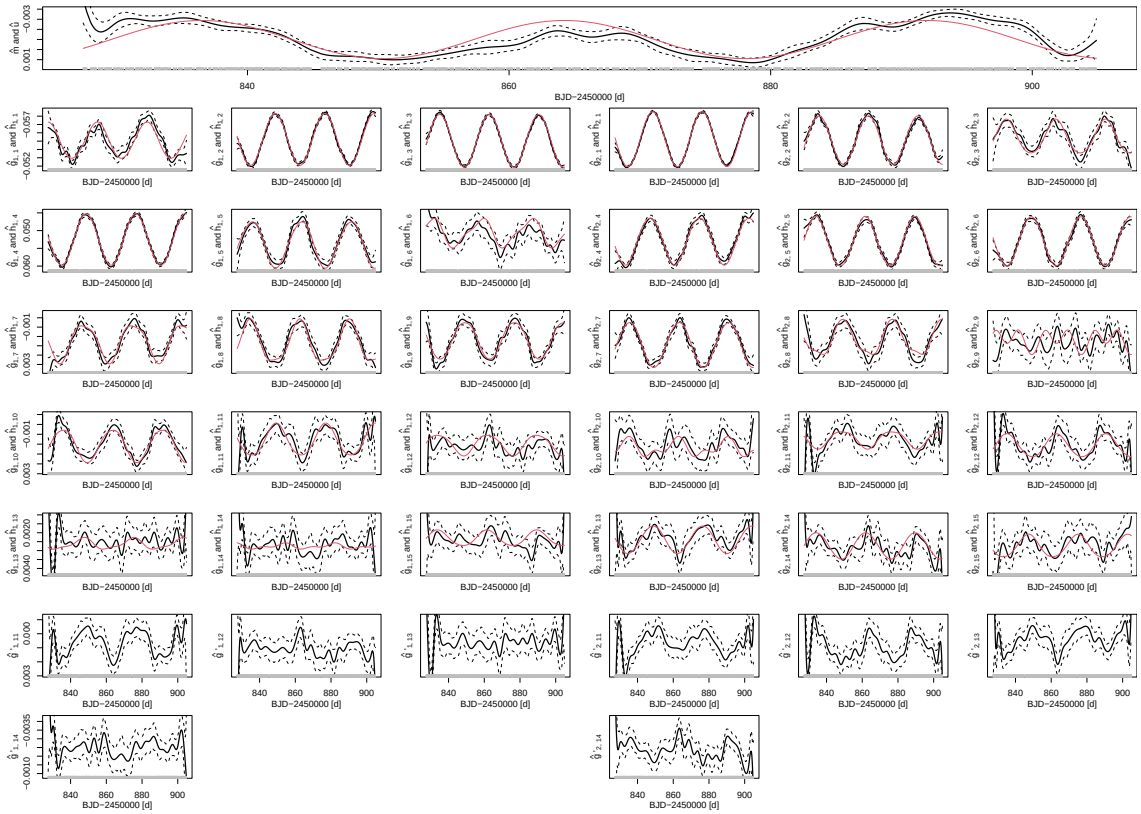

Figure 2.6 Comparing the estimated time-varying trend and amplitudes fitted to the light curve of V783 Cyg studied in Section 2.7.2. Red solid lines: estimates of $u(t)$ and $\{h_{\ell,k}(t),\ \ell = 1, 2,\ k = 1, \ldots, 15\}$ defined in equation (2.7) obtained with the model in equation (2.4). Black solid lines: estimates of $m(t)$, $\{g_{\ell,k}(t),\ \ell = 1, 2,\ k = 1, \ldots, 15\}$, and $\{g'_{\ell,j}(t),\ \ell = 1, 2,\ j = 11, \ldots, 14\}$ obtained with our novel model in equation (2.1). Black dashed lines: 95% confidence intervals for $m(t)$, $g_{\ell,k}(t)$ and $g'_{\ell,j}(t)$ obtained according to equations (2.14) and (2.15).

2.8 Conclusions

In this chapter, we introduced a model for time series observations of variable stars that are modulated by smoothly time-varying mean magnitudes, amplitudes, and phases. Previous approaches assume that the underlying parameters are either time-invariant or piecewise-constant functions. From the modeling viewpoint, our approach is more flexible because it avoids assumptions about the functional form of the mentioned time-dependent quantities.

From the computational viewpoint, estimating our time-varying curves translates into the estimation of time-invariant parameters that can be performed by ordinary least squares.

An important challenge when dealing with astronomical time series is that observations are unequally spaced in time. In some cases, observations are unevenly spaced due to missing values. Missing values are sometimes handled via imputation, that is, the gap generated by the missing value is "filled in" by an estimated value. Our novel approach, which involves the classical periodogram, has the advantage of not relying on any imputation method.

We study the performance of our approach under several simulation scenarios. Finally, we apply our method to V783 Cyg (KIC 5559631), a well-known RR Lyrae star presenting the Blazhko effect. In this case, the effect is characterized by a sinusoidal amplitude and frequency modulation. When comparing the time-varying fit obtained with our novel model with the time-invariant fit obtained with the model proposed by Benkő (2018), we found that both amplitude and frequency modulations are well captured and fitted by our novel model, and also that our time-varying method outperforms the time-invariant fit. Indeed the estimation error obtained with our fit is significantly smaller than the error obtained with the time-invariant fit. In addition, the residuals obtained with our novel method are compatible with the assumption of stationary and uncorrelated errors, whereas the residuals obtained with the time-invariant model by Benkő (2018) exhibit a time-dependent trend and some significant spectral peaks.

Chapter 3

Autocorrelated and nonstationary errors: a nonparametric approach

3.1 Introduction

In the previous chapter, we review the models proposed by Benkő et al. (2011) and Benkő (2018) to describe the brightness of Blazhko stars and we proposed a novel model based on time-varying parameters to describe these stars from a point of view more flexible. The models proposed by Benkő et al. (2011) and Benkő (2018) are based on modulations models (see Section 1.3.1 for more details). However, assuming that Blazhko stars can be described only through modulations models may not be valid. For instance, Guggenberger et al. (2012) described the light curve of the Blazhko star V445 Lyr by means of the combination of a sinusoidal amplitude modulated cascade and a nonsinusoidal frequency modulated model. V445 Lyr is described by Guggenberger et al. (2012) as a complex star in the sense that rapid and strong changes in the Blazhko modulation have been observed. After fitting the model proposed by Guggenberger et al. (2012) to the light curve V445 Lyr, one can see that the residuals show a pattern (see Figure 16 in Guggenberger et al. (2012)) and do not behave like a stationary process, that is, their mean and variance are not constant over time.

One way to enrich the model proposed in the previous chapter is by allowing the observations to be serially correlated. Serial correlation has been indeed observed in light curves of variable stars in previous studies. For example, Feigelson et al. (2018) fitted ARMA models to interpret astronomical light curves of variable stars KIC 005880320, KIC 004276716, HATS-2b, and RR Hyi, concluding that the variability seen in these stars can be effectively modeled by means of ARMA models. Kelly et al. (2014) obtaining good fits using CARMA models for the light curves of a long period variable star and a RR Lyrae star. Although ARMA and CARMA models help describing, respectively, discrete and continuous stationary time series, they fail at capturing the nonstationarity of the light curves.

Starting from the model proposed in Chapter 2, in this chapter we extend the previous results by taking into account two important facts that characterize the error term of astronomical time series: autocorrelation and nonstationarity. In this chapter we allow the errors to follow an autoregressive (AR) process, and extend the model introduced in Chapter 2 into two directions: we first deal with stationary AR errors, and then with locally stationary AR errors. A possible future research line is represented by a model with locally stationary ARMA (autoregressive and moving average) errors.

We divide the chapter into four main sections. Section 3.2 defines the model and illustrates how the serial correlation is introduced. Section 3.3 presents the method we adopt to estimate the time-varying parameters, and Section 3.4 provides simulation results under different scenarios. Finally, Section 3.5 summarizes the main results of this chapter.

3.2 Model definition

Let $\{y_i \equiv y_{t_i},\, i = 1, \ldots, N\}$ be a set of observations occurring at certain discrete times $t_1, \ldots, t_N$, with $t_i = t_0 + i\Delta$ where i is an integer, $\Delta > 0$ is the constant data spacing, and $t_0 = 0$. We decompose the observation y_i into the sum of a deterministic part $\mu(t_i)$ and a random part ε_i. The deterministic part $\mu(t_i)$ consists of a trend $m(t_i)$ and a seasonal part. The seasonal part is a linear combination of K cosines and sines weighted by amplitudes

$g_{\ell,k}(t_i)$ as

$$\begin{aligned} y_i &= \mu(t_i) + \varepsilon_i, \qquad i = 1, \ldots, N, \\ \mu(t_i) &= m(t_i) + \sum_{k=1}^{K} \{g_{1,k}(t_i)\cos(w_k t_i) + g_{2,k}(t_i)\sin(w_k t_i)\}, \end{aligned} \tag{3.1}$$

or in matrix notation

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y} = (y_1, \ldots, y_N)^\top$ is the vector of observations at time $\boldsymbol{t} = (t_1, \ldots, t_N)^\top$, $\boldsymbol{\mu} = [\mu(t_1), \ldots, \mu(t_N)]^\top$ is the expectation of $\boldsymbol{y}$, $m(t_i)$ is the smooth trend at time t_i, $g_{1,k}(t_i)$ and $g_{2,k}(t_i)$ are smooth time-varying amplitudes of the cosine and sine waves at time t_i, respectively, $w_k = 2\pi f_k$ is the angular frequency, and f_k is the ordinary frequency.

In this chapter, an important role is played by the inverse $\boldsymbol{\Gamma}_\varepsilon^{-1}$ of the matrix $\boldsymbol{\Gamma}_\varepsilon = \mathbb{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top\right]$, the covariance matrix of the errors $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)^\top$. Both $\boldsymbol{\Gamma}_\varepsilon$ and $\boldsymbol{\Gamma}_\varepsilon^{-1}$ depend on the underlying model we assume for the vector $\boldsymbol{\varepsilon}$. These two matrices depend on the order p of the underlying process (see Appendices D and E).

The error vector $\boldsymbol{\varepsilon}$ can follow two processes, an autoregressive process or a locally stationary autoregressive process. An autoregressive process of order p (or AR(p), see Box and Jenkins, 1970) is defined by

$$\varepsilon_i = \sum_{j=1}^{p} \phi_j \varepsilon_{i-j} + z_i, \qquad \{z_i\} \sim WN(0, \sigma_z^2), \tag{3.2}$$

and a locally stationary autoregressive process of order p (or LSAR(p), see Dahlhaus, 1996) is defined by

$$\varepsilon_{i,N} = \sum_{j=1}^{p} \phi_j\left(\tfrac{i}{N}\right) \varepsilon_{i-j,N} + z_i, \quad \{z_i\} \sim WN(0, \sigma_z^2), \tag{3.3}$$

where z_i is a white noise process with zero mean and variance σ_z^2, and the coefficients $\phi_j\left(\frac{i}{N}\right)$, $j = 1, \ldots, p$, depend on the location t_i. In the case of LSAR errors, time t_i in the model in equation (3.1) is rescaled to i/N and the frequency w_k is replaced by Nw_k (see

equation (1.18)). Notice that when $p = 0$ both AR(p) and LSAR(p) reduce to the same model, and the estimation of time-varying trend and amplitudes in equation (3.1) can be performed by (penalized) least squares according to the estimator given in Section 2.4 by equation (2.10).

In the next section, we propose a novel penalized version of the method presented in Brockwell and Davis (2016) to estimate the trend, $m(\cdot)$, and amplitudes, $g_{\ell,k}(\cdot)$, in the model in equation (3.1), the autoregressive coefficients, ϕ_j or $\phi_j(\cdot)$, and variance, σ_z^2, of the errors defined in equations (3.2) or (3.3).

3.3 Estimation

In Section 3.3.1 we define estimators of the unknown trend $m(\cdot)$ and amplitudes $\{g_{\ell,k}(\cdot),\ \ell = 1, 2,\ k = 1, \ldots, K\}$, of the model in equation (3.1) in presence of autocorrelation assuming that the inverse of the covariance matrix of the error term $\mathbf{\Gamma}_\varepsilon^{-1}$ is known. In Section 3.3.2 we address the estimation of trend and amplitudes when $\mathbf{\Gamma}_\varepsilon^{-1}$ is unknown. Finally, in Section 3.3.3 we explain how to select the tuning parameters associated with the B-splines, the penalization used in the estimation method, and the order p of the autoregressive process associated with the errors. We denote by N the sample size, $T = t_N - t_1$ the time span, J the number of B-splines that form the basis associated with the time-varying trend and amplitudes, H the number of B-splines that form the basis associated with the time-varying autoregressive coefficients (in the case of LSAR(p) errors), d the degree of the B-splines, K the number of harmonics components, r the order of the penalty, and M the number of replications in Monte Carlo simulations.

We performed our calculations using the R Language for Statistical Computing (R Core Team, 2021). Our codes combine existing functions (available as part of `R` packages) with our own development.

3.3.1 Penalized weighted least squares

As in the previous Section 2.4, we use B-splines to estimate the trend and amplitudes of the model given by equation (3.1) (see Appendix A for the definition of B-splines). The smooth trend and amplitude functions, $m(t_i)$ and $g_{\ell,k}(t_i)$, $\ell = 1, 2$, are modeled as a linear combination of B-splines basis as

$$\begin{aligned} m(t_i) &= \sum_{j=1}^{J} \alpha_j B_j(t_i), \\ g_{1,k}(t_i) &= \sum_{j=1}^{J} \beta_{k,j} B_j(t_i), \qquad k = 1, \ldots, K, \\ g_{2,k}(t_i) &= \sum_{j=1}^{J} \gamma_{k,j} B_j(t_i), \qquad k = 1, \ldots, K. \end{aligned}$$

We define the $N \times N$ matrices $\mathbf{C}_k$ and $\mathbf{S}_k$, $k = 1, \ldots, K$, as

$\mathbf{C}_k = \text{diag}\{\cos(w_k t_1), \cos(w_k t_2), \ldots, \cos(w_k t_N)\}$ and $\mathbf{S}_k = \text{diag}\{\sin(w_k t_1), \sin(w_k t_2), \ldots, \sin(w_k t_N)\}$.

The model for the expected value of $\boldsymbol{y}$ is

$$\boldsymbol{\mu} = \mathcal{B}\boldsymbol{\theta},$$

where $\mathcal{B}$ is the $N \times c$ design matrix given by

$$\mathcal{B} = [\mathbf{B}|\mathbf{C}_1\mathbf{B}| \ldots |\mathbf{C}_K\mathbf{B}|\mathbf{S}_1\mathbf{B}| \ldots |\mathbf{S}_K\mathbf{B}],$$

with $c = J(2K + 1)$, and

$$\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_K^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_K^\top)^\top, \tag{3.4}$$

with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)^\top$, $\boldsymbol{\beta}_k = (\beta_{k,1}, \ldots, \beta_{k,J})^\top$, $\boldsymbol{\gamma}_k = (\gamma_{k,1}, \ldots, \gamma_{k,J})^\top$, $k = 1, \ldots, K$, is the vector of regression coefficients of length c.

The *generalized least squares* (GLS) estimator of $\boldsymbol{\theta}$ is the value $\widehat{\boldsymbol{\theta}}_{\text{GLS}}$ which minimizes the sum of squares

$$M_{\boldsymbol{\theta}} = ||\mathbf{V}\boldsymbol{y} - \mathbf{V}\mathcal{B}\boldsymbol{\theta}||^2, \tag{3.5}$$

where the $N \times N$ matrix $\mathbf{V}$ satisfies $\mathbf{VV} = \sigma_z^2 \mathbf{\Gamma}_\varepsilon^{-1}$, $\sigma_z^2 = \mathbb{V}\mathrm{ar}[z_i]$, and $\mathbf{\Gamma}_\varepsilon = \mathbb{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top\right]$ are the variance and the covariance matrix of z_i and $\boldsymbol{\varepsilon}$, respectively. The matrix $\mathbf{V}$ is not uniquely determined. Using the spectral decomposition

$$\mathbf{VV} = \mathbf{R}\mathrm{diag}\{\lambda_1, \ldots, \lambda_N\}\mathbf{R}^\top$$

of $\mathbf{VV}$, where $\mathbf{R}$ is the orthogonal matrix whose columns are the orthonormal eigenvectors corresponding to the nonnegative eigenvalues $\lambda_1, \ldots, \lambda_N$ of $\mathbf{VV}$, the matrix $\mathbf{V}$ can be obtained as

$$\mathbf{V} = \mathbf{R}\mathrm{diag}\{\lambda_1^{1/2}, \ldots, \lambda_N^{1/2}\}\mathbf{R}^\top. \tag{3.6}$$

Equating to zero the partial derivatives with respect to each component of $\boldsymbol{\theta}$ in equation (3.5), the estimator of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}}_{\text{GLS}} = (\mathcal{B}^\top \mathbf{\Gamma}_\varepsilon^{-1} \mathcal{B})^{-1} \mathcal{B}^\top \mathbf{\Gamma}_\varepsilon^{-1} \boldsymbol{y}. \tag{3.7}$$

Using the penalty on the finite differences of the coefficients proposed by Eilers and Marx (1996) and described in Section 2.4, the estimator of $\boldsymbol{\theta}$ is calculated by minimizing the following expression

$$M_{\boldsymbol{\theta}}^* = ||\mathbf{V}\boldsymbol{y} - \mathbf{V}\mathcal{B}\boldsymbol{\theta}||^2 + \tau_1 ||\mathbf{D}_r \boldsymbol{\alpha}||^2 + \sum_{k=1}^{K} \left\{ \tau_{2k} ||\mathbf{D}_r \boldsymbol{\beta}_k||^2 + \tau_{2k+1} ||\mathbf{D}_r \boldsymbol{\gamma}_k||^2 \right\},$$

where $\{\tau_k,\ k = 1, \ldots, 2K+1\}$ are positive regularization parameters that control the smoothness of the curve, penalizing the coefficients that are far apart from one another, and the matrix $\mathbf{D}_r$ constructs rth order differences of a vector $\boldsymbol{\eta}$ as

$$\mathbf{D}_r \boldsymbol{\eta} = \Delta^r \boldsymbol{\eta}. \tag{3.8}$$

The penalties can be represented as $\boldsymbol{\theta}^\top \mathbf{P} \boldsymbol{\theta}$ with the block-diagonal matrix $\mathbf{P} = \mathbf{T} \otimes \mathbf{D}_r^\top \mathbf{D}_r$ and $\mathbf{T} = \mathrm{diag}\{\tau_1, \tau_2, \tau_3, \ldots, \tau_{2K+1}\}$. Then, minimizing

$$M_{\boldsymbol{\theta}}^* = ||\mathbf{V}\boldsymbol{y} - \mathbf{V}\mathcal{B}\boldsymbol{\theta}||^2 + \boldsymbol{\theta}^\top \mathbf{P} \boldsymbol{\theta},$$

with respect to $\boldsymbol{\theta}$, the *penalized generalized least squares estimator* (PGLS) of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}}_{\text{PGLS}} = \sigma_z^2(\sigma_z^2\mathcal{B}^\top\boldsymbol{\Gamma}_\varepsilon^{-1}\mathcal{B} + \mathbf{P})^{-1}\mathcal{B}^\top\boldsymbol{\Gamma}_\varepsilon^{-1}\boldsymbol{y}. \tag{3.9}$$

Regarding the prediction, in the case of AR errors, the prediction of y at time t_i is given by

$$\widehat{y}_i = \widehat{\mu}(t_i) = \boldsymbol{\mathcal{B}}(t_i)^\top\widehat{\boldsymbol{\theta}}_{\text{PGLS}}, \tag{3.10}$$

and in the case of LSAR errors is

$$\widehat{y}_i = \widehat{\mu}\left(\tfrac{i}{N}\right) = \boldsymbol{\mathcal{B}}\left(\tfrac{i}{N}\right)^\top\widehat{\boldsymbol{\theta}}_{\text{PGLS}}, \tag{3.11}$$

where $\boldsymbol{\mathcal{B}}(t_i)$ and $\boldsymbol{\mathcal{B}}\left(\frac{i}{N}\right)$ are the ith row of $\mathcal{B}$, the residuals are $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$, with $\widehat{\boldsymbol{y}} = (\widehat{y}_1, \ldots, \widehat{y}_N)^\top$, and the mean square error (MSE) is $\text{MSE} = N^{-1}\sum_{i=1}^N(y_i - \widehat{y}_i)^2$. The estimators of the trend $\boldsymbol{m}$ and amplitudes $\{\boldsymbol{g}_{\ell,k},\ \ell = 1, 2,\ k = 1, \ldots, K\}$, are

$$\widehat{\boldsymbol{m}} = \mathbf{B}\widehat{\boldsymbol{\alpha}}, \quad \widehat{\boldsymbol{g}}_{1,k} = \mathbf{B}\widehat{\boldsymbol{\beta}}_k, \quad \widehat{\boldsymbol{g}}_{2,k} = \mathbf{B}\widehat{\boldsymbol{\gamma}}_k. \tag{3.12}$$

Another parameter of interest is the variance σ_z^2. In the case of AR(p) errors, the estimator is given by

$$\widehat{\sigma}_z^2 = (N-p)^{-1}\sum_{i=p+1}^N\left\{y_i - \boldsymbol{\mathcal{B}}(t_i)^\top\widehat{\boldsymbol{\theta}}_{\text{PGLS}} - \sum_{j=1}^p\phi_j\left[y_{i-j} - \boldsymbol{\mathcal{B}}(t_{i-j})^\top\widehat{\boldsymbol{\theta}}_{\text{PGLS}}\right]\right\}^2,$$

and in the case of LSAR(p) errors is

$$\widehat{\sigma}_z^2 = (N-p)^{-1}\sum_{i=p+1}^N\left\{y_i - \boldsymbol{\mathcal{B}}(\tfrac{i}{N})^\top\widehat{\boldsymbol{\theta}}_{\text{PGLS}} - \sum_{j=1}^p\phi_j\left(\tfrac{i}{N}\right)\left[y_{i-j} - \boldsymbol{\mathcal{B}}(\tfrac{i-j}{N})^\top\widehat{\boldsymbol{\theta}}_{\text{PGLS}}\right]\right\}^2. \tag{3.13}$$

In addition to the point estimate, interval estimation for $\widehat{y}_i$ is often of interest and is easy to construct. Assuming that the error terms $\{\varepsilon_i,\ i = 1, \ldots, N\}$ follow an AR or LSAR process and the errors associated with $\{\varepsilon_i,\ i = 1, \ldots, N\}$, $\{z_i,\ i = 1, \ldots, N\}$, follow a Gaussian distribution with zero mean and variance σ_z^2, the $(1-\alpha)\times 100\%$ prediction interval for $\mu(t_i)$, with $i = 1, \ldots, N$, is

$$\boldsymbol{\mathcal{B}}^\top(t_i)\widehat{\boldsymbol{\theta}}_{\text{PGLS}} \pm z(1-\alpha/2)\sqrt{\boldsymbol{\mathcal{B}}^\top(t_i)\mathbb{V}\text{ar}\left[\widehat{\boldsymbol{\theta}}_{\text{PGLS}}\right]\boldsymbol{\mathcal{B}}(t_i)}, \tag{3.14}$$

where $z(1-\alpha/2)$ denotes the $(1-\alpha/2)$ quantile of the standard Gaussian distribution, and

$$\mathbb{V}\mathrm{ar}\left[\widehat{\boldsymbol{\theta}}_{\text{PGLS}}\right] = \sigma_z^4(\sigma_z^2\mathcal{B}^\top\boldsymbol{\Gamma}_\varepsilon^{-1}\mathcal{B}+\mathbf{P})^{-1}\mathcal{B}^\top\boldsymbol{\Gamma}_\varepsilon^{-1}\mathcal{B}(\sigma_z^2\mathcal{B}^\top\boldsymbol{\Gamma}_\varepsilon^{-1}\mathcal{B}+\mathbf{P})^{-1}.$$

The $(1-\alpha)\times 100\%$ confidence interval for the trend $m(t_i)$ is

$$\mathscr{B}(t_i)^\top\widehat{\mathbf{Q}}_m\widehat{\boldsymbol{\theta}}_{\text{PGLS}} \pm z(1-\alpha/2)\sqrt{\mathscr{B}(t_i)^\top\widehat{\mathbf{Q}}_m\mathbb{V}\mathrm{ar}\left[\widehat{\boldsymbol{\theta}}_{\text{PGLS}}\right]\widehat{\mathbf{Q}}_m^\top\mathscr{B}(t_i)}, \tag{3.15}$$

and the $(1-\alpha)\times 100\%$ confidence intervals for the amplitudes $g_{\ell,k}(t_i)$, $\ell=1,2$, $k=1,\ldots,K$, are

$$\mathscr{B}(t_i)^\top\widehat{\mathbf{Q}}_g(\ell,k)\widehat{\boldsymbol{\theta}}_{\text{PGLS}} \pm z(1-\alpha/2)\sqrt{\mathscr{B}(t_i)^\top\widehat{\mathbf{Q}}_g(\ell,k)\mathbb{V}\mathrm{ar}\left[\widehat{\boldsymbol{\theta}}_{\text{PGLS}}\right]\widehat{\mathbf{Q}}_g(\ell,k)^\top\mathscr{B}(t_i)}, \tag{3.16}$$

where $\mathscr{B}(t_i)$ is the ith row of the matrix $\mathscr{B}$, and $\mathscr{B}=[\mathbf{B}|\mathbf{B}|\ldots|\mathbf{B}]$ is a matrix of dimension $N\times c$. The $c\times c$ matrices $\widehat{\mathbf{Q}}_m$, $\{\widehat{\mathbf{Q}}_g(\ell,k),\ \ell=1,2,\ k=1,\ldots,K\}$ satisfy $\widehat{\mathbf{Q}}_m\widehat{\boldsymbol{\theta}}_{\text{PGLS}}=(\widehat{\boldsymbol{\alpha}}^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top$ and

$$\begin{aligned}
&\widehat{\mathbf{Q}}_g(1,1)\widehat{\boldsymbol{\theta}}_{\text{PGLS}}=(\mathbf{0}_J^\top,\widehat{\boldsymbol{\beta}}_1^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top, && \widehat{\mathbf{Q}}_g(2,1)\widehat{\boldsymbol{\theta}}_{\text{PGLS}}=(\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top,\widehat{\boldsymbol{\gamma}}_1^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top,\\
&\widehat{\mathbf{Q}}_g(1,2)\widehat{\boldsymbol{\theta}}_{\text{PGLS}}=(\mathbf{0}_J^\top,\mathbf{0}_J^\top,\widehat{\boldsymbol{\beta}}_2^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top, && \widehat{\mathbf{Q}}_g(2,2)\widehat{\boldsymbol{\theta}}_{\text{PGLS}}=(\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top,\mathbf{0}_J^\top,\widehat{\boldsymbol{\gamma}}_2^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top,\\
&\qquad\vdots && \qquad\vdots\\
&\widehat{\mathbf{Q}}_g(1,K)\widehat{\boldsymbol{\theta}}_{\text{PGLS}}=(\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top,\widehat{\boldsymbol{\beta}}_K^\top,\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top)^\top, && \widehat{\mathbf{Q}}_g(2,K)\widehat{\boldsymbol{\theta}}_{\text{PGLS}}=(\mathbf{0}_J^\top,\ldots,\mathbf{0}_J^\top,\widehat{\boldsymbol{\gamma}}_K^\top)^\top.
\end{aligned}$$

Equations (4.14), (4.15), and (4.16) are useful to obtain the confidence intervals when only one time series is available. The execution time needed to compute these confidence intervals depends on the number of observations – for example, with $N=2000$ the algorithm runs in about 3 minutes.

3.3.2 Penalized GLS estimation with autocorrelated errors

The GLS and PGLS estimators in equations (3.7) and (3.9) depend on the unknown matrix $\boldsymbol{\Gamma}_\varepsilon^{-1}$ which in turn depends on the parameters $\{\phi_j,\ j=1,\ldots,p\}$ or $\{\phi_j(\cdot),\ j=1,\ldots,p\}$,

and σ_z^2. In Appendices D and E, we provide the definition of $\boldsymbol{\Gamma}_{\varepsilon}^{-1}$ for AR(p) and LSAR(p) processes, respectively.

To deal with the case where the errors follow an AR(p) process, Cochrane and Orcutt (1949) proposed an iterative technique to estimate the parameter $\boldsymbol{\theta}$ in equation (3.4) transforming the model in equation (3.1) into a model with uncorrelated, zero mean, and constant-variance errors, so that the OLS can be used to compute estimates of $\boldsymbol{\theta}$ in terms of $y_i^* = \phi(B)y_i$, $i = p+1, \ldots, N$. The main advantage of this approach is that eliminates the need to compute $\mathbf{V}$ in equation (3.6). However, one limitation is that $\boldsymbol{y}^* = (y_{p+1}^*, \ldots, y_N^*)^\top$ does not contain all the information in $\boldsymbol{y}$. Brockwell and Davis (2016, page 184) give an extension of this iterative scheme, proposed by Cochrane and Orcutt (1949). It is based on the observation that for a fixed $\boldsymbol{\phi}$, the value of $\boldsymbol{\theta}$ that minimizes the sum of squares in equation (3.5) is $\widehat{\boldsymbol{\theta}}_{\text{GLS}}$ in equation (3.7). The regression model considered in Brockwell and Davis (2016, page 184) does not include a penalization of the parameters $\boldsymbol{\theta}$. In this section, we extend the approach in Brockwell and Davis (2016, page 184) for estimating $\boldsymbol{\theta}$ in equation (3.4) in two different directions. Firstly, we extend the iterative scheme in Brockwell and Davis (2016, page 184) for AR(p) errors, considering $\widehat{\boldsymbol{\theta}}_{\text{PGLS}}$ in equation (3.9) instead of $\widehat{\boldsymbol{\theta}}_{\text{GLS}}$ in equation (3.7), and secondly, we adapted the iterative scheme in Brockwell and Davis (2016, page 184) to the case of LSAR(p) errors.

Penalized estimator when the errors follow an AR(p) process

The Cochrane-Orcutt algorithm is commonly used to estimate regression models with AR(p) autocorrelated errors (Cochrane and Orcutt, 1949). In order to deal with AR(p) errors and a penalization of the parameter $\boldsymbol{\theta}$ in equation (3.4), in this section we introduce a modified version of the iterative scheme in Brockwell and Davis (2016, page 184), replacing $\widehat{\boldsymbol{\theta}}_{\text{GLS}}$ in equation (3.7) with $\widehat{\boldsymbol{\theta}}_{\text{PGLS}}$ in equation (3.9). The scheme is as follows.

(i) Compute the ordinary least square estimator $\widehat{\boldsymbol{\theta}}_{\text{OLS}} = (\mathcal{B}^\top\mathcal{B})^{-1}\mathcal{B}^\top\boldsymbol{y}$ and the corresponding residuals $\widehat{\boldsymbol{\varepsilon}}^{(0)} = \boldsymbol{y} - \mathcal{B}\widehat{\boldsymbol{\theta}}_{\text{OLS}}$.

(ii) Fit an AR(p) model by maximum Gaussian likelihood to the estimated residuals $\widehat{\boldsymbol{\varepsilon}}^{(0)}$,

and obtain $\widehat{\boldsymbol{\phi}}^{(0)}$ and $\widehat{\sigma}_z^{(0)}$.

(iii) For the fitted AR model compute the corresponding estimator $\widehat{\boldsymbol{\theta}}^{(0)}_{\text{PGLS}}$ in equation (3.9).

(iv) Compute the residuals $\widehat{\boldsymbol{\varepsilon}}^{(1)} = \boldsymbol{y} - \mathcal{B}\widehat{\boldsymbol{\theta}}^{(0)}_{\text{PGLS}}$

(v) Iterate (ii)-(iv) until the estimators have stabilized. Our estimator will be $\widehat{\boldsymbol{\theta}}^{(r)}_{\text{PGLS}}$ if $\|\widehat{\boldsymbol{\phi}}^{(r+1)} - \widehat{\boldsymbol{\phi}}^{(r)}\| < \epsilon$, for a given small positive ϵ.

Penalized estimator when the errors follow a LSAR(p) process

We now adapt the previous estimation procedure of $\widehat{\boldsymbol{\theta}}_{\text{PGLS}}$ to the case where the errors follow a LSAR(p) process. The algorithm is as follows.

(i) Compute the ordinary least square estimator $\widehat{\boldsymbol{\theta}}_{\text{OLS}} = (\mathcal{B}^\top\mathcal{B})^{-1}\mathcal{B}^\top\boldsymbol{y}$ and the corresponding residuals $\widehat{\boldsymbol{\varepsilon}}^{(0)}_N = \boldsymbol{y} - \mathcal{B}\widehat{\boldsymbol{\theta}}_{\text{OLS}}$.

(ii) Fit a LSAR(p) model to the estimated residuals $\widehat{\boldsymbol{\varepsilon}}^{(0)}_N$ as

$$\widehat{\varepsilon}^{(0)}_{i,N} = \sum_{j=1}^{p} \phi_j\left(\tfrac{i}{N}\right)\widehat{\varepsilon}^{(0)}_{i-j,N} + z_i, \quad \{z_i\} \sim WN(0, \sigma_z^2),$$

through the following procedure. Assume that the LSAR(p) coefficients, $\boldsymbol{\phi}(u) = \left[\phi_1(u), \ldots, \phi_p(u)\right]^\top$, vary smoothly over time and they can be modeled as a linear combination of B-splines basis:

$$\phi_k(u) = \sum_{j=1}^{H} \ell_{k,j} B_j(u), \quad k = 1, \ldots, p. \tag{3.17}$$

We are looking for the estimator $\widehat{\boldsymbol{\ell}}^\top = \left(\widehat{\boldsymbol{\ell}}_1^\top, \ldots, \widehat{\boldsymbol{\ell}}_p^\top\right)$, with $\widehat{\boldsymbol{\ell}}_k^\top = \left(\widehat{\ell}_{k,1}, \ldots, \widehat{\ell}_{k,H}\right)$, $k = 1, \ldots, p$, of $\boldsymbol{\ell}^\top = \left(\boldsymbol{\ell}_1^\top, \ldots, \boldsymbol{\ell}_p^\top\right)$, with $\boldsymbol{\ell}_k^\top = \left(\ell_{k,1}, \ldots, \ell_{k,H}\right)$, $k = 1, \ldots, p$, which minimizes the sum of squares

$$||\widehat{\boldsymbol{\varepsilon}}^{(0)}_N - \mathbf{E}\boldsymbol{\ell}||^2 + \sum_{j=1}^{p} \tau_j^{\phi} ||\mathbf{D}_r \boldsymbol{\ell}_j||^2,$$

where $\widehat{\boldsymbol{\varepsilon}}_{N}^{(0)} = \left(\widehat{\varepsilon}_{p+1,N}^{(0)}, \ldots, \widehat{\varepsilon}_{N,N}^{(0)}\right)^{\top}$, $\left\{\tau_{k}^{\phi},\ k = 1, \ldots, p\right\}$ are the smoothing parameter associated to the time-varying coefficients $\phi_{j}(u),\ j = 1, \ldots, p$, $\mathbf{D}_{r}$ is defined in equation (3.8), r is the order of the penalty, and $\mathbf{E}$ is the design matrix of dimension $(N-p) \times Hp$ given by

$$\mathbf{E} = \begin{pmatrix} B_1\left(\frac{p+1}{T}\right)\widehat{\varepsilon}_{p,N}^{(0)} & \cdots & B_H\left(\frac{p+1}{N}\right)\widehat{\varepsilon}_{p,N}^{(0)} & \cdots & B_1\left(\frac{p+1}{N}\right)\widehat{\varepsilon}_{1,N}^{(0)} & \cdots & B_H\left(\frac{p+1}{N}\right)\widehat{\varepsilon}_{1,N}^{(0)} \\ B_1\left(\frac{p+2}{N}\right)\widehat{\varepsilon}_{p+1,N}^{(0)} & \cdots & B_H\left(\frac{p+2}{N}\right)\widehat{\varepsilon}_{p+1,N}^{(0)} & \cdots & B_1\left(\frac{p+2}{T}\right)\widehat{\varepsilon}_{2,N}^{(0)} & \cdots & B_H\left(\frac{p+2}{N}\right)\widehat{\varepsilon}_{2,N}^{(0)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ B_1\left(\frac{N}{N}\right)\widehat{\varepsilon}_{N-1,N}^{(0)} & \cdots & B_H\left(\frac{N}{N}\right)\widehat{\varepsilon}_{N-1,N}^{(0)} & \cdots & B_1\left(\frac{N}{N}\right)\widehat{\varepsilon}_{N-p,N}^{(0)} & \cdots & B_H\left(\frac{N}{N}\right)\widehat{\varepsilon}_{N-p,N}^{(0)} \end{pmatrix}. \tag{3.18}$$

Equating to zero the partial derivatives with respect to each component of $\boldsymbol{\ell}$, we obtain the penalized estimator of $\boldsymbol{\ell}$

$$\widehat{\boldsymbol{\ell}}^{(0)} = (\mathbf{E}^{\top}\mathbf{E} + \mathbf{P}_{\phi})^{-1}\mathbf{E}^{\top}\widehat{\boldsymbol{\varepsilon}}_{N}^{(0)}, \tag{3.19}$$

where

$$\mathbf{P}_{\phi} = \mathbf{T} \otimes \mathbf{D}_{r}^{\top}\mathbf{D}_{r} \tag{3.20}$$

and $\mathbf{T} = \text{diag}\{\tau_{1}^{\phi}, \ldots, \tau_{p}^{\phi}\}$ is a $p \times p$ diagonal matrix.

We plug the estimator $\widehat{\boldsymbol{\ell}}^{(0)}$ in equation (3.19) into $\boldsymbol{\ell}$ of equation (3.17) and obtain the initial estimated LSAR(p) coefficients, $\widehat{\boldsymbol{\phi}}_{j}^{(0)}(u)$, $j = 1, \ldots, p$, that is,

$$\widehat{\phi}_{k}^{(0)}(u) = \sum_{j=1}^{H} \widehat{\ell}_{k,j}^{(0)} B_{j}(u), \quad k = 1, \ldots, p. \tag{3.21}$$

Using the estimated coefficients $\widehat{\boldsymbol{\phi}}_{j}^{(0)}(u)$, $j = 1, \ldots, p$, calculate the estimator of the variance given in equation (3.13) and denote it by $\widehat{\sigma}_{z}^{2(0)}$.

(iii) Using the estimated LSAR(p) coefficients $\left\{\widehat{\boldsymbol{\phi}}_{j}^{(0)}(u),\ j = 1, \ldots, p\right\}$ and the estimated variance $\widehat{\sigma}_{z}^{2(0)}$ obtained in the previous step, we construct the inverse covariance matrix of the errors $\boldsymbol{\Gamma}_{\varepsilon}^{-1}$ according with Appendix E. If we denote by $\widehat{\mathbf{I}}_{\varepsilon}^{(0)} = \widehat{\boldsymbol{\Gamma}}_{\varepsilon}^{-1}$ this estimated inverse covariance matrix, the updated value of $\boldsymbol{\theta}$ in equation (3.9) is

$$\widehat{\boldsymbol{\theta}}_{\text{PGLS}}^{(0)} = \widehat{\sigma}_{z}^{2(0)}(\widehat{\sigma}_{z}^{2(0)}\mathcal{B}^{\top}\widehat{\mathbf{I}}_{\varepsilon}^{(0)}\mathcal{B} + \mathbf{P})^{-1}\mathcal{B}^{\top}\widehat{\mathbf{I}}_{\varepsilon}^{(0)}\boldsymbol{y}. \tag{3.22}$$

(iv) Compute the residuals $\widehat{\boldsymbol{\varepsilon}}_N^{(1)} = \boldsymbol{y} - \mathcal{B}\widehat{\boldsymbol{\theta}}_{\text{PGLS}}^{(0)}$.

(v) Iterate (ii)-(iv) until the estimators have stabilized. Our estimator will be $\widehat{\boldsymbol{\theta}}_{\text{PGLS}}^{(r)}$ if $\sup_{u\in[0,1]} \|\widehat{\boldsymbol{\phi}}^{(r+1)}(u) - \widehat{\boldsymbol{\phi}}^{(r)}(u)\| < \epsilon$, for a given small positive ϵ.

3.3.3 Automatic selection of the tunable parameters

Before calculating the estimator in equation (3.9), it is necessary to select the tuning parameters $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_{2K+1})^\top$ associated to the trend and amplitudes in the model in equation (3.1) and $\boldsymbol{\tau}^\phi = (\tau_1^\phi, \ldots, \tau_p^\phi)^\top$ associated to the coefficients in the case of LSAR(p) errors in equation (3.3). To choose the tuning parameters, we propose to use the Akaike information criterion (AIC). The AIC penalizes the log-likelihood of a fitted model by considering the effective number of parameters (see Hastie et al., 2004).

For the model in equation (3.1) with AR(p) or LSAR(p) errors, we adapt the AIC tuning parameters selector of Hastie et al. (2004). In the case of fitting the model in equation (3.1) with AR(p) errors, we propose to use the AIC given by

$$\text{AIC}(\boldsymbol{\tau}) = \frac{1}{N-p}\sum_{i=p+1}^{N}\left\{y_i - \boldsymbol{\mathcal{B}}(t_i)^\top\widehat{\boldsymbol{\theta}}_{\text{PGLS}} - \sum_{j=1}^{p}\widehat{\phi}_j\left[y_{i-j} - \boldsymbol{\mathcal{B}}(t_{i-j})^\top\widehat{\boldsymbol{\theta}}_{\text{PGLS}}\right]\right\}^2 + 2\frac{(\text{df}_1 + p)}{N-p}\widehat{\sigma}_0^2, \quad (3.23)$$

where $\widehat{\sigma}_0^2$ is given by the variance of the residuals from the $\widehat{y}_i$ that are computed when $\boldsymbol{\tau} = \mathbf{0}_{2K+1}$, df_1 is the effective number of parameters obtained fitting the model in equation (3.1), and is calculated as

$$\begin{aligned}\text{df}_1 =& \text{tr}\left[\mathcal{B}(\mathcal{B}^\top\widehat{\mathbf{V}}\widehat{\mathbf{V}}\mathcal{B} + \mathbf{P})^{-1}\mathcal{B}^\top\widehat{\mathbf{V}}\widehat{\mathbf{V}}\right],\\ \widehat{\mathbf{V}} =& \widehat{\mathbf{R}}\left[\text{diag}\left\{\sqrt{\widehat{\lambda}_1}, \ldots, \sqrt{\widehat{\lambda}_N}\right\}\right]\widehat{\mathbf{R}}^\top,\end{aligned} \quad (3.24)$$

and $\widehat{\mathbf{R}}$ and $\{\widehat{\lambda}_1, \ldots, \widehat{\lambda}_N\}$ are, respectively, the eigenvectors and eigenvalues of $\widehat{\sigma}_z^2\widehat{\boldsymbol{\Gamma}}_\varepsilon^{-1}$, that is,

$$\widehat{\mathbf{R}}\left[\text{diag}\{\widehat{\lambda}_1, \ldots, \widehat{\lambda}_N\}\right]\widehat{\mathbf{R}}^\top = \widehat{\sigma}_z^2\widehat{\boldsymbol{\Gamma}}_\varepsilon^{-1}.$$

In the case of fitting the model in equation (3.1) with LSAR(p) errors, we propose to use

the following AIC function

$$\text{AIC}(\boldsymbol{\tau}, \boldsymbol{\tau}^{\phi}) = \frac{1}{N-p} \sum_{i=p+1}^{N} \left\{ y_i - \boldsymbol{\mathcal{B}}(\tfrac{i}{N})^{\top} \widehat{\boldsymbol{\theta}}_{\text{PGLS}} - \sum_{j=1}^{p} \widehat{\phi}_j \left(\tfrac{i}{N}\right) \left[y_{i-j} - \boldsymbol{\mathcal{B}}(\tfrac{i-j}{N})^{\top} \widehat{\boldsymbol{\theta}}_{\text{PGLS}} \right] \right\}^2 + 2 \frac{(\text{df}_1 + \text{df}_2)}{N-p} \widehat{\sigma}_0^2, \tag{3.25}$$

where $\widehat{\sigma}_0^2$ is given by the variance of the residuals from the $\widehat{y}_i$ that are computed when $\boldsymbol{\tau} = \mathbf{0}_{2K+1}$ and $\boldsymbol{\tau}^{\phi} = \mathbf{0}_p$, df_1 is given in equation (3.24), df_2 is the effective number of parameters obtained fitting the LSAR(p) process in equation (3.3), and is calculated as

$$\text{df}_2 = \text{tr}\left[\mathbf{E}(\mathbf{E}^{\top}\mathbf{E} + \mathbf{P}_{\phi})^{-1} \mathbf{E}^{\top} \right],$$

where $\mathbf{E}$ is defined in equation (3.18) and $\mathbf{P}_{\phi}$ in equation (3.20).

The AIC given by equations (3.23) and (3.25) can also be used for selecting the number of B-splines J, the degree d of the B-splines, the number of harmonic components K, and the order p of the model associated to errors.

In the first column of Figure 3.1, we have generated $N = 500$ equally spaced observations from the model described in equation (3.1) with $\mu(t_i)$ as

$$\mu(t_i) = 0.00001 t_i^2 + (-2 + 0.00002 t_i^2)\cos(0.03\pi t_i) + (-0.00001 t_i^2)\sin(0.03\pi t_i).$$

Time is given by $t_i = t_0 + i\Delta$ with $t_0 = 0$ and $\Delta = 1$. From the first to the fourth row, the error term ε_i follows an AR(1) process with coefficient $\phi = 0.1$, an AR(2) process with coefficients $\phi_1 = 0.1$ and $\phi_2 = 0.2$, a LSAR(1) process with coefficient $\phi(u) = 0.1\sin(\pi u)$, $u \in [0,1]$, and a LSAR(2) process with coefficients $\phi_1(u) = 0.1\sin(\pi u)$ and $\phi_2(u) = 0.2\cos(2\pi u)$, $u \in [0,1]$, respectively. The term z_i follows a Gaussian distribution with zero mean and variance $\sigma_z^2 = 1$. In the first column of Figure 3.1, we show the observations and predictions of them for different values of the tuning parameters. The observations y_i, $i = 1, \ldots, N$, are represented by the gray points, and $\mu(t_i)$ (in the case of AR errors) and $\mu\left(\frac{i}{N}\right)$ (in the case of LSAR errors), $i = 1, \ldots, N$, by the black curve. The orange, blue and green curves illustrate three possible estimates for $\boldsymbol{y}$ obtained using the method described in Section 3.3.1 with increasing smoothing parameters. The orange curve is the fit obtained with $\tau_j = 0$, $j = 1, 2, 3$, the blue curve is obtained using the

smoothing parameters $\tau_j = 6$, $j = 1, 2, 3$, in the case of AR(1) and LSAR(1), and $\tau_j = 5$, $j = 1, 2, 3$, in the case of AR(2) and LSAR(2) errors, and the green curve is obtained using $\tau_j = 40$, $j = 1, 2, 3$. In the second column of Figure 3.1, we observe that the optimal tuning parameters are $\tau_j = 6$, $j = 1, 2, 3$, in the case of AR(1) and LSAR(1) errors, and $\tau_j = 5$, $j = 1, 2, 3$, in the case of AR(2) and LSAR(2) errors, and as the values of $\boldsymbol{\tau}$ increase the obtained curve fits the observed data less closely.

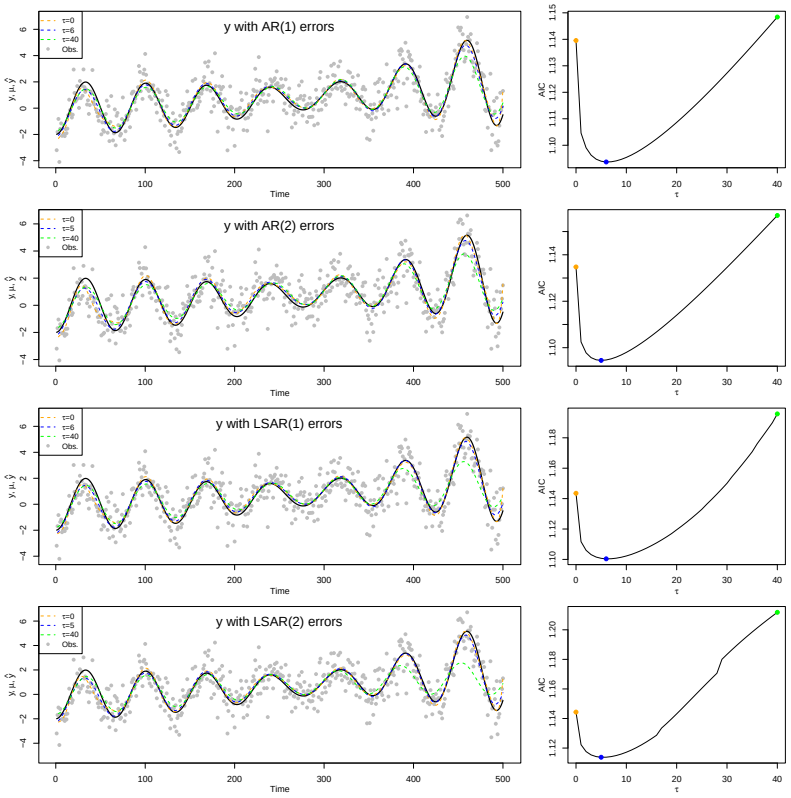

Figure 3.1 Automatic selection of the tunable parameters presented in Section 3.3.3. First column: Data $\boldsymbol{y}$ (gray dots) simulated according to the model in equation (3.1), and $\boldsymbol{\mu}$ (black line). From the first to the fourth row: the errors follow an AR(p) a LSAR(p) process with $p = 1, 2$. The first column illustrates three estimates of $\boldsymbol{y}$, corresponding to three different specifications of τ_k, with $k = 1, 2, 3$,: $\tau_k = 0$ (orange curves), $\tau_k = 6$ and 5 (blue curves), $\tau_k = 40$ (green curves). Second column: Values of the AIC in equations (3.23) and (3.25). The three points on the AIC curve correspond to the three fits presented in the left plot of the figure.

3.4 Simulation results

In this section, we provide Monte Carlo simulations to illustrate the performance of the estimators $\widehat{\mu}(t_i)$, $\widehat{m}(t_i)$, $\widehat{g}_{\ell,k}(t_i)$, $\ell = 1, 2$, $k = 1, \ldots, K$, defined in equations (3.10), (3.11) and (3.12), and the estimators $\widehat{\phi}_j(u)$, $j = 1, \ldots, p$, $u \in [0, 1]$, defined in Section 3.3.2. We also provide simulations to exemplify the method described in Section 3.3.3 to select the order p of the autoregressive process associated with errors. In order to evaluate the estimators $\widehat{\mu}(t_i)$, $\widehat{m}(t_i)$, $\widehat{g}_{\ell,k}(t_i)$, $\ell = 1, 2$, $k = 1, \ldots, K$, in Sections 3.4.1 we simulate the model in equation (3.1) with AR(p), $p = 1, 2$, errors, whereas in Section 3.4.2, we simulate the model in equation (3.1) with LSAR(p), $p = 1, 2$, errors. Finally, in Section 3.4.3 we simulate the model in equation (3.1) with AR(p) and LSAR(p) errors, $p = 3, 4$, to select the optimal order p.

3.4.1 Simulating the time-varying model with stationary errors

We generate the data according to the model described by equation (3.1) with $K = 2$ and $N = 500$. We simulate the trend and amplitudes as $m(t_i) = 0.2t_i - 0.005t_i^2$, $g_{1,1}(t_i) = 0.5\cos(0.2\pi t_i + \pi/2)$, $g_{2,1}(t_i) = 0.7\sin(0.3\pi t_i)$, $g_{1,2}(t_i) = 0.5 - 0.01t_i^2$, $g_{2,2}(t_i) = -0.1t_i + 0.02t_i^2$, with frequencies $w_1 = 20\pi$, and $w_2 = 30\pi$. Time is equally spaced and is defined as $t_i = t_0 + i\Delta$, with $t_0 = 0$ and $\Delta = 0.02$. For the error term ε_i in the model given by equation (3.1), we assume it follows an AR(p) with $\{z_i,\, i = 1, \ldots, N\}$ a Gaussian white noise sequence with zero mean and variance $\sigma_z^2 = 2$. For the order p, we consider two cases: $p = 1$ and 2. Firstly, the error term follows the AR(1) process with coefficient $\phi_1 = 0.1$, and secondly, the error term follows the AR(2) process with coefficients $\phi_1 = 0.1$ and $\phi_2 = 0.3$.

We simulate $M = 200$ times each case, that is, the model in equation (3.1) with the true values previously mentioned. For each $m = 1, \ldots, M$, we compute the estimate $\widehat{\boldsymbol{\theta}}^{(m)}$ defined according to the procedure presented in Section 3.3.2. In the first and second case, the smoothing parameter is $\boldsymbol{\tau} = (1000, 0.06, 0.01, 1000, 1000)^\top$, the total number of B-splines is $J = 7$ of order $d = 3$, and the order penalty is $r = 3$. The selected

values of $\boldsymbol{\tau}, J, d$, and r perform well in our simulation scenario. For this example, we choose these parameters ad-hoc because we are interested in evaluating the performance of our estimators $\widehat{m}(t_i)$ and $\widehat{g}_{\ell,k}(t_i)$, $\ell = 1, 2$, $k = 1, \ldots, K$, rather than selecting the smoothing parameters. In Section 3.3.3 we propose a method to select the smoothing parameters. In Figure 3.2, we show our estimates of $\boldsymbol{\mu}$, $\boldsymbol{m}$, $\boldsymbol{g}_{\ell,k}$, $\ell = 1, 2$, $k = 1, 2$, and their 95% confidence intervals calculated according with Appendix B. Figure 3.2 shows that the iterative method proposed in Section 3.3.2 estimates well the simulated data in both scenarios, that is, with sinusoidal and polynomial trends and amplitudes.

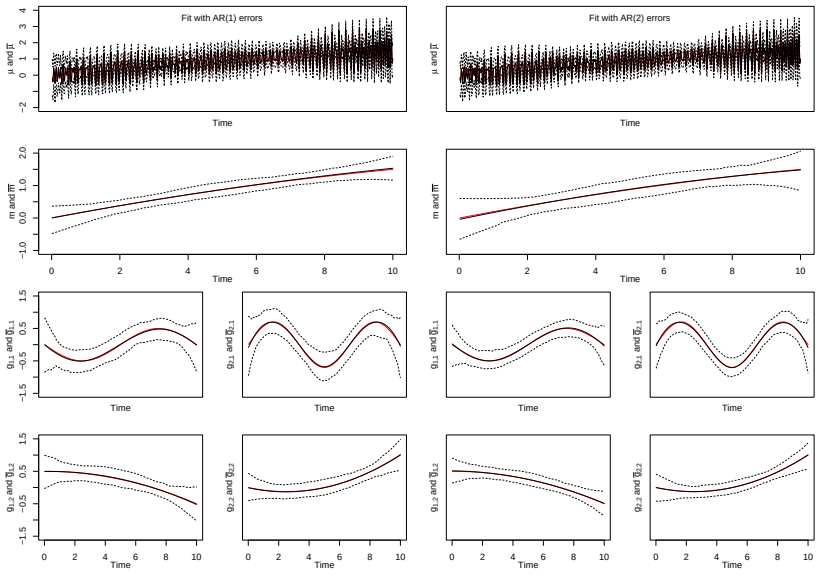

Figure 3.2 Simulation scenarios of Section 3.4.1: estimation of trend and amplitudes, $\boldsymbol{\mu}$, $\boldsymbol{m}$, $\boldsymbol{g}_{\ell,k}$, $\ell = 1, 2$, $k = 1, 2$, of the model in equation (3.1) with AR(p) errors. The first column shows the fit of the model in equation (3.1) with AR(1) errors, whereas the second column shows the fit with AR(2) errors. The first row shows the true $\mu(t_i)$ (red solid line), together with the average $\overline{\mu}(t) = \frac{1}{M}\sum_{j=1}^{M}\widehat{\mu}^{(j)}(t_i)$ of the estimates $\widehat{\mu}^{(j)}(t_i)$ (black solid line). The second row shows the true trend $m(t_i)$ (red solid line) together with the average $\overline{m}(t_i) = \frac{1}{M}\sum_{j=1}^{M}\widehat{m}^{(j)}(t_i)$ of the estimates $\widehat{m}^{(j)}(t_i)$ (black solid line). The third and fourth rows show the true amplitudes $g_{\ell,k}(t_i)$ (red solid line), $\ell = 1, 2$, $k = 1, 2$, together with the average $\overline{g}_{\ell,k}(t_i) = \frac{1}{M}\sum_{j=1}^{M}\widehat{g}_{\ell,k}^{(j)}(t_i)$ of the estimates $\widehat{g}_{\ell,k}^{(j)}(t_i)$ (black solid line). The nonparametric quantiles (black dashed lines) are the confidence intervals corresponding to the 2.5th and 97.5th order statistics, see Appendix B.

3.4.2 Simulating the time-varying model with locally stationary errors

The data is generated according to the model described by equation (3.1) with $K = 2$, $N = 500$ and time is equally spaced defined as $t_i = t_0 + i\Delta$, with $t_0 = 0$ and $\Delta = 0.02$. The trend and amplitudes are considered as in the previous Section (3.4.1), that is, $m(u) = 0.2Nu - 0.005(Nu)^2$, $g_{1,1}(u) = 0.5\cos(0.2\pi Nu + \pi/2)$, $g_{2,1}(u) = 0.7\sin(0.3\pi Nu)$, $g_{1,2}(u) = 0.5 - 0.01(Nu)^2$, $g_{2,2}(u) = -0.1Nu + 0.02(Nu)^2$, $u \in [0, 1]$, with frequencies $w_1 = 20\pi$, and $w_2 = 30\pi$. The errors $\{\varepsilon_{i,N},\, i = 1, \ldots, N\}$ follow a LSAR(p) process with $\{z_i,\, i = 1, \ldots, N\}$ a Gaussian white noise sequence with zero mean and variance $\sigma_z^2 = 2$. For the order p, we consider two cases: $p = 1$ and 2. Firstly, the error term follows a LSAR(1) process with time-varying coefficient $\phi_1(u) = 0.5\sin(2.5\pi u)$, $u \in [0, 1]$, and secondly, the error term follows a LSAR(2) with time-varying coefficients $\phi_1(u) = -0.4\cos(2.5\pi u)$ and $\phi_2(u) = -0.5\sin(\pi u)$, $u \in [0, 1]$.

We simulate $M = 200$ times each scenario and for each $m = 1, \ldots, M$, we compute the estimators $\widehat{\boldsymbol{\theta}}^{(m)}$ and $\widehat{\boldsymbol{\phi}}_j^{(m)}$, $j = 1, 2$, according to the procedure presented in Section 3.3.2. To estimate the parameters $\widehat{\boldsymbol{\theta}}^{(m)}$ and $\widehat{\boldsymbol{\phi}}_1^{(m)}$ in the model given by equation (3.1) with LSAR(1) errors, we select a total number of B-splines $J = 7$ (for the trend and amplitude functions) and $H = 13$ (for the time-varying autoregressive coefficient) of degree $d = 3$, a second-order penalty $r = 3$, and the smoothing parameters $\boldsymbol{\tau} = (1000, 0.06, 0.01, 1000, 1000)^\top$ and $\tau_1^\phi = 1$, whereas to estimate $\widehat{\boldsymbol{\theta}}^{(m)}$, $\widehat{\boldsymbol{\phi}}_1^{(m)}$ and $\widehat{\boldsymbol{\phi}}_2^{(m)}$ in the model given by equation (3.1) with LSAR(2) errors, we choose a total number of B-splines $J = 7$ (for the trend and amplitude functions) and $H = 13$ (for the time-varying autoregressive coefficients) of degree $d = 3$, a second-order penalty $r = 3$, and the smoothing parameters $\boldsymbol{\tau} = (1000, 0.06, 0.01, 1000, 1000)^\top$ and $\boldsymbol{\tau}^\phi = (1, 1)^\top$. For this example, we choose these parameters ad-hoc because we are interested in evaluating the performance of our estimators $\widehat{m}(t_i)$, $\widehat{g}_{\ell,k}(t_i)$, $\ell = 1, 2$, and $\widehat{\phi}_j(u)$, $j = 1, \ldots, p$, rather than selecting the smoothing parameters. In Section 3.3.3 we propose a method to select the smoothing parameters. Figure 3.3 reports the estimation of $\boldsymbol{\mu}$, $\boldsymbol{m}$, $\boldsymbol{g}_{\ell,k}$, $\ell = 1, 2$,

$k = 1, 2$, $\boldsymbol{\phi}_j$, $j = 1, 2$, and their 95% confidence intervals calculated according with Appendix B. Figure 3.3 shows that the iterative method proposed in Section 3.3.2 estimates well the simulated data in both cases, that is, when the trend and amplitudes have a sinusoidal and polynomial form, and also when the time-varying autoregressive coefficients are sinusoidal.

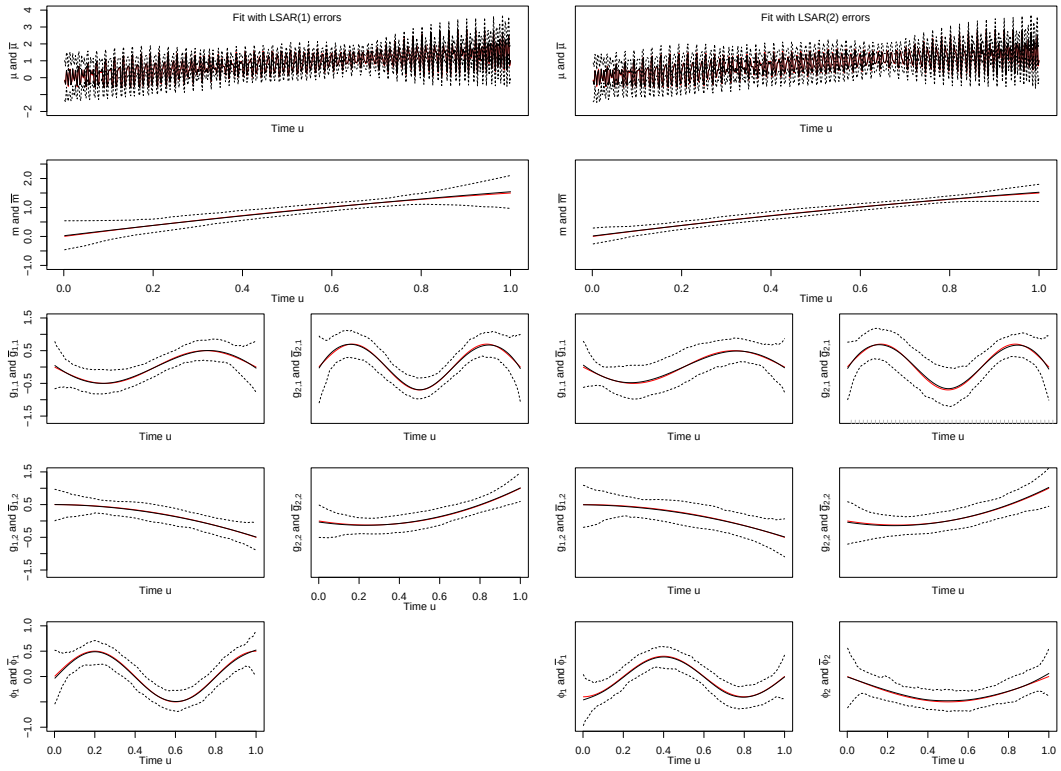

Figure 3.3 Simulation scenarios of Section 3.4.2: estimation of trend and amplitudes, $\boldsymbol{\mu}$, $\boldsymbol{m}$, $\boldsymbol{g}_{\ell,k}$, $\ell = 1, 2$, $k = 1, 2$, and the autoregressive coefficients $\boldsymbol{\phi}(u)$ of the model in equation (3.1) with LSAR(p) errors. The first column corresponds to our model in equation (3.1) with LSAR(1) errors, whereas the second column corresponds to our model in equation (3.1) with LSAR(2) errors. For the description from the first to the fourth row see Figure 3.2. The fifth row shows the true LSAR(1) and LSAR(2) coefficients $\phi_j(u)$ (red solid line), $j = 1, 2$, the estimated coefficients $\overline{\phi}_j(u) = \frac{1}{M}\sum_{m=1}^{M}\widehat{\phi}_j^{(m)}(u)$ (black solid line), $j = 1, 2$, and the nonparametric quantiles confidence intervals (black dashed lines). The nonparametric quantiles are the confidence intervals corresponding to the 2.5th and 97.5th order statistics, see Appendix B.

3.4.3 Estimating the order p of the autoregressive errors

In this section, we evaluate the performance of the AIC function in equations (3.23) and (3.25), through simulations and majority vote, to select the optimal order p of an AR and a LSAR process. We give four examples, two for AR(p) errors, and two for LSAR(p) errors.

We generate the data according to the model in equation (3.1) with $K = 1$, $N = 500$, and time is equally spaced defined as $t_i = t_0 + i\Delta$ with $t_0 = 0$ and $\Delta = 0.002$. We simulate the trend and amplitudes as $m(t_i) = 0.2t_i - 5t_i^2 + 5.5t_i^3$, $g_{1,1}(t_i) = -5t_i^2 + 4t_i^3$, $g_{2,1}(t) = -0.5 - 0.5t_i + 2.5t_i^2 - 0.5t_i^3$ and the frequency $w_1 = 30\pi$.

In the first two examples, the error term ε_i in the model in equation (3.1) follows an AR(p) process given in equation (3.2), whereas in the second two examples, the error term follows a LSAR(p) process given in equation (3.3). For both process, we consider $p = 3, 4$. The coefficients of the AR(3) process are $\phi_1 = -0.2$, $\phi_2 = 0.1$, and $\phi_3 = -0.3$, and the coefficients of the AR(4) process are $\phi_1 = -0.2$, $\phi_2 = 0.1$, $\phi_3 = 0.06$, $\phi_4 = 0.3$. The time-varying coefficients of the LSAR(3) process are $\phi_1(u) = -0.2u + 0.001$, $\phi_2(u) = 0.5u + 0.001$ and $\phi_3(u) = -0.3u + 0.001$, and the time-varying coefficients of the LSAR(4) process are $\phi_1(u) = -0.2u + 0.001$, $\phi_2(u) = 0.1u + 0.001$ and $\phi_3(u) = -0.3u + 0.001$, $\phi_4(u) = 0.35u + 0.01$. The error term $\{z_i,\, i = 1, \ldots, N\}$ is a Gaussian white noise sequence with zero mean and variance $\sigma_z^2 = 2$.

We simulate $M = 200$ times the model in equation (3.1) with AR(3), AR(4), LSAR(3) and LSAR(4) errors, and for each $m = 1, \ldots, M$, we calculate the AIC criterion in equation (3.23), in the case of AR(p) errors, and the AIC function in equation (3.25), in the case of LSAR(p) errors. Both AICs are evaluated over the grid of values $p = 1, \ldots, 6$, fixing $d = 3$, $J = 7$, $H = 8$, $r = 2$, in all examples, and $\tau_j^{\phi} = \tau_k^{\phi} = 1$, for all j, k, in the second two examples (LSAR(3) and LSAR(4) errors). Table 3.1 shows, for the model in equation (3.1) with AR(3), AR(4), LSAR(3) and LSAR(4) errors and for the $M = 200$ replicates, how many times (and in percentages) the orders $p = 1, \ldots, 6$ were selected minimizing the AIC functions. In all the cases, and observing the majority vote, the true order p coincides with the order p most commonly majority selected by the AIC.

AIC criterion for simulated data with AR and LSAR errors

	Model in equation (3.1) with errors:							
	AR(3)		AR(4)		LSAR(3)		LSAR(4)	
Order p	#	%	#	%	#	%	#	%
1	0	0	0	0	0	0	0	0
2	0	0	0	0	7	3.5	0	0
3	121	60.5	0	0	145	72.5	19	9.5
4	26	13	148	74	22	11	138	69
5	29	14.5	29	14	15	7.5	24	12
6	24	12	23	12	11	5.5	19	9.5

Table 3.1 Simulation scenarios of Section 3.4.3: automatic selection of the order p of autoregressive errors. The numbers represent how many times (column #), and in percentages (column %), of the total of 200 simulations, the orders $p = 1, \ldots, 6$ were selected by minimizing the AIC functions in equations (3.23) and (3.25). The columns called AR(p) with $p = 3, 4$, show to the results for data simulated according to the model in equation (3.1) with AR(3) and AR(4) errors, whereas the columns called LSAR(p) with $p = 3, 4$, show to the results for data simulated according to the model in equation (3.1) with LSAR(3) and LSAR(4) errors.

Figure 3.4 shows the AIC values for one simulation of the AR(p) process, with $p = 3$ (a)) and $p = 4$ (b)), and for one simulation of the LSAR(p) process, with $p = 3$ (c)) and $p = 4$ (d)). In Figure 3.4 the order p selected by the AIC functions matches the true order.

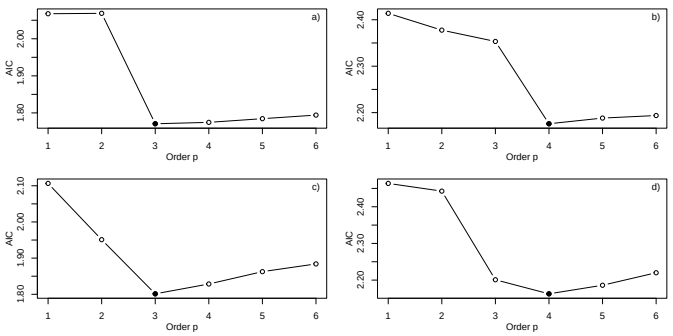

Figure 3.4 Simulation scenarios of Section 3.4.3: Selection of the order p for simulated data according to the model in equation (3.1) with AR errors, a) and b), and with LSAR errors, c) and d). The black points on these lines indicate the smallest value for the AIC functions.

3.5 Conclusions

In this chapter, we extended the model for time series observations of variable stars that are modulated by time-varying magnitudes. In the previous chapter, the errors associated with the model follow a white noise process, whereas in this chapter the autocorrelation is considered by fitting autoregressive processes. Among the autoregressive processes, we considered stationary and locally stationary autoregressive processes. The time-varying coefficients of locally stationary processes are assumed to be smooth. Thus, from the modeling viewpoint, our approach is flexible in the sense that we do not assume any form for the time-dependent quantities (trend, amplitudes, and autoregressive coefficients). From the computational viewpoint, the estimation of the time-varying quantities is translated to time-invariant parameters. We proposed an iterative method performed by generalized least squares to estimate these time-varying parameters and recover the time-varying quantities. Finally, our simulation studies show that the proposed methods deliver good estimates of the time-varying curves.

As we mentioned in Chapter 2, an important challenge in astronomical time series is that observations are unequally spaced in time. Therefore, one limitation of our model is that it can only be applied to equally spaced time series.

Chapter 4

A nonparametric approach for the time-varying spectral density

4.1 Introduction

X-ray binaries are important objects for understanding the physical mechanisms of accretion and jet formation. An X-ray binary system contains either a compact object as a neutron star or a black hole accreting material from a companion star. The matter accreted from the companion star flows on the compact star under the influence of its gravitational potential and it spirals forming an accretion disk. A detailed description of X-ray binaries can be found in Lewin et al. (1997).

Astronomers are interested in modeling the PSD of X-ray binary systems. It is carried out in the frequency domain using Fourier analysis, Lorentzian functions and power-law functions. In general, the procedure is as follows. Firstly, the periodogram is calculated over different segments of the light curve, getting an estimator of the PSD. Then, Lorentzian and power-law functions are fitted to the curve obtained with the periodogram. The periodogram computed over different segments of the light curve allows to track the evolution of of the PSD through time (see Section 1.3.2 for more details). The PSD of black hole

X-ray binaries exhibits five states: quiescence, low/hard, intermediate, high/soft, and very high (see Esin et al., 1997, for more details about these states). These states were also observed in GX 339-4 by Homan and Belloni (2005), in the black hole XTE J1650-500 by Kalemci et al. (2003), and in the black hole Cygnus X-1 by Axelsson, M. et al. (2005). However, the evolution mechanism governing the transitions of the light curve among different states remain not fully understood. Indeed, the approach based on Fourier analysis and Lorentzian functions requires to (i) divide the time into segments and calculate the Fourier transform on each segment, and (ii) fit a model (based on Lorentzian and power-law functions) on each segment. Performing (i) and (ii) *separately* on each segment, forbids to observe *smooth* changes (of PSD, Lorentzian and power-law functions) over time. To overcome this problem we propose to use locally stationary ARMA processes, having in mind two goals. First, we need to prove that the PSD of stationary ARMA processes is suitable to describe the PSD of X-ray binary systems (as suitable as, e.g., the PSD of CARMA processes Kelly et al., 2014). Secondly, we have to extend this result to locally stationary processes and propose a method to estimate its time-varying coefficients. A successful estimation method should be able to handle the huge number of observations of astronomical time series.

In this chapter we introduce, for both stationary and locally stationary ARMA models, an important tool for the study of time series observations coming from X-ray binaries. More precisely, we prove that the PSD of a stationary ARMA(p, q) process can be expressed as a sum of p functions. We provide a mathematical description of these p functions and find a closed form for the frequency at which these functions have a local maximum or minimum. This result makes ARMA models suitable for fitting X-rays binary systems and, in some cases, it avoids having to calculate the frequency at which the PSD has a peak. This is because instead of focusing on the global PSD, it is possible to focus on some of the p functions that describe it. We extend this result to LSARMA processes and, assuming that the time-varying coefficients are smooth, we establish an estimation method (for the time-varying coefficients) that combines B-splines with the Hannan-Rissanen algorithm. Our approach is flexible because it avoids assumptions about the functional form of

the time-varying coefficients, and translates the estimation of the time-varying coefficient into estimating time-invariant parameters. The estimation is performed by ordinary least squares, with the important advantage of avoiding initial values that are needed for other estimation methods – such as for the maximization of the log-likelihood.

We shall divide the present study into five main sections. In Section 4.2 we introduce the class of ARMA(p, q) processes and deliver a novel decomposition of the PSD as the sum of p functions. In Section 4.3 we extended the result in Section 4.2 to the class of LSARMA processes, and we also present a new method to estimate the time-varying coefficients and compute confidence intervals. In Section 4.4 we provide simulation results. Finally, in Section 4.5 we illustrate the decomposition of the PSD presented in Section 4.5 through an application to an X-ray binary system, whereas in Section 4.6 the main conclusions are summarized.

We performed our calculations using the R Language for Statistical Computing (R Core Team, 2021). Our codes combine existing functions (available as part of `R` packages) with our own development.

4.2 Decomposition of the spectral density of an ARMA process

Let $\{x_i,\, i = 1, \ldots, N\}$ be a set observations occurring at certain discrete time $t_1, \ldots, t_N$, with $t_i = t_0 + i\Delta$, i an integer, $\Delta > 0$ is the constant data spacing and $t_0 = 0$. An ARMA(p, q) process $\{x_i,\, i = 1, \ldots, N\}$ with autoregressive coefficients $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$, and moving-average coefficients $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$, is defined to be a solution of the equation

$$x_i = \sum_{j=1}^{p} \phi_j x_{i-j} + z_i + \sum_{j=1}^{q} \theta_j z_{i-j}, \quad \{z_i\} \sim WN(0, \sigma_z^2), \quad i = 1, \ldots, N. \tag{4.1}$$

In order to have a causal solution, we assume that the polynomial $\phi(v) = 1 - \phi_1 v - \cdots - \phi_p v^p$ is non-zero for all complex v such that $|v| \leq 1$.

The ARMA(p, q) process in equation (4.1) has the PSD

$$P_x(f) = \frac{\sigma_z^2}{2\pi} \frac{\left| \sum_{j=0}^{q} \theta_j \exp(-i2\pi f\Delta)^j \right|^2}{\left| -\sum_{j=0}^{p} \phi_j \exp(-i2\pi f\Delta)^j \right|^2}, \quad -\frac{1}{2\Delta} < f < \frac{1}{2\Delta}, \tag{4.2}$$

were $\phi_0 = -1$ and $\theta_0 = 1$. Since the PSD is even, it suffices to confine attention to the values of the PSD over the interval $[0, 1/(2\Delta)]$.

For some values of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, the PSD of an ARMA process has well-defined peaks. We denote by $f_{\max}$ the frequency at which these peaks occur, that is,

$$f_{\max} = \underset{f\in[0,1/(2\Delta)]}{\operatorname{argmax}} P_x(f).$$

These frequencies can be found by differentiating the PSD with respect to $\cos(2\pi f\Delta)$ and setting the derivative equal to zero.

As we mentioned in Section 1.1, the PSD of a CARMA process can be expressed as the sum of Lorentzian functions. This property makes CARMA modeling applicable to many classes of astronomical variables (Kelly et al., 2014). Similarly, the following proposition states that it is possible to express the PSD of an ARMA(p, q) process as the sum of p functions.

Proposition 2 *Let* $\{x_i\}$ *be a causal ARMA*(p, q) *process satisfying equation* (4.1) *with* $q < p$ *and the polynomials* $1 - \phi_1 v - \ldots \phi_p v^p$ *and* $1 + \theta_1 v + \cdots + \theta_q v^q$ *have no common factors, then the PSD in equation* (4.2) *can be expressed as a sum of* p *functions,* $P_{x,j}(f)$*,* $j = 1, \ldots, p$*, as*

$$P_x(f) = \sum_{j=1}^{p} P_{x,j}(f), \tag{4.3}$$

with $P_{x,j}(f)$ *given by*

$$P_{x,j}(f) = -\frac{\sigma_z^2}{2\pi} Re\left[\frac{Q_j}{1 + s_j^2 - 2s_j \cos(2\pi f\Delta)}\right], \tag{4.4}$$

where

$$Q_j = \frac{s_j(1-s_j^2)\sum_{k=0}^{q}\theta_k s_j^k \sum_{k=0}^{q}\theta_k s_j^{-k}}{\left(-\sum_{k=0}^{p}\phi_k s_j^k\right)\left(-\sum_{k=0}^{p}k\phi_k s_j^{-k+1}\right)}, \tag{4.5}$$

and s_j is a root of equation

$$1-\phi_1 v^{-1}-\cdots-\phi_p v^{-p}=0. \tag{4.6}$$

The proof of Proposition 2 is given in Appendix C. The roots of equation (4.6), s_j, $j=1,\ldots,p$, can be reals or complex. Thus, the function $P_{x,j}(f)$ in equation (4.4) can take different form depending on these roots. We describe the component $P_{x,j}(f)$ when the root s_j is real (case 1) and complex (case 2).

Case 1: The root s_j is real

When s_j is real, the function $P_{x,j}(f)$ in equation (4.4) is called Wrapped Lorentzian function. This function is defined by parameters that control the location, the concentration of the function, and a parameter for the normalization (see Appendix F for more details). In the case of $P_{x,j}(f)$, $|s_j|$ controls the concentration and the normalization parameter is $Q_j/(1-s_j^2)$. When $s_j>0$, the centroid of $P_{x,j}(f)$ is equal to 0, whereas when $s_j<0$ the centroid is equal to $1/(2\Delta)$. The function $P_{x,j}(f)$ has the frequencies 0 and $1/(2\Delta)$ as local extremums and depending on the values of Q_j in equation (4.5), they will be local minima or maxima.

Case 2: The roots s_j, s_k are complex, $s_j=a_j+ib_j$, $s_k=a_j-ib_j$

When the roots are complex and take the values $s_j=a_j+ib_j$ and $s_k=a_j-ib_j$, the functions $P_{x,j}(f)$ and $P_{x,k}(f)$ in equation (4.4) are the same and take the form

$$P_{x,\ell}(f) = -\frac{\sigma_z^2}{2\pi}\frac{\mathrm{Re}(Q_\ell)[1+a_\ell^2-b_\ell^2-2a_\ell\cos(2\pi f\Delta)]+2b_\ell\,\mathrm{Im}(Q_\ell)[a_\ell-\cos(2\pi f\Delta)]}{(1+a_\ell^2-b_\ell^2)^2+4a_\ell^2b_\ell^2-4a_\ell(1+a_\ell^2+b_\ell^2)\cos(2\pi f\Delta)+4(a_\ell^2+b_\ell^2)\cos^2(2\pi f\Delta)} \tag{4.7}$$

where $\ell=j,k$. The function $P_{x,\ell}(f)$ in equation (4.7) has local extremums at the frequen-

cies 0 and $1/(2\Delta)$, and at the frequency

$$f_{\max,\ell} = \frac{1}{2\pi\Delta} \arccos\left(\frac{-K_{2,\ell} \pm \sqrt{K_{2,\ell}^2 - 4K_{1,\ell}K_{3,\ell}}}{2K_{1,\ell}} \right), \quad \ell = j, k, \tag{4.8}$$

where

$$\begin{aligned} K_{1,\ell} =& -4(a_\ell^2 + b_\ell^2)[2a_\ell \,\mathrm{Re}(Q_\ell) + 2b_\ell \,\mathrm{Im}(Q_\ell)], \\ K_{2,\ell} =& 8(a_\ell^2 + b_\ell^2)[\mathrm{Re}(Q_\ell)(1 + a_\ell^2 - b_\ell^2) + 2a_\ell b_\ell \,\mathrm{Im}(Q_\ell)], \\ K_{3,\ell} =& [2a_\ell \,\mathrm{Re}(Q_\ell) + 2b_\ell \,\mathrm{Im}(Q_\ell)][(1 + a_\ell^2 - b_\ell^2)^2 + 4a_\ell^2 b_\ell^2] \\ & - 4a_\ell(1 + a_\ell^2 + b_\ell^2)[\mathrm{Re}(Q_\ell)(1 + a_\ell^2 - b_\ell^2) + 2a_\ell b_\ell \,\mathrm{Im}(Q_\ell)], \end{aligned} \tag{4.9}$$

with $K_{2,\ell}^2 - 4K_{1,\ell}K_{3,\ell} \geq 0$, and $\left| \left(-K_{2,\ell} \pm \sqrt{K_{2,\ell}^2 - 4K_{1,\ell}K_{3,\ell}} \right) / (2K_{1,\ell}) \right| < 1$, $\ell = j, k$. Proofs of equations (4.8) and (4.7) are given in Appendix C.

In Figures 4.1-4.3, we exemplify the cases 1 and 2 studying an AR(p) process with $p = 1, 2$, and an ARMA$(3, 2)$ process, all process with data spacing $\Delta = 1$. In Figure 4.1, we give two examples of Case 1 using an AR(1) process with two different values for the coefficient ϕ_1, 0.5 and -0.5. In the left panel of Figure 4.1, the AR(1) process is given by the coefficient $\phi_1 = 0.5$, and the root of equation (4.6) is $s_1 = 0.5$. Thus, its PSD is centered and has local maximum at frequency 0 and local minimum at 0.5. In the right panel of Figure 4.1, the coefficient is $\phi_1 = -0.5$, and the root of equation (4.6) is $s_1 = -0.5$. Therefore, its PSD is centered and has local maximum at 0.5 and a local minimum at 0.

In Figure 4.2, we give three examples, two for Case 1 and one for Case 2 using an AR(2) process. The PSD of the AR(2) process is given by the solid black line, and the functions $P_{x,j}(f)$, $j = 1, 2$, associated to the roots s_1, s_2, are the dashed red and blue lines, respectively. In the left panel of Figure 4.2, the AR(2) coefficients are $\phi_1 = 1.1$, $\phi_2 = -0.27$, and the roots of equation (4.6) are $s_1 = 0.3697$, $s_2 = 0.7303$. Therefore, the functions $P_{x,j}(f)$, $j = 1, 2$, are centered around 0, $P_{x,1}(f)$ has a local maximum at 0.5 and local minimum at 0 (Case 1), and $P_{x,2}(f)$ has a local maximum at 0, and local minimum at 0.5 (Case 1). In the middle panel of Figure 4.2, the coefficients are $\phi_1 = 0.1$, $\phi_2 = 0.2$,

and the roots of equation (4.6) are $s_1 = 0.5$, $s_2 = -0.4$. Thus, $P_{x,1}(f)$ is centered and has a local maximum at 0 and local minimum at 0.5 (Case 1), and the $P_{x,2}(f)$ is centered and has a local maximum at 0.5 and a local minimum 0 (Case 1). In the left panel of Figure 4.2, the AR(2) coefficients are $\phi_1 = 0.5$, $\phi_2 = -0.45$, and the roots of equation (4.6) are $s_1 = 0.25 + 0.6225i$, $s_2 = 0.25 - 0.6225i$. Hence, $P_{x,1}(f) = P_{x,2}(f)$, and $P_{x,\ell}(f)$, $\ell = 1, 2$, have a peak at the frequency 0.184 given by equation (4.8), and local minima at 0 and 0.5 (Case 2).

In Figure 4.3, we give three examples, one for Case 1 and two for Case 2 using an ARMA$(3, 2)$ process. The PSD of the ARMA$(2, 3)$ process is represented by the solid black line, and the functions $P_{x,j}(f)$, $j = 1, 2, 3$, associated to the roots s_j, $j = 1, 2, 3$, respectively, are represented by the dashed red, blue, and green lines, respectively. In the left panel of Figure 4.3, the coefficients of the ARMA$(3, 2)$ process are $\phi_1 = -0.0015$, $\phi_2 = 0.5$, $\phi_3 = -0.1$, $\theta_1 = 0.5$ and $\theta_2 = 0.3$, and the roots of equation (4.6) are $s_1 = 0.2220$, $s_2 = -0.7921$, and $s_3 = 0.5686$. Thus, $P_{x,1}(f)$ is centered and has a local minimum at 0 and local maximum at 0.5 (Case 1), $P_{x,2}(f)$ is centered and has a local maximum at 0.5 and has a local minimum at 0 (Case 1), and $P_{x,3}(f)$ is centered and has a local maximum at 0 and local minimum at 0.5 (Case 1). In the middle panel of Figure 4.3, the coefficients are $\phi_1 = 0.04$, $\phi_2 = -0.1$, $\phi_3 = 0.3$, $\theta_1 = 0.5$ and $\theta_2 = 0.3$, and the roots of equation (4.6) are $s_1 = 0.6323$, $s_2 = -0.2961 + 0.6219i$, and $s_3 = -0.2961 - 0.6219i$. Therefore, $P_{x,1}(f)$ is centered and has a local maximum at 0 and local minimum at 0.5 (Case 1), $P_{x,2}(f) = P_{x,3}(f)$ with a peak at the frequency 0.312 given by equation (4.8) and local minima at frequencies 0 and 0.5 (Case 2). In the right panel of Figure 4.3, the coefficients are $\phi_1 = 0.04$, $\phi_2 = -0.1$, $\phi_3 = -0.3$, $\theta_1 = 0.5$ and $\theta_2 = 0.3$, and the roots of equation (4.6) are $s_1 = -0.6077$, $s_2 = 0.3239 + 0.6235i$, and $s_3 = 0.3239 - 0.6235$. Hence, $P_{x,1}(f)$ is centered and has local maximum at 0.5 and a local minimum at 0 (Case 1), and $P_{x,2}(f) = P_{x,3}(f)$ has a peak at the frequency 0.16 given by equation (4.8) and local minima at 0 and 0.5 (Case 2).

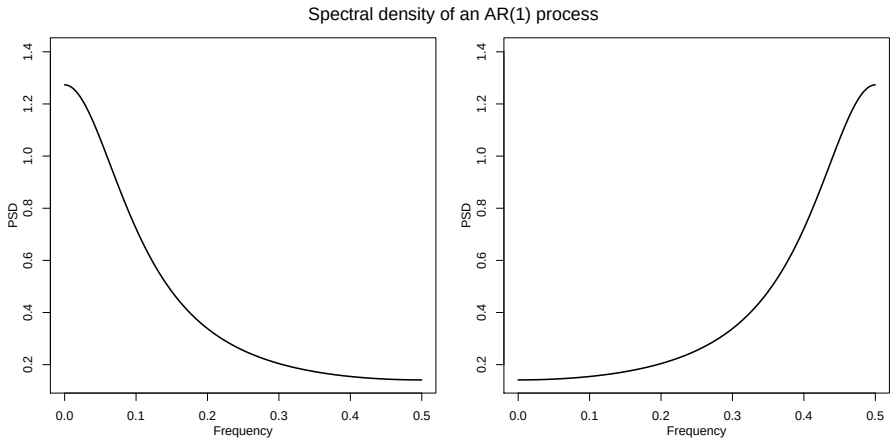

Figure 4.1 PSD of an AR(1) process for different values of the root s_1 of equation $1 - \phi_1 v^{-1} = 0$ studied in Section 4.2. The left panel shows the PSD in equation (4.2) when $s_1 > 0$, and the right panel shows the PSD when $s_1 < 0$.

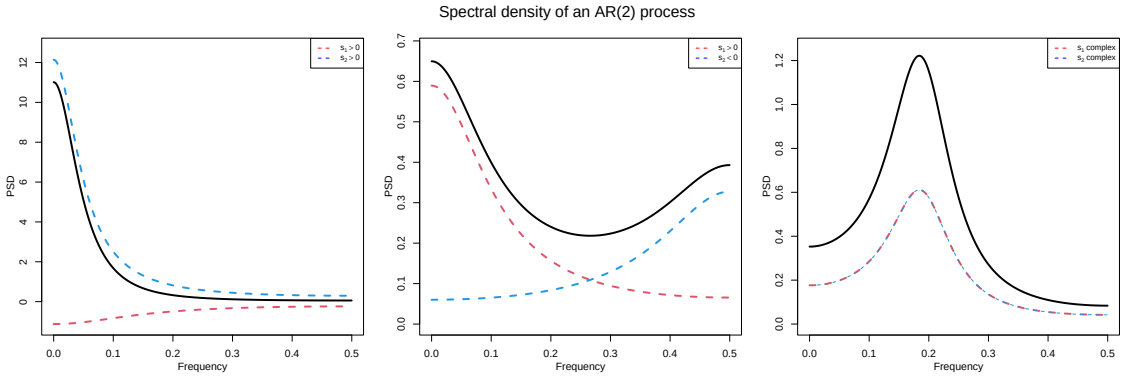

Figure 4.2 Decomposition of the PSD of an AR(2) process for different values of the roots s_1, s_2 of equation $1 - \phi_1 v^{-1} - \phi_2 v^{-2} = 0$ studied in Section 4.2. The PSD in equation (4.2) is the solid black line, whereas the functions $P_{x,1}(f)$ and $P_{x,2}(f)$ in equation (4.4) associated to the roots s_1, s_2 are the dashed red and blue lines, respectively. The left panel shows the case when the roots s_1, s_2 are real and positive, the middle panel shows the case when the roots s_1, s_2 are real, positive and negative, and the right panel shows the case when s_1, s_2 are complex.

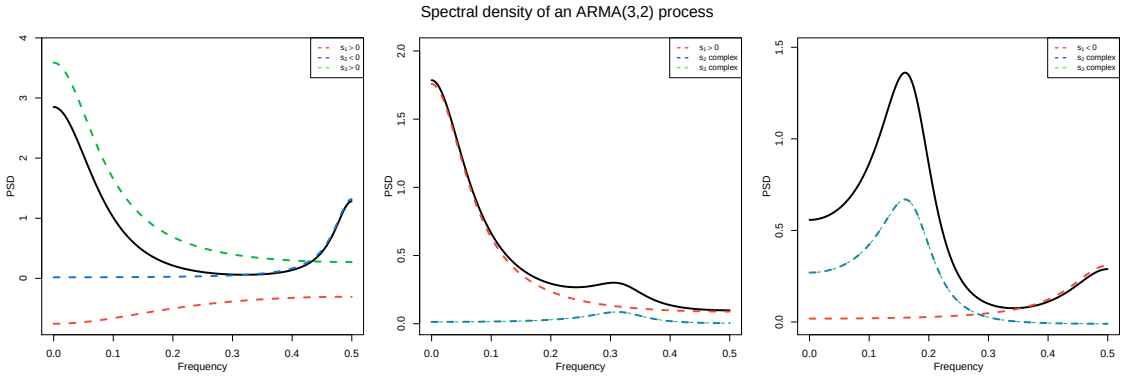

Figure 4.3 Decomposition of the PSD of an ARMA$(3, 2)$ process for different values of the roots s_j, $j = 1, 2, 3$, of equation $1 - \phi_1 v^{-1} - \phi_2 v^{-2} - \phi_3 v^{-3} = 0$ studied in Section 4.2. The PSD in equation (4.2) is the solid black line, whereas the functions $P_{x,j}(f)$, $j = 1, 2, 3$, in equation (4.4) associated to the roots s_j, $j = 1, 2, 3$, are the dashed red, blue and green lines, respectively. The left panel shows the case when the roots s_j, $j = 1, 2, 3$, are real, positive and negative, the middle panel shows the case when the roots s_j, $j = 1, 2, 3$, are real, positive and complex, and the right panel shows the case when s_j, $j = 1, 2, 3$, are real, negative and complex.

4.3 Locally stationary autoregressive moving-average process

In Section 4.3.1 we define LSARMA processes, and extend to their PSD the same decomposition as in Section 4.2. In Section 4.3.2 we provide a method to estimate the time-varying coefficients $\{\phi_1(u), \ldots, \phi_p(u), \theta_1(u), \ldots, \theta_p(u), u \in [0, 1]\}$, as well as the error variance σ_z^2, whereas in Section 4.3.3 we present a method to construct the confidence intervals.

We denote by N the sample size, J the number of B-splines that form the basis associated with the time-varying coefficients, d the degree of the B-splines, r the order of the penalty, and M the number of replications in Monte Carlo simulations.

4.3.1 Definition and decomposition of the spectral density

Let $\{x_{i,N},\, i = 1, \ldots, N\}$ be a set observations occurring at certain discrete time $t_1, \ldots, t_N$, with $t_i = t_0 + i\Delta$, i an integer, $\Delta > 0$ is the constant data spacing, and $t_0 = 0$. An LSARMA(p, q) process $\{x_{i,N},\, i = 1, \ldots, N\}$ with time-varying autoregressive coefficients $\boldsymbol{\phi}(u) = [\phi_1(u), \ldots, \phi_p(u)]^\top$, time-varying moving-average coefficients $\boldsymbol{\theta}(u) = [\theta_1(u), \ldots, \theta_p(u)]^\top$, $u \in [0, 1]$, and WN errors, is defined to be a solution of the equation

$$x_{i,N} = \sum_{j=1}^{p} \phi_j\left(\frac{i}{N}\right) x_{i-j,N} + z_i + \sum_{j=1}^{q} \theta_j\left(\frac{i}{N}\right) z_{i-j}, \quad \{z_i\} \sim WN(0, \sigma_z^2), \quad i = 1, \ldots, N. \tag{4.10}$$

We assume that $1 - \sum_{j=1}^{p} \phi_j(u) v^j \neq 0$ for all $|v| \leq 1 + c$ with $c > 0$ uniformly in u and the coefficients functions $\phi_j(u)$ are continuous in u.

The LSARMA(p, q) process in equation (4.10) has the PSD

$$P_x(u, f) = \frac{\sigma_z^2}{2\pi} \frac{\left| 1 + \sum_{j=1}^{q} \theta_j(u) \exp(-i2\pi f\Delta)^j \right|^2}{\left| 1 - \sum_{j=1}^{p} \phi_j(u) \exp(-i2\pi f\Delta)^j \right|^2}, \quad -\frac{1}{2\Delta} \leq f \leq \frac{1}{2\Delta}, \quad u \in [0, 1], \tag{4.11}$$

Thus, for a fixed time u, we observe the form of the PSD at time u over the frequency interval $[-1/(2\Delta), 1/(2\Delta)]$, and for a fixed frequency f, we observe the time evolution of the PSD at frequency f over time interval $[0, 1]$. As in the case of ARMA processes, for some values of $\boldsymbol{\phi}(u)$ and $\boldsymbol{\theta}(u)$, the PSD of a LSARMA process has well-defined peaks. For a fixed u, we denote by $f_{\max}(u)$, $u \in [0, 1]$, the frequency at which these peaks occur at time u, that is,

$$f_{\max}(u) = \underset{f \in [0, 1/(2\Delta)]}{\operatorname{argmax}} P_x(u, f).$$

These frequencies can be found by differentiating the PSD with respect to $\cos(2\pi f\Delta)$ and setting the derivative equal to zero.

Extending the result in Proposition 2 to LSARMA processes, we can see that the PSD of a causal LSARMA(p, q) process with $q < p$ can be expressed as a sum of p time-varying

function given by

$$P_{x}(u,f) = \sum_{j=1}^{p} P_{x,j}(u,f),$$

with $P_{x,j}(u,f)$ given by

$$P_{x,j}(u,f) = -\frac{\sigma_z^2}{2\pi}\text{Re}\left[\frac{Q_j(u)}{1+s_j^2(u)-2s_j(u)\cos(2\pi f\Delta)}\right], \tag{4.12}$$

where

$$Q_j(u) = \frac{s_j(u)[1-s_j^2(u)]\sum_{k=0}^{q}\theta_k(u)s_j^k(u)\sum_{k=0}^{q}\theta_k(u)s_j^{-k}(u)}{\left[-\sum_{k=0}^{p}\phi_k(u)s_j^k(u)\right]\left[-\sum_{k=0}^{p}k\phi_k(u)s_j^{-k+1}(u)\right]},$$

and, for a fixed $u \in [0,1]$, $s_j(u)$ is a root of equation

$$1-\phi_1(u)v^{-1}-\cdots-\phi_p(u)v^{-p}=0. \tag{4.13}$$

When the roots $s_j(u)$ and $s_k(u)$ of equation (4.13) are complex, that is, $s_j(u)=a_j(u)+ib_j(u)$, $s_k(u)=a_j(u)-ib_j(u)$, the components $P_{x,j}(u,f)$ and $P_{x,k}(u,f)$ in equation (4.12) are the same with the form

$$P_{x,\ell}(u,f) = -\frac{\sigma_z^2}{2\pi}\frac{\text{Re}[Q_\ell(u)][1+a_\ell^2(u)-b_\ell^2(u)-2a_\ell(u)\cos(2\pi f\Delta)]+2b_\ell(u)\,\text{Im}[Q_\ell(u)][a_\ell(u)-\cos(2\pi f\Delta)]}{[1+a_\ell^2(u)-b_\ell^2(u)]^2+4a_\ell^2(u)b_\ell^2(u)-4a_\ell(u)[1+a_\ell^2(u)+b_\ell^2(u)]\cos(2\pi f\Delta)+4[a_\ell^2(u)+b_\ell^2(u)]\cos^2(2\pi f\Delta)}, \tag{4.14}$$

where $\ell = j,k$. The function $P_{x,\ell}(u,f)$ in equation (4.14) has local extremums at the frequencies 0 and $1/(2\Delta)$, and at the frequency

$$f_{\max,\ell}(u) = \frac{1}{2\pi\Delta}\arccos\left(\frac{-K_{2,\ell}(u)\pm\sqrt{K_{2,\ell}^2(u)-4K_{1,\ell}(u)K_{3,\ell}(u)}}{2K_{1,\ell}(u)}\right), \quad \ell=j,k,$$

where

$$\begin{aligned}
K_{1,\ell}(u) =& -4[a_\ell^2(u)+b_\ell^2(u)]\{2a_\ell(u)\,\text{Re}[Q_\ell(u)]+2b_\ell(u)\,\text{Im}[Q_\ell(u)]\},\\
K_{2,\ell}(u) =& 8[a_\ell^2(u)+b_\ell^2(u)]\{\text{Re}[Q_\ell(u)][1+a_\ell^2(u)-b_\ell^2(u)]+2a_\ell(u)b_\ell(u)\,\text{Im}[Q_\ell(u)]\},\\
K_{3,\ell}(u) =& \{2a_\ell\,\text{Re}[Q_\ell(u)]+2b_\ell(u)\,\text{Im}[Q_\ell(u)]\}\{[1+a_\ell^2(u)-b_\ell^2(u)]^2+4a_\ell^2(u)b_\ell^2(u)\}\\
&-4a_\ell(u)[1+a_\ell^2(u)+b_\ell^2(u)]\{\text{Re}[Q_\ell(u)][1+a_\ell^2(u)-b_\ell^2(u)]+2a_\ell(u)b_\ell(u)\,\text{Im}[Q_\ell(u)]\},
\end{aligned}$$

with

$$K_{2,\ell}^2(u) - 4K_{1,\ell}(u)K_{3,\ell}(u) \geq 0, \quad \ell = j,k,$$
$$\left|\left[-K_{2,\ell}(u) \pm \sqrt{K_{2,\ell}^2(u) - 4K_{1,\ell}(u)K_{3,\ell}(u)}\right]/[2K_{1,\ell}(u)]\right| < 1, \quad \ell = j,k,$$

for all $u \in [0,1]$.

4.3.2 Estimation of a LSARMA process

The Hannan-Rissanen method (Hannan and Rissanen, 1982) is an algorithm to estimate the parameters of an ARMA process. In this section we extend the Hannan-Rissanen method to estimate jointly the time-varying autoregressive coefficients $\boldsymbol{\phi}(u)$ and the time-varying moving-average coefficients $\boldsymbol{\theta}(u)$ of a LSARMA process. We assume that the smooth time-varying coefficients can be approximated using B-splines, and then we estimate them using ordinary least squares. Below we detail the proposed estimation method, first for LSAR processes and then for LSARMA processes.

Estimation of LSAR(p) processes

Assuming that the autoregressive coefficients, $\phi_j(u)$, $j = 1,\ldots,p$, $u \in [0,1]$, are smooth and can be modeled as a linear combination of B-splines basis, that is,

$$\phi_k(u) = \sum_{j=1}^{J} \alpha_{\phi,k,j} B_j(u), \quad k = 1,\ldots,p, \tag{4.15}$$

the POLS (penalized ordinary least squares) estimator of $\boldsymbol{\alpha}_\phi = (\boldsymbol{\alpha}_{\phi,1}^\top,\ldots,\boldsymbol{\alpha}_{\phi,p}^\top)^\top$, with $\boldsymbol{\alpha}_{\phi,k} = (\alpha_{\phi,k,1},\ldots,\alpha_{\phi,k,J})^\top$, $k = 1,\ldots,p$, is $\widehat{\boldsymbol{\alpha}} = (\widehat{\boldsymbol{\alpha}}_{\phi,1}^\top,\ldots,\widehat{\boldsymbol{\alpha}}_{\phi,p}^\top)^\top$, with $\widehat{\boldsymbol{\alpha}}_{\phi,k} = (\widehat{\alpha}_{\phi,k,1},\ldots,\widehat{\alpha}_{\phi,k,J})^\top$, $k = 1,\ldots,p$, given by minimizing the sum of squares

$$||\boldsymbol{x}_N - \mathbf{Z}_N\boldsymbol{\alpha}_\phi||^2 + \sum_{j=1}^{p} \tau_j^\phi ||\mathbf{D}_r\boldsymbol{\alpha}_{\phi,j}||^2, \tag{4.16}$$

where $\boldsymbol{x}_N = (x_{p+1,N},\ldots,x_{N,N})^\top$, τ_j^ϕ is positive regularization parameters that control the smoothness of the time-varying coefficient $\phi_j(u)$, $\mathbf{D}_r$ is the matrix that constructs rth order

differences of a vector $\boldsymbol{\eta}$ as $\mathbf{D}_r\boldsymbol{\eta} = \Delta^r\boldsymbol{\eta}$, and $\mathbf{Z}_N$ is the design matrix of dimension $(N-p)\times Jp$ given by

$$\mathbf{Z}_N = \begin{pmatrix} B_1\left(\frac{p+1}{N}\right)x_{p,N} & \cdots & B_J\left(\frac{p+1}{N}\right)x_{p,N} & \cdots & B_1\left(\frac{p+1}{N}\right)x_{1,N} & \cdots & B_J\left(\frac{p+1}{N}\right)x_{1,N} \\ B_1\left(\frac{p+2}{N}\right)x_{p+1,N} & \cdots & B_J\left(\frac{p+2}{N}\right)x_{p+1,N} & \cdots & B_1\left(\frac{p+2}{N}\right)x_{2,N} & \cdots & B_J\left(\frac{p+2}{N}\right)x_{2,N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ B_1\left(\frac{N}{N}\right)x_{N-1,N} & \cdots & B_J\left(\frac{N}{N}\right)x_{N-1,N} & \cdots & B_1\left(\frac{N}{N}\right)x_{N-p,N} & \cdots & B_J\left(\frac{N}{N}\right)x_{N-p,N} \end{pmatrix}.$$

Equating to zero the partial derivatives with respect to each component of $\boldsymbol{\alpha}_\phi$ in equation (4.16), the POLS estimator of $\boldsymbol{\alpha}_\phi$ is

$$\widehat{\boldsymbol{\alpha}}_\phi = (\mathbf{Z}_N^\top\mathbf{Z}_N + \mathbf{P})^{-1}\mathbf{Z}_N^\top\boldsymbol{x}_N, \tag{4.17}$$

where $\mathbf{P} = \mathbf{T}\otimes\mathbf{D}_r^\top\mathbf{D}_r$ and $\mathbf{T} = \text{diag}\{\tau_1^\phi,\ldots,\tau_p^\phi\}$.

Substituting $\boldsymbol{\alpha}_\phi$ in equation (4.15) with its estimator in equation (4.17), $\widehat{\boldsymbol{\alpha}}_\phi$, we get the estimated LSAR(p) coefficients, $\widehat{\boldsymbol{\phi}}_j(u)$, $j=1,\ldots,p$, that is,

$$\widehat{\phi}_k(u) = \sum_{j=1}^{J}\widehat{\alpha}_{\phi,k,j}B_j(u), \quad k=1,\ldots,p.$$

The Hannan-Rissanen estimate of the white noise variance is

$$\widehat{\sigma}_z^2 = (N-p)^{-1}\sum_{i=p+1}^{N}\left\{x_{i,N} - \sum_{j=1}^{p}\widehat{\phi}_j\left(\tfrac{i}{N}\right)x_{i-j,N}\right\}^2.$$

Estimation of LSARMA(p,q) processes

Assuming that the autoregressive coefficients, $\phi_k(u)$, $k=1,\ldots,p$, and the moving-average coefficients $\theta_k(u)$, $k=1,\ldots,q$, $u\in[0,1]$, are smooth and can be approximated using B-splines as

$$\begin{aligned} \phi_k(u) &= \sum_{j=1}^{J}\alpha_{\phi,k,j}B_j(u), \quad k=1,\ldots,p, \\ \theta_k(u) &= \sum_{j=1}^{J}\alpha_{\theta,k,j}B_j(u), \quad k=1,\ldots,q. \end{aligned} \tag{4.18}$$

the estimators of $\phi_k(u)$, $k = 1, \ldots, p$, and $\theta_k(u)$, $k = 1, \ldots, q$, are calculated in three steps:

Step 1: A high-order LSAR(m) model ($m > \max(p, q)$) is fitted using the methodology previously mentioned for LSAR(p). If $[\widehat{\phi}_{m1}(u), \ldots, \widehat{\phi}_{mm}(u)]^\top$ if the vector of estimated time-varying coefficients, then the estimated residuals are computed from the equation

$$\widehat{z}_i^{(0)} = x_{i,N} - \widehat{\phi}_{m1}\left(\tfrac{i}{N}\right) x_{i-1,N} - \cdots - \widehat{\phi}_{mm}\left(\tfrac{i}{N}\right) x_{i-m,N}, \quad i = m+1, \ldots, N.$$

Step 2: Once the estimated residuals $\widehat{z}_i^{(0)}$, $i = m+1, \ldots, N$, have been computed as in Step 1, the vector of parameters, $\boldsymbol{\alpha}_{\phi,\theta} = (\boldsymbol{\alpha}_\phi^\top, \boldsymbol{\alpha}_\theta^\top)^\top$, with $\boldsymbol{\alpha}_\phi = (\boldsymbol{\alpha}_{\phi,1}^\top, \ldots, \boldsymbol{\alpha}_{\phi,p}^\top)^\top$, $\boldsymbol{\alpha}_{\phi,k} = (\alpha_{\phi,k,1}, \ldots, \alpha_{\phi,k,J})^\top$, $k = 1, \ldots, p$, and $\boldsymbol{\alpha}_\theta = (\boldsymbol{\alpha}_{\theta,1}^\top, \ldots, \boldsymbol{\alpha}_{\theta,q}^\top)^\top$, $\boldsymbol{\alpha}_{\theta,k} = (\alpha_{\theta,k,1}, \ldots, \alpha_{\theta,k,J})^\top$, $k = 1, \ldots, q$, are estimated by minimizing the following expression

$$||\boldsymbol{x}_N - \mathbf{Z}_N \boldsymbol{\alpha}_{\phi,\theta}||^2 + \sum_{j=1}^{p} \tau_j^\phi ||\mathbf{D}_r \boldsymbol{\alpha}_{\phi,j}||^2 + \sum_{j=1}^{q} \tau_j^\theta ||\mathbf{D}_r \boldsymbol{\alpha}_{\theta,j}||^2,$$

with respect to $\boldsymbol{\alpha}_{\phi,\theta}$, where $\boldsymbol{x}_N = (x_{m+1+q,N}, \ldots, x_{N,N})^\top$ and $\mathbf{Z}_N$ is the $(N-m-q)\times J(p+q)$ matrix

$$\mathbf{Z}_N = \begin{pmatrix}
B_1\left(\frac{m+1+q}{N}\right) x_{m+q,N} & \cdots & B_J\left(\frac{m+1+q}{N}\right) x_{m+q,N} & \cdots & B_1\left(\frac{m+1+q}{N}\right) x_{m+q+1-p,N} & \cdots & B_J\left(\frac{m+1+q}{N}\right) x_{m+q+1-p,N} \\
B_1\left(\frac{m+2+q}{N}\right) x_{m+q+1,N} & \cdots & B_J\left(\frac{m+2+q}{N}\right) x_{m+q+1,N} & \cdots & B_1\left(\frac{m+2+q}{N}\right) x_{m+q+2-p,N} & \cdots & B_J\left(\frac{m+2+q}{N}\right) x_{m+q+2-p,N} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
B_1\left(\frac{N}{N}\right) x_{N-1,N} & \cdots & B_J\left(\frac{N}{N}\right) x_{N-1,N} & \cdots & B_1\left(\frac{N}{N}\right) x_{N-p,N} & \cdots & B_J\left(\frac{N}{N}\right) x_{N-p,N}
\end{pmatrix}$$

$$\begin{pmatrix}
B_1\left(\frac{m+1+q}{N}\right) \widehat{z}_{m+q}^{(0)} & \cdots & B_J\left(\frac{m+1+q}{N}\right) \widehat{z}_{m+q}^{(0)} & \cdots & B_1\left(\frac{m+1+q}{N}\right) \widehat{z}_{m+1}^{(0)} & \cdots & B_J\left(\frac{m+1+q}{N}\right) \widehat{z}_{m+1}^{(0)} \\
B_1\left(\frac{m+2+q}{N}\right) \widehat{z}_{m+q+1}^{(0)} & \cdots & B_J\left(\frac{m+2+q}{N}\right) \widehat{z}_{m+q+1}^{(0)} & \cdots & B_1\left(\frac{m+2+q}{N}\right) \widehat{z}_{m+2}^{(0)} & \cdots & B_J\left(\frac{m+2+q}{N}\right) \widehat{z}_{m+2}^{(0)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
B_1\left(\frac{N}{N}\right) \widehat{z}_{N-1}^{(0)} & \cdots & B_J\left(\frac{N}{N}\right) \widehat{z}_{N-1}^{(0)} & \cdots & B_1\left(\frac{N}{N}\right) \widehat{z}_{N-q}^{(0)} & \cdots & B_J\left(\frac{N}{N}\right) \widehat{z}_{N-q}^{(0)}
\end{pmatrix},$$

τ_k^ϕ, $k = 1, \ldots, p$, and τ_j^θ, $j = 1, \ldots, q$, are positive regularization parameters that control the smoothness of $\boldsymbol{\phi}_j(u)$, $j = 1, \ldots, p$, and $\boldsymbol{\theta}_j(u)$, $j = 1, \ldots, q$, respectively. Writing the penalties as $\boldsymbol{\alpha}_{\phi,\theta}^\top \mathbf{P} \boldsymbol{\alpha}_{\phi,\theta}$ with $\mathbf{P} = \mathbf{T} \otimes \mathbf{D}_r' \mathbf{D}_r$ and $\mathbf{T} = \text{diag}\{\tau_1^\phi, \ldots, \tau_p^\phi, \tau_1^\theta, \ldots, \tau_q^\theta\}$, the POLS estimator of $\boldsymbol{\alpha}_{\phi,\theta}$ is

$$\widehat{\boldsymbol{\alpha}}_{\phi,\theta} = (\mathbf{Z}_N^\top \mathbf{Z}_N + \mathbf{P})^{-1} \mathbf{Z}_N^\top \boldsymbol{x}_N. \tag{4.19}$$

Substituting $\boldsymbol{\alpha}_{\phi,\theta}$ in equation (4.18) with its estimator in equation (4.19), $\widehat{\boldsymbol{\alpha}}_{\phi,\theta}$, we get the estimated time-varying coefficients, $\widehat{\phi}_j(u)$, $j = 1, \ldots, p$, and $\widehat{\theta}_j(u)$, $j = 1, \ldots, q$, as

$$\begin{aligned} \widehat{\phi}_k^{(0)}(u) &= \sum_{j=1}^{J} \widehat{\alpha}_{\phi,k,j} B_j(u), \quad k = 1, \ldots, p, \\ \widehat{\theta}_k^{(0)}(u) &= \sum_{j=1}^{J} \widehat{\alpha}_{\theta,k,j} B_j(u), \quad k = 1, \ldots, q. \end{aligned} \tag{4.20}$$

Step 3: Compute the residuals

$$\widehat{z}_i^{(1)} = x_{i,N} - \sum_{j=1}^{p} \widehat{\phi}_j^{(0)}\left(\tfrac{i}{N}\right) x_{i-j,N} + \sum_{j=1}^{q} \widehat{\theta}_j^{(0)}\left(\tfrac{i}{N}\right) \widehat{z}_{i-j}^{(0)},$$

with $\widehat{\phi}_j^{(0)}(u)$, $j = 1, \ldots, p$, and $\widehat{\theta}_j^{(0)}(u)$, $j = 1, \ldots, q$, in equation (4.20).

Step 4: Iterate Step 2. - Step 3. until the estimators have stabilized. Our estimators will be $\widehat{\boldsymbol{\phi}}^{(r)}$ and $\widehat{\boldsymbol{\theta}}^{(r)}$ if $\sup_{u\in[0,1]} \|\widehat{\boldsymbol{\phi}}^{(r+1)}(u) - \widehat{\boldsymbol{\phi}}^{(r)}(u)\| < \epsilon$ and $\sup_{u\in[0,1]} \|\widehat{\boldsymbol{\theta}}^{(r+1)}(u) - \widehat{\boldsymbol{\theta}}^{(r)}(u)\| < \epsilon$, for a given small positive ϵ.

The Hannan-Rissanen estimate of the white noise variance is

$$\widehat{\sigma}_z^2 = (N - m - q)^{-1} \sum_{i=m+q+1}^{N} \left\{ x_{i,N} - \sum_{j=1}^{p} \widehat{\phi}_j\left(\tfrac{i}{N}\right) x_{i-j,N} - \sum_{j=1}^{q} \widehat{\theta}_j\left(\tfrac{i}{N}\right) \widehat{z}_{i-j} \right\}^2.$$

4.3.3 Confidence intervals

Dahlhaus (1997) studied the asymptotic distribution of the estimators $\widehat{\boldsymbol{\phi}}(u)$ and $\widehat{\boldsymbol{\theta}}(u)$ given by minimizing the Whittle log-likelihood. This asymptotic distribution can be used for constructing confidence intervals. However, this result is asymptotic, and is not hold when the sample size is small, because the normal approximation is not valid. Even if we want to use the asymptotic distribution to construct the confidence intervals, we need to obtain expressions for the variance and bias. For the estimator in equation (4.20), finding the asymptotic distribution could imply a lot of effort. In order to face this problem, we propose to use bootstrap, which is a tool that approximates certain characteristics like the

variance, mean, among others. For ARMA processes, Kreiss and Franke (1992) proposed a bootstrap method based on the bootstrap principle and the M-estimator. This method allows obtaining bootstrap samples from an ARMA process that can be used to approximate confidence intervals and variance for the parameter estimate. In the same way, we extend the procedure to LSARMA processes and calculate the confidence intervals using the asymptotic normality and the quantile of the bootstrap histogram of the parameter estimate (for more details about these confidence intervals see Efron and Tibshirani (1994, page 168)).

Let $\{x_{i,N}\}$ be a LSARMA process satisfying equation (4.10), and suppose that the time-varying coefficients $\phi_j(u)$, $j = 1, \ldots, p$, and $\theta_j(u)$, $j = 1, \ldots, q$, satisfy the conditions

$$\phi(u, v) = 1 - \sum_{j=1}^{p} \phi_j(u) v^j \neq 0 \text{ for all } |v| \leq 1 + c \text{ with } c > 0 \text{ uniformly in } u$$

$$\theta(u, v) = 1 + \sum_{j=1}^{q} \theta_j(u) v^j \neq 0 \text{ for all } |v| \leq 1 + c \text{ with } c > 0 \text{ uniformly in } u,$$

and $\phi(u, v)$ and $\theta(u, v)$ have no common zero for all u. Also, we suppose that $\phi_p(u) \neq 0$ and $\theta_q(u) \neq 0$ for all $u \in [0, 1]$. The steps to calculate the intervals are:

(i) Set $x_{i,N} = 0$, for $i = p - 1, \ldots, 0$, and fit a LSARMA(p, q) model to $\{x_{i,N}\}$, $i = 1, \ldots, N$. Obtain $\widehat{\boldsymbol{\phi}}(u) = [\widehat{\phi}_1(u), \ldots, \widehat{\phi}_p(u)]^\top$ and $\widehat{\boldsymbol{\theta}}(u) = [\widehat{\theta}_1(u), \ldots, \widehat{\theta}_q(u)]^\top$.

(ii) Define the residuals $\widehat{z}_i$, $i = 1, \ldots, N$, as

$$\widehat{z}_i = \sum_{j=1}^{i} \widehat{\rho}_{j-1}(i, N) \left[- \sum_{k=0}^{p} \widehat{\phi}_k \left(\tfrac{i}{N} \right) x_{i+1-j-k,N} \right],$$

where $\widehat{\phi}_0(u) = -1$ for all $u \in [0, 1]$, and $\widehat{\rho}_j(i, N)$, $j = 0, \ldots, N - 1$, satisfies

$$\left[\widehat{\theta} \left(\tfrac{i}{N}, r \right) \right] \left[\sum_{j=0}^{\infty} \widehat{\rho}_j(i, N) x^j \right] = 1, \quad i = 1, \ldots, N,$$

with $\widehat{\theta}(u, v) = 1 + \widehat{\theta}_1(u) v + \cdots + \widehat{\theta}_q(u) v^q$.

(iii) Center the residuals $\widehat{z}_i$ as $\tilde{z}_i = \widehat{z}_i - \bar{z}$, where $\bar{z} = N^{-1}\sum_{i=1}^{N}\widehat{z}_i$.

(iv) Generate independent and identically distributed bootstrap error variables z_i^*, $N \geq i \geq 1 - \max\{p, q\}$, by sampling randomly with replacement from $\{\tilde{z}_i,\ 1 \leq i \leq N\}$.

(v) Define the bootstrap observations by using the recursion relation

$$x_{i,N}^* = \sum_{j=1}^{p} \widehat{\phi}_j\left(\tfrac{i}{N}\right) x_{i-j,N}^* + \sum_{j=1}^{q} \widehat{\theta}_j\left(\tfrac{i}{N}\right) z_{i-j}^*, \quad i \geq 1 - \max\{p, q\},$$

where, for $i \leq -\max\{p, q\}$, define $x_{i,N}^* = 0$ and $z_i^* = 0$.

(vi) Use the constructed $\{x_{i,N}^*,\ i = 1, \ldots, N\}$ to estimate $\widehat{\boldsymbol{\phi}}^{(1)}(u)$ and $\widehat{\boldsymbol{\theta}}^{(1)}(u)$.

(vii) Repeat the steps (iv)-(vi) B times to generate $\widehat{\boldsymbol{\phi}}^{(i)}(u)$ and $\widehat{\boldsymbol{\theta}}^{(i)}(u)$, $i = 1, 2, \ldots, B$, sets of parameter estimates.

(viii) For a fixed $u \in [0, 1]$, the $100 \cdot (1-\alpha)\%$ confidence intervals for $\phi_j(u)$, $j = 1, \ldots, p$, and $\theta_j(u)$, $j = 1, \ldots, q$, based on the asymptotic normality are

$$\begin{aligned} &\left[\widehat{\phi}_j(u) - z(1-\alpha/2)\widehat{\sigma}_{\phi_j(u)}, \widehat{\phi}_j(u) + z(1-\alpha/2)\widehat{\sigma}_{\phi_j(u)}\right], \quad j = 1, \ldots, p, \\ &\left[\widehat{\theta}_j(u) - z(1-\alpha/2)\widehat{\sigma}_{\theta_j(u)}, \widehat{\theta}_j(u) + z(1-\alpha/2)\widehat{\sigma}_{\theta_j(u)}\right], \quad j = 1, \ldots, q, \end{aligned} \tag{4.21}$$

respectively, where $z(1-\alpha/2)$ denotes the $(1-\alpha/2)$ quantile of the standard Gaussian distribution, and

$$\widehat{\sigma}_{\phi_j(u)} = \sqrt{\frac{1}{B}\sum_{i=1}^{B}\left[\widehat{\phi}_j^{(i)}(u) - \bar{\phi}_j(u)\right]^2}, \quad j = 1, \ldots, p,$$

$$\widehat{\sigma}_{\theta_j(u)} = \sqrt{\frac{1}{B}\sum_{i=1}^{B}\left[\widehat{\theta}_j^{(i)}(u) - \bar{\theta}_j(u)\right]^2}, \quad j = 1, \ldots, q,$$

$$\bar{\phi}_j(u) = \frac{1}{B}\sum_{i=1}^{B}\widehat{\phi}_j^{(i)}(u), \quad j = 1, \ldots, p,$$

$$\bar{\theta}_j(u) = \frac{1}{B}\sum_{i=1}^{B}\widehat{\theta}_j^{(i)}(u), \quad j = 1, \ldots, q,$$

and the $100 \cdot (1-\alpha)\%$ confidence intervals for $\phi_j(u)$, $j = 1, \ldots, p$, and $\theta_j(u)$, $j = 1, \ldots, q$, using the quantiles are

$$\begin{aligned} \left[\widehat{\phi}_{\text{lo},j}(u), \widehat{\phi}_{\text{up},j}(u)\right] &= \left[\widehat{G}^{-1}_{\phi_j(u)}(\alpha/2), \widehat{G}^{-1}_{\phi_j(u)}(1-\alpha/2)\right], \quad j = 1, \ldots, p, \\ \left[\widehat{\theta}_{\text{lo},j}(u), \widehat{\theta}_{\text{up},j}(u)\right] &= \left[\widehat{H}^{-1}_{\theta_j(u)}(\alpha/2), \widehat{H}^{-1}_{\theta_j(u)}(1-\alpha/2)\right], \quad j = 1, \ldots, q, \end{aligned} \tag{4.22}$$

respectively, where $\widehat{G}_{\phi_j(u)}$ is the empirical cumulative distribution of $\left\{\widehat{\phi}_j^{(1)}(u), \ldots, \widehat{\phi}_j^{(B)}(u)\right\}$, $j = 1, \ldots, p$, at time u, $\widehat{H}_{\theta_j(u)}$ is the empirical cumulative distribution of $\left\{\widehat{\theta}_j^{(1)}(u), \ldots, \widehat{\theta}_j^{(B)}(u)\right\}$, $j = 1, \ldots, q$, at time u, $\widehat{G}^{-1}_{\phi_j(u)}(\alpha)$ is the $100 \cdot \alpha$th quantile of $\widehat{G}_{\phi_j(u)}$ at time u, and $\widehat{H}^{-1}_{\theta_j(u)}(\alpha)$ is the $100 \cdot \alpha$th quantile of $\widehat{H}_{\theta_j(u)}$ at time u.

For more details about the residuals $\widehat{z}_i$ and the form of coefficients $\widehat{\rho}_j(i, N)$ in step ii) see Appendix G.

4.4 Simulation results

In this section, we provide Monte Carlo simulations to illustrate the performance of the estimators $\widehat{\boldsymbol{\phi}}(u)$ and $\widehat{\boldsymbol{\theta}}(u)$ given in Section 4.3.2, and the method presented in Section 4.3.3 to construct their confidence intervals. In Section 4.4.1, we simulate observations from a LSAR(2) and a LSARMA$(2, 2)$ process. The forms for the time-varying coefficients are sinusoidal, polynomial and exponential. In Section 4.4.2, we simulate observations from a AR(2) and a LSAR(2) process, and compare the spectral density for both processes. Finally, in Section 4.4.3 we evaluate the performance of coefficients intervals simulating a LSAR(2) process.

4.4.1 Simulating LSARMA processes

To evaluate the estimators of $\boldsymbol{\phi}(u)$ and $\boldsymbol{\theta}(u)$, we consider two simulation examples. In the first example, we generate data from a LSAR(2) process with time-varying coefficients $\phi_1(u) = 0.3\cos(3\pi u)$, and $\phi_2(u) = -0.4u + 0.8u^2$, $u \in [0,1]$. In the second example, we generate data according to a LSARMA$(2,2)$ process with time-varying coefficients $\phi_1(u) = 0.3\cos(3\pi u)$, $\phi_2(u) = 0.6\sin(\pi u)$, $\theta_1(u) = -0.4u + 0.8u^2$, and $\theta_2(u) = 1 - 0.5\exp(u)$, $u \in [0,1]$. For both examples, the errors $\{z_i,\, i = 1,\ldots,N\}$ are a Gaussian white noise sequence with zero mean and variance $\sigma_z^2 = 1$. We generate $N = 500, 1000$ and 5000 observations for each process, and time is given by $t_i = t_0 + i\Delta$ with $t_0 = 0$ and data spacing $\Delta = 1$.

For each $N = 500, 1000$ and 5000, we simulate $M = 200$ each process, and for each $m = 1,\ldots,M$, we compute the estimates $\widehat{\boldsymbol{\phi}}^{(m)}(u)$ and $\widehat{\boldsymbol{\theta}}^{(m)}(u)$ using the methods described in Section 4.3.2. To fit the LSARMA process in the first and second example, we consider the parameters of the B-splines as $J = 6$, $d = 3$, for the penalty parameters as $r = 3$ in both examples, $\tau^{\phi} = (0, 100)^{\top}$ in the first example, and $\boldsymbol{\tau}^{\phi} = (0,0)^{\top}$, $\boldsymbol{\tau}^{\theta} = (100, 100)^{\top}$ in the second example. For these examples, we choose these parameters ad-hoc because we are interested in evaluating the performance of our estimators of $\widehat{\boldsymbol{\phi}}(u)$ and $\widehat{\boldsymbol{\theta}}(u)$, rather than selecting the smoothing parameters. Figure 4.4 shows the estimates of $\boldsymbol{\phi}(u)$ for the first scenario, LSAR(2) process, whereas Figure 4.5 shows the estimates of $\boldsymbol{\phi}(u)$ and $\boldsymbol{\theta}(u)$ for the second scenario, LSARMA(2,2) process, and their 95 % confidence intervals calculated according with Appendix B. Figures 4.4 and 4.5 show that the method proposed in Section 4.3.2 based on the Hannan-Rissanen method estimates well the simulated data in both examples. We also observe that when the number of observations, N, increases, the estimates are closer to the true value.

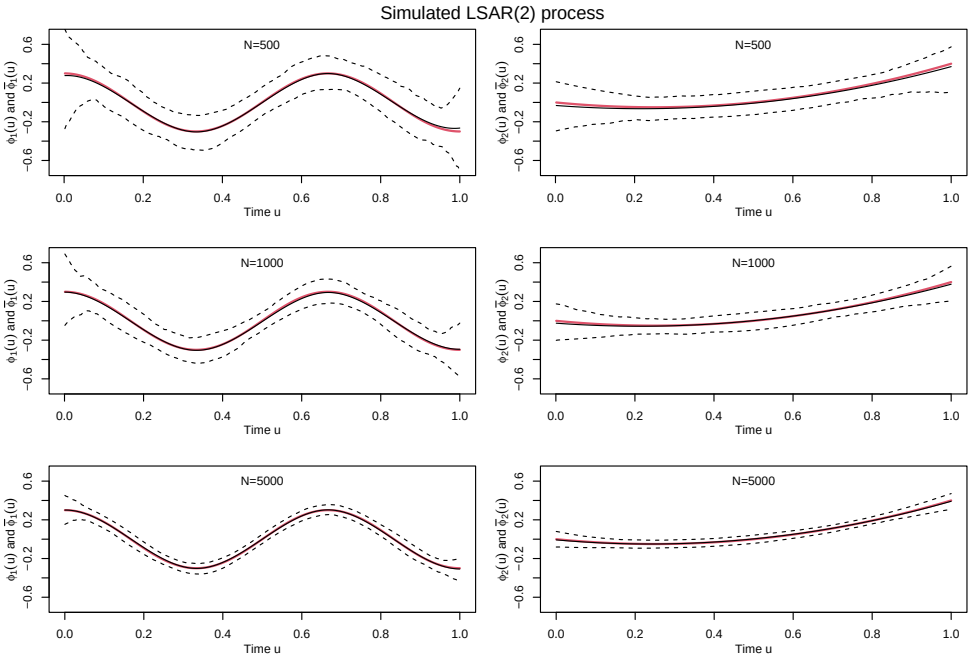

Figure 4.4 Simulation scenarios of Section 4.4.1: estimation of time-varying coefficients of the LSAR(2) process. From the $M = 200$ realizations of our estimators and for each fixed u, we compute two averages and confidence intervals. The left and right panels show the true time-varying coefficients $\phi_1(u)$ and $\phi_2(u)$ (red solid line), together with the averages $\overline{\phi}_1(u) = \frac{1}{M}\sum_{m=1}^{M} \widehat{\phi}_1^{(m)}(u)$ and $\overline{\phi}_2(u) = \frac{1}{M}\sum_{m=1}^{M} \widehat{\phi}_2^{(m)}(u)$ of the estimates $\widehat{\phi}_1^{(m)}(u)$ and $\widehat{\phi}_2^{(m)}(u)$ (black solid line), and their confidence intervals (dashed black line) based on the 0.025 and 0.975 quantiles, see Appendix B. The first, second and third rows show the result for $N = 500, 1000$ and 5000, respectively.

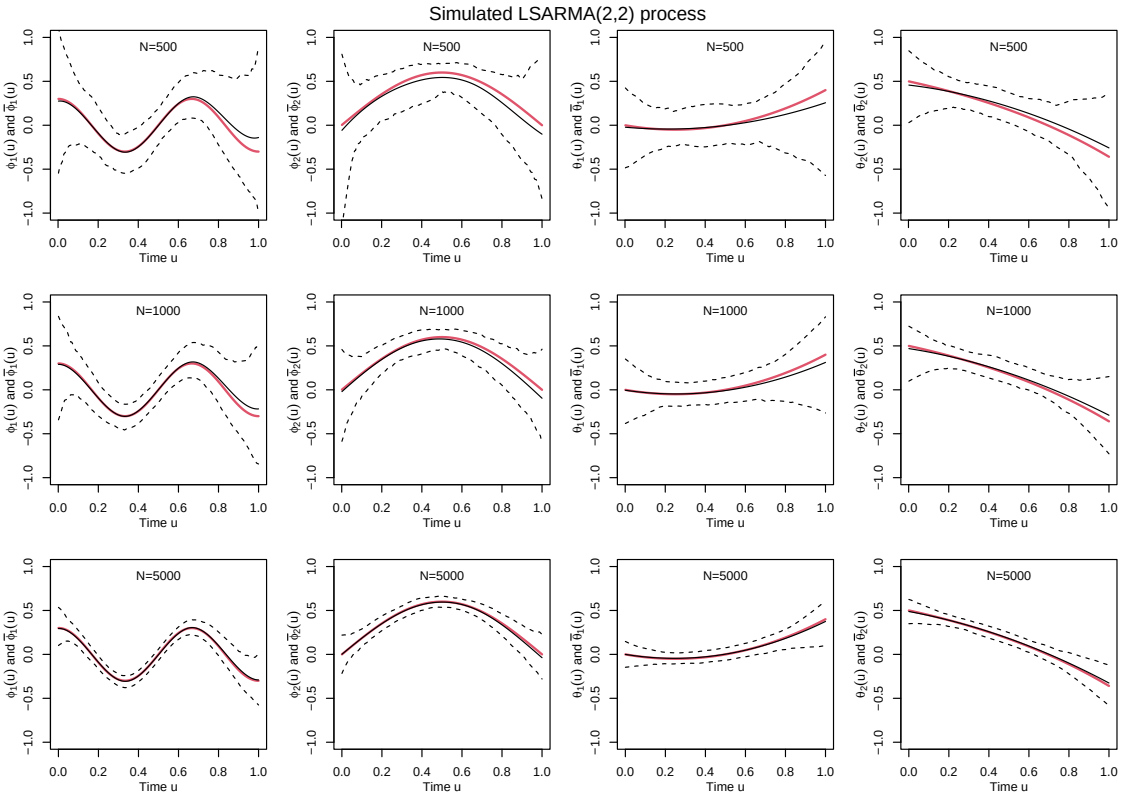

Figure 4.5 Simulation scenarios of Section 4.4.1: estimation of time-varying coefficients of the LSARMA(2,2) process. From the $M = 200$ realizations of our estimators and for each fixed u, we compute four averages and confidence intervals. The first two columns show the true autoregressive time-varying coefficients $\phi_k(u)$, $k = 1, 2$ (red solid line), together with the averages $\overline{\phi}_k(u) = \frac{1}{M}\sum_{m=1}^{M} \widehat{\phi}_k^{(m)}(u)$, $k = 1, 2$, of the estimates $\widehat{\phi}_k^{(m)}(u)$, $k = 1, 2$, (black solid line). The second two columns show the true moving-average time-varying coefficients $\theta_k(u)$, $k = 1, 2$ (red solid line), together with the averages $\overline{\theta}_k(u) = \frac{1}{M}\sum_{m=1}^{M} \widehat{\theta}_k^{(m)}(u)$, $k = 1, 2$, of the estimates $\widehat{\theta}_k^{(m)}(u)$, $k = 1, 2$, (black solid line). Their confidence intervals (dashed black line) are based on the 0.025 and 0.975 quantiles, see Appendix B. The first, second and third rows show the result for $N = 500, 1000$ and 5000, respectively.

4.4.2 Simulating and comparing ARMA and LSARMA processes

In this section, we provide Monte Carlo simulations to illustrate the difference between ARMA and LSARMA processes through the estimation of the coefficients and the PSD. We also show that it is possible to recover the time-invariant coefficients of an ARMA process assuming a LSARMA process as in equation (4.10) and using the estimation method proposed in Section 4.3.2 to estimate them. To do this, we give two scenarios. In the first scenario, we simulate observations from an AR(2) process and we fit AR(2) and LSAR(2) processes, and in the second scenario, we simulate observations from a LSAR(2) process and we fit AR(2) and LSAR(2) processes.

We generate data according to the AR(2) process with time-invariant coefficients $\phi_1 = 1.318$, $\phi_2 = -0.634$, and in the second scenario, we generate data according to the LSAR(2) process with time-varying coefficients $\phi_1(u) = 0.7\cos(3\pi u)$, $\phi_2(u) = -0.81u^2$, $u \in [0, 1]$. For both simulated processes, AR(2) and LSAR(2), the errors $\{z_i,\, i = 1, \ldots, N\}$ are a Gaussian white noise sequence with zero mean and variance $\sigma_z^2 = 289.2$ and $\sigma_z^2 = 1$, respectively, and time is generated as $t_i = t_0 + i\Delta$, with $t_0 = 0$ and $\Delta = 1$, and $N = 1000$.

We simulate $M = 200$ times each process, and for each $m = 1, \ldots, M$, we estimate the time-invariant coefficients ϕ_1 and ϕ_2, and the time-varying coefficients $\phi_1(u)$ and $\phi_2(u)$. The estimation of ϕ_1 and ϕ_2 is made using the `R` function `arima`, whereas the estimation of $\phi_1(u)$ and $\phi_2(u)$ is made using the method proposed in Section 4.3.2. For the observations generated from the AR(2) process, we fit an AR(2) process and a LSAR(2) process. To estimate the time-varying coefficients $\phi_1(u)$ and $\phi_2(u)$ of the LSAR(2) process we consider the following parameters associated with the B-splines, $J = 4, d = 2$, and the following parameters associated with the penalization, $r = 1, \tau^{\phi} = (100000, 100000)^{\top}$. For the observations generated from the LSAR(2) process, we also fit an AR(2) process and a LSAR(2) process. To estimate the time-varying coefficients $\phi_1(u)$ and $\phi_2(u)$ of the LSAR(2) process we consider the following parameters $J = 6, d = 3$ related to the B-splines, and $r = 3, \tau^{\phi} = (0, 100)^{\top}$, related to the penalization. For these examples, we choose these parameters ad-hoc because we are interested in evaluating the performance of our estimator of $\widehat{\boldsymbol{\phi}}(u)$, rather than selecting the smoothing parameters. Figures 4.6 and

4.7 show the estimates of coefficients ϕ_1 and ϕ_2 (fitting an AR(2) and a LSAR(2) process), $\phi_1(u)$ and $\phi_2(u)$ (fitting an AR(2) and a LSAR(2) process), and the estimate of the PSDs. Figure 4.6 shows that even if the true coefficients are time-invariant (red lines), we can recover the constant behavior of them fitting locally stationary processes and using B-splines. On the other hand, Figure 4.7 shows the consequences of fitting stationary processes when the true process is locally stationary. The coefficients and the PSD are estimated inappropriately, been not able to capture the variability over time.

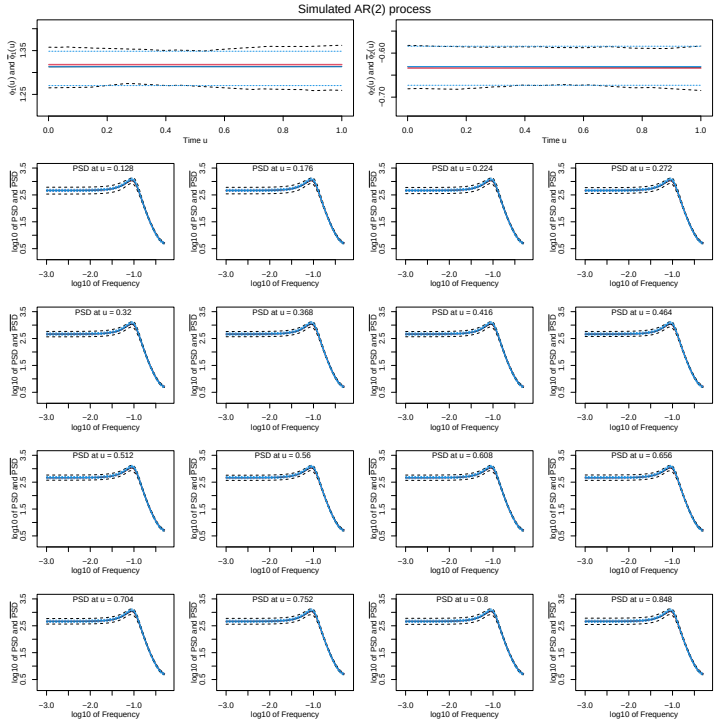

Figure 4.6 Simulation scenarios of Section 4.4.2: data generated from an AR(2) process. The first row shows the true coefficient ϕ_j, $j = 1, 2$, (red lines), together with the averages $\overline{\phi}_j = \frac{1}{M}\sum_{m=1}^{M}\widehat{\phi}_j^{(m)}$, $j = 1, 2$, (blue solid lines), and $\overline{\phi}_j(u) = \frac{1}{M}\sum_{m=1}^{M}\widehat{\phi}_j^{(m)}(u)$, $j = 1, 2$, (black solid lines). The last four rows show the true PSD (red lines), the averages of the estimates of the time-invariant PSD given by $\overline{P}(\omega) = \frac{1}{M}\sum_{m=1}^{M}\widehat{P}^{(m)}(\omega)$ (solid blue lines), and the average of the estimates of the time-varying PSD given by $\overline{P}(u,\omega) = \frac{1}{M}\sum_{m=1}^{M}\widehat{P}^{(m)}(u,\omega)$ (solid black lines), for different instants of time u and in log10 scale. The confidence intervals (dashed black and blue lines) are based on the 0.025 and 0.975 quantiles, see Appendix B.

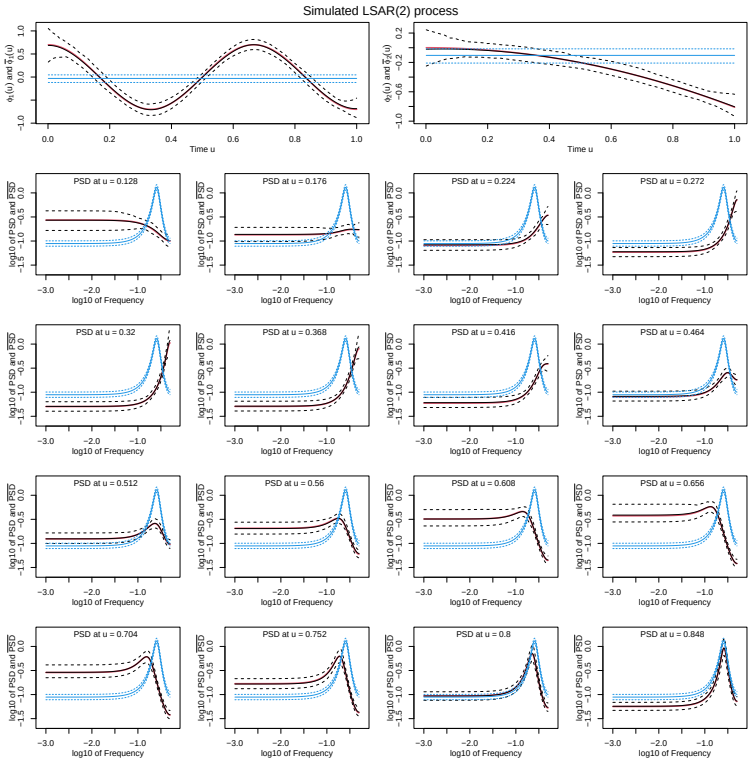

Figure 4.7 Simulation scenarios of Section 4.4.2: data generated from an LSAR(2) process. The first row shows the true coefficient $\phi_j(u)$, $j = 1, 2$, (red lines), together with the averages $\overline{\phi}_j = \frac{1}{M}\sum_{m=1}^{M} \widehat{\phi}_j^{(m)}$, $j = 1, 2$, (blue solid lines), and $\overline{\phi}_j(u) = \frac{1}{M}\sum_{m=1}^{M} \widehat{\phi}_j^{(m)}(u)$, $j = 1, 2$, (black solid lines), of the estimates $\widehat{\phi}_j^{(m)}$ and $\widehat{\phi}_j^{(m)}(u)$, respectively. The last four rows show the true PSD (red lines), the averages of the estimates of the time-invariant PSD given by $\overline{P}(\omega) = \frac{1}{M}\sum_{m=1}^{M} \widehat{P}^{(m)}(\omega)$ (solid blue lines), and the average of the estimates of the time-varying PSD given by $\overline{P}(u, \omega) = \frac{1}{M}\sum_{m=1}^{M} \widehat{P}^{(m)}(u, \omega)$ (solid black lines), for different instants of time u and in log10 scale. The confidence intervals (dashed black and blue lines) are based on the 0.025 and 0.975 quantiles, see Appendix B.

4.4.3 Estimating the confidence intervals

In this section, we evaluate the method proposed in Section 4.3.3 to construct the 95% confidence intervals for the time-varying coefficients $\boldsymbol{\phi}(u)$ through the coverage probability.

The simulation method for estimating the coverage probability of the confidence intervals, based on (i) the asymptotic normality and (ii) the empirical quantiles, is implemented in three steps:

- First, we simulate $M = 200$ samples of size $N = 2000$ from the LSAR(2) process mentioned in Section 4.3.3.
- Then, for each sample, we compute the 95% confidence interval based on the asymptotic normality and the empirical quantiles defined in equations (4.21) and (4.22), respectively, using $B = 500$ bootstrap replicates.
- Finally, for each fixed u, we compute the proportion of samples for which the true parameters $\phi_j(u)$, $j = 1, 2$, are contained in the confidence intervals (based on the asymptotic normality and the empirical quantiles).

Figure 4.8 shows the empirical coverage of the confidence intervals (i.e., the proportion of confidence intervals containing the true values of the parameters) based on the asymptotic normality and the empirical quantiles, for the coefficients $\phi_1(u)$ (left panel) and $\phi_2(u)$ (right panel). In Figure 4.8, left and right panels, we observe that the empirical coverage of both confidence intervals (based on $M = 200$ samples) is approximately 0.95 for each u, which is very close to the theoretical value 0.95. The first column of Figure 4.9 shows the confidence intervals for $\phi_j(u)$, $j = 1, 2$, based on the 0.025 and 0.975 quantiles of the empirical distribution of the estimates of the coefficients $\left\{\widehat{\phi}_j^{(1)}(u), \ldots, \widehat{\phi}_j^{(M)}(u)\right\}$ obtained from the $M = 200$ samples. The second column of Figure 4.9 shows the 95% confidence intervals obtained using the asymptotic normality, whereas the third column shows the 95% confidence intervals obtained using the empirical quantiles for one sample. In Figure 4.9, we observe that the confidence intervals based on bootstrap (second and third

columns) are similar to the confidence intervals construct using the $M = 200$ replications (first column).

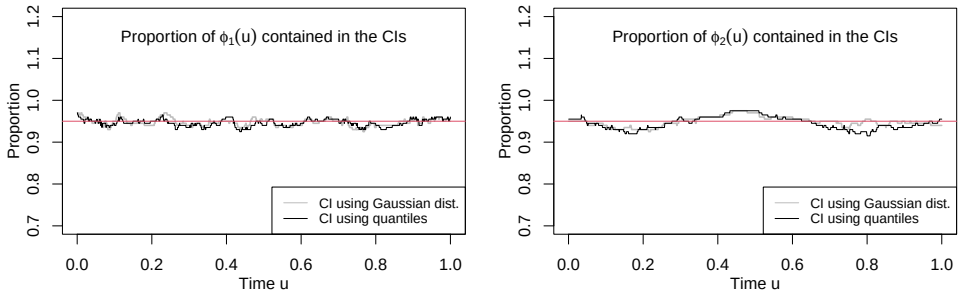

Figure 4.8 Simulation scenario of Section 4.4.3: coverage probability of confidence intervals using bootstrap for $\phi_1(u)$ (left panel) and $\phi_2(u)$ (right panel), from a LSAR(2) process. The gray and black lines represent the proportion of the $M = 200$ samples for which the true parameter $\phi_j(u)$ is contained in the 95% confidence intervals using the asymptotic normality defined in equation (4.21) and the quantiles in equation (4.22), respectively. The red line represents the proportion equal to 0.95 which corresponds to the theoretical value.

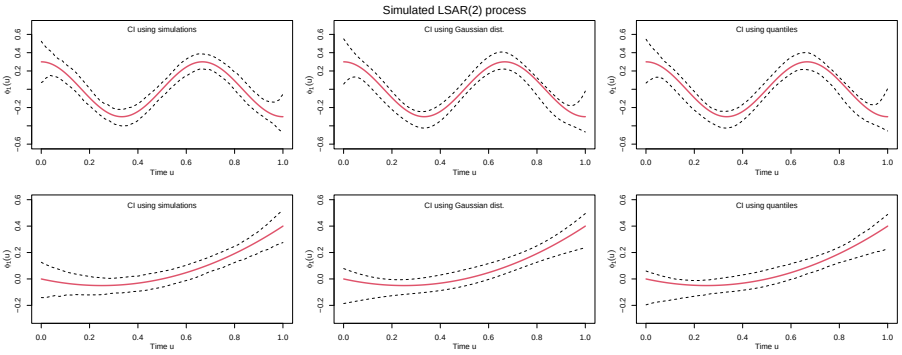

Figure 4.9 Simulation scenario of Section 4.4.3: 95% confidence intervals for $\phi_j(u)$, $j = 1, 2$, using data generated from a LSAR(2) process. The true values of $\phi_j(u)$, $j = 1, 2$, are the red lines. In the first column, the black dashed lines are the confidence intervals based on the 0.025 and 0.975 quantiles of the empirical distribution of the estimates of the coefficients $\left\{\widehat{\phi}_j^{(1)}(u), \ldots, \widehat{\phi}_j^{(M)}(u)\right\}$, with $M = 200$ replicates. In the second and third columns, and for the $B = 500$ bootstrap replications, the black dashed lines are the 95% confidence intervals based on the asymptotic normality defined in equation (4.21) and on the quantiles defined in equation (4.22), respectively.

4.5 Application to real data

In this section we apply the decomposition of the PSD of ARMA models introduced in Section 4.2 by fitting these models to astronomical time series. We study the PSD of a Rossi X-ray Timing Explorer (RXTE) light curve of the X-ray binary XTE 1550-564. This light curve was studied by Kelly et al. (2014) and was chosen because it is regularly sampled every $\Delta = 1/128$ s, and also because it has a complex and well-measured PSD. We analyze a ≈ 39.06 s segment of this light curve, with $N = 5000$ observations, from $t = 3372.742$ s to $t = 3411.797$ s. Due to X-ray binary light curves having a lognormal distribution (Uttley et al., 2005), we transform the data using the logarithm function so to obtain Gaussian data. The transformed light curve is shown in Figure 4.10 and its periodogram in Figure 4.12 (gray line).

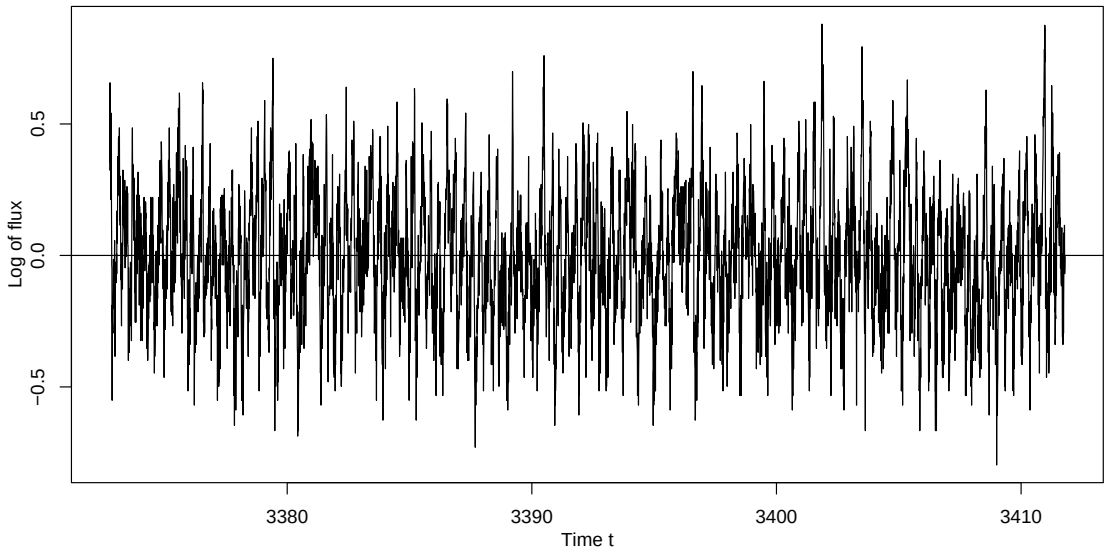

Figure 4.10 RXTE light curve of the X-ray binary XTE 1550-564 studied in Section 4.5. The time series has $N = 5000$ observations.

We fit ARMA models of orders $p = 1, \ldots, 6$, and $q = 1, \ldots, 5$, with $q < p$, to the light curve and select the orders p and q observing the fitted curve in the frequency domain. Comparing the periodogram of the light curve and its estimate of the PSD, we

select the ARMA$(5, 3)$ model. Table 4.1 reports the parameter estimated using the `R` function `arima`. The standard deviation (SD) is obtained with the `arima` function in `R`, the 95% confidence intervals using the `confint` function in `R`, and the p-values using the `coeftest` function in the `R` package `lmtest`. In this table, we observe that the parameters ϕ_j, $j = 1, \ldots, 5$, and θ_j, $j = 1, 2, 3$, are statistically significant at the 5% level.

Parameters estimated fitting an ARMA$(5, 3)$ model to the light curve

Parameter	Estimate	SD	CI	p-value
ϕ_1	2.3323	0.0175	(2.2981,2.3665)	$<$ 2.2e-16
ϕ_2	-1.3956	0.0376	(-1.4693,-1.3219)	$<$ 2.2e-16
ϕ_3	-0.1337	0.0367	(-0.2056,-0.0619)	0.0002661
ϕ_4	0.0764	0.0366	(0.0047,0.1481)	0.0367284
ϕ_5	0.1164	0.0151	(0.0869,0.1459)	1.138e-14
θ_1	-1.8190	0.0215	(-1.8612,-1.7768)	$<$ 2.2e-16
θ_2	0.7234	0.0400	(0.6449,0.8018)	$<$ 2.2e-16
θ_3	0.1159	0.0200	(0.0767,0.1551)	6.777e-09

Table 4.1 Parameters estimated obtained fitting an ARMA$(5, 3)$ model to the light curve studied in Section 4.5. The first column shows the parameters, the second column the estimates, the third column the standard deviations, the fourth column the 95% confidence intervals, and the fifth column the p-values for the estimated parameters.

The standardized residuals of the model are shown in Figure 4.11 (first panel) along with the sample ACF (second and third panels), the partial ACF (PACF) (fourth panel), and the Ljung-Box tests up to lag 30 (fifth panel). From the sample ACF and PACF, it seems that there are no significant autocorrelations in the residuals. This suggests that the ARMA$(5, 3)$ model has captured the correlation structure in the light curve. In addition, from the Ljung-Box tests, we observe that the null hypothesis of white noise is not rejected at the 5% level for all the lags considered, supporting the hypothesis of no significant autocorrelations in the residuals.

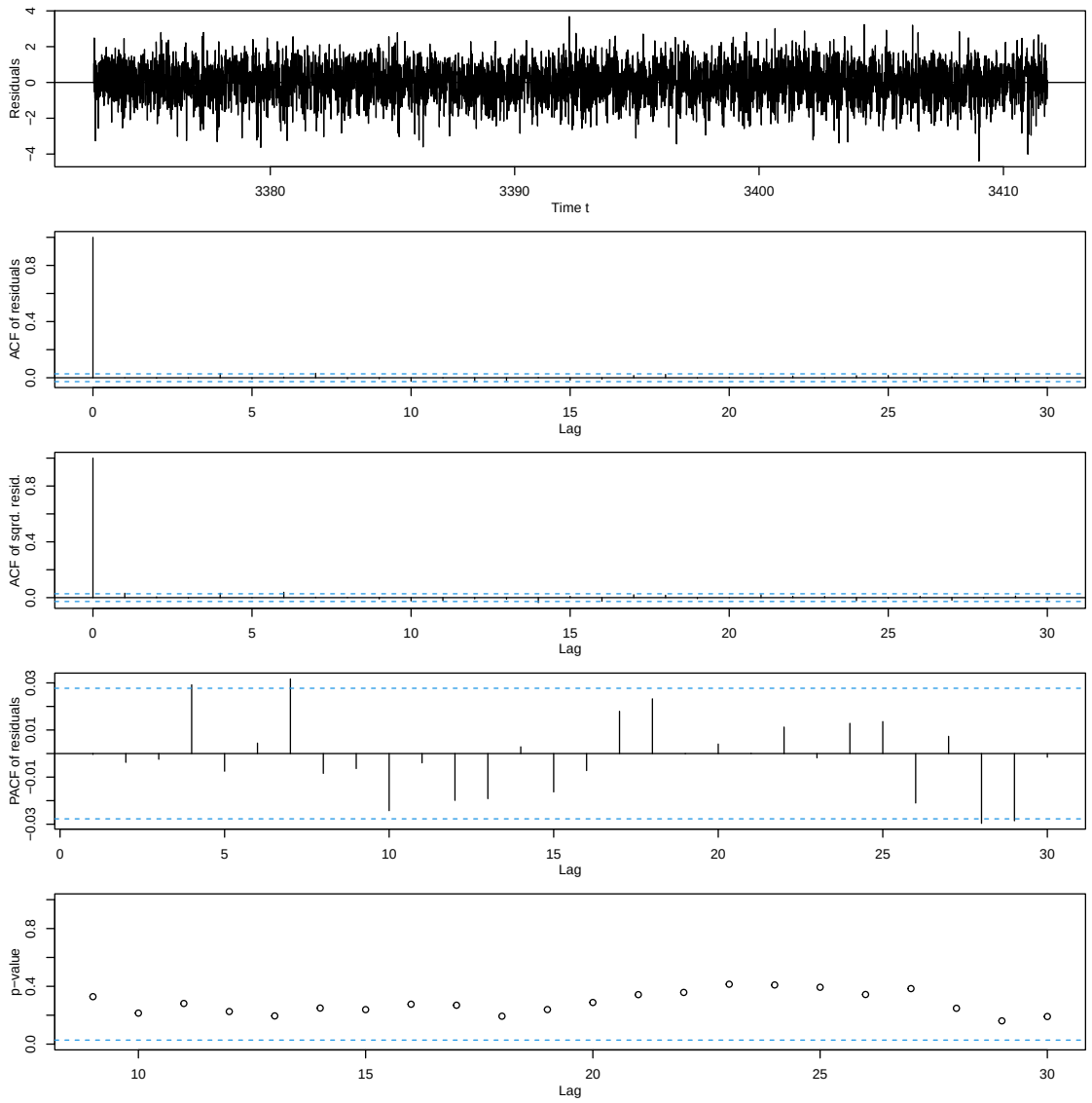

Figure 4.11 Residual diagnostic of the ARMA$(5, 3)$ model fitted to the light curve of the X-ray binary XTE 1550-564 in Section 4.5. The first panel shows standardized residuals, the second panel shows the sample ACF of the standardized residuals, the third panel shows the sample ACF of the squared standardized residuals, the fourth panel shows the PACF of the standardized residuals, and the fifth panel shows the p-values of the Ljung-Box tests.

Finally, we analyzed the fit obtained with the ARMA$(5,3)$ model in the frequency domain. In order to have a better estimation of the PSD of the light curve based on the periodogram, we average five periodograms calculated over five segments of the light curve. The segments are given by observations from $t = 3372.742$ s to $t = 3380.547$ s (segment S_1), observations from $t = 3380.555$ s to $t = 3388.359$ s (segment S_2), observations from $t = 3388.367$ s to $t = 3396.172$ s (segment S_3), observations from $t = 3396.18$ s to $t = 3403.984$ s (segment S_4), and observations from $t = 3403.992$ s to $t = 3411.797$ s (segment S_5), and each of them has 1000 observations. Thus, the average of the periodograms is given by

$$\frac{\overline{I}_x}{N} = \frac{\displaystyle\sum_{i=1}^{5} \sum_{k \text{ if } x_k \in S_i} \sum_{j \text{ if } x_j \in S_i} x_k x_j \exp(i\lambda[t_k - t_j])}{N}, \quad \lambda = 2\pi f. \tag{4.23}$$

Figure 4.12 shows the average of the five periodograms (gray solid line), $\overline{I}_x/N$, the estimate of the PSD (black solid line), $\widehat{P}_x(f)$, obtained replacing the coefficients ϕ_j, $j = 1, \ldots, 5$, and θ_j, $j = 1, 2, 3$, with the estimates of the coefficients given in Table 4.1, and its 95% confidence intervals based on the delta method (black dashed line). This figure shows that the estimated PSD fits well the data and recovers the mean peak observed in the periodogram. Analyzing the estimated PSD according to the decomposition of Section 4.2, we conclude that the PSD of the ARMA(5,3) model fitted to the light curve is described by $p = 5$ components: $\widehat{P}_{x,j}$, $j = 1, \ldots, 5$. These components are obtained replacing the true roots s_j, $j = 1, \ldots, 5$, with the estimates of the roots $\widehat{s}_j$, $j = 1, \ldots, 5$, in equation (4.4). The roots of equation

$$1 - \widehat{\phi}_1 v^{-1} - \widehat{\phi}_2 v^{-2} - \widehat{\phi}_3 v^{-3} - \widehat{\phi}_4 v^{-4} - \widehat{\phi}_5 v^{-5} - \widehat{\phi}_6 v^{-6} = 0$$

are

$$\begin{aligned} &\widehat{s}_1 = -0.2549 + 0.2638i, \quad \widehat{s}_2 = -0.2549 - 0.2638i \\ &\widehat{s}_3 = 0.8787, \quad \widehat{s}_4 = 0.9817 - 0.1446i, \quad \widehat{s}_5 = 0.9817 + 0.1446i. \end{aligned} \tag{4.24}$$

The components associated with these roots are shown in Figure 4.12 (dashed lines). The components $\widehat{P}_{x,1}$ (red dashed line) and $\widehat{P}_{x,2}$ (blue dashed line), given by the complex roots

$\widehat{s}_1$ and $\widehat{s}_2$, are the same and are centered at a frequency other than 0 s^{-1} and $1/(2\Delta)$ s^{-1}. The same happens with the components $\widehat{P}_{x,4}$ (green dashed line) and $\widehat{P}_{x,5}$ (purple dashed line), given by the complex roots $\widehat{s}_4$ and $\widehat{s}_5$, whereas the component $\widehat{P}_{x,3}$, given by the real root $\widehat{s}_3$, is centered at the frequency 0 s^{-1}.

Astronomers are interested in finding a closed form for the frequency at which the periodogram of the light curve exhibits the main peak. If the fitted model is correct (that is, the assumptions are valid) and the estimated PSD describes well the periodogram, finding this frequency is equivalent to finding the frequency at which the estimated PSD is maximized. For ARMA models, this frequency is obtained by differentiating the PSD given in equation (4.2) with respect to f. If we differentiate the PSD corresponding to the ARMA$(5, 3)$ model with respect to f and set to zero, we obtain

$$\begin{aligned} &A_7 \cos^7(2\pi f\Delta) + A_6 \cos^6(2\pi f\Delta) + A_5 \cos^5(2\pi f\Delta) + A_4 \cos^4(2\pi f\Delta) + \\ &A_3 \cos^3(2\pi f\Delta) + A_2 \cos^2(2\pi f\Delta) + A_1 \cos(2\pi f\Delta) + A_0 = 0, \end{aligned} \tag{4.25}$$

where $\{A_j,\ j = 1, \ldots, 7\}$, are given in Appendix H. Replacing the coefficients ϕ_j, $j = 1, \ldots, 5$, and θ_j, $j = 1, 2, 3$, with the estimates given in Table 4.1 in equations (4.25) and (H.1), and solving the equation (4.25), the frequency at which the estimate of the PSD of the ARMA$(5, 3)$ model has a peak is $f_{\max} = 0.02321474$ s^{-1}. As we can see in Figure 4.12, the main peak of the estimated PSD is given by the components $\widehat{P}_{x,4}$ or $\widehat{P}_{x,5}$. Thus, using the expression in equation (4.8), the components $\widehat{P}_{x,4}$ and $\widehat{P}_{x,5}$ have a peak at the frequency $f_{\max,4} = f_{\max,5} = 0.02345531$ s^{-1}. The mean peak in the PSD is very clear and sharp. As a result, the frequencies $f_{\max}$ and $f_{\max,5}$ are very close to each other. In the vertical axis of the graph in Figure 4.12 we observe the position of $f_{\max}$ and $f_{\max,5}$.

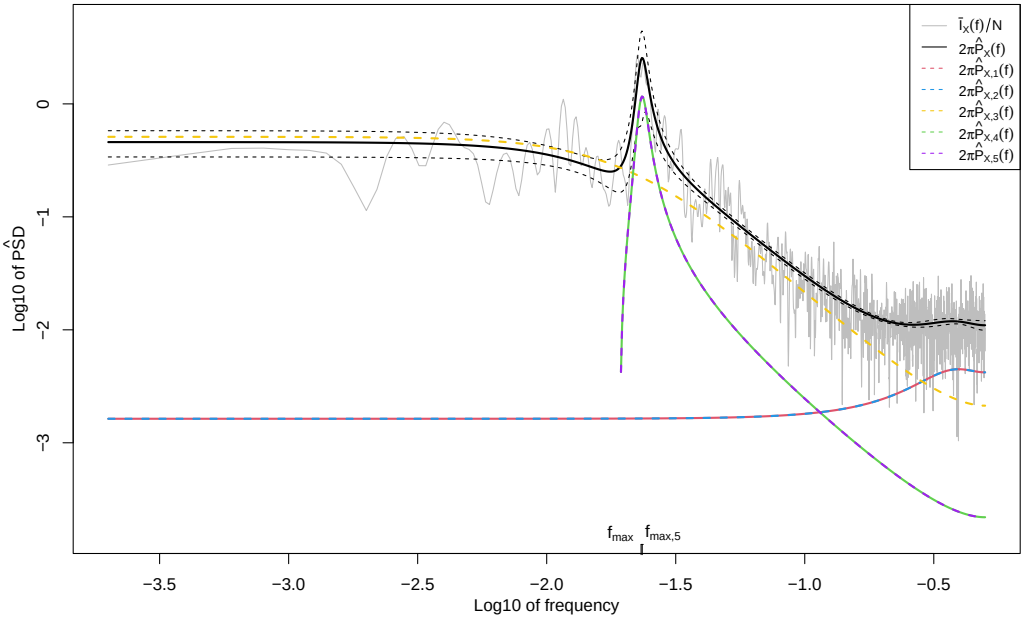

Figure 4.12 PSD for the light curve of the X-ray binary XTE 1550-564 studied in Section 4.5. The gray solid line corresponds to the average of the five periodograms defined in equation (4.23), the black solid line corresponds to the estimate of the PSD given in equation (4.2), the black dashed lines correspond to its 95% confidence intervals, the red, blue, yellow, green and purple dashed lines corresponds to the components of the PSD given in equation (4.4) associated with the roots $\widehat{s}_j$, $j = 1, \ldots, 5$, defined in equation (4.24).

4.6 Conclusions

In this chapter we proved that the PSD of an ARMA process can be decomposed into a sum of some components, and we found a closed form for the frequency at which these components exhibit peaks. We also extended this result to LSARMA process and introduced a nonparametric estimation of its time-varying coefficients. Writing the overall PSD as the sum of individual components allows LSARMA processes to be useful for describing the time-varying PSD of X-ray binary systems with equally spaced time. Previous approaches studied the variability of the PSD of X-ray binary system in discrete time, that it, the periodogram is calculated over different segments of the light curve and the curves obtained are fitted using Lorentzian and power law functions to get an estimation of the PSD. The

frequencies at which the estimated PSD have peaks are then observed in discrete time. Fitting LSARMA process to light curves of X-ray binaries has the advantage of observing the estimated PSD and these frequencies smoothly over time. Concerning the nonparametric LSARMA process proposed, we emphasize that 1) our approach is flexible for four reasons. First, it includes any functional form (such a polynomial) for the time-varying coefficients. Second, in the case that the coefficients are polynomials or time-invariant (such as ARMA process), they can be recovered using B-splines and penalization. Third, the estimation method based on the Hannan-Rissanen algorithm translates the estimation of the time-varying parameters into the estimation of time-invariant parameters that is performed by ordinary least squares. Finally, we propose a bootstrap method for ARMA process to calculate the confidence intervals for the estimated time-varying coefficients.

LSARMA processes are defined for equally spaced time series. As we mentioned in Chapters 2 and 3, in some cases astronomical observations are unequally spaced in time, in which case other methods are implemented. The implications of fitting LSARMA processes to equally spaced time series with missing values goes beyond the scopes of this chapter.

We study the performance of our nonparametric estimation and confidence intervals proposed for LSARMA process under several simulation scenarios. Finally, we apply the decomposition of the PSD of ARMA processes to a Rossi X-ray Timing Explorer (RXTE) light curve of the X-ray binary XTE 1550-564. We fit an ARMA$(5, 3)$ process to the light curve, and thus the estimate of the PSD is represented by the sum of five components, capturing well the curve obtained with the periodogram. The residuals obtained by fitting the ARMA$(5, 3)$ model are compatible with the assumption of stationary and uncorrelated errors. The periodogram of the light curve is characterized by a well defined peak. Therefore, we calculate the frequency at which the estimated PSD and the component of the estimated PSD have this peak, finding values very close. We have not fitted any locally stationary model to these light curves. Nevertheless, fitting LSARMA processes is promising because it allows monitoring the evolution over time of the peak and the corresponding argmax frequency.

Chapter 5

Conclusions and Future Research

In this thesis we introduce three novel models with time-varying parameters to describe light curves of periodic variable stars and X-ray binaries. The time-varying parameters are assumed to be smooth, which allows B-splines to be used for their modeling. From the modeling viewpoint, our approaches presented are flexible because they avoid assumptions about the form of the time-varying parameters.

The first model introduced in Chapter 2 is used for the description of modulated light curves. Our model accounts for a sum of harmonic components with smoothly time-varying trend and amplitudes. The estimation of these time-varying curves is performed by ordinary least squares, translating the estimation of time-varying functions into time-invariant parameters. Astronomical time series are often unequally spaced. Thus, to detect serial correlation in the residuals of our fitted model, we derive the mathematical definition of the spectral density for unequally spaced time series. We studied the performance of our approach under several simulation scenarios and obtained good results. We also applied our model to the light curve of the RR Lyrae star V783 Cyg (KIC 5559631), and compared our results with those obtained fitting the model proposed by Benkő (2018). The estimation error obtained with our fit is significantly smaller than the error obtained with the time-invariant fit. Also, the residuals obtained by fitting the model proposed by Benkő (2018) exhibit a trend and significant spectral peaks, whereas our residuals are compatible

with the assumption of stationary and uncorrelated errors.

In the future, and with respect to the first model introduced, we plan to extend our methodology in four important directions. First, we plan to apply our novel method to the study of a larger sample of Blazhko RR Lyrae stars. Second, our approach can be extended to the analysis of other classes of variable stars presenting long-term changes in their light curve shapes. Third, our fitting method does not require the period(s), amplitude(s), and phase(s) of the Blazhko effect to be determined, as we obtain instead the empirical functions $m(\cdot)$ and $g_{i,k}(\cdot)$. We are currently investigating what kind of (or how much more) information can be obtained from these empirical functions, as compared to conventional approaches. Finally, we aim to study Blazhko light curves characterized by more than one Blazhko frequency – V783 Cyg, which was addressed in some detail in this thesis, is a special case, because this star does not show any additional Blazhko frequencies (Benkő et al., 2014).

The second model introduced in Chapter 3 is an extension of the first model. This second model is proposed for the description of more complex periodic variable stars that, additionally to the modulation, exhibit serial dependence. More precisely, the additional level of complexity is given by the presence of autocorrelation in the errors. Two models are considered: stationary autoregressive (AR) and locally stationary autoregressive (LSAR) models. The main limitation of autoregressive models in astronomy is that it can only be used for equally spaced time series. Finally, we study the performance of the models under several simulation scenarios obtaining good results. In the future, we plan to extend our model to the case of locally stationary autoregressive moving average (LSARMA) errors, that is, by incorporating the moving average component, and apply our model to a light curve.

The last model presented in Chapter 4 corresponds to a nonparametric LSARMA process. Under the assumption of smooth time-varying coefficients, we introduce (i) a nonparametric version of the parametric Hannan-Rissanen algorithm – which was established for stationary ARMA processes, and (ii) a bootstrap method to calculate their confidence intervals. In addition, we showed that the spectral density of ARMA(p, q) processes can be

expressed as the sum of p components, and we extended this result to LSARMA processes. Elaborating on the decomposition, we found a closed form expression for the argmax of the spectrum – the frequency at which the spectral density has a peak. Through an application where we fit an ARMA process to the observed light curve, we showed that the decomposition and the argmax help understanding the spectral density of X-ray binaries. This fit shows that locally stationary processes are good candidates to describe nonstationary astronomical time series. However, as we mentioned before, these processes are only defined for equally spaced data. To address this problem, in the future, we plan to propose an estimation method based on the Kalman filter and fit these models to unequally spaced data sets.

Appendices

A B-splines

In this Appendix we define B-splines and give some details about the estimation method that we used in this manuscript. For more details we refer the reader to the book by de Boor (1978).

A B-spline curve $f(t)$ of degree d is defined as

$$f(t) = \sum_{j=1}^{J} P_j B_{j,d}(t), \tag{A.1}$$

where P_j are the control points and $B_{j,d}(t)$ are the B-spline basis functions. Let $t_{\min}$ and $t_{\max}$ be, respectively, the lower and upper bounds of the domain of interest. In order to build the B-spline basis of degree d, we firstly divide the domain into n intervals, with n being a positive integer, obtaining the $n+1$ knots $\xi_d, \xi_{d+1}, \ldots, \xi_{d+n}$. Each knot satisfies $\xi_j < \xi_{j+1}$, for all j. Secondly, we define $2d$ additional knots $\xi_0, \xi_1, \ldots, \xi_{d-1}, \xi_{n+d+1}, \ldots, \xi_{n+2d-1}, \xi_{n+2d}$. Then, the jth B-spline basis, $B_{j,d}(t)$, can be defined recursively as

$$B_{j,d}(t) = \frac{t - \xi_{j-1}}{\xi_{j+d-1} - \xi_{j-1}} B_{j,d-1}(t) + \frac{\xi_{j+d} - t}{\xi_{j+d} - \xi_j} B_{j+1,d-1}(t), \quad j = 1, \ldots, J, \tag{A.2}$$

with

$$B_{j,0}(t) = \begin{cases} 1 & t \in [\xi_{j-1}, \xi_j), \\ 0 & \text{otherwise} \end{cases} \tag{A.3}$$

being used to initialize the recursion. Thus, to build the B-spline curve given by equation (A.1), we need $n + 2d + 1$ knots, and the total number of B-splines basis functions is $J = n + d$.

To illustrate how to construct a B-spline basis, consider the case of degree $d = 2$ and assume that the domain $[t_{min}, t_{max}]$ has been divided into $n = 3$ intervals, obtaining the knots $\xi_2, \ldots, \xi_5$. In this instance, the $2d = 4$ additional knots are defined as $\xi_0, \xi_1, \xi_6, \xi_7$.

Using equation (A.2), we obtain

$$
\begin{aligned}
B_{1,2}(t) &= \frac{t-\xi_0}{\xi_2-\xi_0}B_{1,1}(t) + \frac{\xi_3-t}{\xi_3-\xi_1}B_{2,1}(t),\\
B_{2,2}(t) &= \frac{t-\xi_1}{\xi_3-\xi_1}B_{2,1}(t) + \frac{\xi_4-t}{\xi_4-\xi_2}B_{3,1}(t),\\
B_{3,2}(t) &= \frac{t-\xi_2}{\xi_4-\xi_2}B_{3,1}(t) + \frac{\xi_5-t}{\xi_5-\xi_3}B_{4,1}(t),\\
B_{4,2}(t) &= \frac{t-\xi_3}{\xi_5-\xi_3}B_{4,1}(t) + \frac{\xi_6-t}{\xi_6-\xi_4}B_{5,1}(t),\\
B_{5,2}(t) &= \frac{t-\xi_4}{\xi_6-\xi_4}B_{5,1}(t) + \frac{\xi_7-t}{\xi_7-\xi_5}B_{6,1}(t),
\end{aligned}
$$

where

$$
\begin{aligned}
B_{1,1}(t) &= \frac{t-\xi_0}{\xi_1-\xi_0}B_{1,0}(t) + \frac{\xi_2-t}{\xi_2-\xi_1}B_{2,0}(t),\\
B_{2,1}(t) &= \frac{t-\xi_1}{\xi_2-\xi_1}B_{2,0}(t) + \frac{\xi_3-t}{\xi_3-\xi_2}B_{3,0}(t),\\
B_{3,1}(t) &= \frac{t-\xi_2}{\xi_3-\xi_2}B_{3,0}(t) + \frac{\xi_4-t}{\xi_4-\xi_3}B_{4,0}(t),\\
B_{4,1}(t) &= \frac{t-\xi_3}{\xi_4-\xi_3}B_{4,0}(t) + \frac{\xi_5-t}{\xi_5-\xi_4}B_{5,0}(t),\\
B_{5,1}(t) &= \frac{t-\xi_4}{\xi_5-\xi_4}B_{5,0}(t) + \frac{\xi_6-t}{\xi_6-\xi_5}B_{6,0}(t),\\
B_{6,1}(t) &= \frac{t-\xi_5}{\xi_6-\xi_5}B_{6,0}(t) + \frac{\xi_7-t}{\xi_7-\xi_6}B_{7,0}(t),
\end{aligned}
$$

and the coefficients $\{B_{j,0}(t),\, j = 1,\ldots,7\}$ are defined in equation (A.3).

Suppose we have N observations $\{t_1,\ldots,t_N\}$, that might be either equally or unequally spaced, with $t_i \in [t_{min}, t_{max}]$ for all $i = 1,\ldots,N$. The B-splines basis matrix evaluated at time $\{t_1,\ldots,t_N\}$, denoted by $\mathbf{B}$, is the $N \times J$ matrix with entries $\{B_{j,d}(t_i),\, i = 1,\ldots,N,\, j = 1,\ldots,J\}$, in a way that each row contains a B-spline basis. The jth B-

spline basis function satisfies the following properties:

$$\begin{cases} B_{j,d}(t) > 0 & \xi_{j-1} < t < \xi_{j+d}, \\ B_{j,d}(t) = 0 & \xi_0 \leq t \leq \xi_{j-1} \text{ and } \xi_{j+d} \leq t \leq \xi_{n+2d}, \\ \sum_{j=1}^{J} B_{j,d}(t) = 1 & \xi_d < t < \xi_{n+d}, \\ \frac{\partial^k B_{j,d}(t)}{\partial t^k}|_{t=\xi_\ell} & 0 \leq k \leq d-1 \text{ are continuous.} \end{cases}$$

For ease of notation, we use, throughout our manuscript, $B_j(t)$ instead of $B_{j,d}(t)$. Let us now consider the example of estimating the mean function $\mu(t)$ of model $y_i = \mu(t_i) + \varepsilon_i$ using B-splines. Let $\boldsymbol{y} = (y_1, \ldots, y_6)'$ be the available $N = 6$ responses observed, respectively, at time $\{t_1, \ldots, t_6\}$, with $t_{min} = t_1$ and $t_{max} = t_6$. Then assume that $\mu(t) = \sum_{j=1}^{J} P_j B_j(t)$, for all $t \in [t_1, t_6]$. We use here B-splines basis functions of degree $d = 2$; in order to construct them, we divide the domain $[t_1, t_6]$ into $n = 3$ intervals. Hence, the total number of knots $\xi_0, \ldots, \xi_7$ is $n + 2d + 1 = 8$, and the total number of B-splines basis functions is $J = n + d = 5$. The 6×5 design matrix $\mathbf{B}$ has entries $B_{ij} = B_j(t_i)$, with $i = 1, \ldots, 6$ and $j = 1, \ldots, 5$, which permits estimating the coefficients $\{P_j,\, j = 1, \ldots, 5\}$ by ordinary least squares. Indeed, if $\boldsymbol{y} = (y_1, \ldots, y_6)^\top$ denotes the response-vector and $\boldsymbol{\theta} = (P_1, \ldots, P_5)^\top$ the parameter-vector, we can rewrite the model as $\boldsymbol{y} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{z}$, where $\boldsymbol{z} = (z_1, \ldots, z_6)^\top$ is the error vector. The estimated parameters are defined as $\widehat{\boldsymbol{\theta}} = (\widehat{P}_1, \ldots, \widehat{P}_5)^\top = (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \boldsymbol{y}$, and the estimated mean as $\widehat{\mu}(t) = \sum_{j=1}^{5} \widehat{P}_j B_j(t)$, for all $t \in [t_1, t_6]$.

B Nonparametric confidence intervals

In this Appendix, we described the confidence intervals constructed in Sections 2.6.1, 3.4.1, 3.4.2 and 4.4 using the quantiles 0.025 and 0.975.

Nonparametric confidence intervals in simulations of Section 2.6.1

We use the quantiles 0.025 and 0.975 to construct the confidence intervals in our simulations of Section 2.6.1. For t fixed, confidence intervals for $\mu(t)$, $m(t)$, and $g_{\ell,k}(t)$, $\ell = 1, 2$, $k = 1, \ldots, K$, are calculated according to the following 3 steps:

1. We estimate the coefficients of interest $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_k$ and $\boldsymbol{\gamma}_k$, $k = 1, \ldots, K$, following the procedure described in Section 2.4.1, and we define the $J \times M$ matrices $\mathbf{A}_M = \left[\widehat{\boldsymbol{\alpha}}^{(1)}, \ldots, \widehat{\boldsymbol{\alpha}}^{(M)}\right]$, $\mathbf{B}_{kM} = \left[\widehat{\boldsymbol{\beta}}_k^{(1)}, \ldots, \widehat{\boldsymbol{\beta}}_k^{(M)}\right]$, and $\mathbf{G}_{kM} = \left[\widehat{\boldsymbol{\gamma}}_k^{(1)}, \ldots, \widehat{\boldsymbol{\gamma}}_k^{(M)}\right]$, where $\widehat{\boldsymbol{\alpha}}^{(j)}$, $\widehat{\boldsymbol{\beta}}_k^{(j)}$, and $\widehat{\boldsymbol{\gamma}}_k^{(j)}$, correspond to the estimators of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_k$, and $\boldsymbol{\gamma}_k$, given by equation (2.10) in the jth Monte Carlo simulation, $j = 1, \ldots, M$.

2. We define the $M \times 1$ vectors $\widehat{\boldsymbol{m}}^{(M)}(t)$, $\widehat{\boldsymbol{g}}_{\ell,k}^{(M)}(t)$, $\ell = 1, 2$, $k = 1, \ldots, K$, and $\widehat{\boldsymbol{\mu}}^{(M)}(t)$ as

$$\begin{aligned}
\widehat{\boldsymbol{m}}^{(M)}(t) &= \boldsymbol{B}(t)^\top \mathbf{A}_M = \left[\widehat{m}^{(1)}(t), \ldots, \widehat{m}^{(M)}(t)\right]^\top, \\
\widehat{\boldsymbol{g}}_{1,k}^{(M)}(t) &= \boldsymbol{B}(t)^\top \mathbf{B}_{kM} = \left[\widehat{g}_{1,k}^{(1)}(t), \ldots, \widehat{g}_{1,k}^{(M)}(t)\right]^\top, \qquad k = 1, \ldots, K \\
\widehat{\boldsymbol{g}}_{2,k}^{(M)}(t) &= \boldsymbol{B}(t)^\top \mathbf{G}_{kM} = \left[\widehat{g}_{2,k}^{(1)}(t), \ldots, \widehat{g}_{2,k}^{(M)}(t)\right]^\top, \qquad k = 1, \ldots, K \\
\widehat{\boldsymbol{\mu}}^{(M)}(t) &= \widehat{\boldsymbol{m}}(t) + \sum_{k=1}^{K} \left\{\widehat{\boldsymbol{g}}_{1,k}(t)\cos(w_k t) + \widehat{\boldsymbol{g}}_{2,k}(t)\sin(w_k t)\right\} = \left[\widehat{\mu}^{(1)}(t), \ldots, \widehat{\mu}^{(M)}(t)\right]^\top,
\end{aligned}$$

where $\widehat{\mu}^{(j)}(t)$, $\widehat{m}^{(j)}(t)$, and $\widehat{g}_{\ell,k}^{(j)}(t)$, $\ell = 1, 2$, $k = 1, \ldots, K$ correspond to the estimators of $\mu(t)$, $m(t)$, $g_{\ell,k}(t)$, given by equations (2.11) and (2.12) in the jth Monte Carlo simulation, and $\boldsymbol{B}(t)$ corresponds to the vector formed by the B-splines evaluated at time t.

3. We calculate the empirical quantiles of order 0.025 and 0.975 of the $M \times 1$ vectors $\widehat{\boldsymbol{m}}^{(M)}(t)$, $\widehat{\boldsymbol{g}}_{\ell,k}^{(M)}(t)$, $\ell = 1, 2$, $k = 1, \ldots, K$, and $\widehat{\boldsymbol{\mu}}^{(M)}(t)$.

Nonparametric confidence intervals in simulations of Sections 3.4.1 and 3.4.2

For t and u fixed, confidence intervals for $\mu(t)$, $m(t)$, $g_{\ell,k}(t)$, $\ell = 1, 2$, $k = 1, \ldots, K$, and $\phi_j(u)$, $j = 1, \ldots, p$, are calculated according to the following 3 steps:

1. We estimate the coefficients of interest $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_k$, $\boldsymbol{\gamma}_k$, $k = 1, \ldots, K$, and $\boldsymbol{\ell}_j$, $j = 1, \ldots, p$, following the procedure described in Section 3.3.2, and we define the $J \times M$ matrices $\mathbf{A}_M = \left[\widehat{\boldsymbol{\alpha}}^{(1)}, \ldots, \widehat{\boldsymbol{\alpha}}^{(M)}\right]$, $\mathbf{B}_{kM} = \left[\widehat{\boldsymbol{\beta}}_k^{(1)}, \ldots, \widehat{\boldsymbol{\beta}}_k^{(M)}\right]$, $\mathbf{G}_{kM} = \left[\widehat{\boldsymbol{\gamma}}_k^{(1)}, \ldots, \widehat{\boldsymbol{\gamma}}_k^{(M)}\right]$, and $\mathbf{P}_{jM} = \left[\widehat{\boldsymbol{\ell}}_j^{(1)}, \ldots, \widehat{\boldsymbol{\ell}}_j^{(M)}\right]$, where $\widehat{\boldsymbol{\alpha}}^{(j)}$, $\widehat{\boldsymbol{\beta}}_k^{(j)}$, and $\widehat{\boldsymbol{\gamma}}_k^{(j)}$, correspond to the estimators of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_k$, and $\boldsymbol{\gamma}_k$, given by equation (3.22), and $\widehat{\boldsymbol{\ell}}_k^{(j)}$ corresponds to the estimator of $\boldsymbol{\ell}_k$ given in equation (3.19), in the jth Monte Carlo simulation, $j = 1, \ldots, M$.

2. We define the $M \times 1$ vectors $\widehat{\boldsymbol{m}}^{(M)}(t)$, $\widehat{\boldsymbol{g}}_{\ell,k}^{(M)}(t)$, $\ell = 1, 2$, $k = 1, \ldots, K$, $\widehat{\boldsymbol{\phi}}_k^{(M)}(u)$, $k = 1, \ldots, p$, and $\widehat{\boldsymbol{\mu}}^{(M)}(t)$ as

$$
\begin{aligned}
\widehat{\boldsymbol{m}}^{(M)}(t) &= \boldsymbol{B}(t)^\top \mathbf{A}_M = \left[\widehat{m}^{(1)}(t), \ldots, \widehat{m}^{(M)}(t)\right]^\top, \\
\widehat{\boldsymbol{g}}_{1,k}^{(M)}(t) &= \boldsymbol{B}(t)^\top \mathbf{B}_{kM} = \left[\widehat{g}_{1,k}^{(1)}(t), \ldots, \widehat{g}_{1,k}^{(M)}(t)\right]^\top, \qquad k = 1, \ldots, K, \\
\widehat{\boldsymbol{g}}_{2,k}^{(M)}(t) &= \boldsymbol{B}(t)^\top \mathbf{G}_{kM} = \left[\widehat{g}_{2,k}^{(1)}(t), \ldots, \widehat{g}_{2,k}^{(M)}(t)\right]^\top, \qquad k = 1, \ldots, K, \\
\widehat{\boldsymbol{\phi}}_k^{(M)}(u) &= \boldsymbol{B}(u)^\top \mathbf{P}_{kM} = \left[\widehat{\phi}_k^{(1)}(u), \ldots, \widehat{\phi}_k^{(M)}(u)\right]^\top, \qquad k = 1, \ldots, p, \\
\widehat{\boldsymbol{\mu}}^{(M)}(t) &= \widehat{\boldsymbol{m}}(t) + \sum_{k=1}^{K} \left\{\widehat{\boldsymbol{g}}_{1,k}(t) \cos(w_k t) + \widehat{\boldsymbol{g}}_{2,k}(t) \sin(w_k t)\right\} = \left[\widehat{\mu}^{(1)}(t), \ldots, \widehat{\mu}^{(M)}(t)\right]^\top,
\end{aligned}
$$

where $\widehat{\mu}^{(j)}(t)$, $\widehat{m}^{(j)}(t)$, and $\widehat{g}_{\ell,k}^{(j)}(t)$, $\ell = 1, 2$, $k = 1, \ldots, K$ correspond to the estimators of $\mu(t)$, $m(t)$, $g_{\ell,k}(t)$, given by equations (3.10), (3.11) and (3.12), and $\widehat{\phi}_k^{(j)}(u)$ corresponds to the estimator of $\phi_k(u)$ given in equation (3.21), in the jth

Monte Carlo simulation, and $\boldsymbol{B}(t)$ and $\boldsymbol{B}(u)$ correspond to the vector formed by the B-splines evaluated at time t and u, respectively.

3. We calculate the empirical quantiles of order 0.025 and 0.975 of the $M \times 1$ vectors $\widehat{\boldsymbol{m}}^{(M)}(t)$, $\widehat{\boldsymbol{g}}_{\ell,k}^{(M)}(t)$, $\ell = 1, 2$, $k = 1, \ldots, K$, $\widehat{\boldsymbol{\phi}}_{k}^{(M)}(u)$, $k = 1, \ldots, p$, and $\widehat{\boldsymbol{\mu}}^{(M)}(t)$.

Nonparametric confidence intervals in simulations of Section 4.4

For u and f fixed, confidence intervals for $\phi_j(u)$, $j = 1, \ldots, p$, $\theta_j(u)$, $j = 1, \ldots, q$, and $P_{X}(u, f)$, are calculated according to the following 3 steps:

1. We estimate the coefficients of interest $\boldsymbol{\alpha}_{\phi,k}$, $k = 1, \ldots, p$, and $\boldsymbol{\alpha}_{\theta,k}$, $k = 1, \ldots, q$, following the procedure described in Section 4.3.2, and we define the $J \times M$ matrices $\mathbf{A}_{kM,\phi} = \left[\widehat{\boldsymbol{\alpha}}_{\phi,k}^{(1)}, \ldots, \widehat{\boldsymbol{\alpha}}_{\phi,k}^{(M)}\right]$ and $\mathbf{A}_{kM,\theta} = \left[\widehat{\boldsymbol{\alpha}}_{\theta,k}^{(1)}, \ldots, \widehat{\boldsymbol{\alpha}}_{\theta,k}^{(M)}\right]$, where $\widehat{\boldsymbol{\alpha}}_{\phi,k}^{(j)}$ and $\widehat{\boldsymbol{\alpha}}_{\theta,k}^{(j)}$, correspond to the estimators of $\boldsymbol{\alpha}_{\phi,k}^{(j)}$ and $\boldsymbol{\alpha}_{\theta,k}^{(j)}$ given by equation (4.17) in the case of LSAR processes, and in equation (4.19) in the case of LSARMA processes, in the jth Monte Carlo simulation, $j = 1, \ldots, M$.

2. We define the $M \times 1$ vectors $\widehat{\boldsymbol{\phi}}_{k}^{(M)}(u)$, $k = 1, \ldots, p$, $\widehat{\boldsymbol{\theta}}_{k}^{(M)}(u)$, $k = 1, \ldots, q$, as

$$\widehat{\boldsymbol{\phi}}_{k}^{(M)}(u) = \boldsymbol{B}(u)^{\top}\mathbf{A}_{kM,\phi} = \left[\widehat{\phi}_{k}^{(1)}(u), \ldots, \widehat{\phi}_{k}^{(M)}(u)\right]^{\top}, \quad k = 1, \ldots, p,$$
$$\widehat{\boldsymbol{\theta}}_{k}^{(M)}(u) = \boldsymbol{B}(u)^{\top}\mathbf{A}_{kM,\theta} = \left[\widehat{\theta}_{k}^{(1)}(u), \ldots, \widehat{\theta}_{k}^{(M)}(u)\right]^{\top}, \quad k = 1, \ldots, q,$$

and the vector $\widehat{\boldsymbol{P}}^{(M)}(u, f) = \left[\widehat{P}^{(1)}(u, f), \ldots, \widehat{P}^{(M)}(u, f)\right]^{\top}$, where $\widehat{P}^{(j)}(u, f)$ is given by replacing $\phi_{k}^{(j)}(u)$, $k = 1, \ldots, p$, $\theta_{k}^{(j)}(u)$, $k = 1, \ldots, q$, and σ_z^2 in equation (4.11) with the estimates $\widehat{\phi}_{k}^{(j)}(u)$, $k = 1, \ldots, p$, $\widehat{\theta}_{k}^{(j)}(u)$, $k = 1, \ldots, q$, and $\widehat{\sigma}_z^2$, obtained in the jth Monte Carlo simulation, and $\boldsymbol{B}(u)$ corresponds to the vector formed by the B-splines evaluated at time u.

3. We calculate the empirical quantiles of order 0.025 and 0.975 of the $M \times 1$ vectors $\widehat{\boldsymbol{\phi}}_{k}^{(M)}(u)$, $k = 1, \ldots, p$, $\widehat{\boldsymbol{\theta}}_{k}^{(M)}(u)$, $k = 1, \ldots, q$, and $\widehat{\boldsymbol{P}}^{(M)}(u, f)$.

C Proofs

In this Appendix we prove the results in Lemma 1 and Proposition 1 (see Section 2.5), and Proposition 2 and proofs of equations (4.7) and (4.8) (see Section 4.2).

C.1 Proof of Lemma 1

Firstly, we prove equation (2.20) and then we prove equation (2.21). Multiplying (2.19) by $\exp(-i\lambda k\Delta)$ on both sides we get

$$\exp(-i\lambda k\Delta)P_\varepsilon(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} r_\varepsilon(h)\exp(i\lambda[h-k]\Delta),$$

and by summing both sides with respect to j, with $j = 1, \ldots, N_\mathcal{I}$

$$\sum_{j=1}^{N_\mathcal{I}} \exp(-i\lambda k\Delta)P_\varepsilon(\lambda) = \frac{1}{2\pi}\sum_{j=1}^{N_\mathcal{I}}\sum_{h=-\infty}^{\infty} r_\varepsilon(h)\exp(i\lambda[h-k]\Delta).$$

Recalling that the sequence $\{r_\varepsilon(h)\}$ is absolutely summable, we obtain

$$\sum_{j=1}^{N_\mathcal{I}} \exp(-i\lambda k\Delta)P_\varepsilon(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} r_\varepsilon(h)\sum_{j=1}^{N_\mathcal{I}}\exp(i\lambda[h-k]\Delta).$$

But, since

$$\sum_{j=1}^{N_\mathcal{I}} \exp(i2\pi j[h-k]\Delta/N_\mathcal{I}) = N_\mathcal{I}\delta_{\{h,k\}},$$

we have

$$\begin{aligned}\sum_{j=1}^{N_\mathcal{I}} \exp(-i\lambda k\Delta)P_\varepsilon(\lambda) &= \frac{N_\mathcal{I}}{2\pi}\sum_{h=-\infty}^{\infty} r_\varepsilon(h)\delta_{\{h,k\}},\\ &= \frac{N_\mathcal{I}}{2\pi}r_\varepsilon(k),\end{aligned}$$

and therefore we get the expression in equation (2.20)

$$r_\varepsilon(k) = \frac{2\pi}{N_\mathcal{I}}\sum_{j=1}^{N_\mathcal{I}} \exp(-i\lambda k\Delta)P_\varepsilon(\lambda).$$

We now prove equation (2.21). Notice that the expectation of the periodogram in equation (2.17) of the observations $\{\varepsilon_i,\ i \in \mathcal{I}\}$ is

$$\mathbb{E}\left[I_\varepsilon(\lambda)\right] = \sum_{k\in\mathcal{I}}\sum_{j\in\mathcal{I}} \mathbb{E}\left[\varepsilon_k \varepsilon_j\right] \exp(i\lambda[t_k - t_j]).$$

If we replace the expectation $\mathbb{E}\left[\varepsilon_k \varepsilon_j\right]$ with the right-hand-side of equation (2.20), we obtain

$$\begin{aligned}
\mathbb{E}\left[I_\varepsilon(\lambda)\right] &= \frac{2\pi}{N_\mathcal{I}} \sum_{j=1}^{N_\mathcal{I}} P_\varepsilon(\omega_j) \sum_{k\in\mathcal{I}}\sum_{j\in\mathcal{I}} \exp(i[\lambda - \omega_j][t_k - t_j]), \\
&= \frac{2\pi}{N_\mathcal{I}} \sum_{j=1}^{N_\mathcal{I}} P_\varepsilon(\omega_j) W_\varepsilon(\lambda - \omega_j), \\
&= \frac{2\pi}{N_\mathcal{I}} P_\varepsilon(\lambda) * W_\varepsilon(\lambda).
\end{aligned}$$

C.2 Proof of Proposition 1

Let $\mathcal{F}\{g_j\}[k]$ denote the Discrete Fourier Transform (DFT) of the sequence of m numbers $g_1, \ldots, g_m$ into another sequence $h_1, \ldots, h_m$, that is,

$$h_k = \mathcal{F}\{g_j\}[k] = \sum_{j=1}^{m} g_j \exp(-ikj2\pi/m), \quad k = 1, \ldots, m,$$

and $\mathcal{F}^{-1}\{h_k\}[j]$ denote the Inverse DFT of the sequence $h_1, \ldots, h_m$ into another sequence $g_1, \ldots, g_m$, that is,

$$g_j = \mathcal{F}^{-1}\{h_k\}[j] = \frac{1}{m}\sum_{k=1}^{m} h_k \exp(ikj2\pi/m), \quad j = 1, \ldots, m.$$

Let $\mathcal{F}\{g_j\}[k]$ and $\mathcal{F}\{\ell_j\}[k]$ be, respectively, the DFTs of the sequences $\{g_j\}$ and $\{\ell_j\}$ into the sequences $\{h_k\}$ and $\{m_k\}$. Then, the Convolution Theorem states that

$$\mathcal{F}\{g_j * \ell_j\}[k] = \mathcal{F}\{g_j\}[k]\mathcal{F}\{\ell_j\}[k]. \tag{C.1}$$

Applying the Convolution Theorem in equation (C.1) to equation (2.21), we obtain

$$\mathcal{F}\{P_\varepsilon(\lambda_j)\}[k] = \frac{N_\mathcal{I}}{2\pi} \frac{\mathcal{F}\{\mathbb{E}\left[I_\varepsilon(\lambda_j)\right]\}[k]}{\mathcal{F}\{W_\varepsilon(\lambda_j)\}[k]}.$$

The Inverse DFT of the last equation gives the result in equation (2.25).

C.3 Proof of Proposition 2

Let $\{x_i\}$ be an ARMA process satisfying the causality condition

$$1 - \sum_{j-1}^{p} \phi_j r^j = 0 \text{ for all } |r| > 1,$$

with autocovariance function $r_x(\cdot)$. The PSD of $\{x_i\}$ is defined by

$$P_x(f) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \exp(-i2\pi f \Delta h) r_x(h), \quad -\frac{1}{2\Delta} < f < \frac{1}{2\Delta}, \tag{C.2}$$

For causal ARMA(p, q) processes, Brockwell (2001) proved that when $q < p$ and the roots s_j of equation (4.6) are distinct, the autocovariance function can be express as

$$r_x(h) = -\sigma_z^2 \sum_{j=1}^{p} \frac{s_j^{|h|+1} \left(\sum_{k=0}^{q} \theta_k s_j^k\right) \left(\sum_{k=0}^{q} \theta_k s_j^{-k}\right)}{\left(-\sum_{k=0}^{p} \phi_k s_j^k\right) \left(-\sum_{k=0}^{p} k\phi_k s_j^{-k+1}\right)}, \tag{C.3}$$

with $\phi_0 = -1$ and $\theta_0 = 1$. Substituting equation (C.3) into equation (C.2), the PSD is given by

$$P_x(f) = -\frac{\sigma_z^2}{2\pi} \sum_{j=1}^{p} \frac{s_j \left(\sum_{k=0}^{q} \theta_k s_j^k\right) \left(\sum_{k=0}^{q} \theta_k s_j^{-k}\right)}{\left(-\sum_{k=0}^{p} \phi_k s_j^k\right) \left(-\sum_{k=0}^{p} k\phi_k s_j^{-k+1}\right)} \sum_{h=-\infty}^{\infty} \exp(-i2\pi f h \Delta) s_j^{|h|}. \tag{C.4}$$

Developing the term $\sum_{h=-\infty}^{\infty} \exp(-i2\pi f h\Delta) s_j^{|h|}$ in equation (C.4), we have

$$\begin{aligned}
\sum_{h=-\infty}^{\infty} \exp(-i2\pi f h \Delta) s_j^{|h|} &= \sum_{h=-\infty}^{0} \exp(-i2\pi f h \Delta) s_j^{-h} + \sum_{h=0}^{\infty} \exp(-i2\pi f h \Delta) s_j^{h} - 1 \\
&= \sum_{h=0}^{\infty} \left[\exp(i2\pi f \Delta) s_j\right]^h + \sum_{h=0}^{\infty} \left[\exp(-i2\pi f \Delta) s_j\right]^h - 1,
\end{aligned}$$

given that the root s_j satisfies $|s_j| < 1$, we use the geometric series formula obtaining

$$\begin{aligned}
\sum_{h=-\infty}^{\infty} \exp(-i2\pi f h \Delta) s_j^{|h|} &= \frac{1}{1 - \exp(i2\pi f \Delta) s_j} + \frac{1}{1 - \exp(-i2\pi f \Delta) s_j} - 1 \\
&= \frac{1 - s_j^2}{1 - 2s_j \cos(2\pi f \Delta) + s_j^2}.
\end{aligned} \tag{C.5}$$

Replacing the expression $\sum_{h=-\infty}^{\infty} \exp(-i2\pi f h\Delta) s_j^{|h|}$ in equation (C.4) with the expression found in equation (C.5), we get the expression in equation (4.3), which is

$$P_x(f) = -\frac{\sigma_z^2}{2\pi} \sum_{j=1}^{p} \left\{ \frac{s_j \left(\sum_{k=0}^{q} \theta_k s_j^k \right) \left(\sum_{k=0}^{q} \theta_k s_j^{-k} \right)}{\left(-\sum_{k=0}^{p} \phi_k s_j^k \right) \left(-\sum_{k=0}^{p} k\phi_k s_j^{-k+1} \right)} \frac{1 - s_j^2}{1 - 2s_j \cos(2\pi f\Delta) + s_j^2} \right\}.$$

C.4 Proof of Equation (4.7)

Replacing s_j with $a_j + ib_j$ in equation (4.4) we get

$$P_{x,j}(f) = -\frac{\sigma_z^2}{2\pi} \text{Re} \left[\frac{\text{Re}(Q_\ell) + i\text{Im}(Q_\ell)}{1 + a_j^2 - b_j^2 - 2a_j \cos(2\pi f\Delta) + i2b_j[a_j - \cos(2\pi f\Delta)]} \right].$$

Multiplying the numerator and denominator by the conjugate of $1+a_j^2-b_j^2-2a_j \cos(2\pi f\Delta)+ i2b_j[a_j - \cos(2\pi f\Delta)]$, and taking the real part of the complex number gotten, we have the result in equation (4.7).

C.5 Proof of Equation (4.8)

Differentiating the component $P_{x,j}(f)$ in equation (4.7) with respect to f, setting the derivative equal zero, and replacing f with $f_{\max,\ell}$, we get the equation

$$K_{3,\ell} + K_{2,\ell} \cos(2\pi f_{\max,\ell}\Delta) + K_{1,\ell} \cos^2(2\pi f_{\max,\ell}\Delta) = 0, \tag{C.6}$$

with $K_{j,\ell}$, $j = 1, 2, 3$, defined in equation (4.9). The expression above corresponds to a quadratic equation in $\cos(2\pi f_{\max,\ell}\Delta)$. Thus, the solution of equation (C.6) is

$$\cos(2\pi f_{\max,\ell}\Delta) = \left(\frac{-K_{2,\ell} \pm \sqrt{K_{2,\ell}^2 - 4K_{1,\ell}K_{3,\ell}}}{2K_{1,\ell}} \right).$$

Assuming that $K_{2,\ell}^2 - 4K_{1,\ell}K_{3,\ell} \geq 0$, and $\left| \left(-K_{2,\ell} \pm \sqrt{K_{2,\ell}^2 - 4K_{1,\ell}K_{3,\ell}} \right) / (2K_{1,\ell}) \right| < 1$, $\ell = j, k$, the frequency $f_{\max,\ell}$ is given by

$$f_{\max,\ell} = \frac{1}{2\pi\Delta} \arccos \left(\frac{-K_{2,\ell} \pm \sqrt{K_{2,\ell}^2 - 4K_{1,\ell}K_{3,\ell}}}{2K_{1,\ell}} \right),$$

which agrees with the result in equation (4.8).

D Covariance matrix of stationary autoregressive processes

Let $\{\varepsilon_i\}$ be a causal AR(p) process occurring at certain discrete time $t_i = t_0 + i\Delta$, with i an integer, $\Delta > 0$ and $t_0 = 0$, given by

$$\varepsilon_i = \sum_{j=1}^{p} \phi_j \varepsilon_{i-j} + z_i, \qquad \{z_i\} \sim WN(0, \sigma_z^2).$$

Then the spectral density is

$$P_\varepsilon(f) = \frac{\sigma_z^2}{2\pi} \frac{1}{\left| 1 - \sum_{j=1}^{p} \phi_j \exp(-i2\pi f \Delta j) \right|^2}, \quad -\frac{1}{2\Delta} < f < \frac{1}{2\Delta},$$

the elements of covariance matrix $\mathbf{\Gamma}_\varepsilon = \mathbb{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top\right]$ are

$$\gamma_{\varepsilon,j,k} = 2\pi\Delta \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} \exp(i2\pi f[j-k]\Delta) P_\varepsilon(f) df,$$

and the element of the inverse covariance matrix $\mathbf{\Gamma}_\varepsilon^{-1}$ are

$$\eta_{\varepsilon,j,k} = \frac{\Delta}{2\pi} \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} \exp(i2\pi f[j-k]\Delta) P_\varepsilon^{-1}(f) df.$$

The following examples show the elements of the inverse covariance matrix of AR(p) processes with $p = 1, \ldots, 4$. For an AR(1) process the elements are

$$\eta_{\varepsilon,i,k} = \sigma_z^{-2} \times \begin{cases} 1 + \phi_1^2, & \text{if } i = k, \\ -\phi_1, & \text{if } i = k \pm 1, \\ 0, & \text{otherwise,} \end{cases}$$

for an AR(2) process

$$\eta_{\varepsilon,i,k} = \sigma_z^{-2} \times \begin{cases} 1 + \phi_1^2 + \phi_2^2, & \text{if } i = k, \\ \phi_1\phi_2 - \phi_1, & \text{if } i = k \pm 1, \\ -\phi_2, & \text{if } i = k \pm 2, \\ 0, & \text{otherwise,} \end{cases}$$

for an AR(3) process

$$\eta_{\varepsilon,i,k} = \sigma_z^{-2} \times \begin{cases} 1 + \phi_1^2 + \phi_2^2 + \phi_3^2, & \text{if } i = k, \\ \phi_1\phi_2 + \phi_2\phi_3 - \phi_1, & \text{if } i = k \pm 1, \\ \phi_1\phi_3 - \phi_2, & \text{if } i = k \pm 2, \\ -\phi_3, & \text{if } i = k \pm 3, \\ 0, & \text{otherwise,} \end{cases}$$

and for an AR(4) process

$$\eta_{\varepsilon,i,k} = \sigma_z^{-2} \times \begin{cases} 1 + \phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2, & \text{if } i = k, \\ \phi_1\phi_2 + \phi_2\phi_3 + \phi_3\phi_4 - \phi_1, & \text{if } i = k \pm 1, \\ \phi_1\phi_3 + \phi_2\phi_4 - \phi_2, & \text{if } i = k \pm 2, \\ \phi_1\phi_4 - \phi_3, & \text{if } i = k \pm 3, \\ -\phi_4, & \text{if } i = k \pm 4, \\ 0, & \text{otherwise.} \end{cases}$$

E Covariance matrix of locally stationary autoregressive processes

Let $\{\varepsilon_{i,N}\}$ be a LSAR(p) process, satisfying the definition given in Section 1.4, and occurring at certain discrete time $t_i = t_0 + i\Delta$, with i an integer, $\Delta > 0$ and $t_0 = 0$, given

by

$$\varepsilon_{i,N} = \sum_{j=1}^{p} \phi_j\left(\tfrac{i}{N}\right)\varepsilon_{i-j,N} + z_i, \qquad \{z_i\} \sim WN(0, \sigma_z^2).$$

Then the spectral density is

$$P_{\varepsilon}(u,f) = \frac{\sigma_z^2}{2\pi}\frac{1}{\left|1 - \sum_{j=1}^{p}\phi_j(u)\exp(-i2\pi f\Delta j)\right|^2}, \quad -\frac{1}{2\Delta} < f < \frac{1}{2\Delta}.$$

For a LSAR(p) processes, Dahlhaus (1996) defined the elements of the covariance matrix $\mathbf{\Gamma}_{\varepsilon} = \mathbb{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}\right]$ as

$$\gamma_{\varepsilon,j,k} = 2\pi\Delta\int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} \exp(i2\pi f[j-k]\Delta)A^0_{j,N}(f)\overline{A^0_{k,N}(f)}df,$$

where $A^0_{j,N}(f) = A\left(\frac{j}{N}, f\right)$ is the transfer function defined in the interval $[-1/(2\Delta), 1/(2\Delta)]$ and given by

$$A\left(\tfrac{j}{N}, f\right) = \frac{\sigma_z}{\sqrt{2\pi}}\frac{1}{1 - \sum_{k=1}^{p}\phi_k\left(\frac{j}{N}\right)\exp(-i2\pi f\Delta k)}.$$

Satisfying the assumptions about A and P_{ε}^{-1} in Proposition 2.4 in Dahlhaus (2000), the element of the inverse covariance matrix $\mathbf{\Gamma}_{\varepsilon}^{-1}$ are approximated by

$$\eta_{\varepsilon,j,k} = \frac{\Delta}{2\pi}\int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} \exp(i2\pi f[j-k]\Delta)P_{\varepsilon}^{-1}\left(\tfrac{1}{N}\left\lfloor\tfrac{j+k}{2}\right\rfloor, f\right)df$$

The following examples show the elements of the inverse covariance matrix of LSAR(p) processes with $p = 1, \ldots, 4$. For a LSAR(1) process the elements are

$$\eta_{\varepsilon,i,k} = \sigma_z^{-2} \times \begin{cases} 1 + \phi_1^2\left(\frac{i}{N}\right), & \text{if } i = k, \\ -\phi_1\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right), & \text{if } i = k \pm 1, \\ 0, & \text{otherwise,} \end{cases}$$

for a LSAR(2) process

$$\eta_{\epsilon,i,k} = \sigma_z^{-2} \times \begin{cases} 1 + \phi_1^2\left(\frac{i}{N}\right) + \phi_2^2\left(\frac{i}{N}\right), & \text{if } i = k, \\ \phi_1\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right)\phi_2\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right) - \phi_1\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right), & \text{if } i = k \pm 1, \\ -\phi_2\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right), & \text{if } i = k \pm 2, \\ 0, & \text{otherwise,} \end{cases}$$

for a LSAR(3) process

$$\eta_{\epsilon,i,k} = \sigma_z^{-2} \times \begin{cases} 1 + \phi_1^2\left(\frac{i}{N}\right) + \phi_2^2\left(\frac{i}{N}\right) + \phi_3^2\left(\frac{i}{N}\right), & \text{if } i = k, \\ \phi_1\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right)\phi_2\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right) + \phi_2\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right)\phi_3\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right) - \phi_1\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right), & \text{if } i = k \pm 1, \\ \phi_1\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right)\phi_3\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right) - \phi_2\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right), & \text{if } i = k \pm 2, \\ -\phi_3\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right), & \text{if } i = k \pm 3, \\ 0, & \text{otherwise,} \end{cases}$$

and for a LSAR(4) process

$$\eta_{\epsilon,i,k} = \sigma_z^{-2} \times \begin{cases} 1 + \phi_1^2\left(\frac{i}{N}\right) + \phi_2^2\left(\frac{i}{N}\right) + \phi_3^2\left(\frac{i}{N}\right) + \phi_4^2\left(\frac{i}{N}\right), & \text{if } i = k, \\ \sum_{j=1}^{4}\left\{\phi_j\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right)\phi_{j+1}\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right)\right\} - \phi_1\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right), & \text{if } i = k \pm 1, \\ \phi_1\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right)\phi_3\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right) + \phi_2\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right)\phi_4\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right) - \phi_2\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right), & \text{if } i = k \pm 2, \\ \phi_1\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right)\phi_4\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right) - \phi_3\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right), & \text{if } i = k \pm 3, \\ -\phi_4\left(\frac{1}{N}\left\lfloor\frac{i+k}{2}\right\rfloor\right), & \text{if } i = k \pm 4, \\ 0, & \text{otherwise.} \end{cases}$$

F Wrapped Lorentzian function

The Wrapped Lorentzian function is defined by

$$W(\omega) = \frac{N}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\omega-\omega_c)}, \quad -\pi \leq \omega < \pi,$$

where $\omega_c \in [-\pi, \pi]$ is the centroid, $\rho \in [0, 1)$ controls the concentration of the function, and N is the normalization factor. See Kato and Jones (2013) for more details.

G Infinite-order locally stationary moving average process

In this appendix, we clarify some aspects of step ii) of the method to construct the confidence intervals for LSARMA process proposed in Section 4.3.3.

Let $\{x_{_{i,N}}\}$ be the LSARMA process defined in equation (4.10) satisfying

$$x_{_{i,N}} = \sum_{j=1}^{p} \phi_j \left(\frac{i}{N}\right) x_{_{i-j,N}} + z_i + \sum_{j=1}^{q} \theta_j \left(\frac{i}{N}\right) z_{_{i-j}}, \quad \{z_i\} \sim WN(0, \sigma_z^2), \quad i = 1, \ldots, N,$$

and suppose that the following power series expansion exists

$$\left[\theta\left(\tfrac{i}{N}, v\right)\right]^{-1} = \sum_{j=0}^{\infty} \rho_j(i, N) v^j. \tag{G.1}$$

Then, by equations (4.10) and (G.1) we may express the $x_{_{i,N}}$'s as a infinite order LSAR process as

$$\begin{aligned}
z_{_i} &= \sum_{j=0}^{\infty} \rho_j(i,N) \left[- \sum_{k=0}^{p} \widehat{\phi}_k \left(\tfrac{i}{N}\right) x_{_{i-j-k,N}} \right] \\
&= \sum_{j=1}^{i} \rho_{j-1}(i,N) \left[- \sum_{k=0}^{p} \widehat{\phi}_k \left(\tfrac{i}{N}\right) x_{_{i+1-j-k,N}} \right] + \sum_{j=i+1}^{\infty} \rho_j(i-1,N) \left[- \sum_{k=0}^{p} \widehat{\phi}_k \left(\tfrac{i}{N}\right) x_{_{i+1-j-k,N}} \right] \\
&= \sum_{j=1}^{i} \rho_{j-1}(i,N) \left[- \sum_{k=0}^{p} \widehat{\phi}_k \left(\tfrac{i}{N}\right) x_{_{i+1-j-k,N}} \right] + \sum_{j=i+1}^{\infty} \rho_{j-1}(i,N) \left[\sum_{k=0}^{q} \widehat{\theta}_k \left(\tfrac{i}{N}\right) z_{_{i+1-j-k}} \right] \\
&= \sum_{j=1}^{i} \rho_{j-1}(i,N) \left[- \sum_{k=0}^{p} \widehat{\phi}_k \left(\tfrac{i}{N}\right) x_{_{i+1-j-k,N}} \right] + \sum_{k=0}^{q} \sum_{s=k}^{\infty} \widehat{\theta}_k \left(\tfrac{i}{N}\right) \rho_{i+s-k}(i,N) z_{_{-s}}.
\end{aligned}$$

Note that, for a fixed i, $\rho_j(i,N) \to 0$ as $j \to \infty$. Therefore, we can concentrate on the first term only and define an approximation to $z_{_i}$ by

$$z_{_i} = \sum_{j=1}^{i} \rho_{j-1}(i,N) \left[- \sum_{k=0}^{p} \widehat{\phi}_k \left(\tfrac{i}{N}\right) x_{_{i+1-j-k,N}} \right].$$

For a fixed i, the coefficients $\rho_j(i,N)$, $j = 0, 1, \ldots, i-1$, in equation (G.1), satisfy the recursion $\rho_{_j}(i,N) = -\sum_{k=1}^{j} \theta_{_k}\left(\tfrac{i}{N}\right) \rho_{_{j-k}}(i,N)$, with $\rho_{_0}(i,N) = 1$.

H Values of A_j, $j = 1, \ldots, 7$

The values of A_j, $j = 1, \ldots, 7$, in equation (4.25) are

$$\begin{aligned}
A_0 &= b_7 b_6 - b_9 b_5 \\
A_1 &= b_7 b_5 + b_8 b_6 - 2b_9 b_4 - b_{10} b_5 \\
A_2 &= b_7 b_4 + b_8 b_5 - 3b_9 b_3 - 2b_{10} b_4 - b_{11} b_5 + 24\theta_3 h \\
A_3 &= b_7 b_3 + b_8 b_4 - 4b_9 b_2 - 3b_{10} b_3 - 2b_{11} b_4 + 16\theta_3 b_5 \\
A_4 &= b_7 b_2 + b_8 b_3 - 5b_9 b_1 - 4b_{10} b_2 - 3b_{11} b_3 + 8\theta_3 b_4 \\
A_5 &= b_7 b_1 + b_8 b_2 - 5b_{10} b_1 - 4b_{11} b_2 \\
A_6 &= b_8 b_1 - 5b_{11} b_1 - 8\theta_3 b_2 \\
A_7 &= 16\theta_3 b_1
\end{aligned} \tag{H.1}$$

where b_j, $= 1, \ldots, 11$, are

$$\begin{aligned}
b_1 &= -32\phi_5 \\
b_2 &= 16(-\phi_4 + \phi_1\phi_5) \\
b_3 &= 8(-\phi_3 + \phi_1\phi_4 + \phi_2\phi_5) + 40\phi_5 \\
b_4 &= 4(-\phi_2 + \phi_1\phi_3 + \phi_2\phi_4 + \phi_3\phi_5) - 16(-\phi_4 + \phi_1\phi_5) \\
b_5 &= 2(-\phi_1 + \phi_1\phi_2 + \phi_2\phi_3 + \phi_3\phi_4 + \phi_4\phi_5) - 6(-\phi_3 + \phi_1\phi_4 + \phi_2\phi_5) - 10\phi_5 \\
b_6 &= 1 + \phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 + \phi_5^2 - 2(-\phi_2 + \phi_1\phi_3 + \phi_2\phi_4 + \phi_3\phi_5) + 2(-\phi_4 + \phi_1\phi_5) \\
b_7 &= 2(\theta_1 - 3\theta_3 + \theta_1\theta_2 + \theta_2\theta_3) \\
b_8 &= 8(\theta_2 + \theta_1\theta_3) \\
b_9 &= 1 + \theta_1^2 + \theta_2^2 + \theta_3^3 - 2\theta_2 - 2\theta_1\theta_3 \\
b_{10} &= 2(\theta_1 - 3\theta_3 + \theta_1\theta_2 + \theta_2\theta_3) \\
b_{11} &= 4(\theta_2 + \theta_1\theta_3).
\end{aligned}$$

Bibliography

Aue, A. and L. Horváth (2013). Structural breaks in time series. *Journal of time series Analysis 34*(1), 1–16.

Axelsson, M., Borgonovo, L., and Larsson, S. (2005). Evolution of the 0.01-25 hz power spectral components in cygnus x-1. *A&A 438*(3), 999–1012.

Benkő, J. M. (2018, January). On the connection between almost periodic functions and Blazhko light curves. *MNRAS 473*(1), 412–418.

Benkő, J. M., E. Plachy, R. Szabó, L. Molnár, and Z. Kolláth (2014). Long-timescale Behavior of the Blazhko Effect from Rectified Kepler Data. *ApJS 213*(2), 31.

Benkő, J. M., R. Szabó, and M. Paparó (2011, October). Blazhko RR Lyrae light curves as modulated signals. *MNRAS 417*(2), 974–991.

Blažko, S. (1907, August). Mitteilung über veränderliche Sterne. *Astronomische Nachrichten 175*, 325.

Box, G. and G. Jenkins (1970). *Time series analysis: forecasting and control*. Holden-Day series in time series analysis. Holden-Day.

Brockwell, P. and R. Davis (2016). *Introduction to Time Series and Forecasting* (2nd ed.). Springer Texts in Statistics. Springer International Publishing.

Brockwell, P. J. (2001). *Handbook of Statistics: Stochastic Processes, Theory and Methods*, Chapter Continuous-time ARMA Processes, pp. 249–276. Elsevier.

Buchler, J. R. and Z. Kolláth (2011, April). On the Blazhko Effect in RR Lyrae Stars. *ApJ 731*(1), 24.

Catelan, M. and H. A. Smith (2015). *Pulsating Stars*. Wiley.

Chattopadhyay, A. K. (2017). *Incomplete Data in Astrostatistics*, pp. 1–12. American Cancer Society.

Cochrane, D. and G. H. Orcutt (1949). Application of least squares regression to relationships containing auto- correlated error terms. *Journal of the American Statistical Association 44*(245), 32–61.

Dahlhaus, R. (1996). On the kullback-leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications 62*, 139–168.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics 25*, 1–37.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics 28*(6), 1762–1794.

Dahlhaus, R. and M. Neumann (2001). Locally adaptive fitting of semiparametric models to nonstationary time series. *Stochastic Processes and their Applications 91*, 277–308.

Dahlhaus, R., M. H. Neumann, and R. V. Sachs (1999). Nonlinear wavelet estimation of time-varying autoregressive processes. *Bernoulli 5*(5), 873 – 906.

de Boor, C. (1978). *A practical guide to splines*. Springer.

Deeming, T. J. (1975, August). Fourier Analysis with Unequally-Spaced Data. *Ap&SS 36*(1), 137–158.

Efron, B. and R. Tibshirani (1994). *An Introduction to the Bootstrap*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Eilers, P. H. C., J. Gampe, B. D. Marx, and R. Rau (2008). Modulation models for seasonal time series and incidence tables. *Statistics in Medicine 27*(17), 3430–3441.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical Science 11*(2), 89–102.

Elzhov, T. V., K. M. Mullen, A.-N. Spiess, and B. Bolker (2016). *minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds*. R Foundation for Statistical Computing. R package version 1.2-1.

Esin, A. A., J. E. McClintock, and R. Narayan (1997, November). Advection-dominated accretion and the spectral states of black hole x-ray binaries: Application to nova muscae 1991. *ApJ 489*(2), 865–889.

Fan, J. and Q. Yao (2003). *Nonlinear time series : nonparametric and parametric methods*. Springer series in statistics. New York (N.Y.) : Springer.

Feigelson, E. D., G. J. Babu, and G. A. Caceres (2018). Autoregressive times series methods for time domain astronomy. *Frontiers in Physics 6*(80), 1–13.

Feigelson, E. D., R. S. de Souza, E. E. Ishida, and G. J. Babu (2021). Twenty-first-century statistical and computational challenges in astrophysics. *Annual Review of Statistics and Its Application 8*(1), 493–517.

Gama, J. (2016). *NISTunits: Fundamental Physical Constants and Unit Conversions from NIST*. R Foundation for Statistical Computing. R package version 1.0.1.

Gillet, D. (2013, June). Atmospheric dynamics in RR Lyrae stars. The Blazhko effect. *A&A 554*, A46.

Gillet, D., B. Mauclaire, T. Lemoult, P. Mathias, J. S. Devaux, T. de France, and T. Garrel (2019, March). Dynamical structure of the pulsating atmosphere of RR Lyrae. I. A typical pulsation cycle. *A&A 623*, A109.

Graham, M. J., S. G. Djorgovski, D. Stern, A. J. Drake, A. A. Mahabal, C. Donalek, E. Glikman, S. Larson, and E. Christensen (2015, October). A systematic search for close supermassive black hole binaries in the Catalina Real-time Transient Survey. *MNRAS 453*(2), 1562–1576.

Grenier, Y. (1983). Time-dependent arma modeling of nonstationary signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing 31*(4), 899–911.

Guggenberger, E., K. Kolenberg, J. M. Nemec, R. Smolec, J. M. Benkő, C.-C. Ngeow, J. G. Cohen, B. Sesar, R. Szabó, M. Catelan, P. Moskalik, K. Kinemuchi, S. E. Seader, J. C. Smith, P. Tenenbaum, and H. Kjeldsen (2012, July). The complex case of V445 Lyr observed with Kepler: two Blazhko modulations, a non-radial mode, possible triple mode RR Lyrae pulsation, and more. *MNRAS 424*(1), 649–665.

Guo, H., J. Wang, Z. Cai, and M. Sun (2017, October). How Far Is Quasar UV/Optical Variability from a Damped Random Walk at Low Frequency? *ApJ 847*(2), 132.

Hannan, E. J. and J. Rissanen (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika 69*(1), 81–94.

Hastie, T., R. Tibshirani, J. Friedman, and J. Franklin (2004, November). The elements of statistical learning: Data mining, inference, and prediction. *Math. Intell. 27*, 83–85.

Homan, J. and T. Belloni (2005). *The evolution of black hole states*, pp. 107–117. Dordrecht: Springer Netherlands.

Kalemci, E., J. A. Tomsick, R. E. Rothschild, K. Pottschmidt, S. Corbel, R. Wijnands, J. M. Miller, and P. Kaaret (2003, March). X-ray temporal properties of XTE j1650-500 during outburst decay. *ApJ 586*(1), 419–426.

Kato, S. and M. Jones (2013, February). An extended family of circular distributions related to wrapped cauchy distributions via brownian motion. *Bernoulli 19*, 154–171.

Kelly, B. C., A. C. Becker, M. Sobolewska, A. Siemiginowska, and P. Uttley (2014, June). Flexible and Scalable Methods for Quantifying Stochastic Variability in the Era of Massive Time-domain Astronomical Data Sets. *ApJ 788*(1), 33.

Koch, D. G., W. J. Borucki, G. Basri, N. M. Batalha, T. M. Brown, D. Caldwell, J. Christensen-Dalsgaard, W. D. Cochran, E. DeVore, E. W. Dunham, I. Gautier, Thomas N., J. C. Geary, R. L. Gilliland, A. Gould, J. Jenkins, Y. Kondo, D. W. Latham, J. J. Lissauer, G. Marcy, D. Monet, D. Sasselov, A. Boss, D. Brownlee, J. Caldwell, A. K. Dupree, S. B. Howell, H. Kjeldsen, S. Meibom, D. Morrison, T. Owen, H. Reitsema, J. Tarter, S. T. Bryson, J. L. Dotson, P. Gazis, M. R. Haas, J. Kolodziejczak, J. F. Rowe, J. E. Van Cleve, C. Allen, H. Chandrasekaran, B. D. Clarke, J. Li, E. V. Quintana, P. Tenenbaum, J. D. Twicken, and H. Wu (2010, April). Kepler Mission Design, Realized Photometric Performance, and Early Science. *ApJL 713*(2), L79–L86.

Kreiss, J.-P. and J. Franke (1992). Bootstrapping stationary autoregressive moving-average models. *Journal of Time Series Analysis 13*(4), 297–317.

Lewin, W. H. G., J. van Paradijs, and E. P. J. van den Heuvel (1997). *X-ray Binaries*. Cambridge Astrophysics Series.

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science 39*, 447–462.

Motta, G., D. Soto, and M. Catelan (2022, January). Periodic variable stars modulated by time-varying parameters. *ApJ 925*(1), 73.

Netzel, H., R. Smolec, I. Soszyński, and A. Udalski (2018, October). Blazhko effect in the first overtone RR Lyrae stars of the OGLE Galactic bulge collection. *MNRAS 480*(1), 1229–1246.

Plachy, E., L. Molnár, A. Bódi, M. Skarka, P. Szabó, R. Szabó, P. Klagyivik, Á. Sódor, and B. J. S. Pope (2019, October). Extended Aperture Photometry of K2 RR Lyrae stars. *ApJS 244*(2), 32.

Pottschmidt, K., J. Wilms, M. Nowak, G. Pooley, T. Gleissner, W. Heindl, D. Smith, R. Remillard, and R. Staubert (2003, September). Long term variability of cygnus x-1. i. x-ray spectral-temporal correlations in the hard state. *A&A 407*, 1039–1058.

Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rao, T. S. (1970). The fitting of non-stationary time-series models with time-dependent parameters. *Journal of the Royal Statistical Society. Series B (Methodological) 32*(2), 312–322.

Richards, J. W., D. L. Starr, N. R. Butler, J. S. Bloom, J. M. Brewer, A. Crellin-Quick, J. Higgins, R. Kennedy, and M. Rischard (2011, May). On Machine-learned Classification of Variable Stars with Sparse and Noisy Time-series Data. *ApJ 733*(1), 10.

Smith, H. A. (2004). *RR Lyrae Stars*. Cambridge Astrophysics Series.

Stothers, R. B. (2006, November). A New Explanation of the Blazhko Effect in RR Lyrae Stars. *ApJ 652*(1), 643–649.

Uttley, P., E. M. Cackett, A. C. Fabian, E. Kara, and D. R. Wilkins (2014, August). X-ray reverberation around accreting black holes. *A&A Rev 22*, 72.

Uttley, P., I. M. McHardy, and S. Vaughan (2005, May). Non-linear X-ray variability in X-ray binaries and active galaxies. *MNRAS 359*(1), 345–362.

Wong, R. K. W., V. L. Kashyap, T. C. M. Lee, and D. A. van Dyk (2016). Detecting abrupt changes in the spectra of high-energy astrophysical sources. *The Annals of Applied Statistics 10*(2), 1107–1134.

Xu, C., H. M. Günther, V. L. Kashyap, T. C. M. Lee, and A. Zezas (2021, April). Change-point Detection and Image Segmentation for Time Series of Astrophysical Images. *AJ 161*(4), 184.