Applied Condition Monitoring 3

Fakher Chaari · Jacek Leskow
Antonio Napolitano · Radoslaw Zimroz
Agnieszka Wylomanska · Anna Dudek
Editors

Cyclostationarity: Theory and Methods - II

Contributions to the 7th Workshop on Cyclostationary Systems and Their Applications, Grodek, Poland, 2014

Springer

Applied Condition Monitoring

Volume 3

About this Series

The book series Applied Condition Monitoring publishes the latest research and developments in the field of condition monitoring, with a special focus on industrial applications. It covers both theoretical and experimental approaches, as well as a range of monitoring conditioning techniques and new trends and challenges in the field. Topics of interest include, but are not limited to: vibration measurement and analysis; infrared thermography; oil analysis and tribology; acoustic emissions and ultrasonics; and motor current analysis. Books published in the series deal with root cause analysis, failure and degradation scenarios, proactive and predictive techniques, and many other aspects related to condition monitoring. Applications concern different industrial sectors: automotive engineering, power engineering, civil engineering, geoengineering, bioengineering, etc. The series publishes monographs, edited books, and selected conference proceedings, as well as textbooks for advanced students.

More information about this series at http://www.springer.com/series/13418

Fakher Chaari · Jacek Leskow
Antonio Napolitano · Radoslaw Zimroz
Agnieszka Wylomanska · Anna Dudek
Editors

Cyclostationarity: Theory and Methods - II

Contributions to the 7th Workshop on Cyclostationary Systems and Their Applications, Grodek, Poland, 2014

Springer

Editors
Fakher Chaari
National School of Engineers of Sfax
Sfax
Tunisia

Jacek Leskow
Institute of Mathematics
Cracow University of Technology
Kraków
Poland

Antonio Napolitano
Department of Engineering
University of Napoli "Parthenope"
Napoli
Italy

Radoslaw Zimroz
Institute of Mining Engineering
Wrocław University of Technology
Wrocław
Poland

Agnieszka Wylomanska
Department of Mathematics
Wrocław University of Technology
Wrocław
Poland

Anna Dudek
Department of Applied Mathematics
University of Science and Technology
Kraków
Poland

ISSN 2363-698X ISSN 2363-6998 (electronic)
Applied Condition Monitoring
ISBN 978-3-319-16329-1 ISBN 978-3-319-16330-7 (eBook)
DOI 10.1007/978-3-319-16330-7

Library of Congress Control Number: 2014930090

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Contents

Introduction

This book is designed for students, researchers and practitioners interested in the analysis of real datasets with nonstationary behaviour. The mentioned nonstationarity can be expressed in different ways. One of the properties of the process which is strongly related to nonstationarity is the cyclostationarity. In the past decades cyclostationary processes have become very important from a practical point of view and their theoretical analysis has become a challenge for many research groups. For 7 years researchers from different centres are meeting in Gródek nad Dunajcem to analyse various aspects of practical applications and theoretical properties of nonstationary processes, especially with characteristics adequate to cyclostationary systems.

The 7th Workshop on Cyclostationary Systems and Their Applications was organized during 9–12 February 2014 in Grodek nad Dunajcem (Poland). As each year, it gathered researches from Europe (France, Italy, Poland, Ukraine) and the USA. The main idea of this meeting was to bring together pure and applied researchers from different disciplines and start discussion on the possible cooperation in the field of the analysis and applications of cyclostationary systems. The main focus of the event was on the analysis of nonstationary signals with special importance to cyclostationary systems and their possible applications in various areas. The presentations covered recent theoretical developments for such systems and their possible applications.

Periodically correlated (or cyclostationary) processes are random systems that exhibit periodic behaviour but still they remain random. Therefore, they are mixed of randomness and periodicity and are likely to have periodic or almost periodic structure. Over the last 60 years, thousands of papers in this field have been published and the class of processes which exhibits behaviour adequate to this phenomenon is very rich and is growing rapidly. The growing interest of researchers interested in this field is caused by a variety of different applications in many areas like mechanics, vibroacoustics, economics, medicine … . Thus, understanding the needs of practitioners and simultaneously present them the new theoretical results is the aim of the meetings devoted to cyclostationary systems.

The chapters of this book can be divided into two groups. In the first group we can find various papers devoted to theoretical properties of cyclostationary models and processes. Most of these papers are related to problems of estimation of systems which exhibit cyclostationary properties. The reader can find the description of the procedure of fitting proper PARMA (periodic ARMA) model, main time series used to description of cyclostationary processes as well as the theoretical aspects of parametric estimation for a time-periodic signal with a periodic noise. The PARMA models (and their special case PAR) are also examined in the context of the limiting distribution of the appropriate coefficients. There is also a chapter in which the influence of different signal characteristics to PAR model stability is examined. In the group of chapters related to the theoretical aspects of cyclostationary models there is also a paper devoted to properties of estimators for main characteristics of the annual and daily rhythmic parametric model. In two chapters, the reader can find also various aspects of cyclostationary systems with heavy-tailed distributions. In the recent years these systems more frequently appear in the literature and can be regarded as an extension of the classical cyclostationary processes. The last paper which can be assigned to the first group is devoted to the problem of estimation for almost periodically (APC) time series. In all of the mentioned papers the theoretical results related to various problems adequate to cyclostationary systems are supported by simulations or real data analysis. The second group of chapters is more applicable oriented. The reader can find here applications of cyclostationary characteristics for gearbox and bearing fault diagnosis as well as cyclostationary processes for damage assessment of rolling element bearing. The last paper contains the multidimensional analysis of ARMA models which are strongly related to the mentioned PARMA systems. The multidimensional time series are applied to the real data from the New Zealand electricity market.

Simulation Study of Performance of MBB in Overall Mean Estimation Problem for APC Time Series

Anna E. Dudek and Jakub Uzar

1 Introduction

In this paper we focus on the class of nonstationary processes called almost periodically correlated (APC). Time series $\{X_t, t \in \mathscr{Z}\}$ with finite second moments is almost periodically correlated if it has an almost periodic mean and covariance functions. The function f is called almost periodic if for every $\varepsilon > 0$, there exists a number l_ε such that for any interval of length greater than l_ε, there exists a number p_ε in this interval such that

$$\sup_{t \in \mathscr{Z}} |f(t + p_\varepsilon) - f(t)| < \varepsilon.$$

Equivalently, almost periodic functions can be defined as uniform limits of trigonometric polynomials. For more details and some properties of almost periodic functions we refer the reader to [2]. APC time series are frequently met in different settings e.g. vibroacoustics, communications, mechanics, signal processing and economics (see e.g. [1, 11, 14, 16]). However, modeling such data is not easy. The asymptotic variance is very difficult to estimate and in practice construction of the confidence intervals for the parameters of interest is almost impossible. In the recent years an alternative approach based on the resampling techniques was proposed. The general idea is to substitute the asymptotic quantiles by their bootstrap or subsampling versions. Bootstrap was introduced by Efron in [9] and was designed for iid data. Sampling with replacement observations from the original sample one gets its bootstrap counterparts that are used to approximate the sampling distribution. The most known modification of IID bootstrap for dependent data called the Moving Block Bootstrap (MBB) was proposed independently by Künsch in [12] and

A.E. Dudek (✉) · J. Uzar
AGH University of Science and Technology, al. Mickiewicza 30,
30-059 Krakow, Poland
e-mail: aedudek@agh.edu.pl

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods - II*,
Applied Condition Monitoring 3, DOI 10.1007/978-3-319-16330-7_1

Liu and Singh in [15]. Instead of the single observations the block of observations are sampled. The method was first applied for the stationary time series (see e.g. [13]) but in the recent years it turned out that it can also be used for the APC time series. Proving bootstrap consistency is very technical and simultaneously difficult task. Moreover, MBB is the only block bootstrap technique, which validity was shown for APC type data. Other methods like the Seasonal Block Bootstrap ([17]), the Periodic Block Bootstrap ([3]) and the Generalized Seasonal Block Bootstrap ([8]) are designed for periodic data. Thus, they can be used only for the subclass of the APC time series, that is for periodically correlated time series.

The overall mean that we consider in this paper can be estimated directly in the time domain or as the cyclic mean at the frequency 0. For the first case, which is more natural, we have so far only one bootstrap consistency result that belongs to Synowiecki, who in [18] obtained MBB validity for APC time series. For the latter case the results presented in [7] (MBB for APC time series), [19] (subsampling for APC time series), [5] (subsampling for continuous APC processes) and [4] (MBB for continuous APC processes that are not observed continuously) can be applied.

Unfortunately, in any of the mentioned papers the authors did not propose any method of the block length choice. This problem is indeed very difficult and so far there is no approach that can provide any indication of reasonable choice. Thus, we decided to perform a simulation study in which we look at the actual probabilities curves. In this paper, we focus on the overall mean estimation problem.

Paper is organized as following. In Sect. 2 we recall the essential definitions and theorems. In Sect. 3 the simulation study results together with some conclusions are presented, while the final comments are in Sect. 4.

2 Theoretical Background

In this section we recall the MBB method and some consistency results from [18].

Let $(X_1, \ldots, X_n)$ be a sample from an APC time series $\{X_t, t \in \mathscr{Z}\}$. By μ and $\widehat{\mu}_n$ we denote respectively the overall mean (which is the mean value of the expectation function EX_t) and its estimator i.e. $\widehat{\mu}_n = 1/n \sum_{i=1}^{n} X_i$. Moreover, we assume that X_t is α-mixing i.e. $\alpha_X(k) \to 0$ as $k \to \infty$, where

$$\alpha_X(k) = \sup_t \sup_{\substack{A \in \mathscr{F}_X(-\infty, t) \\ B \in \mathscr{F}_X(t+k, \infty)}} |P(A \cap B) - P(A)P(B)|$$

and $\mathscr{F}_X(t_1, t_2)$ is a σ-algebra generated by $\{X_t : t_1 \leq t \leq t_2\}$. If $\alpha_X(k) = 0$ it means that the observations that are at least k time units apart are independent (see [6]).

Moving Block Bootstrap algorithm

1. Choose block length $b < n, b \in \mathscr{Z}$.
2. Let $B_i := (X_i, \ldots, X_{i+b-1})$ be the block that starts in the observation X_i and has the length b. Thus, the set of all possible blocks is of the form

$\mathscr{B} = \{B_1, \ldots, B_{n-b+1}\}$. Moreover, take $l = \lfloor n/b \rfloor$. To obtain the bootstrap sample select randomly with replacement $l+1$ block from the set $\mathscr{B}$. In each step probability of choosing any block is always the same and equal to $1/(n-b+1)$ i.e.

$$P^*\left(B_i^* = B_k\right) = \frac{1}{n-b+1}, \quad \text{where} \quad i = 1, \ldots, l+1 \text{ and } k = 1, \ldots, n-b+1.$$

By P^* we denote the conditional probability given the sample $(X_1, \ldots, X_n)$.

3. Join selected blocks $(B_1^*, \ldots, B_{l+1}^*)$ and from the obtained sample take first n observations $(X_1^*, \ldots, X_n^*)$ to get the bootstrap sample of the same length as the original sample.
4. Repeat B times steps 2–3 .

Now we recall MBB consistency result (Corollary 3.2. from [18]) in the overall mean estimation problem.

Theorem 1 *Let $\{X_t, t \in \mathscr{Z}\}$ be an APC, α-mixing time series and let X_t^* be generated by the MBB procedure with $b = o(n)$, but $b \to \infty$. Assume that:*

(i) *the set $\Lambda = \{\lambda \in [0, 2\pi) : M_t(E(X_t e^{-i\lambda t})) \neq 0\}$ is finite;*
(ii) *$\sup_{t \in \mathscr{Z}} E|X_t|^{4+2\delta} < \infty$ for some $\delta > 0$;*
(iii) *$\sum_{\tau=1}^{\infty} \tau \alpha_X^{\delta/(4+\delta)}(\tau) < \infty$;*
(iv) *the CLT holds i.e.*

$$\sqrt{n}\left(\widehat{\mu}_n - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right).$$

Then, the MBB procedure is consistent i.e.

$$\sup_{x \in \mathscr{R}} \left| P\left(\sqrt{n}\left(\widehat{\mu}_n - \mu\right) \leq x\right) - P^*\left(\sqrt{n}\left(\widehat{\mu}_n^* - E^*\left(\widehat{\mu}_n^*\right)\right) \leq x\right) \right| \xrightarrow{p} 0,$$

where by M_t we denote mean over the variable t and $\widehat{\mu}_n^*$ is a bootstrap counterpart of $\widehat{\mu}_n$.

The asymptotic variance σ^2 is very difficult to estimate and hence in practical applications the asymptotic confidence intervals for μ are almost impossible to obtain. However, the cited theorem enables the construction of the bootstrap confidence intervals. Below we present shortly the idea of so called *percentile* bootstrap confidence intervals. For more details we refer to he reader to [10]. Let

$$K_{BOOT}(x) = P^*\left(\sqrt{n}\left(\widehat{\mu}_n^* - E^*\left(\widehat{\mu}_n^*\right)\right) \leq x\right).$$

Then the equal-tailed 95 % bootstrap confidence interval is of the form

$$\left(\widehat{\mu}_n - \frac{K_{BOOT}^{-1}(0.975)}{\sqrt{n}}, \widehat{\mu}_n - \frac{K_{BOOT}^{-1}(0.025)}{\sqrt{n}}\right). \tag{1}$$

3 Simulation Study

In this section we present results of our simulation study. Our aim is to calculate the actual coverage probabilities (ACPs) and to try to point out the optimal choices of the block length b. We use the bootstrap algorithm described in Sect. 2. In our study we consider five groups of time series:

TS1: $X_t = A_1 \cdot (\varepsilon_t + 4\varepsilon_{t-1}/5) + A_2 \cdot \cos(\pi t/8)$,
TS2: $X_t = B_1 \cdot (\varepsilon_t + 3\varepsilon_{t-1}/7) + B_2 \cdot \sin(\sqrt{3}t/4)$,
TS3: $X_t = C_1 \cdot \sin(3\pi\sqrt{t}/32) \cdot X_{t-1} + \varepsilon_t + C_2$,
TS4: $X_t = D_1 \cdot \cos(\pi t/16) \cdot X_{t-1} + \varepsilon_t + D_2$,
TS5: $X_t = E_1 \cdot \cos t \cdot \cos(\sqrt{2}t) \cdot X_{t-1} + E_2 \cdot \sin t + \varepsilon_t$,

where $X_0 = \varepsilon_0$ and random variables $\varepsilon_0, \varepsilon_1, \ldots$ are iid from the standard normal distribution. Taking different forms of parameters $A_1, A_2, \ldots, E_1, E_2$, i.e. constants, periodic or almost periodic functions, we get the time series that have the mean and the autocovariance functions periodic or almost periodic. Note that the first two models (**TS1**, **TS2**) are the moving average type, while the **TS3**–**TS5** are the autoregressive type. Mentioned parameters help us to investigate closer some subgroups of APC time series, like e.g. having null or purely periodic mean functions.

For each example of time series we consider two sample sizes $n_1 = 240$ and $n_2 = 1200$. Moreover, we take the block length $b \in \{1, 2, 4, 6, 10, 12, 15, 20, 24, 40, 60, 120\}$ (for n_1) and $b \in \{1, 2, 4, 6, 10, 12, 15, 20, 24, 40, 60, 120, 200, 300, 600\}$ (for n_2). In each case we construct the equal-tailed 95 % bootstrap confidence intervals for the overall mean taking 500 bootstrap replications. The number of iteration of algorithm is equal to 1000. The calculated ACPs are presented in Tables 1, 2, 3, 4 and 5. In bold are indicated those ACPs that are the closest to the nominal coverage probability level.

We introduced parameters $A_1, A_2, \ldots, E_1, E_2$ not only to change the structure of the data, but also to change the relation between the size of the mean function and noise (i.e. signal to noise ratio). It turns out that this is a key factor influencing the optimal choice of the block size. By optimal block length we mean such b for which the obtained ACP is the closest to the nominal one. In the columns 2–4 of Table 1 the constant $A_1 = 1$, while A_2 takes values $1/4$, 1, 3. In the first case mean function is strongly affected by noise and in the last one we observe a very strong periodic mean component. The larger is the value of A_2, the smaller is the optimal block length, i.e. it is around 24, 20, 15 for n_2 and A_2 equal to $1/4$, 1, 3, respectively. For shorter sample the optimal b choices are around 4, 2, 15. For $b < 15$ and n_1 **TS1** with $A_1 = 1$ and $A_2 = 3$ has ACPs close to or equal to 1 (see Fig. 1). Moreover, one may notice that independently on the sample size for each considered block length the confidence intervals are getting wider when the mean component is more significant. Good example illustrating this phenomena is **TS2** with $B_1 = 1/5$, $B_2 = 1$ (see Fig. 2), in which independently on the sample size for $b \leq 12$ the coverage probability is constantly equal to 1. Simultaneously, for most considered series when we decrease the strength of the mean value component ACPs are usually too low for

Table 1 **TS1** model: the first row of each column contains the chosen parameters A_1 and A_2

$A_1; A_2$	$1; \frac{1}{4}$		1; 1		1; 3		$\frac{7}{2}; \frac{1}{4}$		$\frac{7}{2}; 1$		$\frac{7}{2}; 3$	
n	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200
b = 1	82.7	84	89	87.2	98.9	98.8	83.8	83	83.4	83.6	86.1	86.3
b = 2	92.1	90.3	**94.9**	95.6	100	99.9	90.5	92.4	90.9	91.1	**94.8**	92.9
b = 4	**93.9**	92.8	97.5	98.5	100	100	93.1	93.3	93.8	94.4	97.1	96.1
b = 6	93.4	93.8	97.1	98	100	100	91.7	94	**94**	94.3	97.3	97.9
b = 10	91.4	94.2	96.9	96.8	100	100	**94.1**	94.7	93.9	95.3	95.8	97.3
b = 12	92.6	93.3	94.2	97	99.9	99.7	92.7	95.3	93.8	**95**	93.7	**95.2**
b = 15	92.9	94.8	93	94.5	**94.7**	**94.7**	93.1	93.6	92.1	93.7	93	94.5
b = 20	91.9	94.1	92.3	**95**	98.9	98.9	90.7	94.5	91.6	93.3	91.3	94.6
b = 24	91.7	**94.9**	94.3	95.4	99.6	100	91.3	**95.2**	90.2	92.6	92.4	95.7
b = 40	87.6	92.1	89.6	94.4	97.8	99.3	88.7	93.6	89.9	94.4	90.5	96
b = 60	83.4	93.5	84.1	93.8	93.4	97	80.9	91.8	84.3	94	83.8	93.5
b = 120	66.1	91.3	68.9	92.6	85.2	96.2	65.1	91.5	65.3	92.2	65.4	91.5
b = 200	–	86.2	–	88	–	91.9	–	88.3	–	86.6	–	86
b = 300	–	83.3	–	83.2	–	87.6	–	84.9	–	83.3	–	85.4
b = 600	–	65.6	–	63.7	–	69.5	–	61.4	–	63	–	66.7

(continued)

Table 1 (continued)

$A_1; A_2$	$\frac{1}{2}; \frac{1}{4}$		$\frac{1}{2}; 1$		$\sin \frac{\pi t}{4}; \frac{1}{4}$		$\sin \frac{\pi t}{4}; 1$		$\sin \frac{\sqrt{5}\pi t}{32}; \frac{1}{4}$		$\sin \frac{\sqrt{5}\pi t}{32}; 1$	
n	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200
b = 1	81.9	84.7	95.4	95.6	87.6	85.2	**94.7**	93.8	84.1	84.3	92.9	91.7
b = 2	92.6	93.3	99.4	99	92.9	91.1	98	98.2	90.7	91.5	97.4	97.1
b = 4	92.8	93.6	100	100	**93.9**	94	99.3	99	93.4	94.1	98.9	99.2
b = 6	93.9	95.9	100	99.9	93.1	94.2	98.9	99.9	93.7	94	99.6	99.6
b = 10	**94.4**	95.2	99.4	99.1	93.2	95.5	98.7	99.2	**93.8**	96	98.6	98.4
b = 12	93.9	93.7	98.3	98.1	93.8	93.8	96.9	98	92.3	93.7	96.6	97.8
b = 15	93.3	93.1	92.4	**94.7**	93.1	92.3	93.9	**94.3**	93.2	94.2	92.8	94.6
b = 20	91.5	93.9	96	98.1	92.1	93.8	96.4	97.5	92.3	95.2	93.7	96.9
b = 24	90.7	**95.1**	98	98.8	91.7	**94.9**	97.2	98.2	92.9	**95**	**95.9**	97.1
b = 40	87.7	92.8	**95.1**	98.1	88.7	92.3	94.6	96.5	88.6	94.8	91.6	**95.1**
b = 60	82.3	91.5	86.6	95.6	84.3	93.5	87.9	93.7	83.7	92.8	86.5	93.2
b = 120	66.3	91	76.3	93.3	66.2	91.9	71.8	92.6	66	90.4	72.5	92
b = 200	–	87.5	–	92.4	–	88.6	–	90.5	–	86.7	–	89.1
b = 300	–	85.3	–	88.2	–	83.9	–	84.5	–	84.9	–	86.4
b = 600	–	66.8	–	67	–	65.6	–	68.6	–	64.9	–	67.4

In rows 3–15 on the left- and the right-hand side of each column are presented ACPs (in %) for sample lengths $n_1 = 240$ and $n_2 = 1200$, respectively. The nominal coverage probability is equal to 95%. In bold are indicated those ACPs that are the closest to the nominal coverage probability level

Table 2 **TS2** model: the first row of each column contains the chosen parameters B_1 and B_2

$B_1; B_2$	$1; \frac{1}{3}$		1; 1		$1; \frac{7}{3}$		$2; \frac{1}{3}$		2; 1		$2; \frac{7}{3}$	
n	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200
b = 1	85.8	88.3	89.1	92.2	98.8	98.6	82.7	87.9	88.3	87.9	92.9	92.6
b = 2	92.6	93.3	97.2	96.1	99.8	100	90.9	91.6	92.3	93	97.3	97
b = 4	93.5	94.1	98.2	98.6	100	100	92.2	94.1	**95.6**	95.9	99.1	99.4
b = 6	**94.4**	96.3	98.7	98.7	100	100	93.3	92.9	96	95.9	98.7	99.8
b = 10	93.9	94.8	**95.5**	96.7	99.7	99.7	**93.5**	93.7	94	95.3	97.2	98.2
b = 12	92.3	96.5	92.9	94.3	96.9	97.9	93	94.1	94.3	93.1	94.7	97.1
b = 15	92.2	94.1	90.2	94.9	89.7	**94.4**	92.3	93.2	91.4	94.2	90.7	93.3
b = 20	92.7	95.7	93.9	97.1	98.9	99.7	92.4	92.8	92.3	94.5	**95.1**	97.8
b = 24	90.7	94.1	93.7	96.6	97.4	99.2	91.2	93.4	91.4	**94.8**	94.5	97.1
b = 40	87.8	**94.9**	89.1	**95**	**93.6**	96.8	88.3	93.2	88.7	94.5	88.6	**94.7**
b = 60	82.8	93.1	84	94	80.8	93.3	84.3	**94.2**	84.4	92.6	84.2	92.8
b = 120	64.5	89.8	66.8	92.3	72.7	92.7	65.2	90.9	66.2	89.6	64.3	90
b = 200	–	88.4	–	87.5	–	87.9	–	89.7	–	87.2	–	86.8
b = 300	–	83.7	–	84.2	–	86.4	–	84.3	–	84.8	–	85.3
b = 600	–	64.8	–	65.9	–	69.6	–	64.4	–	65	–	63.2

(continued)

Table 2 (continued)

$B_1; B_2$	$\frac{1}{5}; \frac{1}{3}$		$\frac{1}{5}; 1$		$\cos \frac{\pi t}{4}; \frac{1}{3}$		$\cos \frac{\pi t}{4}; 1$		$\cos \frac{\sqrt{5}\pi t}{32}; \frac{1}{3}$		$\cos \frac{\sqrt{5}\pi t}{32}; 1$	
n	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200
b = 1	96.4	97.5	100	100	89.6	90.6	96.5	95.6	89.2	87.4	94.1	96.1
b = 2	99.2	99.5	100	100	92.8	93	99.5	99.2	93.4	93.7	97.9	98.8
b = 4	99.8	100	100	100	**95**	96	99.9	99.8	**95.3**	95.7	99.5	100
b = 6	99.8	100	100	100	96	96.2	99.8	99.8	94.5	97.4	99.5	99.7
b = 10	98.2	99.5	100	100	93.7	95.7	98.9	99	93.6	**94.9**	98.3	99.1
b = 12	**95.5**	97.4	99.3	100	92.9	**95**	95.7	96	92	95.2	**95.3**	96.8
b = 15	90.1	**94.9**	82.7	**94.8**	92.4	95.4	91.4	**94.6**	94	93.8	91.7	93.8
b = 20	96.8	98.1	100	100	93.5	94.3	96.9	98.3	90.3	93.9	97.8	97.9
b = 24	96.5	98.1	99.8	100	92.5	94.8	**94.5**	98.1	90	94.6	94	97.4
b = 40	88.8	**94.9**	**97**	99.7	87.4	94	91	95.9	89.9	93.9	89.7	94.2
b = 60	83.5	92.8	77.3	95.9	85	94.8	84.2	93.3	84.8	91.7	84	**94.4**
b = 120	69.2	92	77.7	96.8	63.7	92	71.3	91.8	62.6	90.2	69.6	92.8
b = 200	–	88.4	–	92.4	–	87.1	–	87.7	–	87.5	–	88.4
b = 300	–	84.2	–	91.2	–	84.3	–	83.2	–	83.3	–	86.1
b = 600	–	66	–	74.9	–	66.4	–	66.6	–	64.5	–	65.9

In rows 3–15 on the left- and the right-hand side of each column are presented ACPs (in %) for sample lengths $n_1 = 240$ and $n_2 = 1200$, respectively. The nominal coverage probability is equal to 95 %. In bold are indicated those ACPs that are the closest to the nominal coverage probability level

Table 3 **TS3** model: the first row of each column contains the chosen parameters C_1 and C_2

$C_1; C_2$	$\frac{1}{3}; 0$		1; 0		2; 0		$\frac{1}{3}; \cos\frac{t}{2}$		1; $\cos\frac{t}{2}$		$\frac{1}{3}; \frac{1}{5}\sin\frac{\sqrt{5}\pi t}{64}$	
n	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200
b = 1	92	93	33.2	20.8	47.8	0	97.6	97.7	39.3	24.5	92.6	93.7
b = 2	92.1	92.3	36	20.6	26.7	0	98.9	99.4	43.1	21.2	93.1	92.5
b = 4	93	92.9	47.4	26.7	32.5	0	99.6	99.7	57.1	33.2	95.1	94.4
b = 6	**94.2**	94.5	58.6	34.4	43.6	0.1	99.6	99.8	69.6	37.3	**95**	**95**
b = 10	93.7	92.9	81.3	42.9	58.1	89.4	95.9	97.4	82.4	45.3	95.4	95.9
b = 12	93.3	94	87.8	46.9	59.5	99.5	93.3	95.1	88.6	51.3	95.3	96.8
b = 15	93.1	94.1	**94.2**	56.4	65.4	100	**95.6**	**95**	**96**	59	95.2	96.5
b = 20	92.5	94.3	97.5	68.2	66.7	100	96.3	97.1	97.4	69.8	94.5	96.6
b = 24	91.1	**94.8**	98	73.9	69.2	100	93.1	**95**	98.4	74.1	95.3	97
b = 40	89.2	94	98.5	92.3	**70**	100	89.5	94.5	98.9	91.6	91.7	95.3
b = 60	83.1	93.1	98	97.7	66	100	88.3	94.4	98.1	**97**	84.8	92.9
b = 120	66.6	92	91.3	98.4	27.2	100	71.4	92.7	91.8	99	70.5	89.7
b = 200	–	88.7	–	98.1	–	100	–	89.6	–	98.4	–	89.2
b = 300	–	84.3	–	**96.9**	–	100	–	86	–	98	–	86.3
b = 600	–	62.6	–	22.8	–	**92.3**	–	61.7	–	21.3	–	64.5

(continued)

Table 3 (continued)

$C_1; C_2$	$1; \frac{1}{5}\sin\frac{\sqrt{5}\pi t}{64}$		$\frac{1}{3}; \sin\frac{\sqrt{5}\pi t}{64}$		$1; \sin\frac{\sqrt{5}\pi t}{64}$		$2; \sin\frac{\sqrt{5}\pi t}{64}$		$\frac{1}{3}; 4\cdot\sin\frac{\sqrt{5}\pi t}{64}$		$1; 4\cdot\sin\frac{\sqrt{5}\pi t}{64}$	
n	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200
b = 1	32.6	21.7	**96.3**	97.2	14.4	31.5	34.9	0	99.9	100	0	63.3
b = 2	37.4	22.2	99.3	99.3	19.6	38.8	22.7	0	100	100	0	81.9
b = 4	46.1	28.8	99.9	100	36.3	55.6	38.6	0	100	100	12.6	**95.6**
b = 6	61.5	37.7	100	100	66.1	67.6	53	0.3	100	100	79.1	99.4
b = 10	80.6	47.2	100	100	**95.1**	80.1	71.9	91.6	100	100	100	99.9
b = 12	89.2	54.1	100	100	98.8	84.6	75.4	99.9	100	100	100	100
b = 15	**95.4**	59.8	100	100	100	90.1	80.9	100	100	100	100	100
b = 20	98.5	69.5	100	100	100	**95.1**	**82.7**	100	100	100	100	100
b = 24	99.1	78.1	100	100	100	96.7	82	100	100	100	100	100
b = 40	98.4	**93.3**	100	100	100	97.2	81.9	100	100	100	100	100
b = 60	98.4	97.5	83	**94.6**	98.8	**95.1**	81.3	100	77.3	**98.4**	100	97
b = 120	94.3	98.5	74.7	93.9	98.5	97.3	24.9	100	**91.3**	99.7	**99.6**	96.6
b = 200	–	98.5	–	98.1	–	96.6	–	100	–	100	–	96.2
b = 300	–	96.9	–	92.8	–	91.5	–	100	–	100	–	77.4
b = 600	–	22.7	–	82.5	–	34.8	–	**92.5**	–	100	–	68.2

In rows 3–15 on the left- and the right-hand of each column are presented ACPs (in %) for sample lengths $n_1 = 240$ and $n_2 = 1200$, respectively. The nominal coverage probability is equal to 95 %. In bold are indicated those ACPs that are the closest to the nominal coverage probability level

Table 4 **TS4** model: the first row of each column contains the chosen parameters D_1 and D_2

$D_1; D_2$	$\frac{1}{2}; 0$		1; 0		2; 0		$\frac{1}{2}; \sin\frac{t}{2}$		1; $\sin\frac{t}{2}$		$\frac{1}{2}; \frac{1}{5}\cos\frac{\sqrt{7}\pi t}{64}$	
n	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200
b = 1	92.1	90.7	71.6	72.1	72.7	72.4	97.5	96.8	80.8	79.5	90.7	91.9
b = 2	90.5	90.7	74	72.5	73.4	71.1	98.1	97.7	85.9	86.3	92.3	93
b = 4	92.9	93.6	83	81.2	86.6	83.1	99.5	99.3	94.3	93	**94.8**	**94.5**
b = 6	92.5	93.9	87.5	86.8	88.3	91.2	99.5	99.6	96.1	95.6	96.9	97.1
b = 10	92.6	92.4	**91.7**	91.6	**92.8**	92	96.2	96.8	96.3	**95.4**	96.4	97.3
b = 12	**94**	93.2	91.6	91.8	92.7	92.1	94.2	96.6	95.4	96.8	96.8	96.8
b = 15	93	**94.6**	89.9	92.2	91.6	93.7	**94.9**	**95.1**	**95.1**	96.8	96.6	98.4
b = 20	90.9	93.5	91	93.5	89.9	93.2	95.6	96.1	94.4	96.8	97.5	98.7
b = 24	89.1	93.9	87.6	93.1	89.5	92.8	92.6	95.8	93.7	96.5	95.8	98.1
b = 40	87.6	94.1	86.9	**93.6**	86.9	**93.8**	88.4	93.7	89.6	95.9	90.4	93.1
b = 60	84.6	92.6	80.9	92.7	82.5	93.7	84.7	93.9	82.1	92.9	86.5	93.1
b = 120	63.5	90.9	62.6	91.7	63	91.4	72	90.8	67.8	91.4	71.5	93.3
b = 200	–	89.5	–	87	–	88.4	–	88.9	–	88.4	–	87.5
b = 300	–	82.1	–	82.6	–	87.4	–	81.7	–	81.7	–	83.3
b = 600	–	66	–	66.5	–	62.1	–	64.9	–	65.2	–	67.1

(continued)

Table 4 (continued)

D_1; D_2	1; $\frac{1}{5}\cos\frac{\sqrt{7}\pi t}{64}$		$\frac{1}{2}$; $\cos\frac{\sqrt{7}\pi t}{64}$		1; $\cos\frac{\sqrt{7}\pi t}{64}$		2; $\cos\frac{\sqrt{7}\pi t}{64}$		$\frac{1}{2}$; $2\cos\frac{\sqrt{7}\pi t}{64}$		1; $2\cos\frac{\sqrt{7}\pi t}{64}$	
n	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200
b = 1	73.6	75	**94.7**	**95.5**	83.3	88.3	79	88.9	**99.5**	99.2	**95**	**96**
b = 2	75.5	76.3	98.4	98.2	**93.4**	**94.5**	92.9	**95.5**	100	100	99.5	99.9
b = 4	88.6	89.3	99.7	99.5	98.4	98.7	98.4	99	100	100	100	100
b = 6	91.4	92.4	100	100	99.9	99.8	99.2	99.5	100	100	100	100
b = 10	94.1	**95.8**	100	100	100	100	99.5	99.9	100	100	100	100
b = 12	95.7	96.4	99.9	100	100	100	99.6	99.9	100	100	100	100
b = 15	95.7	96.7	100	100	99.9	100	99.7	100	100	100	100	100
b = 20	95.6	96.4	100	100	99.9	100	99	99.9	100	100	100	100
b = 24	**95**	98	100	100	100	100	99	99.9	100	100	100	100
b = 40	91.4	95.9	98.1	99.7	99	100	97.8	100	99.9	100	100	100
b = 60	86	93.8	98.1	99.4	96.8	99.6	**94**	99	99.8	100	100	100
b = 120	71.1	92.3	93.3	98.8	88	98.4	81.5	96.3	99.8	100	97.5	100
b = 200	–	86	–	90.5	–	88.3	–	87.2	–	91.2	–	89.1
b = 300	–	85.1	–	87.5	–	86.2	–	83.5	–	**91.8**	–	89.2
b = 600	–	66.8	–	77	–	76.6	–	73.5	–	89.3	–	87.3

In rows 3–15 on the left- and the right-hand side of each column are presented ACPs (in %) for sample lengths $n_1 = 240$ and $n_2 = 1200$, respectively. The nominal coverage probability is equal to 95 %. In bold are indicated those ACPs that are the closest to the nominal coverage probability level

Table 5 **TS5** model: the first row of each column contains the chosen parameters E_1 and E_2

$E_1; E_2$	$\frac{1}{2}; 0$		1; 0		4; 0		$\frac{1}{2}; \frac{1}{3}$		1; $\frac{1}{3}$		4; $\frac{1}{3}$	
n	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200
b = 1	94.6	94.1	93.5	93.9	100	15.1	100	100	100	100	100	22.9
b = 2	**94.8**	**94.8**	93.1	**95**	100	12.6	100	100	100	100	100	17.5
b = 4	94.4	94.5	94	**95**	100	29	100	100	100	100	100	57.6
b = 6	92.2	94.6	**94.4**	95.4	100	28.6	96	**96.2**	98.5	98.9	100	58
b = 10	92.8	**95.2**	93.6	95.1	100	29.1	99.6	99.9	99.8	99.9	100	51.7
b = 12	92.5	94	94.2	93.5	99.9	24	**95.1**	96.5	96.9	97.5	100	45.3
b = 15	92.8	93.7	94.1	95.3	100	21.1	99.3	99.5	99.4	99.4	100	30.1
b = 20	91.1	93.3	92.4	94.7	100	18.7	95.5	96.6	96.8	97.2	100	21.9
b = 24	90	94.1	90.9	94.6	100	27.3	94.4	97.2	96.3	97.9	100	41.2
b = 40	88.7	95.5	89.8	93.8	**99.1**	90.8	94	97.7	**94.1**	97.2	**99.6**	**98.9**
b = 60	82.5	93.7	83.8	94.2	47.2	99.6	92.2	96.7	92.8	**96.9**	51.6	100
b = 120	68.2	91	63.6	90	0.1	99.4	66.4	92.7	67	90.1	0	100
b = 200	–	87.9	–	89.4	–	**93.8**	–	89.6	–	89.6	–	100
b = 300	–	83.9	–	83.8	–	46.3	–	85.2	–	85.2	–	99.4
b = 600	–	65	–	65.7	–	6.2	–	67.3	–	67.4	–	1.9

(continued)

Table 5 (continued)

E_1; E_2	$\frac{1}{2}$; 1		1; 1		4; 1		$\frac{1}{2}$; 5		1; 5		4; 5	
n	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200
b = 1	97.3	98.4	98.6	98.3	99.9	21.6	100	100	100	100	100	14.5
b = 2	99.3	98.3	99.3	98.8	100	20.6	100	100	100	100	100	7
b = 4	97.8	97.5	97.5	97.8	100	45	100	100	100	100	100	58.5
b = 6	**94.4**	95.2	95.5	94.7	100	43.9	98.7	99.1	100	99.9	100	58.8
b = 10	96.3	97.2	**95.1**	96.4	100	38	100	100	100	100	100	47.8
b = 12	94.3	**95**	94.8	94.3	100	38.1	**97.7**	97.6	98.5	99.3	100	41
b = 15	94.3	96	94.2	97.8	100	32.9	100	100	100	100	100	26.1
b = 20	91.7	94.2	91.1	94.6	100	27.9	99.1	99.1	99.1	99.7	100	14.7
b = 24	92	93.6	92	**95.2**	100	39.2	**97.7**	99	99	99.2	100	33.8
b = 40	90.2	95.1	90.1	92.7	**99.6**	**94.8**	98.7	99.3	98.9	99.6	**99.4**	**99.5**
b = 60	84.8	93.7	85.2	93.1	50.5	99.8	97.8	98.7	**98.2**	98.5	53.8	100
b = 120	64.7	92.2	66.9	91.3	0.2	99.8	66.5	**92.7**	73.8	**92.2**	0.6	100
b = 200	–	89.4	–	89.5	–	93.5	–	89.9	–	90.2	–	100
b = 300	–	86.2	–	82.8	–	74.3	–	86.3	–	85.5	–	100
b = 600	–	62.7	–	67.7	–	5.9	–	72.2	–	72.8	–	0.2

In rows 3–15 on the left- and the right-hand side of each column are presented ACPs (in %) for sample lengths $n_1 = 240$ and $n_2 = 1200$, respectively. The nominal coverage probability is equal to 95 %. In bold are indicated those ACPs that are the closest to the nominal coverage probability level

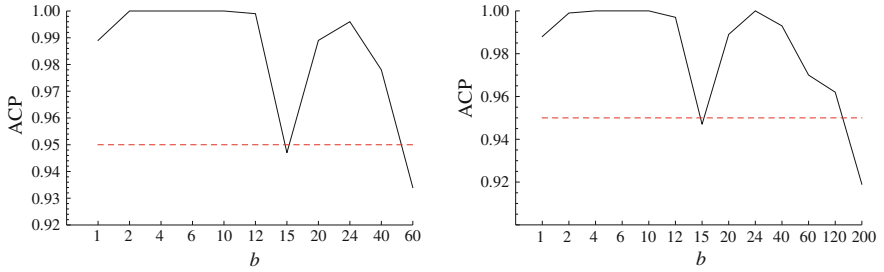


Fig. 1 **TS1** model with $A_1 = 1$ and $A_2 = 3$: ACPs (*solid line*) together with nominal coverage probability (*dashed line*) for n_1 (*left-hand side*) and n_2 (*right-hand side*)

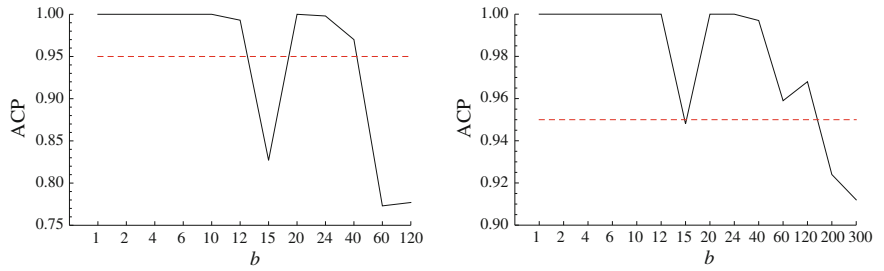


Fig. 2 **TS2** model with $B_1 = \frac{1}{5}$ and $B_2 = 1$: ACPs (*solid line*) together with nominal coverage probability (*dashed line*) for n_1 (*left-hand side*) and n_2 (*right-hand side*)

small values of b. Moreover, one may note that for series **TS3** with $C_1 = 1/3$ and $C_2 = 4\sin(\sqrt{5}\pi t/64)$ we almost always obtain ACP equal to 1. When we take C_2 four times lower ACPs decrease for larger values of b. Finally, multiplying C_2 by the factor $1/20$ provides ACPs too low for most b values (see Fig. 3).

One may note that ACP curves are quite flat for the moving average type models. The contrary can be observed for the autoregressive type examples. The confidence intervals are more affected by the choice of parameters and can be much too narrow or much too wide for majority of the considered block lengths. For time series **TS3** with $C_1 = 1$ and $C_2 = 0$ (see Fig. 4) the larger block lengths result in better ACPs. On the other hand there are examples, like **TS5** with $E_1 = 1/2$ and $E_2 = 0$, in which the opposite situation can be observed, i.e. the small values of b are the best ones (see Fig. 5). Notice that both mentioned series are mean zero and almost periodically correlated. They differ in the form of the autocovariance function.

For majority of the considered models the lowest ACPs are around 70–80 % for most block lengths. The significant exception from this rule are **TS3** and **TS5** with values of the first parameter much higher than the second one. If $E_1 = 4$ and $E_2 \in \{0, 1/3, 1\}$ for $b \leq 24$ ACPs are equal to 1 for n_1 and change from about 20 % to 40–50 % for n_2. In this situation taking not optimal b, especially for the larger sample size, brings dramatic consequences, because confidence intervals are extremely narrow.

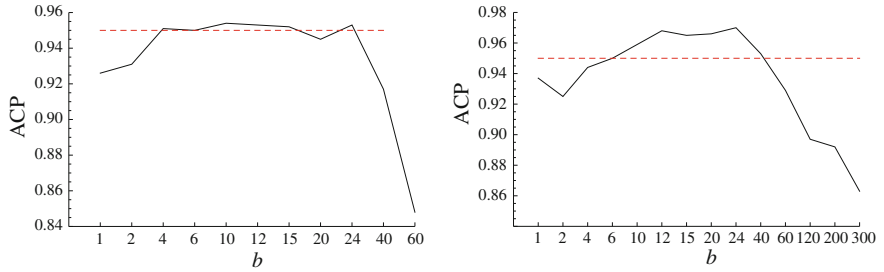


Fig. 3 **TS3** model with $C_1 = \frac{1}{3}$ and $C_2 = \frac{1}{5} \sin \frac{\sqrt{5}\pi t}{64}$: ACPs (*solid line*) together with nominal coverage probability (*dashed line*) for n_1 (*left-hand side*) and n_2 (*right-hand side*)

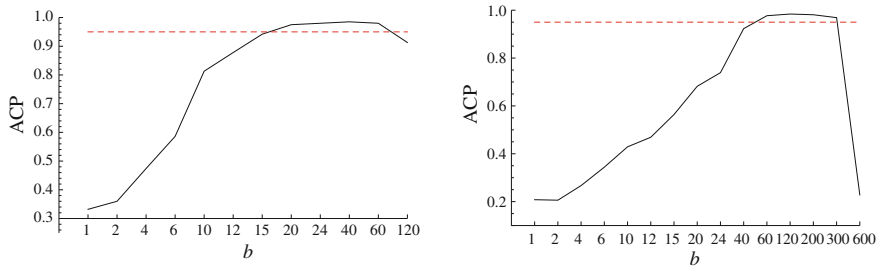


Fig. 4 **TS3** model with $C_1 = 1$ and $C_2 = 0$: ACPs (*solid line*) together with nominal coverage probability (*dashed line*) for n_1 (*left-hand side*) and n_2 (*right-hand side*)

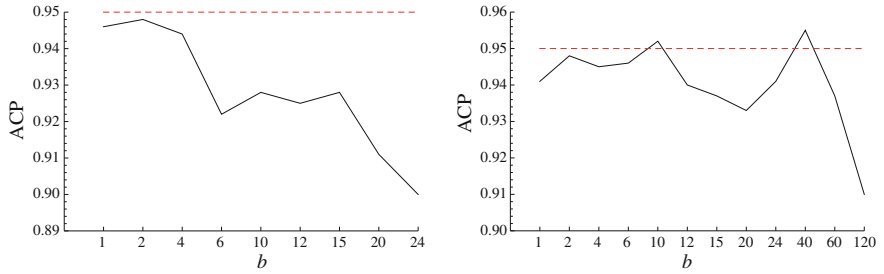


Fig. 5 **TS5** model with $E_1 = \frac{1}{2}$ and $E_2 = 0$: ACPs (*solid line*) together with nominal coverage probability (*dashed line*) for n_1 (*left-hand side*) and n_2 (*right-hand side*)

It seems that the main factor that we should consider in the future looking for a method of the optimal block length choice is a signal to noise ratio. In the stationary case choice of b was determined by the autocovariance function and value at 0 of the spectral density function (see [13]). The nonstationary case that we study in this paper is much more complicated. For APC processes we may expect the dependence on more spectral density functions. Moreover, the autocovariance function for stationary time series does not depend on time variable in contrary to APC case.

4 Conclusions

In the paper we constructed the percentile equal-tailed MBB confidence interval for the overall mean of an APC time series. We investigated the behaviour of the actual coverage probability curves trying to find common patterns and hints that may help in the future to find the method for the optimal block length choice. The considered examples indicated the signal to noise ratio as very important factor influencing the optimal block length.

Acknowledgments Research of Anna E. Dudek was partially supported by the Polish Ministry of Science and Higher Education and AGH local grant.

References

1. Antoni, J. (2009). Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, *23*(4), 987–1036.
2. Besicovitch, A. S. (1932). *Almost Periodic Functions*. Cambridge: Cambridge University Press.
3. Chan, V., Lahiri, S. N., & Meeker, W. Q. (2004). Block bootstrap estimation of the distribution of cumulative outdoor degradation. *Technometrics*, *46*, 215–224.
4. Dehay, D., & Dudek, A. E. (2015). Block bootstrap for Poisson-sampled almost periodic processes. *Journal of Time Series Analysis*. doi:10.1111/jtsa.12115.
5. Dehay, D., Dudek, A., & Leśkow, J. (2014). Subsampling for continuous-time nonstationary stochastic processes. *Journal of Statistical Planning and Inference*, *150*, 142–158.
6. Doukhan, P. (1994). *Mixing: Properties and examples*. Lecture Notes in Statistics. New York: Springer.
7. Dudek, A. E. (2015). Circular block bootstrap for coefficients of autocovariance function of APC time series. *Metrika*, *78*(3), 313–335. doi:10.1007/s00184-014-0505-9.
8. Dudek, A. E., Leśkow, J., Politis, D., & Paparoditis, E. (2014). A generalized block bootstrap for seasonal time series. *Journal of Time Series Analysis*, *35*, 89–114.
9. Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *Annals of Statistics*, *7*, 1–26.
10. Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.
11. Gardner, W. A., Napolitano, A., & Paura, L. (2006). Cyclostationarity: Half a century of research. *Signal Processing*, *86*(4), 639–697.
12. Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, *17*, 1217–1241.
13. Lahiri, S. N. (2003). *Resampling methods for dependent data*. New York: Springer.
14. Lenart, Ł., & Pipień, M. (2013). Seasonality revisited—Statistical testing for almost periodically correlated stochastic processes. *Central European Journal of Economic Modelling and Econometrics*, *5*, 85–102.
15. Liu, R., & Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In R. Lepage & L. Billard (Eds.), *Exploring the limits of bootstrap* (pp. 225–248). New York: Wiley.
16. Napolitano, A. (2012). *Generalizations of cyclostationary signal processing: Spectral analysis and applications*. Chichester: Wiley-IEEE Press.

17. Politis, D. N. (2001). Resampling time series with seasonal components, in Frontiers in Data Mining and Bioinformatics. In *Proceedings of the 33rd Symposium on the Interface of Computing Science and Statistics*. Orange County, California, 13–17 June, pp. 619–621.
18. Synowiecki, R. (2007). Consistency and application of moving block bootstrap for nonstationary time series with periodic and almost periodic structure. *Bernoulli*, *13*(4), 1151–1178.
19. Synowiecki, R. (2008). Metody resamplingowe w dziedzinie czasu dla niestacjonarnych szeregów czasowych o strukturze okresowej i prawie okresowej. *Dissertation*. AGH University of Science and Technology, Krakow, Poland. http://winntbg.bg.agh.edu.pl/rozprawy2/10012/full10012.pdf.

Parametric Estimation Problem for a Time Periodic Signal in a Periodic Noise

Khalil El Waled

1 Introduction

Consider the following periodic signal plus periodic noise model

$$d\zeta_t = f(t,\theta)dt + \sigma(t)dW_t, \tag{1}$$

where $f(\cdot,\cdot) : \mathbb{R} \times \Theta \mapsto \mathbb{R}$, $\sigma(\cdot) : \mathbb{R} \mapsto \mathbb{R}$ are continuous functions and periodic in t with the same period P, $\sigma(\cdot)$ is positive, $\theta \in \Theta$ is an unknown parameter; Θ is a compact of $\mathbb{R}$, $\{W_t\}$ is a Brownian motion defined over a complete probability space $(\Omega, \mathscr{F}, \mathrm{P})$.

As an application, let $\{\xi_t, t \geq 0\}$ be the time-dependent geometric Brownian motion which verifies the following linear stochastic differential equation

$$d\xi_t = f(t)\xi_t dt + \sigma(t)\xi_t dW_t. \tag{2}$$

The connection between these processes is given by

$$d\zeta_t = \frac{d\xi_t}{\xi_t}.$$

Such types of processes appear in many domains, for instance, in finance [13] (Black–Scholes–Merton model), mechanics ([8, 12]) and in biology [1, 9].

The problem of parameter estimation when the process is continuously observed has been studied by many others (see for instance [15, 16]...). Ibragimov and Has'minskil ([11], Sect. 5 in Chap. 3) have considered the general model

$$dX_\varepsilon(t) = S_\varepsilon(t,\theta) + db(t).$$

K. El Waled (✉)
IRMAR, CNRS Umr 6625, Université Rennes 2, Rennes, France
e-mail: khalil.elwaled@gmail.com; khalil.elwaled@uhb.fr

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods - II*,
Applied Condition Monitoring 3, DOI 10.1007/978-3-319-16330-7_2

Although model (1) can be reduced to the well-known model $X_\varepsilon(t)$, we prefer to deal with the first case and we are interested in its properties that are driven by the periodicity of $f(\cdot,\cdot)$ and $\sigma(\cdot)$. These properties will enable us to establish the convergence almost sure, to give an explicit expression of the Fisher information and to prove simply the efficiency of the MLE.

Since in practice it is difficult to obtain a complete continuous time observation of the sample path of a process. So the problem of discretization has to be considered.

Kasonga [14] has used the least square method to show the consistency of the estimators based on the discrete schemas. The case of stationary diffusion models is studied in [3, 5].

Using maximum contrast, and for small variance diffusion models [6], has shown under some classical assumptions, asymptotic results. Harison [7] has used this method to estimate the drift parameter for some one-dimensional nonstationary Gaussian diffusion models.

This paper is organized as follows: First, in Sect. 2 we study the parametric estimation problem of θ from a continuous time observation. We prove the existence of an estimator that converges in probability. When $f(t,\theta) = \theta f(t)$ we get the mean square convergence, the asymptotic normality of this estimator and the asymptotic efficiency.

In the Sect. 3 we assume that the observations have been done at discrete moments. We compute the likelihood function and we propose an estimator of the unknown parameter θ then we establish its convergence in probability. When $f(t,\theta) = \theta f(t)$ we show the mean square convergence and the asymptotic normality property as well as the asymptotic efficiency.

For an easier reading of the paper proofs which are long are put in the appendix.

2 Drift Estimation from a Continuous Time Observation

2.1 Likelihood

In the continuous time observation context to estimate the unknown parameter θ we can use the method of the maximum of likelihood. To compute the likelihood function of this model we apply Theorem 7.18 in [16] and Corollary which follows it to the next two processes (see also Theorem 6.10 in [10])

$$d\zeta_t^\theta = f(t,\theta)dt + \sigma(t)dW_t, \tag{3}$$

$$d\eta_t = \sigma(t)dW_t. \tag{4}$$

These two processes satisfy the four conditions of Theorem 7.18 in [16]. So $\mu_\theta^T \sim \nu^T$, where

$$\mu_\theta^T := \mathscr{L}(\zeta_t^\theta,\ 0 \le t \le T),\ \nu^T := \mathscr{L}(\eta_t,\ 0 \le t \le T).$$

In addition, the conditions of the corollary which follows this Theorem 7.18 are satisfied and we have P_θ-a.s.

$$\frac{d\mu_\theta^T}{d\nu^T}(\zeta^\theta) = \exp\left(\int_0^T \frac{\rho(s,\theta)}{\sigma(s)} d\zeta_s^\theta - \frac{1}{2}\int_0^T \rho^2(s,\theta)ds\right) \tag{5}$$

where $\rho(s,\theta) := \frac{f(s,\theta)}{\sigma(s)}$. Remark that $\sigma(\cdot)$ is positive continuous so $\inf_{t\in[0,P]}\sigma(t) > 0$ and $\rho(\cdot)$ is well defined and continuous.

To simplify the computation we take $T = nP$. Assuming that the real value of the unknown parameter θ is θ_0, the likelihood function is given by

$$L_n(\theta) := \exp\left(\int_0^{nP} \frac{\rho(s,\theta)}{\sigma(s)} d\zeta_s^{\theta_0} - \frac{1}{2}\int_0^{nP} \rho^2(s,\theta)ds\right) \tag{6}$$

Thanks to the continuity of $f(\cdot,\cdot)$ and $\sigma(\cdot)$ and the compactness of Θ there exists $\hat{\theta}_n$ such that

$$\hat{\theta}_n = \arg\sup_{\theta\in\Theta} L_n(\theta).$$

Maximizing $L_n(\theta)$ in θ is identical to maximizing the function $U_n(\theta)$ given by

$$U_n(\theta) := \frac{1}{nP}\int_0^{nP} \frac{\rho(s,\theta)}{\sigma(s)} d\zeta_s^{\theta_0} - \frac{1}{2nP}\int_0^{nP} \rho^2(s,\theta)ds. \tag{7}$$

In the following we need the next separability condition:

(S) Assume that for each $\theta \neq \theta_0$ there exists an s such that $f(s,\theta) \neq f(s,\theta_0)$.

2.2 Consistency of the Estimator

Here we prove the consistency of the MLE $\hat{\theta}_n$. To establish this result we first check that $U_n(\cdot)$ is a contrast process in the sense of [2] (Definition 3.2.7).

Lemma 1 *Under the continuity and the periodicity of $f(\cdot,\cdot)$ and $\sigma(\cdot)$ in addition to the separability condition (S), $U_n(\cdot)$ is a contrast process.*

Proof We show that $U_n(\theta)$ converges in P_{θ_0} to the nonrandom contrast function $K(\theta,\theta_0)$ where

$$K(\theta,\theta_0) := -\frac{1}{2P}\int_0^P (\rho(s,\theta_0) - \rho(s,\theta))^2 ds + \frac{1}{2P}\int_0^P \rho^2(s,\theta_0)ds. \tag{8}$$

Recall that

$$d\zeta_s^{\theta_0} = f(s,\theta_0)ds + \sigma(s)dW_s.$$

So from the periodicity of $f(\cdot,\cdot)$ and $\sigma(\cdot)$ we have

$$\begin{aligned} U_n(\theta) &= \frac{1}{P}\int_0^P \rho(s,\theta_0)\rho(s,\theta)ds - \frac{1}{2P}\int_0^P \rho^2(s,\theta)ds \\ &\quad + \frac{1}{n}\sum_{k=0}^{n-1}\frac{1}{P}\int_0^P \rho(s,\theta)dW_s^{(kP)} \\ &= -\frac{1}{2P}\int_0^P (\rho(s,\theta_0)-\rho(s,\theta))^2 ds + \frac{1}{2P}\int_0^P \rho^2(s,\theta_0)ds \\ &\quad + \frac{1}{n}\sum_{k=0}^{n-1}\frac{1}{P}\int_0^P \rho(s,\theta)dW_s^{(kP)} \end{aligned}$$

where $W_s^{(kP)} = W_{s+kP} - W_{kP}$.

Since the random variables $\frac{1}{P}\int_0^P \rho(s,\theta)dW_s^{(kP)}$, $\{k = 0,\ldots,n-1\}$ are iid then thanks to the Strong Law of Large Numbers

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\frac{1}{P}\int_0^P \rho(s,\theta)dW_s^{(kP)} = 0 \ \ \mathrm{P}_{\theta_0}-\text{a.s.}$$

Therefore $U_n(\theta)$ converges almost surely to $K(\theta,\theta_0)$, so in probability.

Thanks to the fact that for each $\theta \neq \theta_0$ there exists an s such that $f(s,\theta) \neq f(s,\theta_0)$ the function $\theta \mapsto K(\theta,\theta_0)$ has a strict maximum when $\theta = \theta_0$ and

$$K(\theta_0,\theta_0) = \frac{1}{2P}\int_0^P \rho^2(s,\theta_0)ds.$$

This achieves the proof of the lemma.

Theorem 1 *Assume that $f(\cdot,\cdot)$ and $\sigma(\cdot)$ are continuous and periodic in t, $f(\cdot,\cdot)$ is continuously differentiable in θ and the separability condition (S) is satisfied. Then the MLE $\hat{\theta}_n$ is consistent.*

$$\hat{\theta}_n \overset{\mathrm{P}_{\theta_0}}{\to} \theta_0.$$

2.3 Consistency and Efficiency When $f(t,\theta) = \theta f(t)$

The choice of the case $f(t,\theta) = \theta f(t)$ is justified by its linearity which makes it relevant. The modèle (1), in this case, is given as $d\zeta_t = \theta f(t)dt + \sigma(t)dW_t$. In the continuous case, the non-parametric estimation of the function $f(\cdot)$ was studied [4, 11]. For the parametric estimation, thanks to (6), the likelihood is defined as

$$L_n(\theta) = \exp\left(\theta \int_0^{nP} \frac{\rho(s)}{\sigma(s)} d\zeta_s^{\theta_0} - \frac{\theta^2}{2} \int_0^{nP} \rho^2(s) ds\right).$$

The score function is equal to

$$S(\theta) := \frac{\partial \ln L_n(\theta)}{\partial \theta} = \int_0^{nP} \frac{\rho(s)}{\sigma(s)} d\zeta_s^{\theta_0} - \theta \int_0^{nP} \rho^2(s) ds.$$

This implies that the MLE $\hat{\theta}_n$ is equal to

$$\hat{\theta}_n := \frac{\int_0^{nP} \frac{\rho(s)}{\sigma(s)} d\zeta_s^{\theta_0}}{\int_0^{nP} \rho^2(s) ds} = \frac{\int_0^{nP} \frac{\rho(s)}{\sigma(s)} (\theta_0 f(s) ds + \sigma(s) dW_s)}{\int_0^{nP} \rho^2(s) ds}.$$

So we can write $\hat{\theta}_n$ as:

$$\hat{\theta}_n = \theta_0 + \frac{V_n}{I_n},$$

where

$$V_n := \int_0^{nP} \rho(s) dW_s, \; I_n := \int_0^{nP} \rho^2(s) ds.$$

Then $\hat{\theta}_n$ is unbiased. Moreover, we are going to state the almost sure convergence, mean square convergence, the asymptotic normality and the asymptotic efficiency.

2.3.1 Almost Sure Convergence, Mean Square Convergence, Asymptotic Normality

Theorem 2 *$\hat{\theta}_n$ converges almost surely to θ_0.*

Remark 1 Noticing that I_n is the quadratic variation of the martingale V_n we can directly deduce from the martingale convergence the strong convergence of $\hat{\theta}_n$.

Since

$$\mathscr{L}\left(\int_0^{nP} \rho(s) dW_s\right) = \mathscr{N}\left(0, \int_0^{nP} \rho^2(s) ds\right) \text{ and } \hat{\theta}_n - \theta_0 = \frac{\int_0^{nP} \rho(s) dW_s}{\int_0^{nP} \rho^2(s) ds},$$

we deduce that

$$\mathscr{L}\left(\hat{\theta}_n - \theta_0\right) = \mathscr{N}\left(0, I_n^{-1}\right). \tag{9}$$

So we obtain the mean square convergence as well as the asymptotic normality.

Theorem 3 *$\hat{\theta}_n$ converges in mean square to θ_0, and $\bar{\theta}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal: $\bar{\theta}_n \sim \mathcal{N}\left(0, I^{-1}\right)$, where $I = \int_0^P \rho^2(s)ds$ is the Fisher information.*

2.3.2 Asymptotic Efficiency of $\hat{\theta}_n$

To justify the relevance of this estimator we are going to establish that it is asymptotically efficient. For this purpose we use the Hájek-Le Cam inequality (see [15, 17] for further details).

We show first that the family $(\mathrm{P}_\theta^{(n)})_{\theta\in\Theta}$ is locally asymptotically normal (see Definition 1.2.1 in [15]).

Proposition 1 *Using the notations of Definition 1.2.1 in [15] $(\mathrm{P}_\theta^{(n)})_{\theta\in\Theta}$ is locally asymptotically normal at θ_0 for any θ_0.*

Theorem 4 *The estimator $\hat{\theta}_n$ is asymptotically efficient for the square error (see Definition 1.2.2 in [15]).*

Proof From (9)

$$\mathrm{E}\left[I_n(\hat{\theta}_n - \theta_0)^2\right] = 1.$$

Hence $\hat{\theta}_n$ is asymptotically efficient.

3 Drift Estimation from a Discrete Time Observation

In the second section we have assumed that the process is observed continuously throughout the interval $[0, T]$, however it is difficult to record numerically a continuous time process, and the observations generally take place at discrete moments. Hence the motivation of this section.

Recall the model of signal plus noise type that we consider

$$d\zeta_t = f(t,\theta)dt + \sigma(t)dW_t \quad t \in [0,T]. \tag{10}$$

The goal of this section is the estimation of the parameter θ.

Thanks to the simplicity of this model we are able to compute its likelihood.

Assume that the observations take place at the instants $t_i := i\Delta_n, i \in 0 \ldots n-1$ of the interval $[0, T]$, where $\Delta_n = \frac{T}{n}$. Assume also that $T = n\Delta_n = N_n P$ and that $T = n\Delta_n \to \infty$ and $\Delta_n \to 0$ when $n \to \infty$. So $N_n \to \infty$ when $n \to \infty$.

To simplify the representation denote

$$Y_i^\theta := \int_{t_i}^{t_{i+1}} \rho(s,\theta)ds + W_{t_{i+1}} - W_{t_i}. \tag{11}$$

Hence

$$Y_i^{\theta} \sim \mathcal{N}\left(\int_{t_i}^{t_{i+1}} \rho(s,\theta)ds,\, \Delta_n\right). \tag{12}$$

Therefore, the log-likelihood is given by

$$\begin{aligned} \Lambda_n(\theta) &:= \sum_{i=0}^{n-1}\left(-\ln(\sqrt{2\pi\,\Delta_n}) - \frac{1}{2\Delta_n}\left(Y_i^{\theta_0} - \int_{t_i}^{t_{i+1}} \rho(s,\theta)ds\right)^2\right) \\ &= -\frac{n}{2}\ln(2\pi\,\Delta_n) - \frac{1}{2\Delta_n}\sum_{i=0}^{n-1}\left({Y_i^{\theta_0}}^2 - 2Y_i^{\theta_0}\int_{t_i}^{t_{i+1}} \rho(s,\theta)ds\right) \\ &\quad -\frac{1}{2\Delta_n}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}} \rho(s,\theta)ds\right)^2 \end{aligned} \tag{13}$$

where θ_0 is the true value of the unknown parameter θ. Thanks to the continuity of $f(\cdot,\cdot)$ and $\sigma(\cdot)$ and the compactness of Θ there exists $\hat{\theta}_n$ such that

$$\hat{\theta}_n = \arg\sup_{\theta\in\Theta} \Lambda_n(\theta).$$

Since θ does not depend on the observation $Y_i^{\theta_0}$ then

$$\hat{\theta}_n = \arg\sup_{\theta\in\Theta} U_n(\theta),$$

where

$$U_n(\theta) := \frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1} Y_i^{\theta_0}\int_{t_i}^{t_{i+1}} \rho(s,\theta)ds - \frac{1}{2n\Delta_n^2}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}} \rho(s,\theta)ds\right)^2. \tag{14}$$

3.1 Consistency of the Estimator

As in the continuous case, to prove the convergence in P_{θ_0} of $\hat{\theta}_n$ to the true value of the parameter θ_0 we check that $U_n(\theta)$ is a contrast process in the sense of [2] (Definition 3.2.7) and then we apply a version of (Theorem 3.2.4, [2], see also Theorem 5.7 of van der Vaart [17]).

Proposition 2 *Under the continuity and the periodicity of $f(\cdot,\cdot)$ and $\sigma(\cdot)$ and under the separability condition (S), $U_n(\theta)$ is a contrast process.*

Theorem 5 *In addition to the conditions of Proposition 2 assume that $\rho(s,\theta)$ is continuously differentiable, then $\hat{\theta}_n$ converges in probability to the real parameter θ_0.*

3.2 Study of the Case $f(t,\theta) = \theta f(t)$

Now we consider the case $f(t,\theta) = \theta f(t)$ where $f(\cdot)$ is a continuous and periodic function with same period as $\sigma(\cdot)$ and θ is unknown parameter. In this case the expression of the estimator $\hat{\theta}_n$ is explicit and we show that its estimator converges in mean square and it is asymptotically normal.

From (13) the log-likelihood is given by

$$\Lambda_n(\theta) := -\frac{n}{2}\ln(2\pi\Delta_n) - \frac{1}{2\Delta_n}\sum_{i=0}^{n-1}\left(Y_i^{\theta_0} - \theta\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2, \tag{15}$$

where $\rho(s) = \frac{f(s)}{\sigma(s)}$, Y_i^{θ} is given by (11). Hence the score function $S(\theta)$ is

$$S(\theta) = \frac{1}{\Delta_n}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\rho(s)ds\left(Y_i^{\theta_0} - \theta\int_{t_i}^{t_{i+1}}\rho(s)ds\right).$$

So we get the expression of the MLE

$$\hat{\theta}_n = \frac{\sum_{i=0}^{n-1} Y_i^{\theta_0}\int_{t_i}^{t_{i+1}}\rho(s)ds}{\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2} = \theta_0 + \frac{\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\rho(s)ds\left(W_{t_{i+1}} - W_{t_i}\right)}{\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2}.$$

Then $\hat{\theta}_n$ is unbiased estimator. Furthermore, we are going to show that it converges in mean square and it is asymptotically normal.

3.2.1 Mean Square Convergence, Asymptotic Normality

Theorem 6 *$\hat{\theta}_n$ converges in mean square to θ_0, furthermore we have the rate of this convergence*

$$\lim_{n\to\infty}\mathrm{E}_{\theta_0}\left[\left|\sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0)\right|^2\right] = \left(\frac{1}{P}\int_0^P \rho^2(s)ds\right)^{-1}.$$

Since

$$\sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0) \sim \mathcal{N}\left(0, \frac{n\Delta_n^2}{\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2}\right).$$

We deduce that

$$\lim_{n\to\infty} \mathscr{L}\left(\sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0)\right) = \mathscr{N}\left(0, \left(\frac{1}{P}\int_0^P \rho^2(s)ds\right)^{-1}\right).$$

Therefore

Theorem 7 *$\sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal.*

3.2.2 Asymptotic Efficiency of $\hat{\theta}_n$

Here as in the continuous case we show that the estimator $\hat{\theta}_n$ is asymptotically efficient. We establish first that the family $(\mathrm{P}_\theta^{(n)})_{\theta\in\Theta}$ is locally asymptotically normal (see Definition 1.2.1 in [15]).

Proposition 3 *$(\mathrm{P}_\theta^{(n)})_{\theta\in\Theta}$ is locally asymptotically normal at θ_0 for any θ_0.*

Theorem 8 *The estimator $\hat{\theta}_n$ is also asymptotically efficient for the square error (see Definition 1.2.2 in [15]).*

Proof According to (26)

$$\begin{aligned}
\mathrm{E}_{\theta_0}\left[\left|\Phi_n^{-1}(\hat{\theta}_n - \theta_0)\right|^2\right] &= \Phi_n^{-2}\mathrm{E}_{\theta_0}\left[\left|\frac{\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\rho(s)ds(W_{t_{i+1}} - W_{t_i})}{\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2}\right|^2\right] \\
&= \Phi_n^{-2}\frac{\Delta_n}{\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2} \\
&= \Phi_n^{-2}\Phi_n^2 \\
&= 1.
\end{aligned}$$

Hence $\hat{\theta}_n$ is asymptotically efficient.

4 Simulation

As an example we illustrate this estimator $\hat{\theta}_n$ with different values of θ_0, $\theta_0 = 0$, $\theta_1 = 1$ and $\theta = -4$.

Assume that the sample size $T = nP = 1000$, the period $P = 1$. First, we consider the next model

$$d\zeta_t = \theta\cos(2\pi t) + dW_t, \text{ i.e. } f(t) = \cos(2\pi t),\ \sigma(t) = 1,$$

and the discretization step δ .

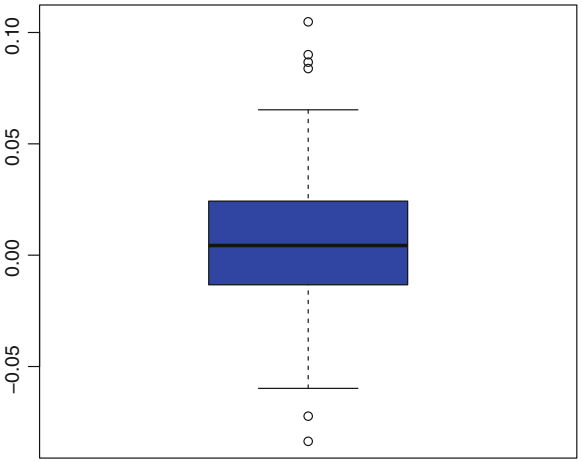


Fig. 1 Boxplot of the values of the estimator $\hat{\theta}_n$ from 100 repetitions

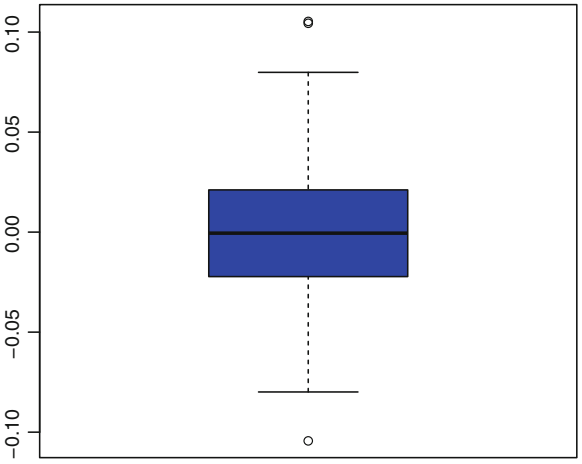


Fig. 2 Boxplot of the values of the estimator $\hat{\theta}_n$ from 1000 repetitions

For $\theta_0 = 0, \delta = 10^{-2}$ in Figs. 1 and 2 we see that all the values are around 0 and the interquartile range is less than 0.2.

For $\theta_0 = 1, \delta = 10^{-3}$ in Fig. 3 we see the importance of the step of discretization δ, the median is equal to 1 and the interquartile range is equal to 0.1.

Now we illustrate this estimator with another model (Fig. 4):

$$f(t) = \cos(2\pi t),\ \sigma(t) = 2 + \sin(2\pi t),\ \theta_0 = -4,\ \delta = 10^{-3}.$$

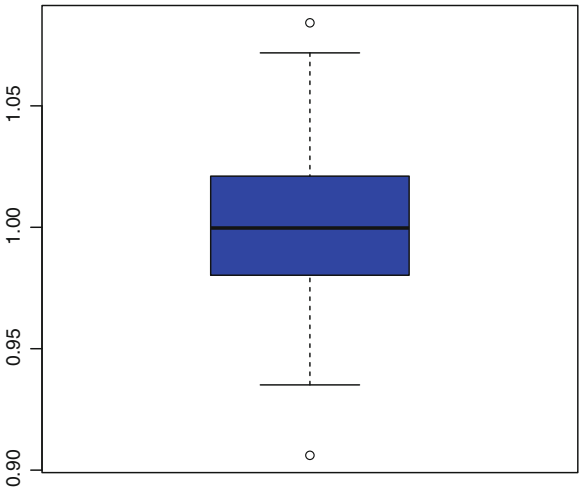


Fig. 3 Boxplot of the values of the estimator $\hat{\theta}_n$ from 1000 repetitions

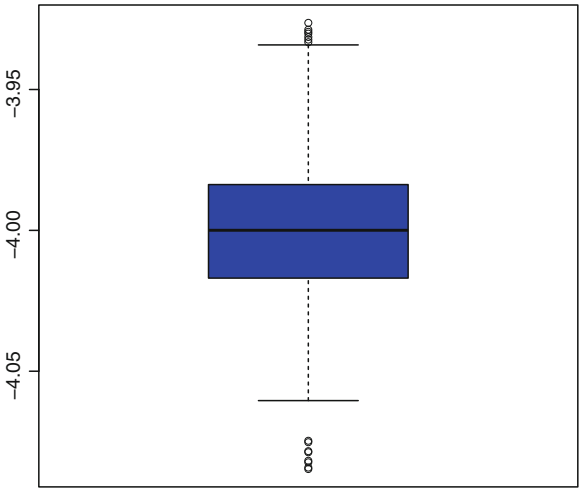


Fig. 4 Boxplot of the values of the estimator $\hat{\theta}_n$ from 1000 repetitions

As we can see the median of the 1000 values is -4 and the interquartile range is around 0.1.

When $f(t) = \sigma(t) = 1, \theta = 1$, Fig. 5 shows that all the values are next to 1. In the previous cases although the dispersion around the median is not very big, the degradation with respect to the case $f(t) = \sigma(t) = 1$ is due to the variations of $f(t)$ and $\sigma(t)$.

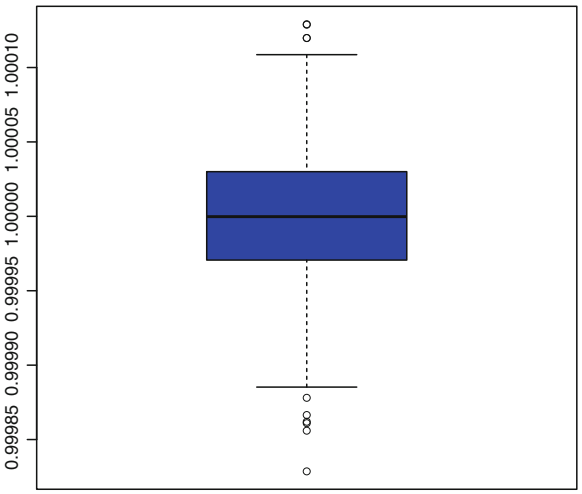


Fig. 5 Boxplot of the values of the estimator $\hat{\theta}_n$ from 1000 repetitions

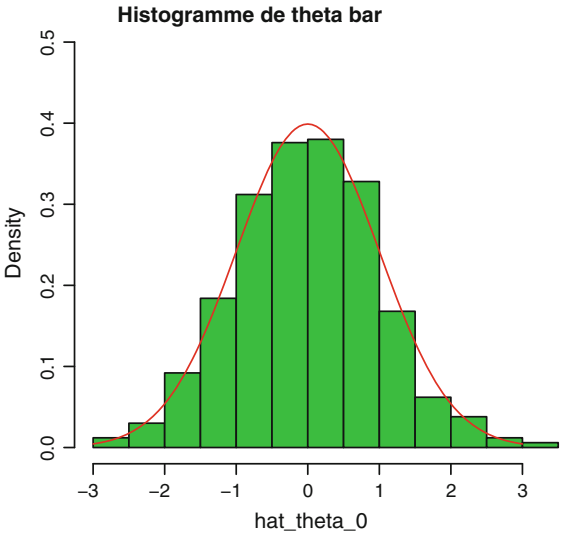


Fig. 6 Histogram $\bar{\theta}_n$, $\theta_0 = 0$ from 1000 repetitions

Now we check the asymptotic normality property. From the Theorem 3

$$\bar{\theta}_n := \sqrt{n}(\hat{\theta}_n - \theta_0), \lim_{n\to\infty} \mathscr{L}\left(\sqrt{n}(\hat{\theta}_n - \theta_0)\right) = \mathscr{N}\left(0, \frac{1}{\int_0^P \rho^2(s)ds}\right).$$

In Figs. 6, 7 and 8 we note that the histograms obtained by the 1000 repetitions of the simulations fit to the Gaussian distribution.

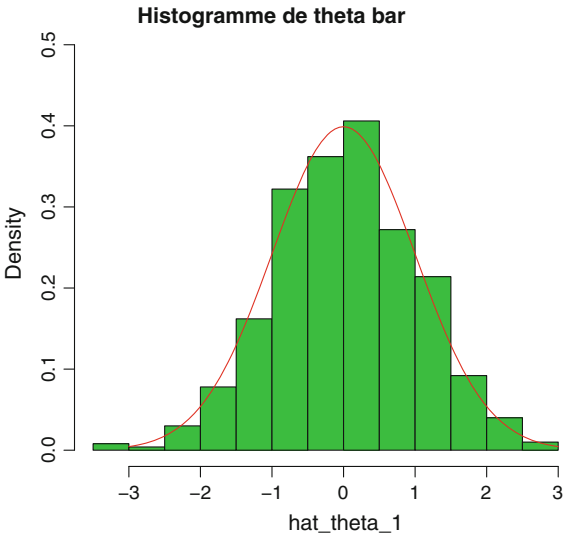


Fig. 7 Histogram of $\bar{\theta}_n$, $\theta_0 = 1$ from 1000 repetitions

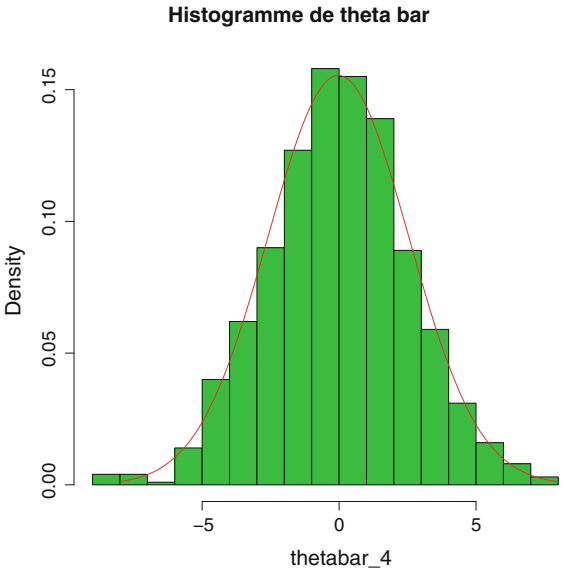


Fig. 8 Histogram of $\bar{\theta}_n$, $\theta_0 = -4$ from 1000 repetitions

5 Conclusion

We have considered a model of type signal plus noise given by

$$d\zeta_t = f(t,\theta)dt + \sigma(t)dW_t.$$

From a continuous time observation throughout the interval $[0, T]$ we have built an estimator $\hat{\theta}_n$ of the unknown parameter θ, using the method of maximum likelihood. We have established that this estimator converges in probability. When $f(t,\theta) = \theta f(t)$, we have given the expression of $\hat{\theta}_n$ and we have established its mean square convergence, almost sure convergence, asymptotic normality and asymptotic efficiency. From a discrete time observation on $[0, T]$ we have computed the likelihood function and we have shown the same convergences as in the continuous case except the strong consistency convergence.

As a prospect I am going to consider model (1) but with $\sigma(\cdot)$ which can take the value 0. I plan also to study the estimation of the parameters θ_i from the observations of ζ_t following the model

$$d\xi_t = (\theta_1 f_1(t) + \theta_2 f_2(t) \ldots \theta_n f_n(t))\, dt + \sigma(t)dW_t.$$

Acknowledgments I would like to thank the region of Brittany for funding my Ph.D. scholarship.

Appendix

Proof **Theorem** 1

We apply (Theorem 3.2.4, [2]) which gives the convergence in probability of the maximum of the contrast process $U_n(\theta)$. See also (Theorem 5.7 of van der Vaart 2005). For this purpose we need to show the two following conditions:

(i) Θ is a compact of $\mathbb{R}$, the functions $\theta \mapsto U_n(\theta)$, $\theta \mapsto K(\theta, \theta_0)$ are continuous.

(ii) for all $\varepsilon > 0$, there exists $\eta > 0$ such that

$$\lim_{n\to\infty} \mathrm{P}_\theta \left(\sup_{|\theta-\theta'|<\eta} \left|U_n(\theta) - U_n(\theta')\right| > \varepsilon \right) = 0.$$

The first condition is readily obtained from the hypothesis on $f(\cdot,\cdot)$, $\sigma(\cdot)$ and Θ, and from the definitions of $U_n(\cdot)$ and $K(\cdot,\cdot)$.

Below we prove the second condition

$$\begin{aligned} U_n(\theta) - U_n(\theta') &= \frac{1}{2nP}\int_0^{nP} \big(\rho(s,\theta) - \rho(s,\theta')\big)\big(2\rho(s,\theta_0) - \rho(s,\theta) - \rho(s,\theta')\big)\,ds \\ &\quad + \frac{1}{nP}\int_0^{nP} (\rho(s,\theta) - \rho(s,\theta'))dW_s \\ &= \frac{1}{2P}\int_0^{P} \big(\rho(s,\theta) - \rho(s,\theta')\big)\big(2\rho(s,\theta_0) - \rho(s,\theta) - \rho(s,\theta')\big)\,ds \\ &\quad + \frac{1}{nP}\int_0^{P} (\rho(s,\theta) - \rho(s,\theta'))dW_s \\ &:= A_1(\theta,\theta') + A_2(\theta,\theta') \end{aligned} \tag{16}$$

where $A_1(\theta,\theta')$ is the nonrandom term of this equality and $A_2(\theta,\theta')$ is the random term.

The absolute value of the nonrandom term of this equality is bounded by a multiple of η where $|\theta - \theta'| \le \eta$ and the second term converges in mean to 0 when $n \to \infty$. Indeed

As $\rho(\cdot,\cdot)$ is continuously differentiable then

$$\rho(s,\theta) - \rho(s,\theta') = (\theta - \theta')\rho'(s,\theta_1),\ \theta_1 \in [\theta,\theta'],$$

where $\rho'(\cdot,\cdot)$ is the derivative of $\rho(\cdot,\cdot)$ with respect to θ. Denote

$$C_0 := \sup_{\theta\in\Theta,\ s\in[0,P]} \big|\rho'(s,\theta)\big|.$$

Thus

$$\begin{aligned} \sup_{\theta,\theta'\in\Theta,\ s\in[0,P]} \big|2\rho(s,\theta_0) - \rho(s,\theta) - \rho(s,\theta')\big| &= 2\sup_{\theta\in\Theta,\ s\in[0,P]} |\rho(s,\theta_0) - \rho(s,\theta)| \\ &\le C_1 \sup_{\theta\in\Theta} |\theta_0 - \theta| \\ &\le C \end{aligned}$$

where C is positive constant because Θ is a compact. So for the nonrandom part

$$\sup_{|\theta-\theta'|<\eta} \big|A_1(\theta,\theta')\big| \le \sup_{|\theta-\theta'|<\eta} \frac{|\theta-\theta'|}{2P}\int_0^P C ds \le \frac{C}{2}\eta \tag{17}$$

Now show that the random part converges in mean to 0. First

$$A_2(\theta,\theta') = \frac{1}{nP}\int_0^P (\rho(s,\theta) - \rho(s,\theta'))dW_s = \int_\theta^{\theta'} \frac{1}{nP}\int_0^P \rho'(s,u)dW_s\,du.$$

The set Θ being compact

$$\begin{aligned}
\mathrm{E}_{\theta_0}\left[\sup_{|\theta-\theta'|<\eta}\left|A_2(\theta,\theta')\right|\right] &\leq \mathrm{E}_{\theta_0}\left[\sup_{|\theta-\theta'|<\eta}\int_{\theta}^{\theta'}\left|\frac{1}{nP}\int_0^P \rho'(s,u)dW_s\right| du\right] \\
&\leq \mathrm{E}_{\theta_0}\left[\int_{\Theta}\left|\frac{1}{nP}\int_0^P \rho'(s,u)dW_s\right| du\right] \\
&\leq \int_{\Theta}\mathrm{E}_{\theta_0}\left[\left|\frac{1}{nP}\int_0^P \rho'(s,u)dW_s\right|\right] du \\
&\leq \int_{\Theta}\left(\mathrm{E}_{\theta_0}\left[\left|\frac{1}{nP}\int_0^P \rho'(s,u)dW_s\right|^2\right]\right)^{\frac{1}{2}} du \\
&= \int_{\Theta}\left(\frac{1}{n^2P^2}\int_0^{nP}\rho'^2(s,u)ds\right)^{\frac{1}{2}} du \\
&= \frac{1}{\sqrt{nP}}\int_{\Theta}\left(\frac{1}{P}\int_0^P \rho'^2(s,u)ds\right)^{\frac{1}{2}} du.
\end{aligned}$$

Then

$$\lim_{n\to\infty}\mathrm{E}_{\theta_0}\left[\sup_{|\theta-\theta'|<\eta}\left|A_2(\theta,\theta')\right|\right] = 0. \tag{18}$$

According to the equality (16) we have for all $\varepsilon > 0$

$$\begin{aligned}
\mathrm{P}_{\theta_0}\left(\sup_{|\theta-\theta'|<\eta}\left|U_n(\theta)-U_n(\theta')\right| > \varepsilon\right) &\leq \mathrm{P}_{\theta_0}\left(\sup_{|\theta-\theta'|<\eta}\left|A_1(\theta,\theta')\right| > \frac{\varepsilon}{2}\right) \\
&\quad + \mathrm{P}_{\theta_0}\left(\sup_{|\theta-\theta'|<\eta}\left|A_2(\theta,\theta')\right| > \frac{\varepsilon}{2}\right)
\end{aligned}$$

Since

$$\lim_{n\to\infty}\mathrm{E}_{\theta_0}\left[\sup_{|\theta-\theta'|<\eta}\left|A_2(\theta,\theta')\right|\right] = 0$$

then

$$\lim_{n\to\infty}\mathrm{P}_{\theta_0}\left(\sup_{|\theta-\theta'|<\eta}\left|A_2(\theta,\theta')\right| > \frac{\varepsilon}{2}\right) = 0.$$

Thanks to (17) is suffice, for all $\varepsilon > 0$, to take $\eta = \frac{\varepsilon}{2C}$ to have

$$\sup_{|\theta-\theta'|<\eta} \left|A_1(\theta,\theta')\right| < \frac{\varepsilon}{2}.$$

So for all $\varepsilon > 0$, there exists $\eta > 0$ such that

$$\lim_{n\to\infty} \mathrm{P}_{\theta_0}\left(\sup_{|\theta-\theta'|<\eta} \left|U_n(\theta) - U_n(\theta')\right| > \varepsilon\right) = 0.$$

Therefore, the second condition is fulfilled. So the MLE $\hat{\theta}_n$ is consistent.

Proof **Theorem 2**

Recall that $f(\cdot)$ and $\sigma(\cdot)$ are continuous and periodic with the same period $P > 0$. On one hand

$$I_n = \int_0^{nP} \rho^2(s)ds = n\int_0^P \rho^2(s)ds.$$

So

$$\lim_{n\to\infty} \frac{I_n}{n} = \int_0^P \rho^2(s)ds.$$

On the other hand

$$V_n = \sum_{k=0}^{n-1} \int_{kP}^{(k+1)P} \rho(s)dW_s = \sum_{k=0}^{n-1} \int_0^P \rho(s)dW_s^{(kP)}.$$

As

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \int_0^P \rho(s)dW_s^{(kP)} = \mathrm{E}_{\theta_0}\left[\int_0^P \rho(s)dW_s^{(kP)}\right] = 0 \ \ \mathrm{P}_{\theta_0} - a.s.$$

we deduce that

$$\lim_{n\mapsto\infty} \hat{\theta}_n - \theta_0 = \lim_{n\mapsto\infty} \hat{\theta}_n - \theta_0 = \lim_{n\mapsto\infty} \frac{\frac{1}{n}V_n}{\frac{1}{n}I_n} = 0 \ \ \mathrm{P}_{\theta_0} - a.s.$$

Proof **Proposition 1**

Denote

$$\Phi_n := I_n^{-\frac{1}{2}} = \left(\int_0^{nP} \rho^2(s)ds\right)^{-\frac{1}{2}},$$

$$\Delta_n(\zeta^\theta) := I_n^{-\frac{1}{2}} \int_0^{nP} \rho(s) dWs.$$

According to (6) we have

$$\begin{aligned} Z_n(u) &:= \frac{L_n(\theta_0 + \Phi_n u)}{L_n(\theta_0)} \\ &= \exp\left(\Phi_n u \int_0^{nP} \frac{\rho(s)}{\sigma(s)} d\zeta_s^{\theta_0} - \frac{(\theta_0 + \Phi_n u)^2 - \theta_0^2}{2} I_n \right) \\ &= \exp\left(\Phi_n u \int_0^{nP} \frac{\rho(s)}{\sigma(s)} d\zeta_s^{\theta_0} - \frac{\Phi_n^2 u^2}{2} I_n - \theta_0 \Phi_n u I_n \right). \end{aligned}$$

Since

$$d\zeta_t^{\theta_0} = \theta_0 f(t)dt + \sigma(t)dW_t$$

then we have

$$\begin{aligned} Z_n(u) &= \exp\left(\theta_0 u I_n^{-\frac{1}{2}} I_n + u I_n^{-\frac{1}{2}} \int_0^{nP} \rho(s) dW_s + \frac{1}{2} u^2 - \theta_0 u I_n^{\frac{1}{2}} \right) \\ &= \exp\left(u I_n^{-\frac{1}{2}} \int_0^{nP} \rho(s) dW_s - \frac{1}{2} u^2 \right) \\ &= \exp\left(u \Delta_n(\zeta^\theta) - \frac{1}{2} u^2 \right). \end{aligned} \tag{19}$$

As

$$\mathscr{L}\left(\int_0^{nP} \rho(s) dW_s \right) = \mathscr{N}(0, I_n)$$

then we have

$$\mathscr{L}\left\{ \Delta_n(\zeta^\theta) | \mathrm{P}_{\theta_0} \right\} = \mathscr{N}(0, 1).$$

So from (19) the family $(\mathrm{P}_\theta^{(n)})_{\theta \in \Theta}$ is locally asymptotically normal at θ_0 for every θ_0.

Proof **Proposition** 2 We show that $U_n(\theta)$ converges in P_{θ_0} to the contrast function $K(\theta, \theta_0)$ defined by (8). To do this we use the next lemma

Lemma 2 *For the continuous and periodic function* $\rho(\cdot, \cdot)$ *and* $\Delta_n \to 0$ *when* $n \to \infty$ *we have*

$$\lim_{n \to \infty} \frac{1}{2n\Delta_n^2} \sum_{i=0}^{n-1} \left(\int_{t_i}^{t_{i+1}} \rho(s, \theta) - \int_{t_i}^{t_{i+1}} \rho(s, \theta_0) \right)^2 ds = \frac{1}{2P} \int_0^P (\rho(s, \theta_0) - \rho(s, \theta))^2 ds.$$

$$\lim_{n\to\infty}\frac{1}{2n\Delta_n^2}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s,\theta_0)ds\right)^2=\frac{1}{2P}\int_0^P\rho^2(s,\theta_0)ds.$$

We continue the proof of Proposition 2, recall that

$$Y_i^{\theta_0}=\int_{t_i}^{t_{i+1}}\rho(s,\theta_0)ds+W_{t_{i+1}}-W_{t_i}.$$

So

$$\begin{aligned}U_n(\theta):=&\frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1}Y_i^{\theta_0}\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds-\frac{1}{2n\Delta_n^2}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds\right)^2\\&=\frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds\int_{t_i}^{t_{i+1}}\rho(s,\theta_0)ds-\frac{1}{2n\Delta_n^2}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds\right)^2\\&\quad+\frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds(W_{t_{i+1}}-W_{t_i})\\&=-\frac{1}{2n\Delta_n^2}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s,\theta)-\int_{t_i}^{t_{i+1}}\rho(s,\theta_0)\right)^2ds\qquad(20)\\&\quad+\frac{1}{2n\Delta_n^2}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s,\theta_0)ds\right)^2+\frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds(W_{t_{i+1}}-W_{t_i}).\end{aligned}$$

To complete the proof and thanks to Lemma 2 it remains to establish the convergence in P_{θ_0} probability of the last part of (20). For this purpose we show the convergence in mean square.

From the independence of the increment of the Brownian motion $W_{t_{i+1}}-W_{t_i}$ we can write

$$\begin{aligned}&\mathrm{E}_{\theta_0}\left[\left|\frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds(W_{t_{i+1}}-W_{t_i})\right|^2\right]\\&=\frac{1}{n^2\Delta_n^4}\sum_{i=0}^{n-1}\mathrm{E}_{\theta_0}\left[\left|\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds(W_{t_{i+1}}-W_{t_i})\right|^2\right]\\&=\frac{1}{n^2\Delta_n^3}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds)^2\right)\\&=\frac{1}{n\Delta_n}\frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds\right)^2.\end{aligned}$$

Thanks to (20) we deduce that

$$\lim_{n\to\infty} \mathrm{E}_{\theta_0}\left[\left|\frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds(W_{t_{i+1}}-W_{t_i})\right|^2\right]=0.$$

So the contrast process $U_n(\theta)$ converges in mean square to the contrast function $K(\theta,\theta_0)$ which has a strict maximum when $\theta=\theta_0$

$$K(\theta_0,\theta_0)=\frac{1}{2P}\int_0^P \rho^2(s,\theta_0)ds.$$

This achieves the proof of Proposition 2.

Proof **Theorem 5**

As in Theorem 1 we apply (Theorem 3.2.4 in [2]). By their definitions, $U_n(\theta)$ and $K(\theta,\theta_0)$ satisfy the first condition

For the second condition

$$\begin{aligned}
&U_n(\theta)-U_n(\theta') \qquad (21)\\
&=\frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\rho(s,\theta_0)ds\int_{t_i}^{t_{i+1}}(\rho(s,\theta)-\rho(s,\theta'))ds-\frac{1}{2n\Delta_n^2}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s,\theta)ds\right)^2\\
&\quad+\frac{1}{2n\Delta_n^2}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s,\theta')ds\right)^2+\frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}(\rho(s,\theta)-\rho(s,\theta'))ds(W_{t_{i+1}}-W_{t_i})\\
&=\frac{1}{2n\Delta_n^2}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}(\rho(s,\theta)-\rho(s,\theta'))ds\int_{t_i}^{t_{i+1}}(2\rho(s,\theta_0)-\rho(s,\theta)-\rho(s,\theta'))ds\\
&\quad+\frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}(\rho(s,\theta)-\rho(s,\theta'))ds(W_{t_{i+1}}-W_{t_i})\\
&:=B_1(\theta,\theta')+B_2(\theta,\theta') \qquad (22)
\end{aligned}$$

As in the continuous case we can establish that the absolute value of the nonrandom part $B_1(\theta,\theta')$ is bounded by a multiple of η and the random part $B_2(\theta,\theta')$ converges in mean to 0. Hence we have the second condition.

So

$$\hat{\theta}_n \overset{\mathrm{P}_{\theta_0}}{\to} \theta_0.$$

Proof **Theorem 6**

$$\begin{aligned}
\mathrm{E}_{\theta_0}\left[\left|\sqrt{n\Delta_n}(\hat{\theta}_n-\theta_0)\right|^2\right] &= n\Delta_n \mathrm{E}_{\theta_0}\left[\left|\frac{\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\rho(s)ds\left(W_{t_{i+1}}-W_{t_i}\right)}{\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2}\right|^2\right] \\
&= \frac{n\Delta_n \mathrm{E}_{\theta_0}\left[\left|\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\rho(s)ds\left(W_{t_{i+1}}-W_{t_i}\right)\right|^2\right]}{\left(\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2\right)^2} \\
&= \frac{n\Delta_n^2\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2}{\left(\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2\right)^2} \\
&= \frac{n\Delta_n^2}{\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2}.
\end{aligned}$$

From equality (20) with $f(s,\theta)=\theta f(s)$ we have

$$\lim_{n\to\infty}\frac{1}{n\Delta_n^2}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2=\frac{1}{P}\int_0^P\rho^2(s)ds.$$

So

$$\lim_{n\to\infty}\mathrm{E}_{\theta_0}\left[\left|\sqrt{n\Delta_n}(\hat{\theta}_n-\theta_0)\right|^2\right]=\left(\frac{1}{P}\int_0^P\rho^2(s)ds\right)^{-1}.$$

Proof **Proposition 3**

Denote

$$\Phi_n:=\sqrt{\Delta_n}\left(\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds\right)^2\right)^{-\frac{1}{2}}, \tag{23}$$

$$\Delta_n(\zeta^\theta):=\frac{\Phi_n}{\Delta_n}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\rho(s)ds(W_{t_{i+1}}-W_{t_i})\right). \tag{24}$$

According to (12) and (13) we have

$$
\begin{aligned}
Z_n(u) &:= \frac{L_n(\theta_0 + \Phi_n u)}{L_n(\theta_0)} \\
&= \frac{\prod_{i=0}^{n-1} \frac{1}{\sqrt{2\pi\Delta_n}} \exp\left(-\frac{1}{2\Delta_n}\left(Y_i^{\theta_0} - (\theta_0 + \Phi_n u)\int_{t_i}^{t_{i+1}} \rho(s)ds\right)^2\right)}{\prod_{i=0}^{n-1} \frac{1}{\sqrt{2\pi\Delta_n}} \exp\left(-\frac{1}{2\Delta_n}\left(Y_i^{\theta_0} - \theta_0 \int_{t_i}^{t_{i+1}} \rho(s)ds\right)^2\right)} \\
&= \prod_{i=0}^{n-1} \exp\left(-\frac{1}{\Delta_n}\left(Y_i^{\theta_0}\Phi_n u \int_{t_i}^{t_{i+1}} \rho(s)ds - \theta_0 \Phi_n u \left(\int_{t_i}^{t_{i+1}} \rho(s)ds\right)^2\right)\right) \\
&\quad \times \prod_{i=0}^{n-1} \exp\left(\frac{1}{2\Delta_n}\Phi_n^2 u^2 \left(\int_{t_i}^{t_{i+1}} \rho(s)ds\right)^2\right)
\end{aligned}
$$

Since

$$
Y_i^{\theta_0} = \theta_0 \int_{t_i}^{t_{i+1}} \rho(s)ds + W_{t_{i+1}} - W_{t_i},
$$

then thanks to (23) and (24) we have

$$
\begin{aligned}
Z_n(u) &= \prod_{i=0}^{n-1} \exp\left(\frac{u\Phi_n}{\Delta_n}\int_{t_i}^{t_{i+1}} \rho(s)ds(W_{t_{i+1}} - W_{t_i}) - \frac{u^2\Phi_n^2}{2\Delta_n}\left(\int_{t_i}^{t_{i+1}} \rho(s)ds\right)^2\right) \\
&= \exp\left(\frac{u\Phi_n}{\Delta_n}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} \rho(s)ds(W_{t_{i+1}} - W_{t_i}) - \frac{u^2\Phi_n^2}{2\Delta_n}\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}} \rho(s)ds\right)^2\right) \\
&= \exp\left(u\Delta_n(\zeta^\theta) - \frac{1}{2}u^2\right) \qquad (25)
\end{aligned}
$$

From the independence of the increments $W_{t_{i+1}} - W_{t_i}$, $i = 0, \ldots, n-1$ and (23)

$$
\begin{aligned}
\mathscr{L}\left(\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} \rho(s)ds(W_{t_{i+1}} - W_{t_i})\right) &= \mathscr{N}\left(0, \Delta_n \sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}} \rho(s)ds\right)^2\right) \\
&= \mathscr{N}\left(0, \Phi_n^{-2}\Delta_n^2\right). \qquad (26)
\end{aligned}
$$

So from (24)

$$
\mathscr{L}\left\{\Delta_n(\zeta^\theta)|\mathrm{P}_{\theta_0}\right\} = \mathscr{N}(0, 1).
$$

Hence from (25) the family $(\mathrm{P}_\theta^{(n)})_{\theta\in\Theta}$ is locally asymptotically normal at θ_0 for every θ_0.

References

1. Collet, P., & Martinez, S. (2008). Asymptotic velocity of one dimensional diffusions with periodic drift. *Journal of Mathematical Biology*, *56*, 765–792.
2. Dacunha-Castelle, D., & Duflo, M. (1983). *Probabilité et Statistique 2, Problèmes à Temps Continu*. Paris: Masson.
3. Dacunha-Castelle, D., & Florens-Zmirou, D. (1986). Estimation of the coefficient of a diffusion from discrete observations. *Stochastics*, *19*, 263–284.
4. Dehay, D., & El Waled, K. (2013). Nonparametric estimation problem for a time-periodic signal in a periodic noise. *Statistics and Probability Letters*, *83*, 608–615.
5. Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, *20*, 547–575
6. Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion process from discrete observations. *Statistics*, *21*, 99–116.
7. Harison, V. (1996). Drift estimation of a certain class of diffusion processes from discrete observations. *Computers and Mathematics with Applications*, *31*(6), 121–133.
8. Has'minskiĭ, R.Z. (1980). Stochastic stability of differential equations. Sijthoff & Noordhoff, Alphaen aan den Rijn (The Netherlands).
9. Höpfner, R. (2007). On a set of data for the membrane in a neuron. *Mathematical Biosciences*, *207*, 275–301.
10. Höpfner, R. (2014). *Asymptotic statistics*. Berlin: De Gruyter
11. Ibragimov, I. A., & Has'minskii, R. Z. (1981). *Statistical estimation, asymptotic theory*. New York: Springer.
12. Jankunas, A., & Khas'minskiĭ, R. Z. (1997). Estimation of parameters of linear homogeneous stochastic differential equations. *Stochastic Processes and Their Applications*, *72*, 205–219.
13. Karatzas, I., & Shreve, S. E. (1991). *Brownian motion and stochastic calculus* (2nd ed.). New York: Springer.
14. Kasonga, R. D. (1988). The consistency of a non-linear least squares estimator from diffusion processes. *Stochastic Processes and Their Applications*, *30*, 263–275.
15. Kutoyants, Yu. (1984). *Parameter estimation for stochastic processes. In Research and Exposition in mathematically, Vol. 6*. Berlin: Heldermann-Verlag.
16. Liptser, R. S., & Shiryaev, A. N. (2001). *Statistics of random processes. I general theory*. Berlin: Springer.
17. van der Vaart, A. W. (2005). *Asymptotic statistics*. Madrid: Cambridge University Press.

Damage Assessment of Rolling Element Bearing Using Cyclostationary Processing of AE Signals with Electromagnetic Interference

David Quezada Acuña and Cristián Molina Vicuña

1 Introduction

Rolling element bearings are the most widespread machine component, present in almost every rotating machine. The normal failure mode of rolling element bearings is due to Hertzian contact fatigue [18]. The failure initiates typically as subsuperficial microcracks that eventually reach the surface and create localized damaged zones. This can occur in the inner race, outer race and/or rolling elements. The interaction between the bearing parts and the damaged zone produce changes in some dynamic variables, such as vibrations and acoustic emissions.

Traditionally, the most used methods for bearing diagnosis are those based on the analysis of the Fourier magnitude spectrum of the vibration envelope signal. Among several procedures leading to obtain the envelope signal, the most studied has been the High Frequency Resonant Technique (HFRT) method ([11, 13]). other similar methods—is the selection of the frequency band where the filter operation is performed. The selected band must present the highest signal-to-noise ratio between the signal component due to the bearing defect and the rest of the signal (noise). The method is actually critically sensitive to the correct selection of the filter band. Erroneous bearing failure assessment can be the result of the incorrect selection of the filter's frequency band. The hypothesis of linear behaviour of the bearing/bearing housing system suggests that the optimal frequency band is around some resonant frequency, which can be detected by means of impact tests or PSD-difference [17, 19]. However, the former needs additional measurements and the latter requires historical data, which can be not available in some cases.

D.Q. Acuña · C.M. Vicuña (✉)
Laboratorio de Vibraciones Mecánicas, Universidad de Concepción,
Edmundo Larenas 270 (int.), Concepción, Chile
e-mail: crimolin@udec.cl

D.Q. Acuña
e-mail: dquezada@udec.cl

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods - II*,
Applied Condition Monitoring 3, DOI 10.1007/978-3-319-16330-7_3

In recent years, several signal processing tools have been developed in order to automate the selection of the optimum frequency band for the filtering stage. In principle, these methods are able to find defect signatures without the need of any additional data than the signal itself. Among these methods are cyclostationary-based signal processing tools, auto-regressive filter, Kurtogram, etc. [15].

Acoustic emission (AE) is another technique which is useful for bearing fault analysis. AE are high-frequency, structure-borne elastic waves, transient in nature, which travel from their sources in all directions within materials and/or on their surfaces. The fundamental theory of AE is that of elastic wave propagation in continuum [1, 9]. Despite the fact that AE were introduced long ago for bearing monitoring [6], the AE technique is not as popular as the vibration technique, even though it can be advantageous in some situations, e.g. when shafts rotate at low speeds. The main drawbacks of the AE technique are the requirements of a continuous path between source and sensor, and difficulties for sensor mounting [10]. On the other hand, the vibration technique is more flexible and does only require the sensor to be firmly attached to the surface, e.g. by using a magnet mounting base.

2 Problem Description

Rolling element bearings are typically installed with the inner ring fixed to the rotating shaft. The outer ring is supported by the bearing housing. The AE sensor is placed on the outer part of the bearing housing. A flat zone can be prepared on the housing, so as to allow a better placement of the sensor. A viscous substance is included between the housing surface and the sensing element of the sensor to improve the AE transmission between them. This situation describes the typical measurement procedure of AE in bearings. Even though much care is taken in the procedure of sensor mounting on the bearing housing, there is still an important discontinuity between the AE source (i.e. the damaged zone of the bearing) and the sensor, which is the interface between the outer ring and the sensor housing. The contact between both surfaces is dry and partially filled with air. Unfortunately, the AE technique is highly sensitive to interfaces of this type, because most part of the AE waves that reach the interface are reflected to the first medium (the outer ring), and only a portion is refracted to the second medium (the bearing housing). That is, the AE waves transmitted to the bearing housing, and from there to the sensor, have lost a significant amount of energy and amplitude due to the interface. The high reflection and low refraction coefficients are a result of the impedance difference between the elements of the dry contact interface [12]. For this reason, some researchers have measured AE with the sensor mounted directly on the outer race [8, 20], but this can hardly be performed in practice.

On the other hand, AE sensors are in practice very prone to receive electromagnetic interference (EMI), with the result of signal contamination. Depending on the intensity of the AE generated at its source, the attenuation due to the outer ring/bearing housing interface, and the intrinsic attenuation due to wave propagation,

this can result in transmitted AE waves whose amplitudes match, or are lower than the amplitudes of the EMI contamination.

In the case presented in the next section, we are interested in studying the possibilities of using AE measured on the bearing housing to estimate the size of a seeded defect on the outer race of one tapered roller bearing. For these purposes, AE waveform analysis is performed. However, it is not possible to directly proceed with waveform analysis, because the measured signal is strongly contaminated with EMI. The EMI source in this case is the drive system, composed of an AC motor and a frequency converter. This is a typical source of EMI to which AE instrumentation would usually be exposed to in practice.

3 Experimental Set-up

The test rig, shown in Fig. 1 (element description in Table 1), has a 35 mm diameter shaft, mounted on two tapered roller bearings type 30207. Radial load is applied on the shaft, in the mid-point between the supports, by using a scissor jack which was calibrated to quantify the effective load applied to the shaft. An AC motor controlled by a frequency converter is connected to the shaft through flexible coupling.

AE were measured using a broadband piezoelectric transducer Piezotron 8152B2, which is located on the free-side bearing housing, as shown in Fig. 2. As observed, vibrations were measured parallel to AE, but we do not refer to them in this paper. Also, it can be seen that a flat surface was milled on the bearing housing to allow for better sensor mounting. Silicon grease was used to build a coupling film between the sensing element and the bearing housing.

The seeded defect is located on the surface of the outer race of the free-side bearing. It has the form of a line orthogonal to the rolling direction of the rolling elements, and extends across the complete width of the outer race (Fig. 3a). The width of the defect line is around 0.4–0.5 mm. Details of the damaged zone can be seen in

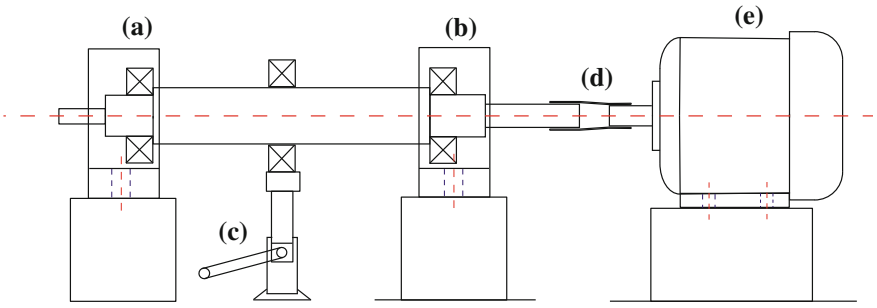


Fig. 1 Bearing test rig schematic

Table 1 Description of test rig components

Component (Fig. 1)	Description
(a)	Free-side bearing, 30207
(b)	Motor-side bearing, 30207
(c)	Calibrated scissor jack
(d)	Flexible coupling
(e)	Three-phase induction motor, 1.5 HP, controlled by a frequency converter

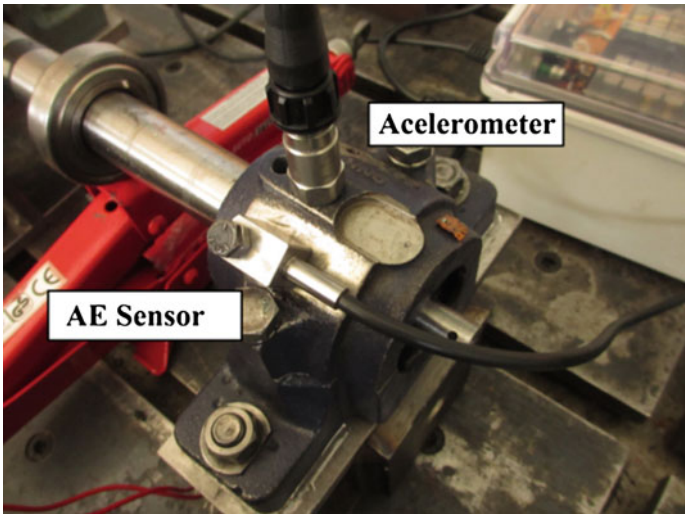


Fig. 2 Sensor setup

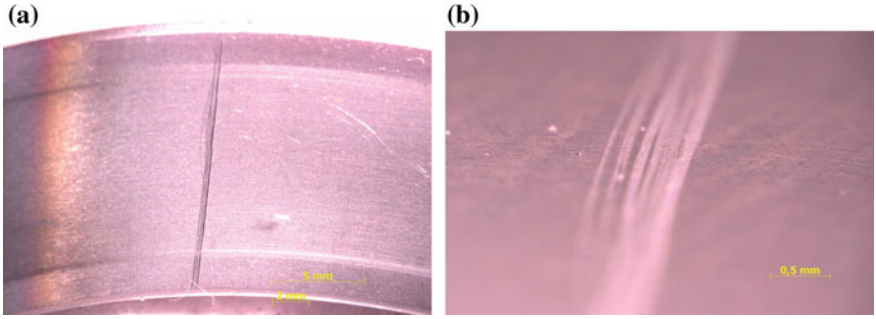


Fig. 3 Seeded defect on the outer ring of the free-side bearing. Amplification factors of **a** 0.65X, **b** 4.0X

the figures obtained with a microscope presented in Fig. 3. According to the bearing kinematics, the fault frequency for stationary outer race is 7.12 times the rotational frequency of the inner race.

4 Results

In the measured signals, the EMI and the AE due to the defect have the same order of magnitude. As explained before, this is because much of the energy of the AE bursts generated by the interaction between the rolling elements and the defect is not transmitted across the outer ring/bearing housing interface. Figure 4 (top) shows a portion of the AE measured with the faulty bearing, where the strong presence of EMI is evidenced. The slightly different zone between 1.8 and 2.7 s corresponds to the AE activity due to the defect. For comparison purposes, Fig. 4 (bottom) shows the AE measured with a non-faulty bearing. Note that the AE portion shown has a total length of 5 ms. This portion was carefully chosen to present the EMI and the AE due to the defect in a joint form. When observing a larger portion, the AE due to the defect are not recognized, because the EMI are dominant. Actually, the

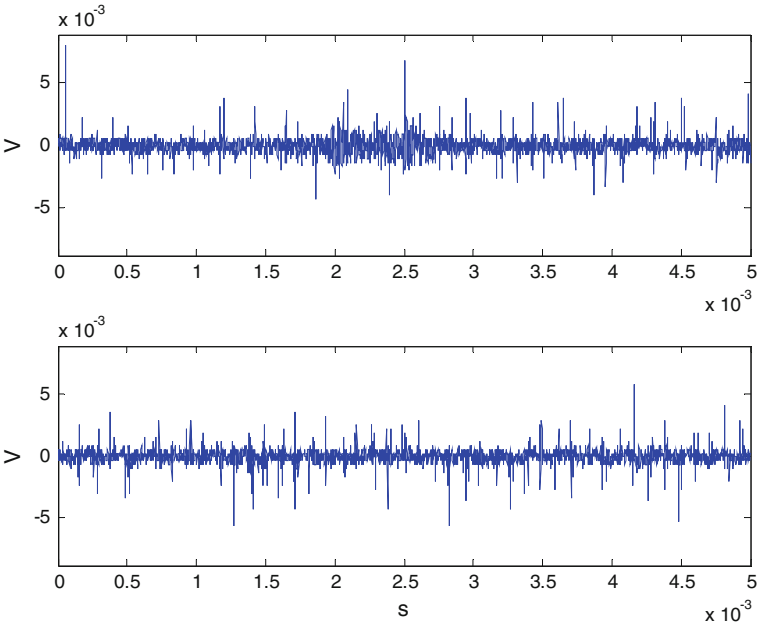


Fig. 4 AE signal with defective bearing (*top*) and with non-faulty bearing (*bottom*), with strong presence of EMI. In both cases the rotational speed is 8 Hz and load is 300 kg

difference between faulty and non-faulty case is not observable when looking at a larger portion.

The possibility of using a filter to increase the signal-to-noise ratio is evaluated. The objective is to find the frequency range where the signal has to be filtered. Inspection of the spectrum does not provide any hint, because the EMI pattern, which has a fundamental frequency of 9 kHz and modulation at 100 Hz and multiples, dominates the complete frequency range. There is not clear evidence of the defect on the spectrum. However, with careful inspection is possible to observe symptoms of the defect, in the form of scattered components in some frequency ranges, but is difficult to define a filter based on this method. Therefore, other methods are considered.

The first method used to determine the filter frequency range is the PSD-difference. This method consists simply in subtracting the PSD of the non-faulty case from the PSD of the faulty case. It points out the frequency band of the PSD which has increased due to the bearing defect.

The second method used is Spectral Kurtosis (SK). This method shows the frequency bands in which the impulsive-like features of the signal are located. Based on the hypothesis that the impulsive activity is due to the defect and is more or less homogeneous, this method has been used increasingly in recent years [4, 5].

The PSD-difference and SK methods are based on the energy and shape of the signal, respectively, but none of them considers the fact that the AE bursts originated by the defect occur in an approximately periodic manner. The last method studied, which is the Cyclostationary Peak Ratio Distribution method, takes advantage of this fundamental fact.

Traditionally, the Peak Ratio (PR) indicator is a descriptive scalar quantity which is computed from the spectrum and is defined as

$$\mathrm{PR} = \frac{N \sum_{k=1}^{K} H_k}{K \sum_{n=1}^{N} A_n} \tag{1}$$

where H_k is the kth harmonic, A_n is the nth spectral line, and K and N are the total number of harmonics and spectral lines to consider, respectively. The so defined Peak Ratio is used as an indicator of the degree of *visibility* of a certain group of harmonics over a specific frequency range.

In consideration of the second-order cyclostationary characteristics of the AE resulting from bearing defects, we consider the computation of the Peak Ratio based on cyclostationary distributions, instead of on any spectrum representation. Accordingly, we calculate one Peak Ratio along the α-axis for each frequency of the Spectral Coherence map, that is,

$$\mathrm{CPR}(f) = \frac{N \sum_{k=1}^{K} \gamma_{xx}^{k\mathrm{BPFO}}(f)}{K \sum_{n=1}^{N} \gamma_{xx}^{n\Delta\alpha}(f)} \tag{2}$$

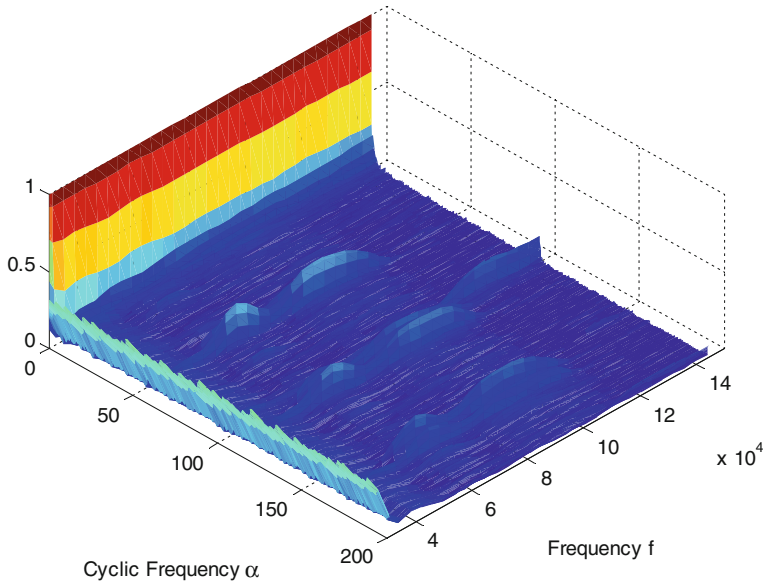


Fig. 5 Spectral coherence map of the AE from faulty bearing at 8 Hz of shaft speed and 300 kg of radial load. Frequency and cyclic frequency axes are in Hz

The result is a Peak Ratio distribution along the frequency domain. Since the harmonics considered for calculation are the harmonics of the expected fault frequency, the obtained distribution highlights the frequency intervals in which the second-order cyclostationary features of these harmonics is stronger.

The spectral coherence of the AE measured with the faulty bearing is shown in Fig. 5. The bearing radial load is 300 kg and the shaft speed is 8 Hz. For this speed, the fault frequency is 57 Hz. The second-order cyclostationary activity produced by the defect is evidenced in the discrete lines that appear at the cyclic frequencies that match the fault frequency and its harmonics.

The results of the three methods applied to the contaminated AE signal from Fig. 4 (top) are shown in Fig. 6. As observed, only the result from the Spectral Kurtosis differs significantly. Differently, the result from the PSD-difference method and from the Cyclostationary Peak Ratio Distribution (CPRD) method, they both point out the same frequency range for filtering.

Next, the AE signal contaminated with EMI is filtered according to the results obtained from the three methods (Fig. 7). As expected, the signal obtained when filtering according to the result obtained from the SK method is highly different from the rest, and does not enhance correctly the symptoms of the bearing defect. The problem is not the SK method as such. The problem is that the hypothesis in which its use is based—that the highest impulsiveness contained in the signal is repetitive and with similar characteristics—is not correct. Although it is a reasonable assumption, this example shows that it does not always hold in practice. In this case,

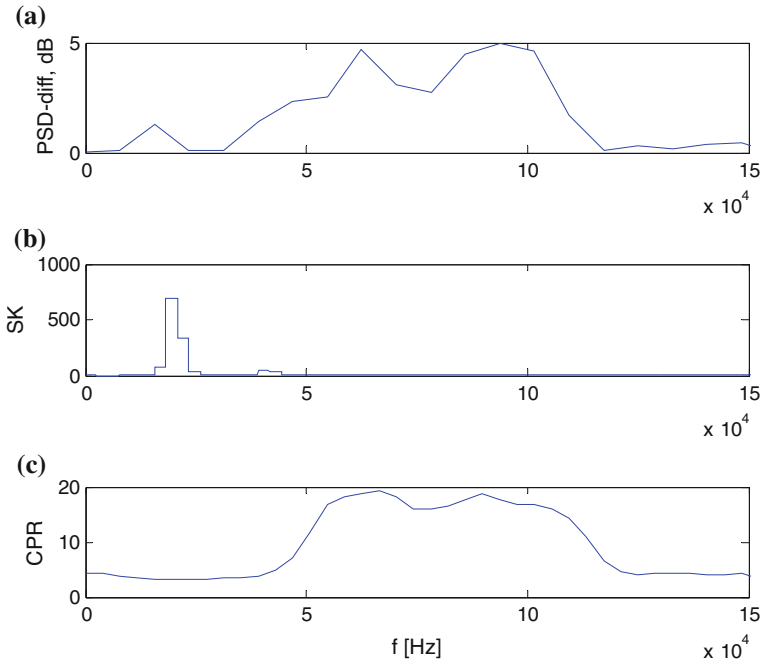


Fig. 6 Results of the three methods for filter determination. **a** PSD-difference. **b** Spectral Kurtosis. **c** Cyclostationary peak ratio distribution

a single burst produced by the defect for some reason was different from the rest and excited a different frequency range. Incidentally, in this range the signal is much more impulsive and therefore the SK focuses on this range. This situation is similar to the case studied in [14]. Although they are not used in this work, it should be considered that some variations of the SK method have been developed to improve the band selection capabilities [7, 16, 21].

On the other hand, the signals filtered with the PSD-difference and CPRD methods look almost the same. Both resultant signals are able to correctly reveal the AE pattern produced by the bearing defect. However, it is important to keep in mind that for the PSD-difference method an additional measurement under non-faulty conditions is required. This is not the case for the CPRD method, being therefore advantageous. Moreover, the selective features of the CPRD method—in terms of the cyclic frequency harmonics that are selected for the calculation of CPR(f)—make it more robust even under the presence of other possible second-order cyclostationary AE sources.

Following, the signal filtered with the CPRD method is further used to estimate the size of the defect. The AE generated when a rolling element rolls over the damaged zone is composed of multiple bursts resulting from the interaction between the rolling element and the multiple irregularities which form the defect [3]. Typically, the

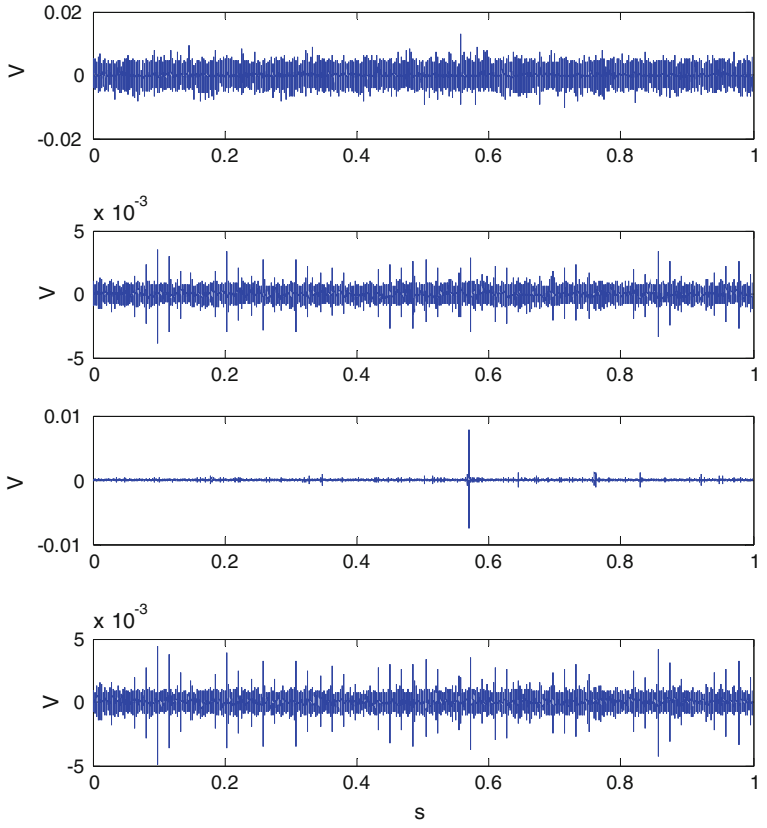


Fig. 7 From *top* to *bottom*: unfiltered AE signal; AE filtered according to PSD-difference; AE filtered according to Spectral Kurtosis; and AE filtered according to CPRD

entrance and exit of the rolling element to/from the damaged zone produce distinctive bursts, but this can be different depending on the shape of the defect. In the case under study, two bursts are dominant, as can be seen in the signal portions presented in Fig. 8.

The damage size estimation is done first in terms of the arc of the outer race, calculated by

$$s = N_c \Delta_t \tag{3}$$

where N_c is the rotational frequency of the cage assembly in rad/s and Δ_t is the time period in seconds between the distinctive peaks inside the AE resulting from the interaction between a single rolling element and the outer race. In consideration of the diameter of the outer race, it is possible to calculate the defect size in mm. Results, presented in Table 2, are in agreement with the real defect size, as presented in Fig. 3. It is noticed that at higher values of shaft speed the Δ_t decreases, while the

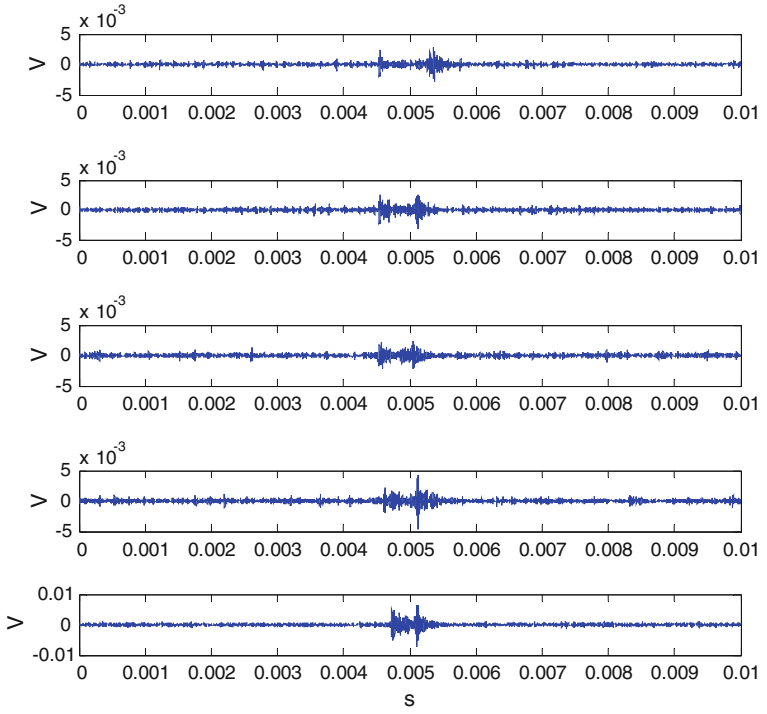


Fig. 8 Filtered AE according to CPRD method for shaft rotating frequencies of 6, 7, 8, 9 and 10 Hz (from *top* to *bottom*)

Table 2 Results of defect size estimation

Shaft speed (Hz)	Estimated defect size Arc (rad)	Width (mm)
6	0.0129	0.405
7	0.0108	0.339
8	0.0101	0.316
9	0.0119	0.375
10	0.00987	0.309

burst decay does not, which results in merging of the two bursts. Therefore, damage assessment by this method is limited to low shaft rotational speeds. For higher speeds other researchers have proposed to estimate the defect size on the basis of the duration of the merged burst [2]. However, the duration of the merged burst can be difficult to determine at higher rotational speeds, because the level of continuous AE activity generated by the interaction between the rolling elements and the non-faulty zone of the race increases with speed. So, even when at high speed the AE bursts have more energy, the signal-to-noise ratio can be lower than for low speed.

5 Conclusion

The problem of size estimation of localized bearing damage based on acoustic emission measurements on the bearing housing has been investigated. The problem is not straightforward due to the strong presence of electromagnetic interference (EMI) in the measured signals. Three methods are used for the definition of a filter aiming to increase the signal-to-noise ratio; namely PSD-difference, Spectral Kurtosis and Cyclostationary Peak Ratio Distribution (CPRD). Results obtained from Spectral Kurtosis method are poor, because the hypothesis in which its use is based does not hold in practice. On the other hand, results from PSD-difference and CPRD methods are similar, both resulting in signals which appropriately reveal the symptoms of the defect. That is, both methods are able to significantly reduce the effect of the EMI, and allow for defect size estimation. However, the CPRD method is advantageous because (i) it does not require a measurement of the non-faulty case, as the PSD-difference does; and (ii) because it allows selecting the cyclic frequencies of interest (i.e. the fault frequency and harmonics), thus making the method apt for situations when other second-order cyclostationary components are also contained in the signal. Finally, CPRD method is used for defect size estimation, obtaining results which are in agreement with the dimension of the actual defect. Since at higher rotational speed merging of AE bursts generated by the interaction between a single rolling element and the damaged zone occurs, the presented method for size estimation is restricted to low rotational speeds.

References

1. Achembach, J. D. (1987). *Wave propagation in elastic solids*. Amsterdam: North-Holland Publishing Company.
2. Al-Dossary, S., Raja-Hamzah, R. I., & Mba, D. (2006). Acoustic emission waveform changes for varying seeded defect sizes. *Advanced Materials Research,13–14*, 427–432.
3. Al-Ghamdi, A. M., Zhechkov, D., & Mba, D. (2004). The use of acoustic emission for bearing defect identification and estimation of defect size. Lecture 38 EWGAE.
4. Antoni, J. (2006). The Spectral Kurtosis: a useful tool for characterising nonstationary signals. *Mechanical Systems and Signal Processing*, *20*(2), 282–307.
5. Antoni, J. (2007). Fast computation of the Kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, *21*(1), 108–124.
6. Balderston, H. L. (1969). The detection of incipient failure in bearings. *Materials Evaluation*, *27*, 121–128.
7. Barszcz, T., & Jablonski, A. (2011). A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram. *Mechanical Systems and Signal Processing*, *25*(1), 431–451.
8. Catlin, J. B. (1983). The use of ultrasonic diagnostic technique to detect rolling element bearing defects. In *Proceeding of Machinery and Vibration Monitoring and Analysis Meeting, Vibration Institute, USA,* pp. 123–130
9. Davis, J. L. (2000). *Mathematics in wave propagation*. Princeton: Princeton University Press.
10. Holroyd, T. (2000). Acoustic emission and ultrasonics. *Machine and systems condition monitoring series*. Oxford: Coxmoor Publishing Company.

11. Ho, D., & Randall, R. B. (2000). Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mechanical Systems and Signal Processing*, *14*(5), 763–788.
12. Matthews, J. R. (1983). *Acoustic emission*. New York: Gordon and Breach Science Publishers Inc.
13. McFadden, P. D., & Smith, J. D. (1984). Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of Sound and Vibration*, *96*(1), 69–82.
14. Obuchowski, J., Wyłomańska, A., & Zimroz, R. (2014). Selection of informative frequency band in local damage detection in rotating machinery. *Mechanical Systems and Signal Processing*, *48*(1–2), 138–152.
15. Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—A tutorial. *Mechanical Systems and Signal Processing*, *25*(2), 485–520.
16. Sawalhi, N., Randall, R. B., & Endo, H. (2007). The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral Kurtosis. *Mechanical Systems and Signal Processing*, *21*(6), 2616–2633.
17. Shiroishi, J., Li, Y., Liang, S., Kurfess, T., & Danyluk, S. (1997). Bearing condition diagnostics via vibration and acoustic emission measurements. *Mechanical Systems and Signal Processing*, *11*(5), 693–705.
18. Tallian, T. E. (1992). *Failure Atlas for Hertz contact machine elements*. New York: ASME.
19. Vicuña, C. M., & Quezada, D. (2014). Cyclostationary processing of vibration and acoustic emissions for machine failure diagnosis. *Cyclostationarity: Theory and methods*. Berlin: Springer.
20. Yoshioka, T., & Fujiwara, T. (1984). *Application of acoustic emission technique to detection of rolling bearing failure* (Vol. 14, pp. 55–76). New York: ASME Production Engineering Division.
21. Zhang, Y., & Randall, R. B. (2009). Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast Kurtogram. *Mechanical Systems and Signal Processing*, *23*(5), 1509–1517.

The Stochastic Recurrence Structure of Geophysical Phenomena

I. Javors'kyj, R. Yuzefovych, I. Matsko and I. Kravets

1 Introduction

Methods of oscillations analysis started to develop by using the ideas of harmonic analysis for periodic and almost periodic functions, the methods of spectral correlation analysis for stationary, periodically and almost periodically non-stationary random processes. In the model in the form of periodic function it is assumed that values of geophysical processes precisely repeat over the time interval, defined by its period. Recurrence of oscillation properties within the stationary model are represented by power spectral density peaks, which is the outcome of covariance function varying by time lag in oscillating form. Stochastic oscillation models in the form of periodically non-stationary random processes (within the theory of second-order periodically correlated random processes (PCRP)) and their generalizations [1–8] represent recurrence of oscillation properties in the form of time-varying moment functions of first and the second order, but requirements formulated for the covariance function do not concern its behaviour by time lag.

Mean of PCRP $m(t) = E\xi(t)$ and its covariance $b(t,u) = E\overset{\circ}{\xi}(t)\overset{\circ}{\xi}(t+u)$, $\overset{\circ}{\xi}(t) = \xi(t) - m(t)$ are periodical functions of time t: $m(t) = m(t+T), b(t,u) = b(t+T,u)$. These characteristics can be represented in Fourier series form:

$$m(t) = \sum_{k\in Z} m_k e^{ik\frac{2\pi}{T}t} = m_0 + \sum_{k\in N}\left(m_k^c \cos k\frac{2\pi}{T}t + m_k^s \sin k\frac{2\pi}{T}t\right),$$

I. Javors'kyj · R. Yuzefovych · I. Matsko (✉) · I. Kravets
Karpenko Physico-Mechanical Institute of National Academy of Sciences of Ukraine,
Naukova Str. 5, L'viv 79060, Ukraine
e-mail: matskoivan@gmail.com

I. Javors'kyj
Institute of Telecommunication, University of Technology and Life Sciences,
Al. Prof. S. Kaliskiego 7, 85789 Bydgoszcz, Poland

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods - II*,
Applied Condition Monitoring 3, DOI 10.1007/978-3-319-16330-7_4

$$b(t,u) = \sum_{k\in Z} B_k(u)\, e^{ik\frac{2\pi}{T}t} = B_0(u) + \sum_{k\in N}\left(B_k^c(u)\cos k\frac{2\pi}{T}t + B_k^s(u)\sin k\frac{2\pi}{T}t\right),$$

where

$$m_k = \frac{1}{2}(m_c - im_s), \quad B_k(u) = \frac{1}{2}\left(B_k^c(u) - iB_k^s(u)\right) \text{ for } k \neq 0.$$

The mean $m(t)$ describes the deterministic periodical oscillations, the variance $b(t,0)$ characterizes the periodicity of the power oscillations around this deterministic part and the covariance function describes the structure of periodical changes in correlations between fluctuation values in different moment of time t separated on lag u. The fluctuation properties in spectral domain are described by time-varying spectral density

$$f(\omega,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} b(t,u)\, e^{-i\omega u}du,$$

which can be represented in Fourier series form

$$f(\omega,t) = \sum_{k\in Z} f_k(\omega)\, e^{ik\frac{2\pi}{T}t},$$

where

$$f_k(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} B_k(u)\, e^{-i\omega u}du.$$

The quantities $B_k(u)$ and $f_k(\omega)$ are called covariance and spectral components, respectively [1]. Zeroth covariance component $B_0(u)$ is a covariance function of stationary approximation of PCRP, and zeroth spectral component $f_0(\omega)$ the spectral power density of this approximation. Using the mean, covariance function, time-varying spectral density and their Fourier coefficients m_k, $B_k(u)$, $f_k(\omega)$ gives us a possibility to fully and qualitatively combine descriptions of recurrence and irregularity. They give us a possibility to develop physically meaningful and mathematically strong theory of stochastic oscillations if some conditions are satisfied. These models in a natural way develop deterministic and probabilistic approaches to oscillation investigations, which contain as special cases the well-known literature representations used for description of interaction between recurrence and irregularity: polyharmonic, additive, multiplicative and quadrature models, and also their combinations and modifications. The possibility of jointly treating the different types of stochastic recurrences on the basis of PCRP and their generalizations consists in representations that take into consideration modulations of main harmonics and the chosen probabilistic characteristics describe the regularity of such modulations.

PCRP models of oscillation and their generalizations can be verified only with the experimental data with the use of the respective processing algorithms. The methods of statistical analysis of the given class of non-stationary processes: coherent [9],

component [10], linear filtration methods [11, 12], least square method [13] are the theoretical bases for the given classes of non-stationary random processes. Being a development of statistical methods for stationary signals, they allow us to pass to the estimations of probability characteristic dependences on time of probable characteristics and also quantities, which characterize these dependences. The examples of PCRP statistic methods used for investigations of annual and daily variations of physical phenomena are given in this paper and the specific properties of their spectral correlation structure are analysed; also the estimation of obtained results reliability is provided. This paper is prepared on the basis of the results obtained by the authors and their colleagues.

2 Annual Variation of Geophysical Processes

The specific property of time variations of many physical processes in annual and daily diapasons is almost recurring and irregular in their values. Recurrence of oscillation properties of rhythmicity is caused by astrophysical factors related to the rotation of the Earth around the Sun and the rotation of the Earth around its own axis. Data on quantitative characteristics of annual and daily rhythmics are necessary for elucidation of the processes of physical nature, developing calculation and predicting methods, preparation of manual and security handbooks for the national economy.

In Fig. 1a–d is shown changes in monthly averaged data of temperature, sea water saltiness, ozone content and angular velocity of the Earth's rotation during the years and in Fig. 1e, f—changes of hourly averaged data of air temperature and modulus of the sea current. The most significant annual variation is represented by the changes in ground, water and air temperatures. In the series of average monthly values, the amounts of annual oscillations are so recurrent that there is an opinion that they are periodically variable.

Ratio between powers of deterministic and fluctuation oscillations and amount of annual variations can be estimated by analysing their experimental data within the stationary model using the covariance function calculation formula

$$\hat{R}(u) = \frac{1}{K}\sum_{n=0}^{K-1}\left[\xi(nh) - \hat{m}\right]\left[\xi(nh+u) - \hat{m}\right], \quad \hat{m} = \frac{1}{K}\sum_{n=0}^{K-1}\xi(nh),$$

and for calculation of power spectral density

$$\hat{f}(\omega) = \frac{\Delta u}{2\pi}\sum_{n=-L}^{L}\hat{R}(n\Delta u)\,k(n\Delta u)\,e^{-i\omega n\Delta u},$$

where $k(n\Delta u)$—correlation window, $h = \theta/K$—sampling step for time, θ—realization length, $\Delta u = u_m/L$—sampling step of covariance function for time lag,

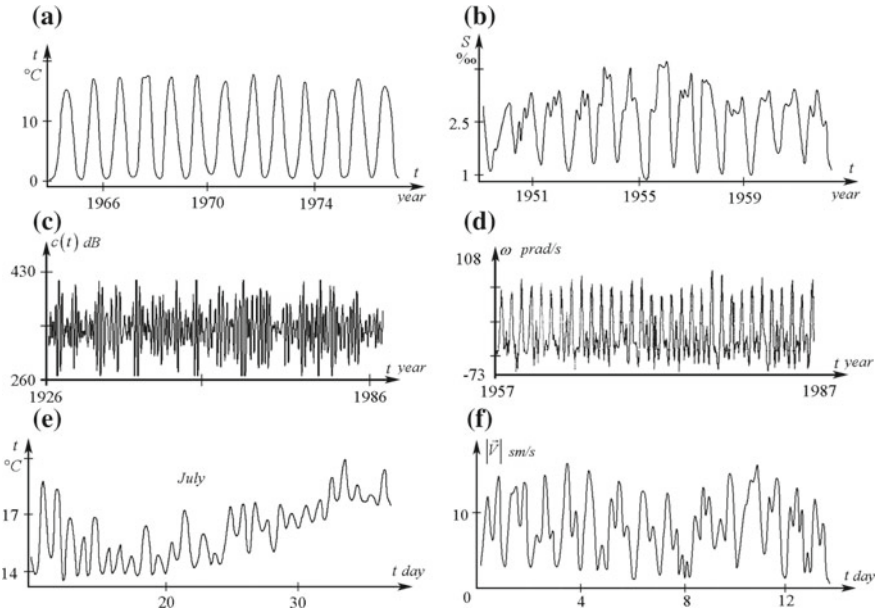


Fig. 1 Time series of geophysical processes: **a** average monthly values of water temperature; **b** average monthly values of sea water saltiness; **c** average monthly values of ozone content; **d** average monthly values of angular velocity of Earth rotation; **e** hourly values of water temperature; **f** hourly values of modulus of sea current

K, L—natural numbers, u_m—point of correlogram cut-off. Graphs for estimators $\hat{R}(u)$ and $\hat{f}(\omega)$ calculated for average monthly data of air temperature, sea water saltiness, sea level, river flows are presented in Fig. 2. Undamped tails of covariance function estimators give evidence of the presence of deterministic periodical components in annual variations of these amounts. Value of estimator $\hat{R}(u)$ at zero defines the common power of oscillations and the undamped tail amplitude gives information about the power of deterministic part. Among the analysed amounts, as it can be seen from Fig. 2, the largest powers of deterministic periodical parts of monthly averaged data relate to water temperature and river flows and the least powers to annual oscillations of sea level. Spectral density estimators, obtained using Hemming window, have significant peaks at annual frequency and frequencies multiple to it. On the basis of spectral density estimators we can evaluate the ratio between powers of separate harmonics. As it can be seen, such ratios have the characteristic property for each of the processes.

Results of analysis of geophysical processes of annual variations within the stationary model became a basis for the assumption that relatively small stochastic changes may be described by stationary random process $\eta(t)$. Then average monthly series can be written in the form

$$\xi(t) = f(t) + \eta(t) \tag{1}$$

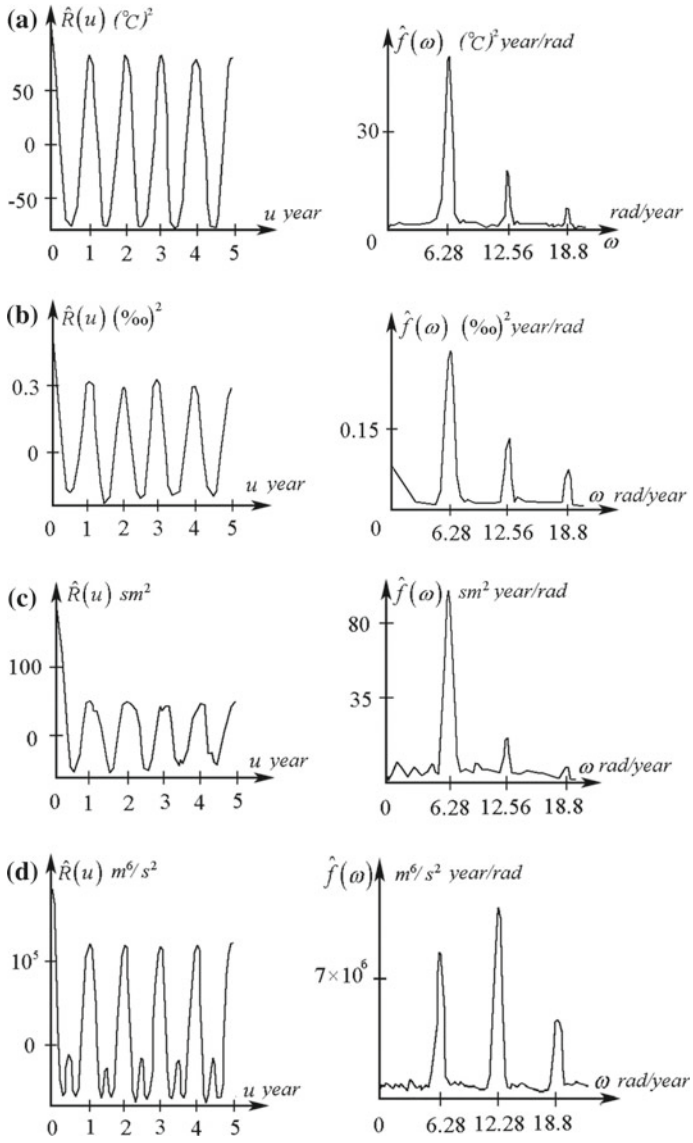


Fig. 2 Covariance function and spectral density estimators for stationary approximation of monthly average series: **a** water temperature; **b** sea water saltiness; **c** sea level oscillation; **d** river flows

It seems that the additive model does not contradict the results of the obtained data statistical processing. Powerful undamped oscillation with annual variation is proper for covariance function estimators calculated in the stationary approximation; the respective power spectral density estimators have peaks on the frequency of annual variation and multiple to it (Fig. 2). After exception of regular annual variation from

the analysed series (norm) correlations the character considerably changes: they rapidly damp and any oscillation appears weak. Exactly this property of covariance function estimators was the reason for conclusion drawn in the absence of periodic second-order irregularity in the time series. However, researchers are skeptical about the model correctness (1) usage. It was noted that at considerable nonlinearity of processes appearing in the atmosphere and oceans, the validity of this representation is problematic, since statistical characteristics of anomalies and the character of their correlations in the different environments have pronounced seasonal character. During verification of the observed data it was found that some separate harmonics of norm $f\ (t)$ vary in time or, more precisely, are modulated by low- and high-frequency processes. Because of this we cannot speak about existence of some "poured" annual variation. Thus, an important problem of this change analysis appears. Such analysis is possible on the basis of the more complete model representation [14–19] than (1). The stochastic modulation of annual variation harmonics should be taken into account in the model. So we arrive at the representation

$$\xi\ (t) = \sum_{k \in Z} \xi_k\ (t)\, e^{ik\frac{2\pi}{T}t}, \tag{2}$$

i.e. to the class of PCRP. In mean function $m(t)$, variance $b(t, 0)$, covariance function $b(t, u)$ of PCRP, which periodically varies in time (in the given case period is 12 months), the character features of geophysical phenomena represent time variety.

It follows from expression (2) that coefficients m_k are means of modulating processes $\xi_k\ (t)$, i.e. $m_k = E\xi_k\ (t)$, and covariance components $B_k\ (u)$ are defined by their cross-covariance functions $R_{lk}\ (u) = E\overline{\overset{\circ}{\xi}_l}\ (t)\, \overset{\circ}{\xi}_k\ (t+u)$, $\overset{\circ}{\xi}_k\ (t) = \xi_k\ (t) - m_k$, i.e.

$$B_k\ (u) = \sum_{l \in Z} R_{l-k,l}\ (u)\, e^{il\frac{2\pi}{T}u}. \tag{3}$$

As it can be seen from the representation (2), the PCRP model covers various simpler models of rhythmic variations [1, 3], including the additive model (1), multiplicative model $\xi\ (t) = \eta\ (t)\ f\ (t)$, additive–multiplicative model $\xi\ (t) = g\ (t) + \eta\ (t)\ f\ (t)$, where $g\ (t)$ is a periodic function and others. Obviously, the probabilistic structure of concrete phenomenon can be investigated only on the basis of experimental data using methods of statistical analysis of PCRP [9–13]. The feature of these methods is what information is required about period value T for use. Calculation of this value can be provided using functionals that look like statistics used in coherent or component methods and also least squares method, but they contain some test value τ instead of period value T. Period calculation in this case has led to searching of such values of test period, which responds to extremes of those functionals. Using the coherent method, functional of mean has the form

$$\hat{m}\ (t, \tau) = \frac{1}{2M+1} \sum_{n=-M}^{M} \xi\ (t + n\tau), \tag{4}$$

and functionals of correlation and covariance functions, respectively,

$$\hat{k}\,(t, u, \tau) = \frac{1}{2M+1} \sum_{n=-M}^{M} \xi\,(t + n\tau)\,\xi\,(t + u + n\tau), \tag{5}$$

$$\hat{b}(t, u, \tau) = \frac{1}{2M+1} \sum_{n=-M}^{M} \left[\xi\,(t + n\tau) - \hat{m}\,(t + n\tau)\right]\left[\xi\,(t + u + n\tau) - \hat{m}(t + u + n\tau)\right], \tag{6}$$

where M is integer number. Such period estimation is a generalization of the well-known Buys-Ballot scheme [20], which is used for calculation of periodic function period [3, 21] for the case of PCRP. Period of oscillation deterministic part can be estimated using functional (4) and the period of covariance function time variety with the help of functional (6). If periodicity in time is proper for the mean and the covariance function, then using functional (5) allows to enhance the estimation efficiency.

Note that analysing annual and daily variations period values [22] may be considered as known. Functionals (4)–(6) in this case can be used for proving the assumption of periodicity of time variety of moments as both the first and the second orders. On the basis of the processing results the correctness of the PCRP approach for annual or daily variation analysis of given processes can be grounded. Dependences of coherent functionals of mean $\hat{m}\,(t, \tau)$, covariance function $\hat{b}\,(t, u, \tau)$, and correlation function $\hat{k}\,(t, u, \tau)$ on test period τ computed, as an example, for air temperature series are shown in Fig. 3. As we can see, in all cases functionals have pronounced extremes at point $\tau = T = 12$ months. Processing realizations of other processes obtain similar results. This proves that observed data has properties of irregularity of both first and second orders.

For annual variation analysis, traditionally the method of interval averaging (month) of initial data is used. This procedure equals filtration of high-frequency components, which characterizes inside daily and synoptic varieties. After such averaging realizations, obviously, we have a view of stochastically changed annual oscillations observed against the slowly changed line. If the processed realization length does not allow to obtain reliable data about the low-frequency changes (it is too short), they should be excepted using acceptable filtration methods. In the simplest case trend removal is achieved using centring of average monthly series values on the respective average annual values.

Monthly average data can be processed using both coherent and component estimation methods [9, 10], but since the last one requires a priori data about the number of harmonics in Fourier series of computing characteristics, so in the first stage we should use the coherent method and further estimator values can be improved by using the component method.

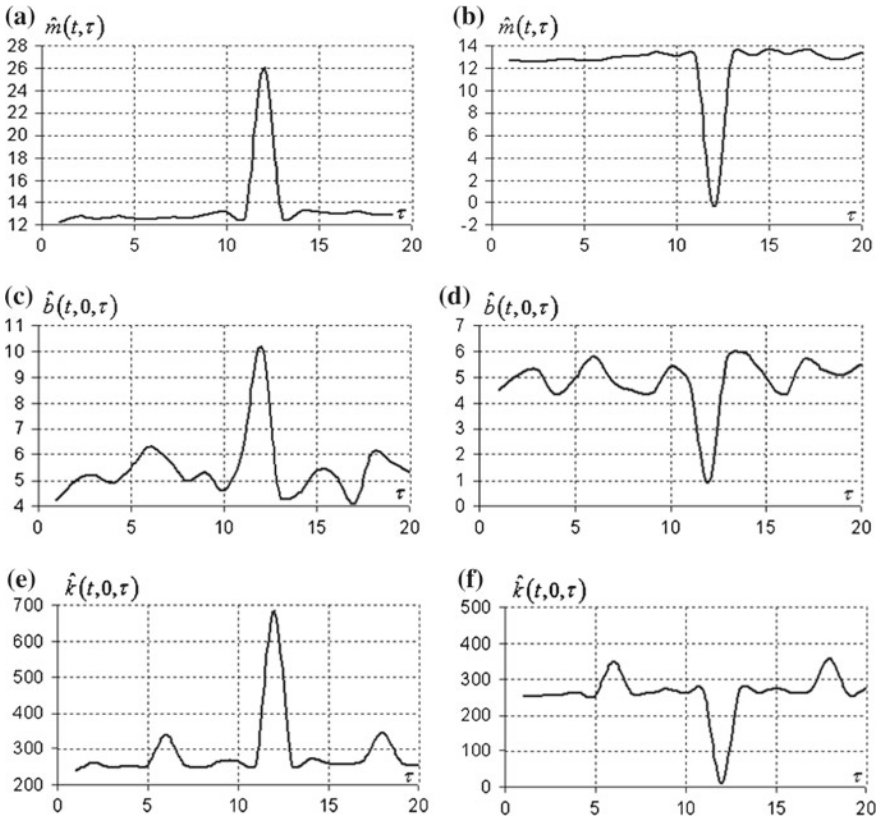


Fig. 3 Coherent functionals of mean $\hat{m}\,(t,\tau)$ (**a**–**b**), covariance function $\hat{b}\,(t,0,\tau)$ (**c**–**d**) and correlation $\hat{k}\,(t,0,\tau)$ (**e**–**f**) of air temperature in dependence on test period τ for the moments of time close to extremes of respective characteristic estimators

Coherent estimators for the mean and covariance function are written as:

$$\hat{m}\,(lh) = \frac{1}{N}\sum_{n=0}^{N-1}\xi\,(lh+nT),$$

$$\hat{b}\,(lh,k\Delta u) = \frac{1}{N}\sum_{n=0}^{N-1}\xi\,(lh+nT)\,\xi\,(lh+k\Delta u+nT) - \hat{m}\,(lh)\,\hat{m}\,(lh+k\Delta u),$$

where N is the number of periods averaged, $h = T/P$—sampling interval for time, Δu—sampling interval for lag, P—natural number. Estimators of mean Fourier components and covariance components can be calculated using formulae

$$m_0 = \frac{1}{P}\sum_{l=0}^{P-1} \hat{m}\,(lh), \quad m_k^c = \frac{1}{P}\sum_{l=0}^{P-1} \hat{m}\,(lh)\cos k\frac{2\pi}{P}l, \quad m_k^s = \frac{1}{P}\sum_{l=0}^{P-1} \hat{m}\,(lh)\sin k\frac{2\pi}{P}l,$$

$$B_0\,(u) = \frac{1}{P}\sum_{l=0}^{P-1} \hat{b}\,(lh, k\Delta u), \quad B_k^c\,(u) = \frac{1}{P}\sum_{l=0}^{P-1} \hat{b}\,(lh, k\Delta u)\cos k\frac{2\pi}{P}l,$$

$$B_k^s\,(u) = \frac{1}{P}\sum_{l=0}^{P-1} \hat{b}\,(lh, k\Delta u)\sin k\frac{2\pi}{P}l.$$

Values $\left|\hat{m}_k\right| = \sqrt{\left(\hat{m}_k^c\right)^2 + \left(\hat{m}_k^s\right)^2}$, $\varphi = arctg\frac{\hat{m}_k^s}{\hat{m}_k^c}$ define amplitude an phase spectra of mean estimator and the values $\left|\hat{B}_k\,(0)\right| = \sqrt{\left(\hat{B}_k^c\,(0)\right)^2 + \left(\hat{B}_k^s\,(0)\right)^2}$, $\varphi = arctg\frac{\hat{B}_k^s(0)}{\hat{B}_k^c(0)}$—amplitude an phase spectra of variance estimator.

Processing results show that the pronounced regular variation described by PCRP are present in the annual variation of all analysed processes (Figs. 4 and 5). The variety of its forms can be easily explained on the basis of physical factors under whose influence every process occurs, quantitatively characterized by amplitude estimators $|m_l|$ (Table 1). The largest is the amplitude of the first harmonic. Correlation between harmonics amplitudes and phases defines the typical features of regular variation forms: location of its maxima and minima, appearance of extremes, sharp vertices or cavities or their widening.

Estimators of process variances, which characterize the power of average perennial scattering of values in their realization relatively to the standard have also a pronounced annual periodicity (Figs. 4 and 5). This property of variance is the main argument that forces us to reject the additive model (1) since its variance is constant. Variance estimators in most cases have more harmonics than mean function estimator, which leads to non-coherence of these characteristic time variations. Estimators of variances for temperature series have two maxima at the points which coincide with points of the fastest increase and the fastest decrease in temperature.

The presented examples show that time variety ranges of variance estimator are very significant and for some concrete months their values can exceed by several times the average power N_{av}^f defined by value of zeroth covariance component estimator at zero $N_{av}^f = B_0\,(0)$ variance of stationary approximation of PCRP (in the figures these values are marked by dashed line). Consider also the values that define the average power of deterministic oscillations $N_{av}^d = \frac{1}{2}\sum_{l=1}^{N_2}\left[\left(m_l^c\right)^2 + \left(m_l^s\right)^2\right]$ and vibrational power of fluctuations $N_{osc}^f = \sum_{l=1}^{N_2}\left[\left[B_l^c\,(0)\right]^2 + \left[B_l^s\,(0)\right]^2\right]$. Estimators of absolute and relative values of these quantities obtained in processing of temperature series, saltiness, sea level, rivers flows, South ocean ice cover (ocean square covered by ice) are given in Table 2. Comparing them we can see that temperature changes

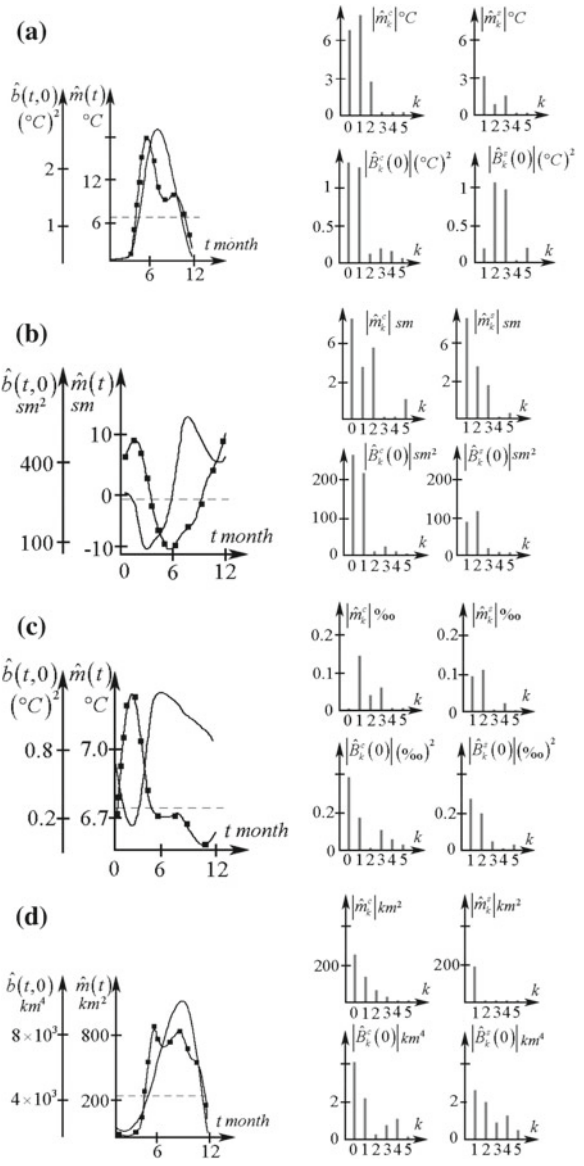


Fig. 4 Estimators of mean function $\hat{m}(t)$, variance $\hat{d}(t) = \hat{b}(t,0)$ (–■–■) and their Fourier coefficients: **a** water temperature; **b** level oscillation; **c** sea water saltiness; **d** ice amount in South ocean

are the most regular processes: average power of deterministic oscillations exceeds by many times the average power of fluctuations. For river flows, as the result of different processing realizations, this value changes within a rather wide range (from 0.15 till 3.0). The regular (deterministic oscillations) variations of saltiness and sea level are comparatively low powered and in this case the ratio N_{av}^d / N_{av}^f for different realizations changes insignificantly. Hence, the high noise level of deterministic

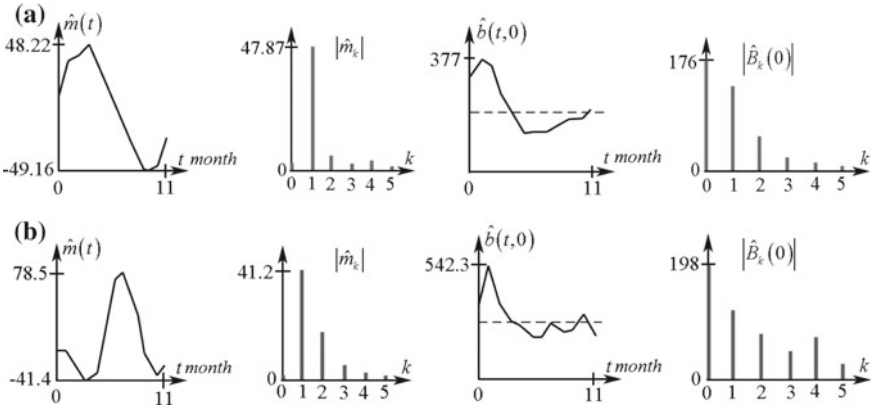


Fig. 5 Estimators of mean function $\hat{m}(t)$, variance $\hat{d}(t) = \hat{b}(t, 0)$ and their Fourier coefficients: **a** ozone concentration; **b** angular velocity of Earth rotation

Table 1 Magnitudes of harmonic components of mean function $|\hat{m}_l|$ and variance $|\hat{B}_l(0)|$

Number of harmonic		1	2	3	4	5
Water temperature (°C)	$\lvert\hat{m}_l\rvert$	9.6	2.5	0.8	0.4	0.3
	$\mathrm{P}\lvert\hat{B}_l(0)\rvert$	1.3	0.8	0.7	0.2	0.2
Saltiness (%)	$\lvert\hat{m}_l\rvert$	0.18	0.12	0.05	0.02	0.01
	$\lvert\hat{B}_l(0)\rvert$	0.30	0.20	0.10	0.06	0.02
Sea level (sm)	$\lvert\hat{m}_l\rvert$	8	5	1	1	2
	$\lvert\hat{B}_l(0)\rvert$	210	46	34	12	35
River flows (sm^3/s)	$\lvert\hat{m}_l\rvert$	324	352	234	165	125
	$\lvert\hat{B}_l(0)\rvert$	6.9×10^4	8.3×10^4	4×10^4	3×10^4	1×10^4
Ice cover (km^2)	$\lvert\hat{m}_l\rvert$	229	43	32	11	4
	$\lvert\hat{B}_l(0)\rvert$	3580	1496	640	1281	338

periodic oscillations does not mean that annual rhythmic weakly appears in the background of the common processes variety. It is proved by comparing the average power of fluctuations $\hat{N}^f_{av}$ and their vibration power $\hat{N}^f_{osc}$. Quantity $\hat{N}^f_{osc}$ can be used as a measure of process periodical irregularity by variance. If $\hat{N}^f_{osc} = 0$, the variance is constant and in this case rhythmic appears only in deterministic changes (in this case model (1) can be used for its description). The ratio N^f_{osc}/N^f_{av} can be used as a relative measure of the second-order irregularity. This ratio, as it can be seen, is rather large for the saltiness, temperature and flows. Annual changes in variances

Table 2 Estimators of power components and their ratios

Quantities	Water temperature	Saltiness	Sea level	River flows	Ice cover
$\hat{N}_{av}^{(d)}$	50	0.02	40	1.64×10^5	2.76×10^4
$\hat{N}_{av}^{(f)}$	1.3	0.29	270	8.28×10^4	4.55×10^3
$\hat{N}_{k}^{(f)}/\hat{N}_{av}^{(f)}$	2.4	2.3	1.3	2.8	1.6
$\hat{N}_{av}^{(d)}/\hat{N}_{av}^{(f)}$	38.2	0.08	0.15	2	6.1
$\hat{N}_{k}^{(f)}/\hat{N}_{av}^{(d)}$	0.1	27.9	8.3	1.4	0.3

for oscillation of sea level and square of ice are observed in the background of higher average level, however, the contribution of annual periodic changes for these processes is significant.

Comparing quantities $\hat{N}_{av}^{f}$ and $\hat{N}_{osc}^{f}$ for different processes (Table 2) we conclude that if the main power of periodic annual variations is proper to regular oscillations of temperature and ice square, then for other series annual rhythmic appears in most cases in periodical changes of fluctuation power.

The examples of geophysical process analysis obtained by PCRP methods show (Table 2) that annual rhythmic is represented by variation of each of them in different ways. At the same time in the properties of mean function and variance these specific features of time variety are represented, which are typical for separate regions of the Earth or the world's oceans.

Using coherent method for mean function and variance estimators computation, such forms of annual variation of the maximal number of harmonic components are taken into account. Magnitudes of higher harmonics are very small in comparison with the first ones, while relative error of their values is large. It is clear that these harmonics should not be taken into account for obtaining the more reliable estimators of annual curves forms. Because of this decision is made about the actuality of use of the component method for characteristics estimation, which allows choosing a number of harmonics to be estimated. Obviously, making a choice, we should proceed from the requirement of description of needed details of the annual curve forms and confidence level of harmonics amplitude estimators, which are neglected.

Properties of variance estimators of analysed series, as it was noted, force us to decline the additive model and lead to the conclusion about their periodical second-order irregularity. Obviously, the analysis of series correlation structure is the next step in investigations. As processing results show (Figs. 6 and 7) correlations between fluctuation values at the moments of time t and $t + u$ change during a year. The main correlogram feature is its relatively rapid damping as u increases. Even if some oscillations are present in the correlogram, they are of low power. So, series centring on mean function estimator significantly changes the processes correlation structure, and it appears that powerful oscillations of covariance function estimators

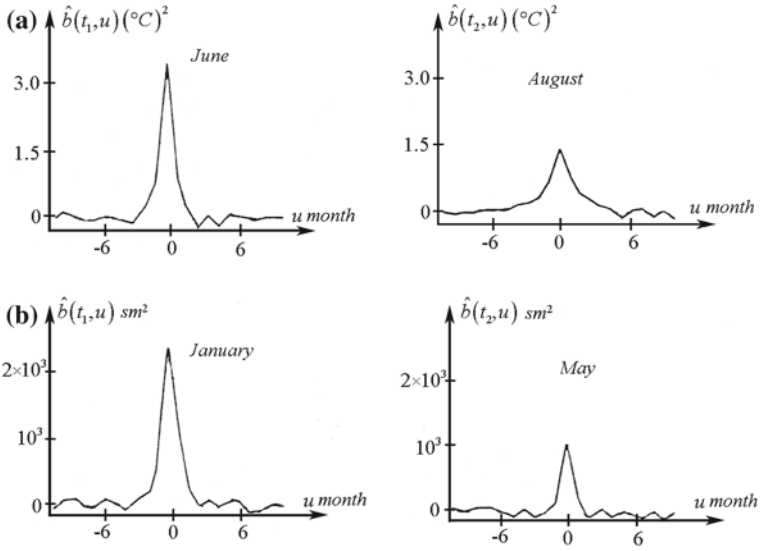


Fig. 6 Dependences of covariance function estimators on time lag for different months of the year: **a** air temperature; **b** sea level oscillations

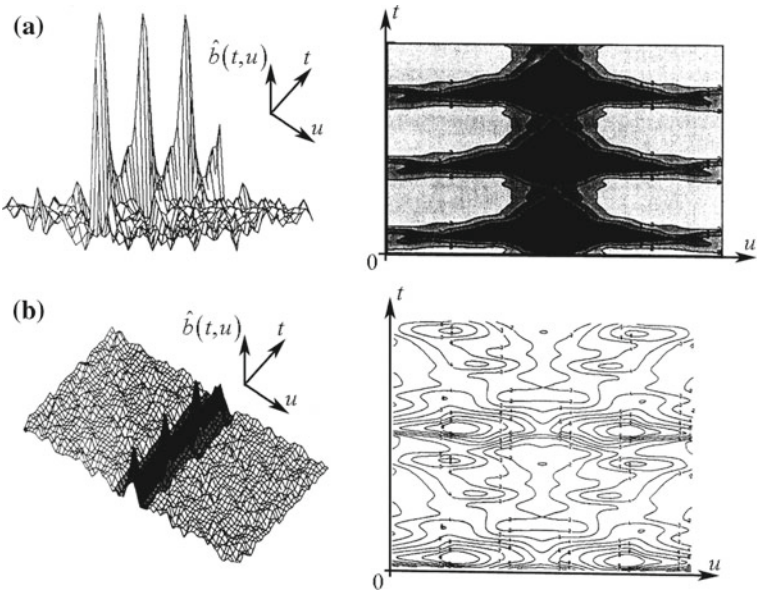


Fig. 7 Covariance function estimators and their cross sections: **a** angular velocity of Earth rotation; **b** ozone concentration

in stationary approximation with period $T = 12$ months are caused by the presence of regular changes. This is where the formal character of the stationary model is used for geophysical process investigation: correlations between process values cannot exist for many years. Because of this the change to PCRP model is natural, it gives the possibility to separate deterministic and stochastic parts of oscillations and on the basis of dependence of stochastic parts covariance function estimator on time and bias to make a decision about the properties of these correlations.

If PCRP covariance function estimator as a function of two variables represents information about process in the form of matrix, then covariance components analysis gives us a possibility to pass this information into the area of one variable function: knowledge of the number of covariance components and their dependences on time lag completely defines changes in covariance function in time and lag. The zeroth covariance component is a covariance function of stationary approximation of PCRP and defines time-averaged correlations of the fluctuation part.

As it follows from expression

$$B_0(u) = \sum_{l \in Z} R_{ll}(u)\, e^{il\omega_0 u}, \tag{7}$$

this quantity is defined by autocovariance functions of all stationary components that form this process. In formula (7) autocovariance functions of all stationary components, except the zeroth, are entered with oscillation products. Graphs of estimators $B_0(u)$, shown in Figs. 8 and 9, do not contain any pronounced oscillations

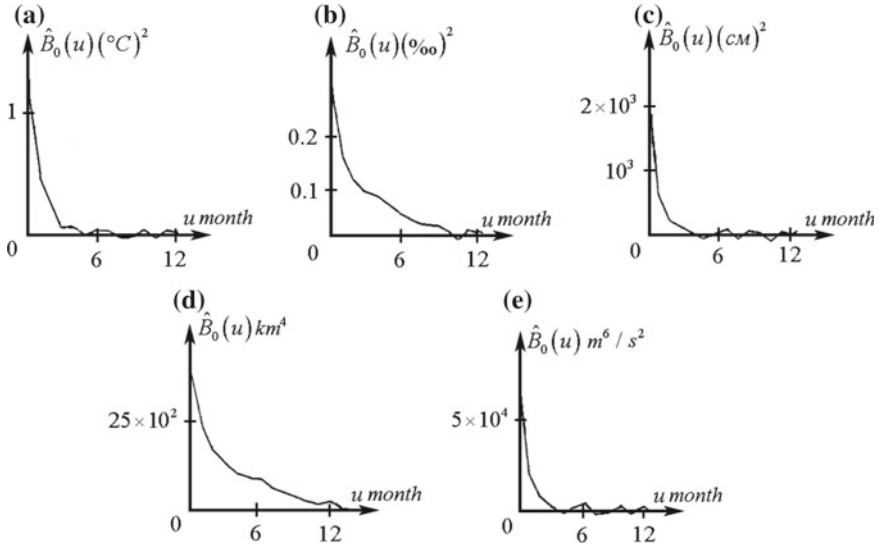


Fig. 8 Estimator of the zeroth covariance components: **a** water temperature; **b** saltiness of sea water; **c** sea level oscillation; **d** ice square; **e** river flow

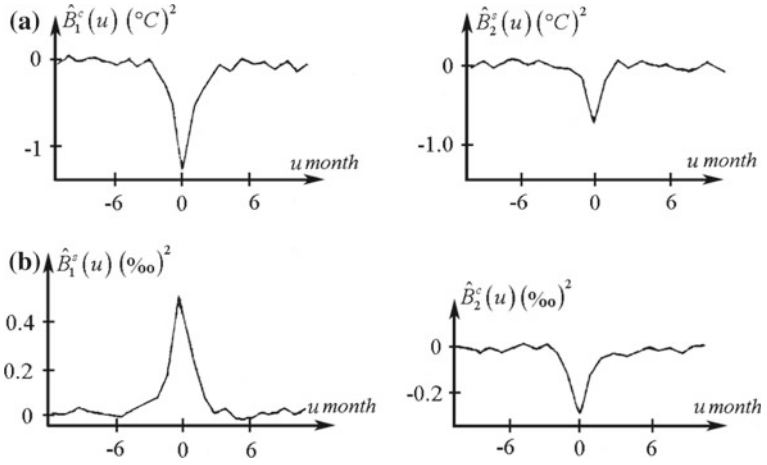


Fig. 9 Estimators of the most significant sine and cosine covariance components: **a** water temperature; **b** saltiness of sea water

of the year's period and its overtones (multiples). It allows us to assume that for the analysed processes every correlation is weak in comparison with $R_{00}(u)$ or all of them rapidly damp with time lag growth and because of it oscillations cannot appear.

Higher covariance components, on one hand, define a time variety form of covariance function dependence for some lag: for given u they are coefficients of Fourier series of $b(t, u)$. Their knowledge allows the quantitative characterizing of such a shape. On the other hand, on the basis of relationship (3) they define correlations between stationary components of PCRP. Here, lth covariance component is determined by the covariance functions of the components whose numbers differ by l. If such correlations are insignificant the covariance components of the respective numbers will be small. Inequality to zero of the covariance components estimators, besides the zeroth, is explained by existing correlations between the stationary components of PCRP. Actually, the consequence of such correlations is a rhythmic, which appears in periodicity of moment functions of the second order. In the case there are no correlations between the stationary components, neither variance nor covariance function will vary in time, i.e. power of fluctuations and correlations between them during a year will be constant. But if we find the second-order periodicities during processing, the processes that modulate harmonics of annual variation are correlated. Covariance component estimators allow the analysis of such correlations. For example, for two covariance components accounting for two components in harmonic decomposition (2) on the basis of (3) we obtain

$$B_1(u) = R_{-1,-2}(u)\, e^{-i\omega_0 u} + R_{-1,0}(u) + R_{0,1}(u)\, e^{i\omega_0 u} + R_{1,2}(u)\, e^{2i\omega_0 u},$$

$$B_2(u) = R_{-2,0}(u) + R_{-1,1}(u)\, e^{i\omega_0 u} + R_{0,2}(u)\, e^{2i\omega_0 u}.$$

If covariance functions $R_{0,1}(u)$, $R_{1,2}(u)$, $R_{-1,1}(u)$, $R_{0,2}(u)$ are slowly damped functions, then oscillations with annual and semiannual periods will appear in the graphs of covariance component estimators and at $R_{1,2}(u) \approx 0$ and $R_{0,2}(u) \approx 0$ only of the annual period. Consider now the processing results. Graphs of the most significant first and second covariance component estimators of air temperature and sea water saltiness are given in Fig. 9. As it can be seen in these cases the covariance component estimators do not have any oscillation properties. So we can suggest that correlations between processes $\xi_1(t)$, $\xi_2(t)$ and the zeroth random process $\xi_0(t)$ contain powerful rapidly damping components, against the background of which the correlations' oscillated changes are slightly visible. Note that the speed of higher component estimators damping is, as a rule, higher than the zeroth component estimators. It is difficult to draw a conclusion about the prevailing correlations between covariance components on the perspective of the base of correlogram even if their appearance does not contradict the fact that such correlations exist between the zeroth and higher stationary components.

Note that taking into account the character of covariance components dependence on lag we can choose such components that are significant for description of the process correlation structure. Neglecting the components which have a view of low-powered chaotic oscillations and using the component method of estimation we can considerably improve the estimator accuracy of covariance function time variety form.

The analysis of estimator for time-varying spectral density of PCRP and its components allows to carry out frequency properties of the annual variation processes. Time-varying spectral density is a complex function. Its real part describes distribution of process power on the area (ω, t) and the imagine part gives a frequency characteristic of relative speed of correlation changes on time and lag [1].

The estimators of time-varying spectral density $\hat{f}(\omega, t)$ and spectral components $\hat{f}_k(\omega)$ can be computed using Blackman-Tukey method, i.e. on the basis of statistics

$$\hat{f}(\omega, lh) = \frac{\Delta u}{2\pi} \sum_{n=-L}^{L} \hat{b}(n\Delta u, lh)\, k(n\Delta u)\, e^{-i\omega n \Delta u},$$

$$\hat{f}_k(\omega) = \frac{\Delta u}{2\pi} \sum_{n=-L}^{L} \hat{B}_k(n\Delta u)\, k(n\Delta u)\, e^{-i\omega n \Delta u},$$

where $k(u)$ is the Hemming correlation window. Since the process correlations rapidly damp, the lag value, which corresponds to point of correlogram cut-off, is chosen small. Calculations show that estimator values decrease with the frequency growths (Fig. 10). Characteristic peaks are represented weakly on graphs. Estimators of $Re\,\hat{f}(\omega, t)$ significantly change during a year and it concerns both their value and the speed of damping with the frequency increases. Annual rhythmic appears within the wide range of frequencies from inter-annual to within-year variety. Form of annual oscillations $Re\,\hat{f}(\omega, t)$ changes insignificantly by frequency

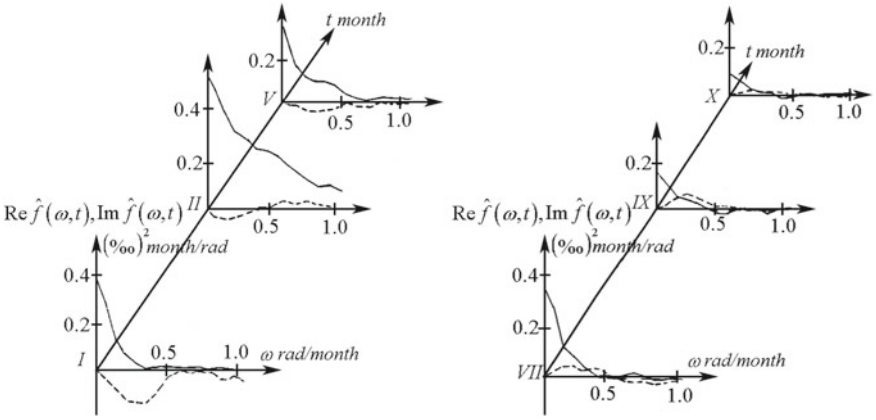


Fig. 10 Estimators of spectral characteristics for sea water saltiness in different months ($-Re\,\hat{f}\,(\omega,t)$; $-\mathrm{Im}\,\hat{f}\,(\omega,t)$)

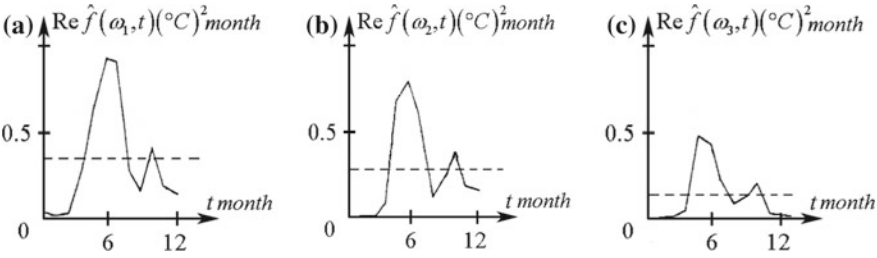


Fig. 11 Estimators of dependences on time of $Re\,\hat{f}\,(\omega,t)$ for water temperature at different frequencies: **a** $-\omega_1 = 0.05$ rad/months; **b** $-\omega_2 = 0.5$ rad/months; **c** $-\omega_3 = 1$ rad/months; (–values of $\hat{f}_0\,(\omega_0)$)

and practically repeats variance form (Fig. 11). Values of $Re\,\hat{f}\,(\omega,t)$ for some months may be greater or less than the values calculated within stationary approximation of PCRP, i.e. values of $\hat{f}_0\,(\omega)$ (the last counted are presented in graphs by dashed-and-dotted lines). Annual changes of estimators $\mathrm{Im}\,\hat{f}\,(\omega,t)$ occur near zero. Their maximal absolute values are significantly lower than the maximal deflection of $Re\,\hat{f}\,(\omega,t)$ from $\hat{f}_0\,(\omega)$ and it is natural, since in the time interval equal to the correlation interval the covariance function changes relatively slowly. Since practically for all (ω,t) estimators of $Re\,\hat{f}\,(\omega,t)$ are greater than zero, so in the first approximation we can accept that in many cases annual variation can be described by locally stationary PCRP and the quantity $\hat{f}\,(\omega) \approx Re\,\hat{f}\,(\omega,t)$ can be interpreted as an instantaneous power spectral density.

Estimators of the zeroth spectral component $\hat{f}_0\,(\omega)$, contrary to spectral density estimators of series obtained in stationary approximation, do not contain any pronounced peaks at annual frequency and frequencies multiple to it (Fig. 12). So,

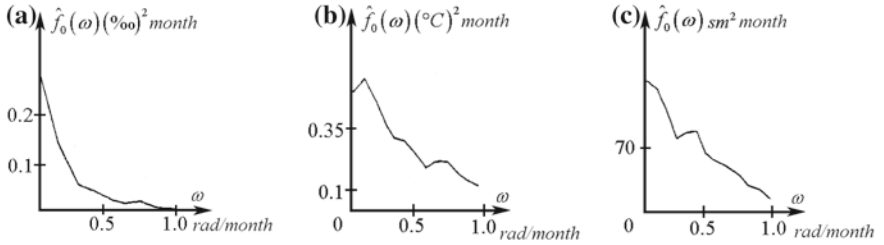


Fig. 12 Estimators of the zeroth spectral component for sea water saltiness (**a**), water temperature (**b**) and sea level oscillation (**c**)

considerable power concentrated at these frequencies is caused by regular annual changes. Values of $\hat{f}_0(\omega)$ decrease as frequency increases, that is, the evidence of slow decrease of the modulated stationary components spectral densities with frequency growth.

Analysis of $Re\hat{f}(\omega, t)$ and $\mathrm{Im}\,\hat{f}(\omega, t)$ estimators proves that change in annual variation has a wideband character. Estimator values gradually decrease with frequency growth, although for some processes some frequency ranges can be separated where the intensity of time variety of annual, semiannual, etc., periods increase or decrease a little.

Hence, the model of annual variation of geophysical processes in the form of PCRP gives us a possibility to analyse the following probabilistic characteristics: mean function $m(t)$ and variance $b(t, 0)$ during an annual period, their amplitude and phase spectra, covariance function $b(t, u)$, covariance components $B_n(u)$, spectral density $f(\omega, t)$, spectral components $f_k(\omega)$. The variety of chosen characteristics opens new possibilities for a detailed analysis of annual rhythmic regularities. Such an approach, by considering periodicity and stochasticity from not alternative but common positions, allows to detect and to describe these properties that result from the interaction between these important components of annual variation and cannot be investigated using harmonic analysis or stationary random process methods.

3 Daily Variation

Daily variations, like annual ones, are clearly seen in the realizations of many geophysical processes (Figs. 1 and 13a). Estimators of their covariance functions have undamped tails (Fig. 13b) and the estimators of spectral density peak at the daily frequency and multiple to it (Fig. 13c). During the PCRP analysis of daily variation it is accepted that $T = 24\,\mathrm{h}$ and it is, as a rule, carried out using hourly values [3, 19, 23]. Choosing the realization length we should account that daily and annual variations interact and the form of the first one changes during a year. Usually, daily variation is analysed on the basis of data observed for a month. Since in this interval the valid results of the influence estimation of synoptic ranges components on

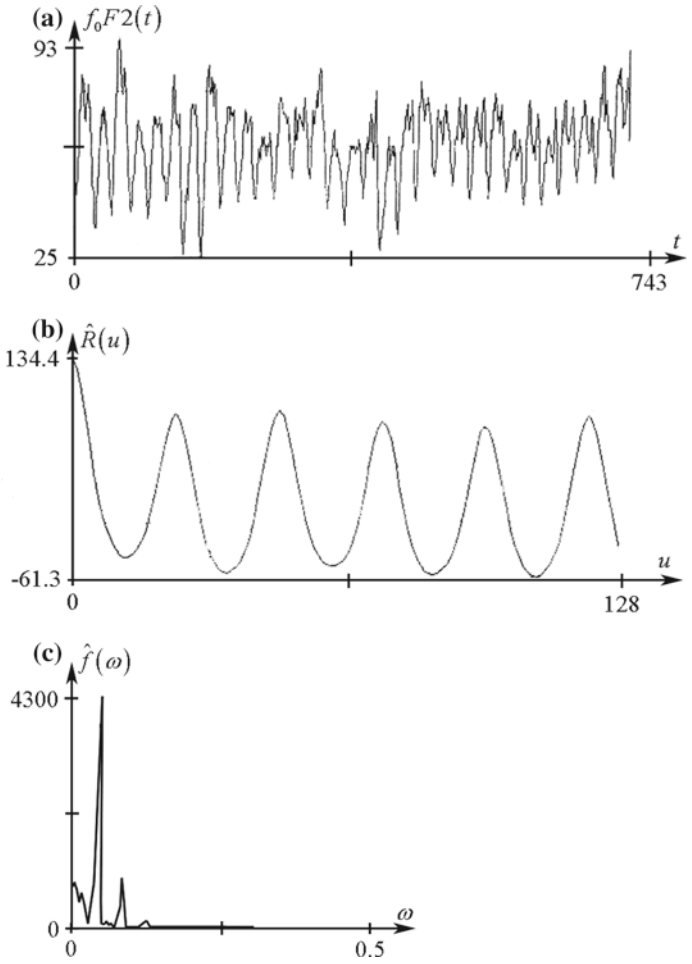


Fig. 13 Changes of critical frequency f_0F2 in time (**a**), covariance function estimator (**b**) and spectral density (**c**) of its stationary approximation

the daily variation cannot be obtained they should not be considered. The accepted filtration methods are used for this procedure [3, 14, 18].

Variety of daily variation properties appear in the relationships between powers of deterministic and fluctuation parts of the processes and also between amplitudes and phases of characteristics harmonic components, which describe them (Figs. 14 and 15).

Harmonic composition of the daily regular part, as it can be seen from the obtained results, is not wide. Only the first 2–3 harmonics have significant values. For air temperature (Fig. 16) the amplitude of the first harmonic significantly exceeds the amplitudes of others.

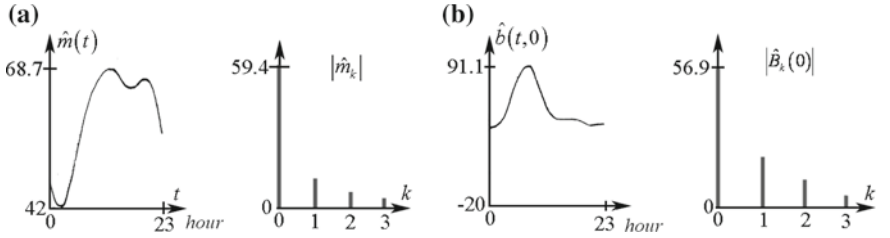


Fig. 14 Estimators of mean function (**a**) and variance (**b**) and absolute value of their Fourier coefficients for critical frequency f_0F2

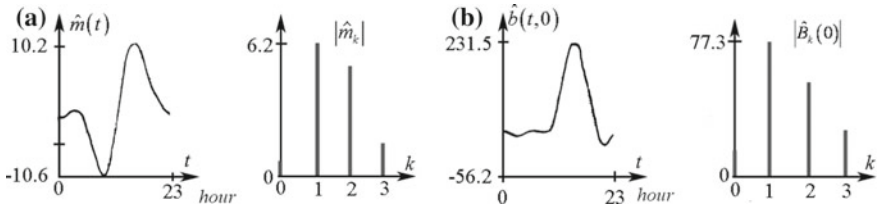


Fig. 15 Estimators of mean function (**a**) and variance (**b**) and absolute value of their Fourier coefficients for vertical component of the Earth magnetic field H_z

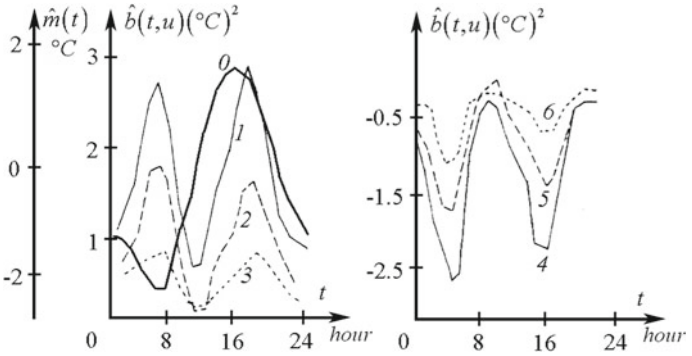


Fig. 16 Estimators of mean function (0), variance (1) and covariance function (2) dependences on time of air temperature daily variety in August for Krasnodar Territory: (2) $u = 24$ h; (3) $u = 48$ h; (4) $u = 12$ h; (5) $u = 36$ h; (6) $u = 60$ h

Variance estimators have the wider harmonic composition and their form is specific both for every process and for the season when the observations were carried out. In the given case for vertical component of the Earth's magnetic field H_z and for the critical frequency f_0F2 they have one maximum and for the first quantity it is close to the point of maximum of the regular part and for the second, to the interval of its growth. Variance estimator of air temperature has two maxima close to the points of extremes of regular variation.

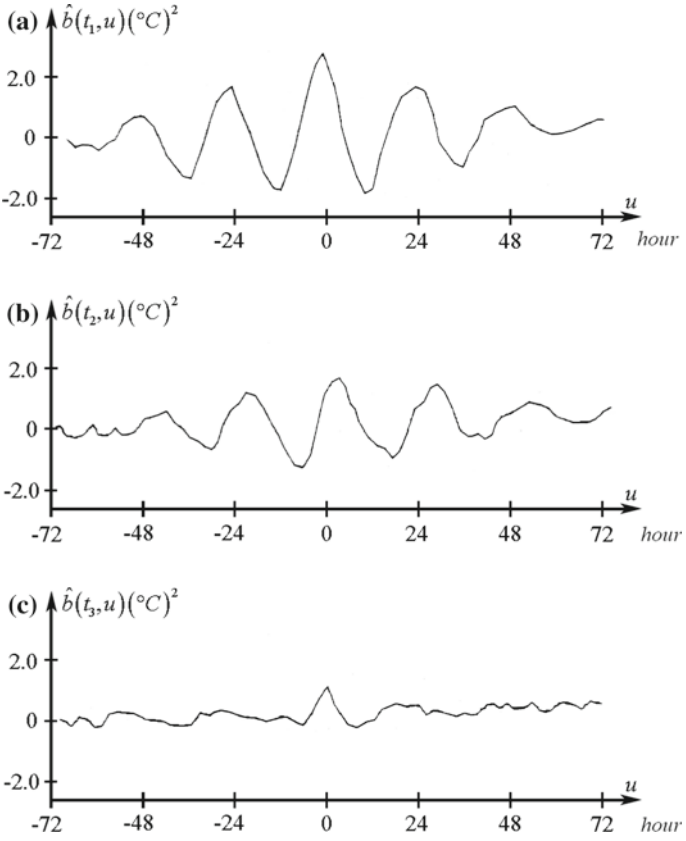


Fig. 17 Dependences of air temperature covariance function estimator on lag for different moments of time during the day: **a** $t_1 = 6\,\text{h}$; **b** $t_2 = 14\,\text{h}$; **c** $t_3 = 22\,\text{h}$

Contrary to estimators of annual variation covariance functions in the case of daily variation these estimators have a dumping-oscillating character. This can clearly be seen from the graphs of covariance functions estimators and their covariance components (Figs. 17, 18, 19 and 20). Values of spectral characteristics estimators for positive frequencies are concentrated in the area of daily frequency (Figs. 21 and 22), thus proving that modulation of main harmonic of PCRP model is of low frequency. Changes of $Re\,\hat{f}\,(\omega, t)$ in time are significant and occur practically synchronously with variance changes in time. Range of $\text{Im}\,\hat{f}\,(\omega, t)$ is significant too and it is a consequence of correlations of slow damping as lag increases in comparison with its changes in time.

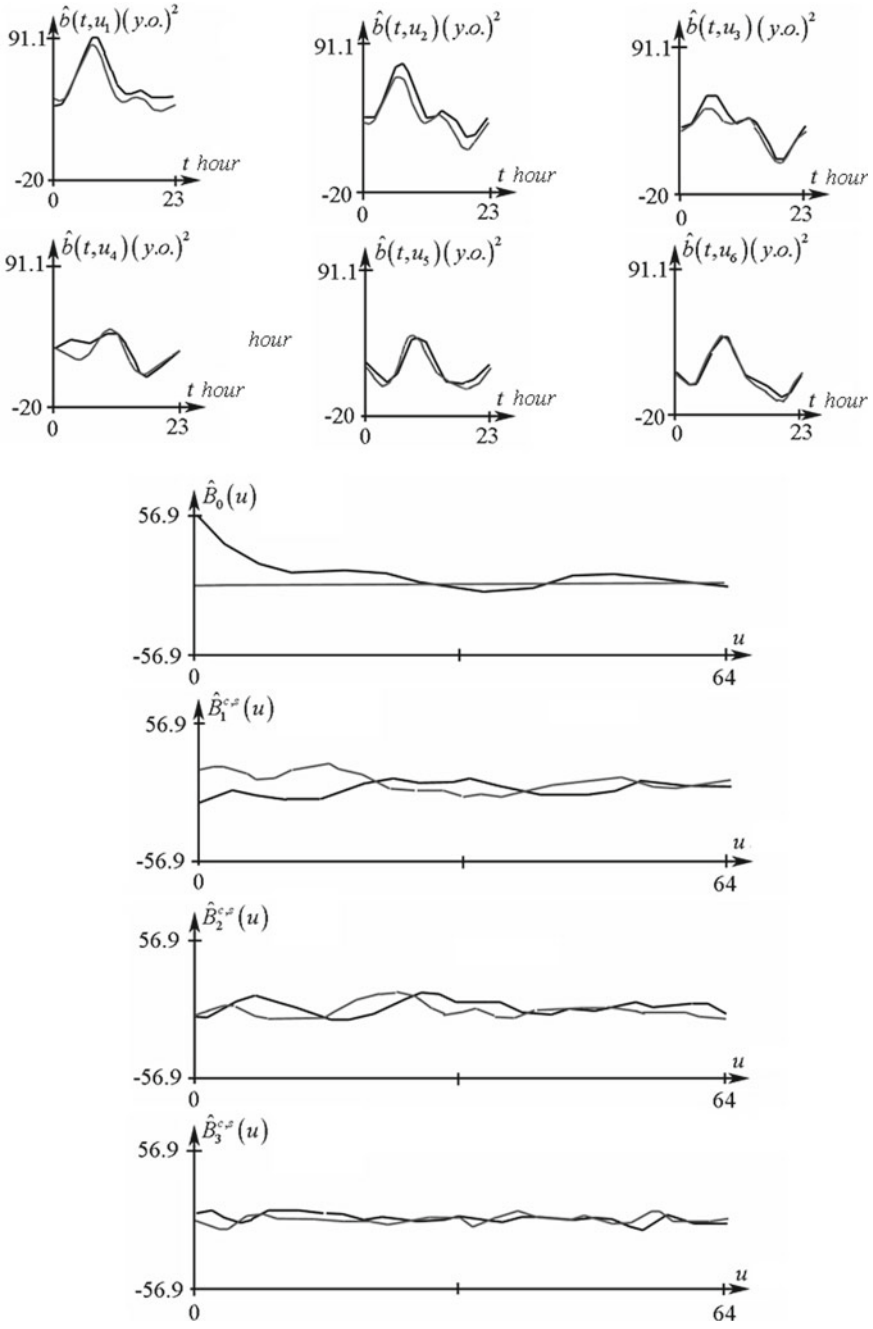


Fig. 18 Estimators of dependences on time for critical frequency f_0F2 covariance function for different lags ($u \in [0;\ 24]$) and estimators of its covariance components

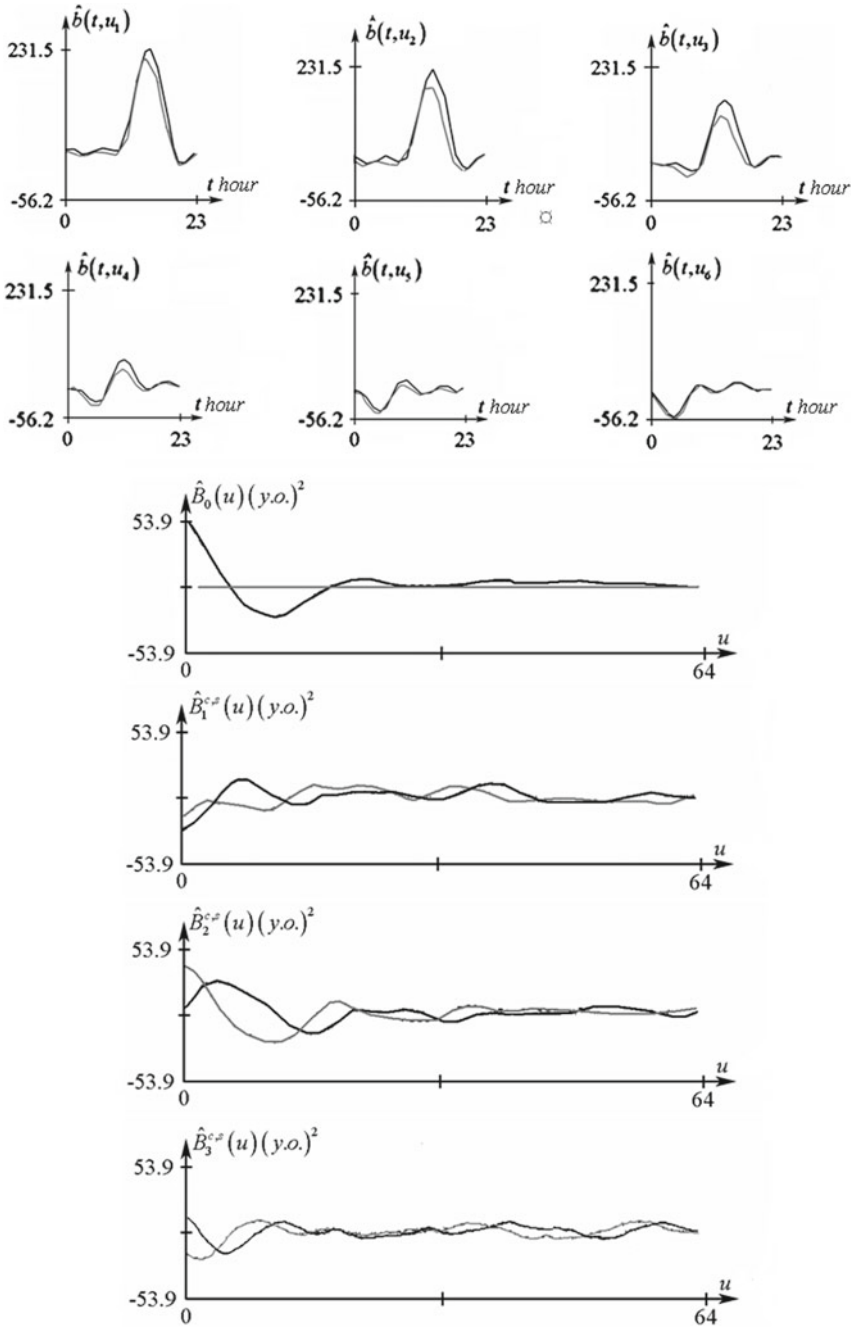


Fig. 19 Dependence of covariance function estimator for vertical component of the Earth's magnetic field H_z on time for different lags ($u \in [0;\ 24]$) and estimators of its covariance components

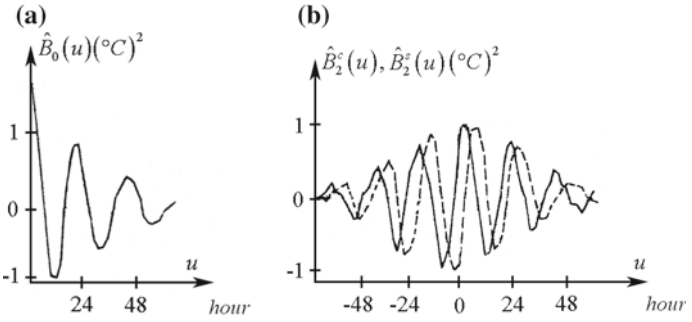


Fig. 20 Estimators of the zeroth (**a**) and the second covariance components (**b**) for air temperature

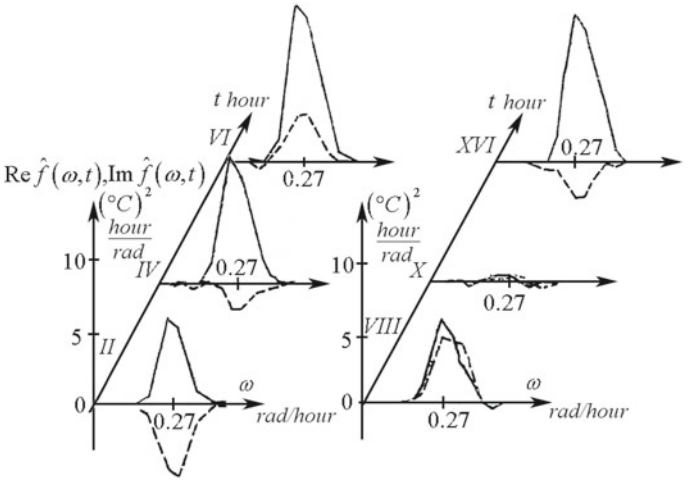


Fig. 21 Spectral characteristic estimators for air temperature daily variety ($-Re\,\hat{f}(\omega, t)$,$-\mathrm{Im}\,\hat{f}(\omega, t)$)

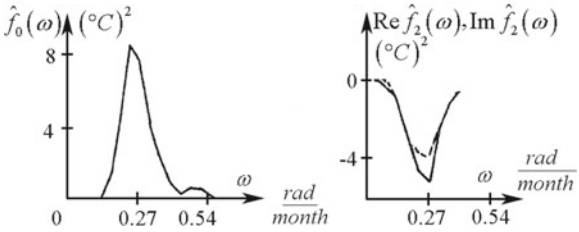


Fig. 22 Estimators of spectral components computed for air temperature daily variety ($-Re\,\hat{f}_2(\omega)$, $-\mathrm{Im}\,\hat{f}_2(\omega)$)

The detected properties of correlation and spectral characteristic estimators allow us to conclude that the structure of daily variation PCRP model is formed generally by the first covariance components in representation (2), i.e.

$$\xi(t) = \xi_{-1}(t)\, e^{-i\omega_0 t} + \xi_1(t)\, e^{i\omega_0 t} = \xi_c(t)\cos\omega_0 t + \xi_s(t)\sin\omega_0 t,$$

where $\xi_1(t) = \frac{1}{2}[\xi_c(t) - i\xi_s(t)]$ and $\xi_{-1}(t) = \overline{\xi_1}(t)$. Autocovariance functions of quadrature components

$$R_c(u) = E\overset{\circ}{\xi}_c(t)\overset{\circ}{\xi}_c(t+u), \quad \overset{\circ}{\xi}_c(t) = \xi_c(t) - m_c,$$

$$R_s(u) = E\overset{\circ}{\xi}_s(t)\overset{\circ}{\xi}_s(t+u), \quad \overset{\circ}{\xi}_s(t) = \xi_s(t) - m_s$$

and their cross-covariance functions slowly damp as lag increases.

Daily variation properties, as mentioned above, depend on the annual seasons. In Fig. 23 graphs of mean function and variance estimators for air and water temperatures in the Saint Petersburg recreation zone for separate months are shown. Regular daily oscillations in November–December are small, they have the highest amplitude in summer. The first harmonic is the main for all months. In the month with small regular oscillations variance changes are also small. The greatest amplitude of daily variation variance is in March–April. For the air temperature variance estimators $\hat{b}(t,0)$ have maxima at the minima of $\hat{m}(t)$ in winter, in spring these maxima shift left to $\hat{m}(t)$ maxima and in summer time changes of $\hat{b}(t,0)$ and $\hat{m}(t)$ occur practically in one phase. In autumn maxima of variances estimators pass ahead of the arrival of average temperature maxima. For the water temperature changes in time of $\hat{b}(t,0)$ and $\hat{m}(t)$ occur more synchronously but it can be seen that the point of variance maximum from spring to winter moves in the reverse direction.

The above results of temperature series processing clearly show that annual and daily variations of the processes cannot be considered separately. The changes occurring in the probabilistic structure of daily rhythmic during the year cannot be explained by interference of annual and daily variations. Modulation effects appear here significantly. It allows us to choose for description of the model of annual and daily rhythmics not in the form of two PCRP sum but in the form of bi-PCRP (BPCRP), which also characterizes modulation interaction between two rhythmics. Mean and covariance function of BPCRP can be represented in the form [3]

$$m(t) = \sum_{l,n\in Z} m_{l,n} e^{i\left(l\frac{2\pi}{T_1} + n\frac{2\pi}{T_2}\right)t},$$

$$b(t,u) = \sum_{l,n\in Z} B_{l,n}(u)\, e^{i\left(l\frac{2\pi}{T_1} + n\frac{2\pi}{T_2}\right)t},$$

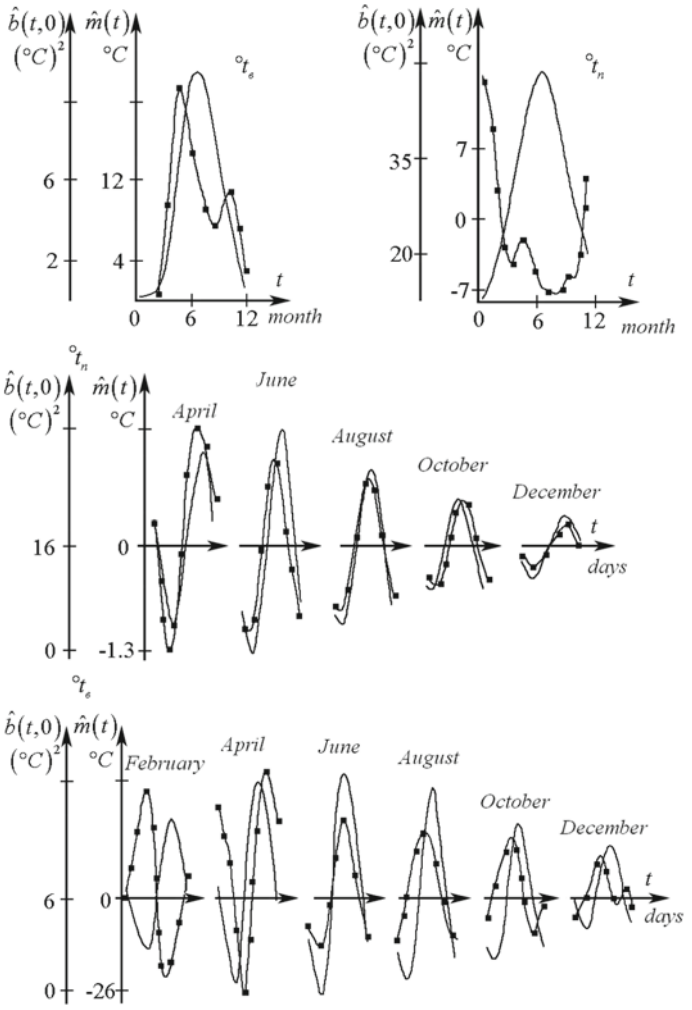


Fig. 23 Characteristic estimators for annual and daily variations of air (t_a) and water (t_w) temperatures in the Saint Petersburg recreation zone ($-\hat{b}\,(t,0)$, $-\blacksquare-\blacksquare\ \hat{m}\,(t)$)

where T_1 and T_2 periods of annual and daily rhythmics. The modulation interaction is quantitatively characterized by mutual components of mean function $m_{l,n}$ and covariance function $B_{l,n}\,(u)$. The estimators of these characteristics were computed on the basis of 14-year series of air temperature in Saint Petersburg using the formulae

$$\hat{m}_{l,n} = \frac{1}{K}\sum_{l=0}^{K-1} \xi\,(lh)\, e^{i\left(l\frac{2\pi}{T_1}+n\frac{2\pi}{T_2}\right)lh},$$

Table 3 Estimators of combinational components amplitudes for mean function and variance

Characteristics	Components	$\lvert\hat{m}_{l\pi}\rvert$				$\lvert\hat{B}_{l\pi}(0)\rvert$			
		n							
	l	0	1	2	3	0	1	2	3
Air temperature	0	4.8	12.8	0.5	0.9	23.2	15.0	10.2	8.0
	1	1.8	0.9	0.1	0.1	0.5	2.3	0. 7	1.1
	2	0.3	0.1	0.1	0	0.2	0.7	0.5	0.5
Water temperature	0	7.7	10.1	2.1	0.8	4.0	3.4	2.8	1.2
	1	0.7	0.5	0.1	0.1	0.7	0.5	0.3	0.1

$$\hat{B}_{l,n}(u) = \frac{1}{K}\sum_{l=0}^{K-1}\left[\xi(lh) - \hat{m}(lh)\right]\left[\xi(lh+u) - \hat{m}(lh+u)\right]e^{i\left(l\frac{2\pi}{T_1}+n\frac{2\pi}{T_2}\right)},$$

where

$$\hat{m}(lh) = \sum_{l=-N_1}^{N_1}\sum_{l=-N_2}^{N_2}\hat{m}_{l,n}e^{i\left(l\frac{2\pi}{T_1}+n\frac{2\pi}{T_2}\right)lh}$$

and N_l are the number of mean harmonics taken into account at estimation ($N_l < 5$). Estimation results are given in Table 3. First, daily and annual components are the most significant. Relative contribution of respective harmonic components to daily oscillations are the highest and especially concerns the variance estimator. These facts are the features of strong nonlinearity of daily and annual variations interaction.

Properties of BPCRP model characteristics estimators demonstrate that synoptic processes have considerable influence on the variety of daily and annual variations. Such influence leads to the appearance of additional maxima in the additive annual part of variance estimator and the powerful oscillations of daily period in its changes. The main feature of time variety of bi-rhythmic characteristics is the fact that neither the maximum nor the minimum ranges of daily oscillations of mean function and variance estimators coincide with moments of the highest and the least values of the additive annual parts of these characteristics.

4 The Properties of Annual and Daily Characteristic Estimators

The found properties of daily and annual rhythmics became a basis for passing to the parametric description of probabilistic characteristic, which represents some general views of the structure of these types of recurrence. A typical feature of geophysical processes' annual rhythmic is that estimators of harmonic component amplitudes for

mean and covariance function of their PCRP model rapidly damp as their number grows. It gives us a possibility to use finite sums in the representation of PCRP in terms of stationary components

$$\xi(t) = \sum_{k=-L}^{L} \xi_k(t)\, e^{ik\omega_0 t},$$

and from here we have

$$B_n(u) = \begin{cases} \sum\limits_{l=n-L}^{L} R_{l-n,l}(u)\, e^{il\omega_0 u}, & n \geq 0, \\ \sum\limits_{l=-L}^{n+L} R_{l-n,l}(u)\, e^{il\omega_0 u}, & n \leq 0. \end{cases}$$

Covariance components structure of every concrete process is determined by auto- and cross-covariance functions of the covariance components forming it. If cross-covariance functions of components whose numbers differing by number n are insignificant, the covariance components with number n will be small. If estimator of any covariance component is not zero, it is a consequence of correlation dependence between some stationary components of PCRP (except of the zeroth covariance components defined by their autocovariance functions).

Note that even if the number of components in representation (3) is not large the choice of approximation expression for covariance components is not an easy task because of the large number of functions $R_{l-n,l}(u)$. It is equal to $2L - n + 1$ for every covariance component. Because of it the analysis of processing results for detection of these components, correlations between which are the strongest, becomes very important. Such an analysis should be done with the use of covariance component estimators and their Fourier transformation spectral components. PCRP spectral components are defined by spectral and cross-spectral densities shifted by values multiple to the main frequency ω_0.

Analysis of correlation and spectral characteristics estimators of annual variation of many geophysical processes shows that all stationary components except of the zeroth are random processes with rapidly damping covariance functions and their spectral densities are significantly less than the spectral density of the zeroth component. So, correlations between other stationary components and the zeroth one are more significant than autocovariations. Views of graphs of covariance and spectral components estimators (Figs. 8, 9 and 12) do not contradict it. So, for the annual rhythmic we can assume that

$$B_0(u) = R_{00}(u) + \sum_{l=L}^{L} \left[\frac{1}{2} \left[R_{ll}^{c}(u) + R_{ll}^{s}(u) \right] \cos l\omega_0 u + N_{ll}^{cs}(u) \sin l\omega_0 u \right], \quad (8)$$

$$B_n^{c}(u) = R_{n0}^{c}(u) + R_{n0}^{c}(u) \cos n\omega_0 u + R_{0n}^{s}(u) \sin n\omega_0 u, \quad (9)$$

$$B_n^s(u) = R_{n0}^s(u) + R_{n0}^s(u)\cos n\omega_0 u - R_{0n}^c(u)\sin n\omega_0 u. \tag{10}$$

As it can be seen, the expression for the zeroth covariance component (8) contains oscillated components with period T/l. If in the graphs of this quantity estimators the oscillations with such periods do not appear, we can suppose what covariances $R_{ll}^{cs}(u)$ and $N_{ll}^{cs}(u)$ are insignificant and can be neglected:

$$B_0(u) = R_{00}(u).$$

Formulae (9)–(10) do not contain monotonous components, so if graphs of higher covariance component estimator oscillations with annual, semiannual, etc., periods are absent, we can assume that they damp before reaching the zeroth level. In this case functions can be represented as

$$R_{n0}^{c,s}(u) = D_n^{c,s} e^{-\alpha_n^{c,s}|u|}.$$

Similar approximation may be assumed for estimators $\hat{B}_0(u)$ if they monotonously damp. If correlograms $\hat{B}_0(u)$ have the oscillated character (periods of oscillations are not equal to annual one or to the values multiple to it) the approximation expressions for the zeroth covariance component can be written in the form

$$B_0(u) = \sum_{p=0}^{M} D_p e^{-\beta_p|u|}\cos\omega_p u.$$

Such oscillations should be taken into account for approximations of higher covariance components, assuming that

$$R_{n0}^{c,s}(u) = \sum_{p=0}^{M} D_{p_n}^{c,s} e^{-\beta_{p_n}^{c,s}|u|}\cos\omega_p u.$$

Damping decrements β_{p_n} are the parameters describing the speed of correlations damping by lag. Calculation of damping decrements using respective dependences of covariance components estimators for many geophysical processes shows that their values vary with a small deviation about unity.

Having an analytical expression for annual variation covariance function in the form

$$b(t,u) = R_{00}(u) + \sum_{n=1}^{L}\left[\begin{array}{l} R_{n0}^c(u)\cos n\omega_0 t + R_{n0}^c(u)\cos n\omega_0(t+u) + \\ +R_{n0}^s(u)\sin n\omega_0 t + R_{n0}^s(u)\cos n\omega_0(t+u) \end{array}\right] \tag{11}$$

we can analyse estimator accuracy of the probabilistic characteristics that describe it. Coherent estimator of mean function is unbiased and its variance is obtained by substituting into expression

$$D\left[\hat{m}\,(t)\right] = \frac{1}{N}\left[b\,(t,0) + 2\sum_{k=1}^{N-1}\left(1-\frac{k}{N}\right)b\,(t,kT)\right]$$

ratio (11). Since covariance functions (11) rapidly damp as lag increases, the sum in the last formula can be neglected. Computation results also prove this. Then

$$D\left[\hat{m}\,(t)\right] \approx \frac{b\,(t,0)}{N}.$$

Periodic change in time of mean estimator variance, as it can be seen, occurs synchronously with the change in variance that describes the given series. Since the last one significantly varies in time, the estimator variance changes during a year. For example, quantity $D\left[\hat{m}\,(t)\right]$ for water temperature in the Saint Petersburg recreation zone has the maximum in June. If $\hat{b}\,(t,0) = 3.62\,^{\circ}\mathrm{C}$ and at $N = 26$ then $D\left[\hat{m}\,(t)\right] = 0.14\,^{\circ}\mathrm{C}$ and the mean square root error $\sigma\left[\hat{m}\,(t)\right] = \sqrt{D\left[\hat{m}\,(t)\right]} = 0.37\,^{\circ}\mathrm{C}$. So, the averaged many-year value temperature in June, which equals to 15.4 °C, was found with an error ±0.37 °C. Relative mean square root error determined by quantity $\gamma = \sigma\left[\hat{m}\,(t)\right]/\hat{m}\,(t)$ is equal to 0.024. In July, when the average annual water temperature is the maximum and equals 18.3 °C, a mean square root error $\sigma\left[\hat{m}\,(t)\right] = 0.28\,^{\circ}\mathrm{C}$, $\gamma = 0.015$. The minimum average annual water temperature in February (0.011 °C) is calculated with a mean square root error $\sigma\left[\hat{m}\,(t)\right] = 0.015\,^{\circ}\mathrm{C}$., i.e. absolute error is greater than the value of temperature for this month.

Thus, processing the realizations of given length estimators of regular annual variation—so-called norm (standard)—will considerably differ in validity for different months. For characteristic of estimators quality for the whole year the average estimator variance $D_T\left[\hat{m}\,(t)\right]$ can be used, which in coherent estimation is determined by the zeroth covariance component $D_T\left[\hat{m}\,(t)\right] = B_0\,(0)/N$. Relation of the square root and regular oscillations range $\sigma_T\left[\hat{m}\,(t)\right]/\Delta m$ is also a very informative characteristic. For the above-mentioned temperature series $\sigma_T\left[\hat{m}\,(t)\right] = 0.22\,^{\circ}\mathrm{C}$ and $\sigma_T\left[\hat{m}\,(t)\right]/\Delta m = 0.01$.

Absolute value of covariance function estimator bias

$$\varepsilon\left[\hat{b}\,(t,u)\right] = -\frac{1}{N}\sum_{n=-N}^{N}\left(1-\frac{|n|}{N}\right)b\,(t,u+nT)$$

at $u = 0$ has the same properties as variance of mean function estimator and the relative value at this point with positive accuracy defined quantity inverse to number of periods N. Increasing lag to value $u_m \geq 6$ months bias absolute value decreases but after it increases to the value, which is a little smaller than the initial one. Relative

bias $\left|\varepsilon\left[\hat{b}\,(t,u)\right]\right|/|b\,(t,u)|$ for given u varies in time slightly and for the values of lag within the interval [0, 6 months] has an order $\mathrm{O}\left(N^{-1}\right)$. Further, it rapidly increases.

Increasing lag to point $u_m \geq 6$ months estimator variances decrease

$$D\left[\hat{b}\,(t,u)\right] = \frac{1}{N}\sum_{k=-N+1}^{N-1}\left(1-\frac{|k|}{N}\right)[b\,(t,kT)\,b\,(t+u,kT) + b\,(t,u+kT)\,b\,(t,u-kT)]\,.$$

However, increase of a relative mean square root error $\gamma = \left[D\left[\hat{b}\,(t,u)\right]\right]^{\frac{1}{2}}/|b\,(t,u)|$ starts immediately after moving from point $u = 0$. It can be explained that damping of function $D\left[\hat{b}\,(t,u)\right]$ by lag is slower than $b\,(t,u)$. To calculate estimator variances at the zeroth lag the approximated formula can be used

$$D\left[\hat{b}\,(t,u)\right] \approx \frac{2b^2\,(t,0)}{N},$$

and respectively $\gamma = \sqrt{2/N}$. So, a relative mean square root error of variance estimators is determined by the realization length.

Note that the dependence of coherent estimators statistical characteristics on damping decrements at their values greater than 0.3 month^{-1} can be neglected but no account of the value correlations when analysing the validity of component estimators can lead to considerable errors. As calculations show, value of $D\left[\hat{m}\,(t)\right]$ during component estimation, for example for damping decrements 0.3 month^{-1} is less than the respective variance of coherent estimator by 25 %. Increasing decrement values' difference between variances of coherent and component estimators for probabilistic characteristics increases and for the real data may exceed 50 %.

Calculating estimators of characteristics Fourier components with usage of average monthly data the overlapping errors will be practically absent for the sampling step values $h \geq 1$ month. Thus, it follows that a number of harmonic components that should be taken into account in Fourier series for respective characteristics estimators in most case is less than five. This means that the use of the average monthly data for components estimation does not lead to visible systematic errors of processing.

Increasing variances of components discrete estimators at $h = 1$ month in relation to variance of continuous estimators is less than 10 %. Variances of continuous and discrete component estimators differ similarly.

The mathematical model of daily rhythmic in the first approximation can be chosen in the form of quadrature components. Modulated processes are correlated and have different covariance functions. Damping decrements values for auto- and cross-covariances are within the interval $\left[0.01\ \mathrm{h}^{-1},\ 0.1\ \mathrm{h}^{-1}\right]$. The properties of coherent and component estimators for the quadrature model were investigated in this chapter

and also in Chap. 3. Obviously, the results of such investigations are easily extended for this concrete case.

As mentioned above, the parameters of daily variation changed during a year and because of it in real data processing we should choose such averaging intervals for which a systematic error will be small. Calculations show that the use of the average monthly series gives relative bias of sliding characteristics of PCRP to annual harmonic is about 1 %, second harmonic—5 %, third—10 %. Within 10 % constitute the values of relative errors for correlation characteristics estimators calculated with the use of these series. Statistic errors of mean function and its components estimators are less by an order.

Changing sampling step within [1 h, 4 h] the validity of daily variation estimators changes a little. So, sampling step value 4 h provides acceptable estimator accuracy. The component method [24] should be used for obtaining temporary values of estimators.

Difference between the accuracy of coherent and component estimators for daily rhythmic, opposed to annual, is not significant. An important feature of the latter is their dependence on the correlation damping speed.

5 Conclusions

Since the main idea of stochastic oscillations model in the form of PCRP and their generalizations consist of the properties of time variation of their probability characteristics, the first task for experimental data analysis is the determination of the time variety regularity using statistical methods for given classes of random processes. Obtained results of time series processing are the basis for model verification. Seasonal rhythmic of geophysical processes is accepted by many researchers but more detailed investigation of its regularity is possible only with the use of PCRP methods. It is not reduced to the additive model. Combination of recurrence and stochasticity features of each process is different because it depends on the conditions in which processes are formed. Specific features of seasonal variation of separate quantities are represented in amplitude and phase spectra of probabilistic characteristics, and also in behaviour of correlation characteristics by lag and spectral ones by frequency. They are represented in the degree of the process periodic non-stationarity, in the relations between time-averaged power of regular oscillations, on one hand, and average power of fluctuations with their oscillated power, on the other. Features of daily rhythmic are lower power of regular part and oscillated behaviour of correlations damping. Probabilistic structure of daily rhythmic significantly changes during a year, because of this we should analyse it together with seasonal. Regularities of such analysis are described by characteristics of BPCRP. The constructed parametric models of seasonal and daily rhythmics are the basis for description of regional features of processes variety, formation of databases, building of physico-mathematical models for the processes and their prediction.

References

1. Yavorskyj, I., Yuzefovych, R., Kravets, I., & Matsko, I. (2014). Methods of periodically correlated random processes and their generalizations. In F. Chaari, J. Leskow, & A. Sanches-Ramirez (Eds.), *Cyclostationarity: Theory and methods*. Lecture Notes in Mechanical Engineering. (pp. 73–93), New York: Springer.
2. Hurd, H. L., & Miamee, A. (2007). *Periodically correlated random sequences. Spectral theory and practice* (p. 353). New Jersey: Wiley-Interscience.
3. Dragan, Ya., Rozhkov, V., & Javorskyj, I. (1987). *Methods of probabilistic analysis of oceanological processes rhythmics*. Leningrad: Gidrometeoizdat (in Russian).
4. Gardner, W. A. (1987). *Statistical spectral analysis: A nonprobabilistic theory* (p. 566). New York: Prentice Hall.
5. Gardner, W. A. (1985). *Introduction to random processes with application to signals and systems* (p. 434). New York: Macmillan.
6. Gardner, W. A. (Ed.). (1994). *Cyclostationarity in communications and signal processing* (p. 504). New York: IEEE Press.
7. Gardner, W. A., Napolitano A., Paural L. (2006). Cyclostationarity: Half century of research. *Signal processing*. (Vol. 86, pp. 639–697).
8. Napolitano, A. (2012). *Generations of cyclostationarity signal processing: Spectral analysis and applications*. Wiley: IEEE Press.
9. Javorskyj, I., Isayev, I., Zakrzewski, Z., & Brooks S. P. (2007). Coherent covariance analysis of periodically correlated random processes. *Signal processing*. (Vol. 87, pp. 13–32).
10. Javorskyj, I., Isayev, I., Majewski, J., & Yuzefovych R. (2010). Component covariance analysis for periodically correlated random processes. *Signal processing*. (Vol. 90, pp. 1083–1102).
11. Javorskyj, I., Leskow, J., Kravets, I., Isayev, I., & Gajecka, E. (2012). Linear filtration methods for statistical analysis of periodically correlated random processes—Part I: Coherent and component methods and their generalization. *Signal processing*. (Vol. 92, pp. 1559–1566).
12. Javorskyj, I., Leskow, J., Kravets, I., Isayev, I., & Gajecka, E. (2011). Linear filtration methods for statistical analysis of periodically correlated random processes—Part II: Harmonic series representation. *Signal processing*. (Vol. 91, pp. 2506–2519).
13. Javorskyy, I., Yuzefovych, R., Krawets, I., & Zakrzewski, Z. (2011). Least squares method in the statistic analysis of periodically correlated random processes. *Radioelectronics and communications systems*. (Vol. 54(1), pp. 45–59).
14. Rozhkov, V. A. (1974). *Probabilistic analysis methods for oceanic processes (in Russian)*. Leningrad: Gidrometeoizdat.
15. Mykhailyshyn, V. Y., & Javorskyj, I. N. (1994). Probabilistic structure of air temperature seasonal variety. *Meteorologia i gidrologia*. No. 2, pp. 20–35 (in Russian).
16. Dragan, Y. P., Rozhkov, V. A., & Javorskyj, I. N. (1984). *CRP methods usage for probabilistic analysis of oceanic time series (in Russian)* (pp. 4–23). Leningrad: Gidrometeoizdat.
17. Rozhkov, V. A., Cherneshov, S. Y., Trapeznikov, Y. A., & Javorskyj, I. N. (1984). Comparison of estimation methods for probabilistic characteristics of oceanic processes seasonal variety (in Russian). *Mode forming factors, information bases and methods of their analysis* (pp. 138–149). Leningrad: Gidrometeoizdat.
18. Myakisheva, N. V., Rozhkov, V. A., Ulyanich, I. G., & Javorskyj, I. N. (1986). Probabilistic characteristics estimators for seasonal and daily varieties hydro-meteorological processes (in Russian). *Mode forming factors, information bases and methods of their analysis* (pp. 140–152). Leningrad: Gidrometeoizdat.
19. Mezentsev, V. P., & Javorskyj, I. N. (1989). Probabilistic methods for analysis of daily and seasonal varieties radiophysical processes. *All-USSR scientific and technical conference "Problems of radiophysical complexes and flight service enhancement"*. Kyiv. pp. 71–72 (in Russian).
20. Javorskyy, I. (1984). The application of Buys-Ballot scheme for statistical analysis of rhythmic signals. *Izvestiya Vysshikh Uchebnykh Zavedenič Radioelectronika*, *27*(11), 31–37. (in Russian).

21. Javorskyj, I., & Mykhajlyshyn, V. (1996). Probabilistic models and investigation of hidden periodicities. *Applied Mathematics Letters*, *9*(2), 21–23.
22. Mezentsev, V. P., & Javorskyj, I. N. (1987). Methods for statistical analysis of natural radiophysical processes daily rhythmic (in Russian). *All-USSR scientific and technical conference "Registration and analysis of very low-frequency oscillations with natural derivation"*. Voronezh. pp. 50–51.
23. Javorskyy, I. (1985). On statistical analysis of periodically correlated random processes. *Radiotekh. Electron.* (Vol. 6, pp. 1096–1104).
24. Yavorskyj, I., Dehay, D., & Kravets, I. (2014). Component statistical analysis of second order hidden periodicities. *Digital signal processing*. (Vol. 26, pp. 50–70).

Influence of Different Signal Characteristics on PAR Model Stability

Agnieszka Wylomanska, Jakub Obuchowski, Radosław Zimroz and Harry Hurd

1 Introduction

The periodic autoregressive model (PAR) is an extension of the autoregressive (AR) time series [2]. This extension is based on time-varying periodic coefficients of PAR time series (instead of fixed coefficients in AR model); therefore, it can be used to describe real signals from machines operating under time-varying conditions. The problem of examining such vibration signals is difficult due to the fact that most of their characteristics (like statistical properties, signal to noise ratio) are time-varying; therefore, the classical models seems to be inappropriate in this case, [1, 8, 19]. The examples of signals acquired from complex mechanical systems for which the classical methods cannot be used are helicopters, wind turbines and mining machines [1, 15, 17].

The PAR model was considered for many different real phenomena. We only mention here two examples: The electricity data [4] and vibration signal of multistage gearbox used in drive unit for driving bucket wheel [19]. In the last mentioned

A. Wylomanska (✉)
Department of Mathematics, Hugo Steinhaus Center,
Wroclaw University of Technology, Janiszewskiego 14a, Wroclaw, Poland
e-mail: agnieszka.wylomanska@pwr.edu.pl

J. Obuchowski · R. Zimroz
Diagnostics and Vibro-Acoustics Science Laboratory,
Wroclaw University of Technology, Na Grobli 15, Wroclaw, Poland
e-mail: jakub.obuchowski@pwr.edu.pl

R. Zimroz
e-mail: radoslaw.zimroz@pwr.edu.pl

H. Hurd
Department of Statistics, The University of North Carolina at Chapel Hill,
Chapel Hill, NC, USA
e-mail: hurd@stat.unc.edu

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods - II*,
Applied Condition Monitoring 3, DOI 10.1007/978-3-319-16330-7_5

example, the PAR model was used not only to describe the real data, but also as the method of local damage detection for machine operating under nonstationary conditions.

One of the most popular methods for estimation of periodic autoregressive model is based on solving of so-called Yule-Walker equations. This method is especially useful for signals with Gaussian noise. However, many real signals exhibit behavior not appropriate to Gaussian distribution, [21, 22]; therefore, there is a need to consider if the most commonly used estimation method for PAR coefficients can be used also for signals with different from Gaussian behavior and what is the influence of the noise distribution for PAR model stability. The PAR model with non-Gaussian distribution was also examined in [13]. In this paper, we consider Pareto distribution of noise [5] that was originally used to describe claim sizes in car insurances. The Pareto distribution found later another interesting applications; for example, it was used for modeling of distributions of diagnostic features in condition monitoring of mining machines [16]. The distribution is considered as a heavy-tailed one; therefore, it can be used for data with large probability of extremely high claims.

The second problem that may appear in the real signal analysis is variable periodicity and small number of period replications. For many data, we observe those properties similar as for non-Gaussian noise; and also in this case, the PAR model stability and the efficiency of estimation method should be checked. We motivate our analysis by a signal that represents vibration acceleration of rotating machinery which operate in an open-pit mine. The results obtained in this paper allow to answer the question whether the PAR model might be applicable to industrial signals and how far the signal might be from the ideal case.

The paper is structured as follows: In Sect. 2 we introduce the PAR time series with Gaussian and non-Gaussian behavior and point our attention to the Pareto distribution of the noise. Next, we motivate our analysis by examining the real signal of rotating machinery operating in an open-pit mine for which we observe non-Gaussian behavior. In Sect. 4, we introduce the method of checking of the PAR model stability for different distributions of the noise. Here we analyze simulated signals corresponding to vibration acceleration of a planetary gearbox used in a bucket wheel excavator. The proposed method is based on the analysis of amplitude of frequency response of examined simulated signals based on the appropriate estimated PAR model. Next, in Sect. 5, we propose the method of checking PAR model stability for different number of period repetitions. This method is based on the behavior of fitted PAR model coefficients. Similar as in Sect. 4, the results are presented for simulated signals corresponding to bucket wheel excavator. Last section contains conclusions.

2 Periodic Autoregressive (PAR) Model

Definition 1 ([11]) The periodic autoregressive time series of order p is defined as follows:

$$X(t) - \sum_{i=1}^{p} a_i(t)X(t-i) = b(t)Z(t), \tag{1}$$

where $\{Z(t)\}$ is a white noise time series, i.e. time series of uncorrelated random variables with mean zero and the same variance and the coefficients $\{a_i(t)\}\, i = 1, 2, \ldots, p$, $\{b(t)\}$ are periodic with the same period T.

The periodic autoregressive time series (PAR) is a special case of PARMA sequence (periodic autoregressive moving average), i.e. time series which is defined as:

$$X(t) - \sum_{i=1}^{p} a_i(t)X(t-i) = b_0(t)Z(t) + \sum_{j=1}^{q} b_j(t)Z(t-i). \tag{2}$$

In the above equation, the sequences of coefficients are also periodic with the same period T. Let us mention, the process {X(t)} defined in (2) is unique under the conditions specified in [20], where also the form of the unique solution is presented.

The PARMA sequence is one of the main time series which can be used to describe periodically correlated (or cyclostationary) processes. We only mention here, that periodically correlated (PC) random processes of second order are random systems for which the mean as well as the covariance (or correlation) functions are periodic with the same period, [12].

Due to their interesting properties, periodically correlated time series, especially PARMA systems have received much attention in the literature and turned out to be an alternative to the conventional stationary time series like classical ARMA (autoregressive moving average) systems. Some applications one can find in [3, 4, 10, 14, 18].

The estimation procedure of PAR model based on the so-called Yule-Walker equations is described in [19]. Further in this paper, we use the same method.

2.1 Non-Gaussian Noise

In the classical definition of PAR model, the white noise time series $\{Z(t)\}$ is considered as a sequence of Gaussian-distributed random variables. We only mention here that the one-dimensional Gaussian distribution is characterized by the probability density function (PDF) which has the following form

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in R.$$

In the above equation, the μ and σ parameters are mean and standard deviation of the Gaussian-distributed random variable, respectively.

In this paper, we consider not only the classical PAR model for which the sequence $\{Z(t)\}$ comes from Gaussian distribution but also white noise time series from Eq. (1) with distribution different than Gaussian. In the experimental and simulation part of the paper, we consider the rich class of distribution called Pareto one (known also as double Pareto). Let us mention the one-sided Pareto distributed random variable X has the PDF which has following form [5]:

$$f(x) = \frac{\alpha\lambda^{\alpha}}{(\lambda + x)^{\alpha+1}}, \quad x > 0\,. \tag{3}$$

Moreover the cumulative distribution function (CDF) is given by:

$$F(x) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^{\alpha}, \quad x > 0\,. \tag{4}$$

In Fig. 1 we present the PDF and CDF of Pareto distribution for $\lambda = 1$ and two different values of α parameter, namely 3.6 and 1.5.

In Eqs. (3) and (4), the shape parameter α and the scale parameter λ are both positive. The one-sided Pareto distribution is very useful in actuarial mathematics,

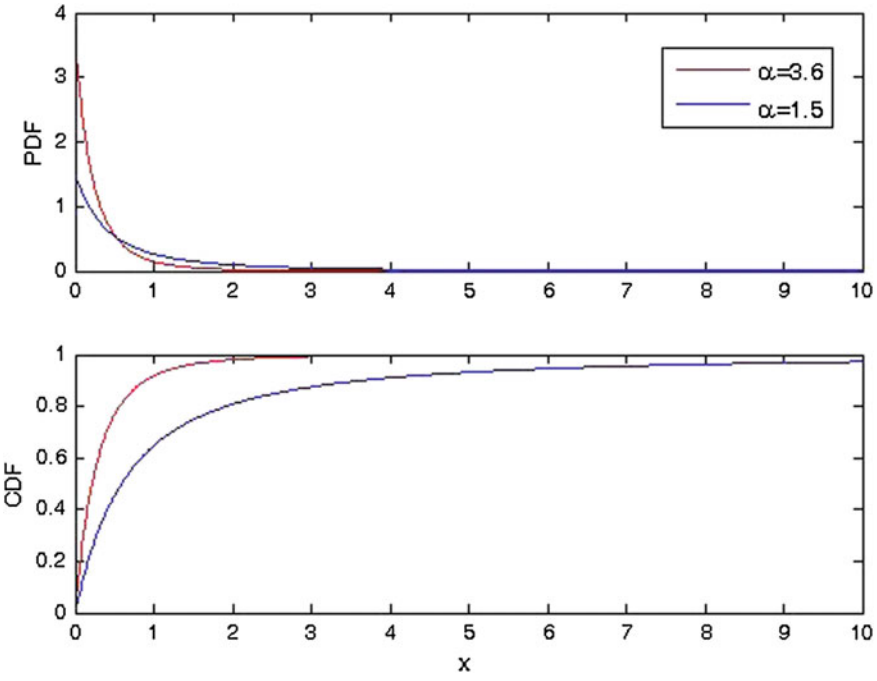


Fig. 1 The PDF (*top panel*) and CDF (*bottom panel*) of Pareto distribution with the parameter $\lambda = 1$ and two different values of α parameter, namely $\alpha = 3.6$ and $\alpha = 1.5$

e.g. to model claim sizes, due to the fact that the heavy tail of the Pareto distribution follows large probability of extremely high claims. Its main drawback lies in its lack of mathematical tractability in some situations [5]. This distribution was also used in the field of heavy-duty machinery maintenance management, i.e. to model diagnostic data [16] as an alternative (instead of Gaussian) distribution of diagnostic features. The main property of Pareto distribution is the asymptotic behavior of the tail (i.e. *1-F(x)*), which exhibits so-called power-law behavior, namely for Pareto distribution we have

$$1 - F(x) \approx x^{-\alpha}. \tag{5}$$

We should mention that for the Gaussian distribution, the tail behaves like an exponential function.

The second important property of the Pareto distribution is the existence of theoretical moments, namely theoretical moment of order k of one-sided Pareto distributed random variable exists only when $k < \alpha$. This important remark will be used in the simulation part of the paper.

In this paper, we consider an extension of the one-sided Pareto distribution, i.e. the double Pareto (called later as Pareto). Let us mention the Pareto distributed random variable Y is defined as follows:

$$Y = X - Z,$$

where X and Z have one-sided Pareto distribution with the same shape and scale parameters, moreover they are independent. And similar to the one-sided Pareto distribution, also for the double Pareto, the kth moment exists if $k < a$ and the theoretical tail behaves like a power function given in (5).

3 Motivation

The motivation of our research is the fact that a lot of considered real vibration signals does not always follow the Gaussian distribution. Moreover, for different signals we can observe different numbers of period repetitions. Therefore, there is need to consider how such facts influence the stability of the PAR model estimation procedure.

Below we present an exemplary signal for which we observe non-Gaussian behavior. The signal represents vibration acceleration of a bearing from a belt conveyor driving system from an open-pit mine.

The signal was acquired in horizontal direction using an accelerometer. Duration of the signal is 5 s and sampling frequency is 19,200 Hz. Since the gearbox operates in industrial conditions, the signal contains not only high-energy contamination from a gearbox located nearby, but also accidental impulses as well.

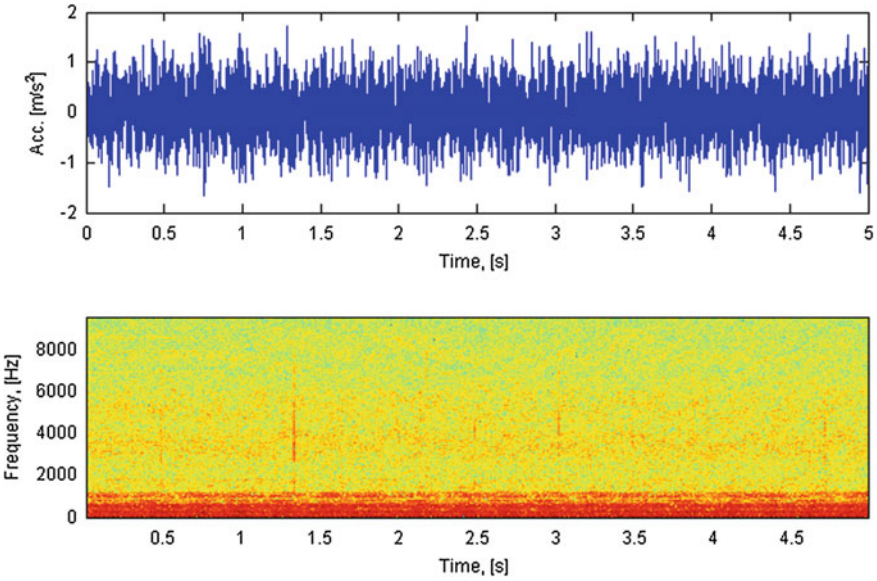


Fig. 2 Real data (*top panel*) and corresponding spectrogram (*bottom panel*)

Thus, the signal might be considered as the motivation of the analysis performed in this paper. In Fig. 2, we present the examined signal and the corresponding spectrogram. Because the non-flat spectrum of the signal can influence the distribution, then we prewhitened the examined signal before the further analysis. The result is presented in Fig. 3, where we show the obtained data and its corresponding spectrogram.

In order to confirm the signal cannot be modeled by using Gaussian distribution in Fig. 3 we present the visual test for gaussianity, namely the QQ plot on which we compare the empirical quantiles of examined signal (in our case after the prewhitening) and quantiles of appropriate Gaussian distribution (Fig. 4).mention, the QQ plot is a visual test on the basis of which we can conclude if given vector of observations can be modeled by a given theoretical distribution. If the empirical quantiles calculated for real data are compatible with theoretical ones, then we can suspect the empirical and theoretical distributions are the same [7]. As we observe, the empirical quantiles do not correspond to Gaussian theoretical quantiles. In order to show the Pareto distribution is more appropriate in the considered case, in Fig. 5, we show the empirical tail of prewhitened signal and the tail corresponding to Pareto distribution (i.e. the fitted power function (5)). As we observe, the function adequate to Pareto tail has similar behavior as the empirical one based on the prewhitened signal.

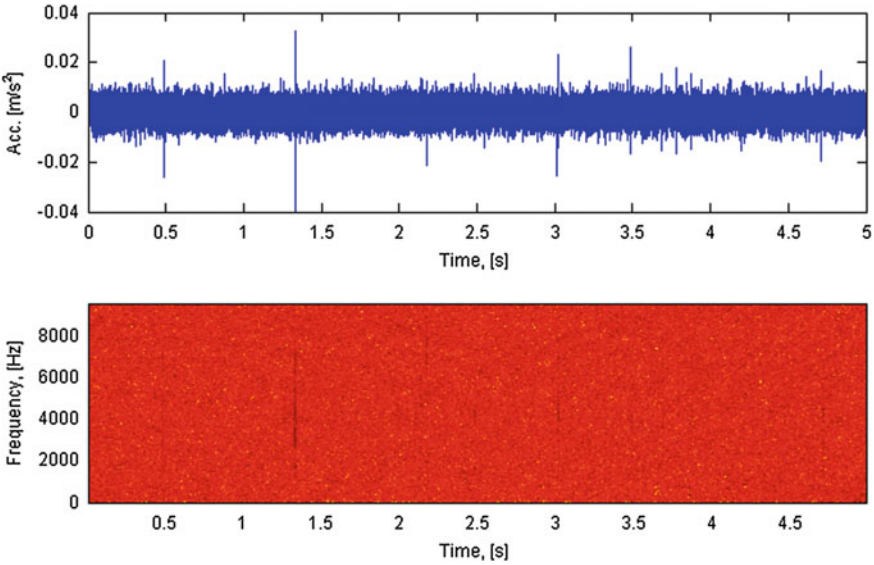
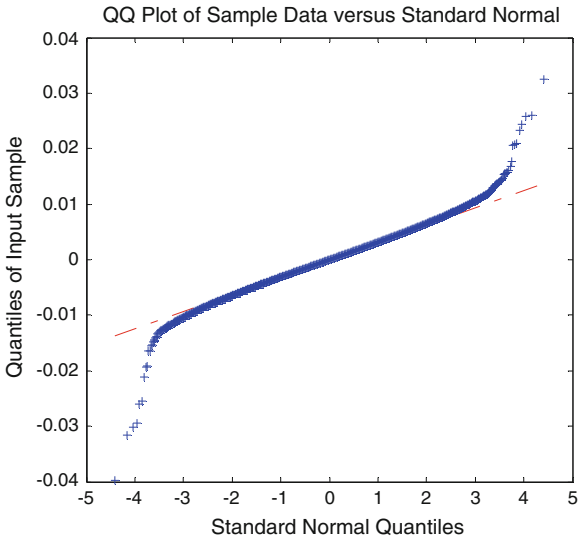


Fig. 3 Prewhitened real data (*top panel*) and corresponding spectrogram (*bottom panel*)

Fig. 4 The QQ plot for prewhitened real data. The visual test indicate the signal cannot be considered as the Gaussian-distributed one



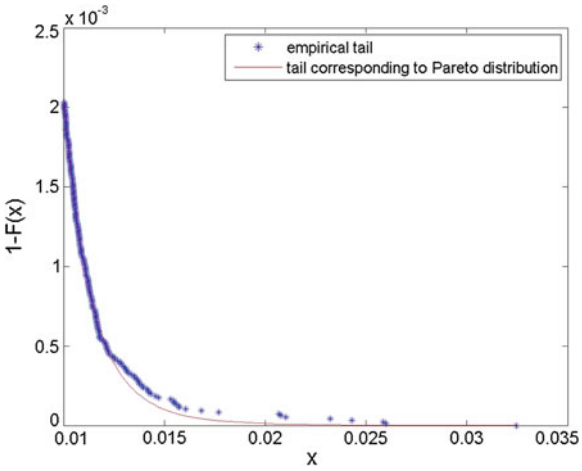


Fig. 5 The empirical tail of prewhitened signal and fitted tail corresponding to Pareto distribution. We observe the Pareto distribution corresponds to considered signal

4 Influence of Non-Gaussian Noise to PAR Estimation

4.1 Methodology

The signals analyzed in this section represent vibration acceleration of a planetary gearbox used in a bucket wheel excavator [9]. Such vibration signal might be simulated as a sum of several sinusoidal components with frequencies that meet the gear mesh frequency and its harmonics and a white noise. Due to the cyclic regime in which the excavator operates, the sinusoidal components are frequency and amplitude modulated with the period corresponding to the period of the bucket wheel operation [6]. Due to industrial environment, we decided to analyze white noises that follow the double Pareto distribution with parameters $\alpha_1 = 1.5$ (denoted later as Pareto 1.5) and $\alpha_2 = 3.6$ (denoted as Pareto 3.6) and compare result with the Gaussian case. Let us point out α_2, the examined Pareto distribution have finite second moment while for α_1 the second moment does not exists. This fact has important influence for the results.

Influence of non-Gaussian noise to PAR estimation is performed using the following algorithm. At first, a lot of signals related to each type of noise are simulated to preserve reliability of results. Energy of each noise sequence is normalized, i.e. time series are divided by its standard deviation to preserve fair comparison. Secondly, parameters of the PAR models are estimated for each signal. The procedure of estimation is based on the Yule-Walker method and is described in [19]. Order of the PAR model is chosen as 15 for every signal. This choice is motivated by the number of sinusoidal components and the fact, that the residual time series are satisfactory [19]. After that, we analyze amplitude response of the models at each $1 \leq t \leq T$,

namely a surface of amplitude response. It is a natural extension of the amplitude response of an autoregressive model. The autoregressive model (AR) of order p and parameters $a = (a_i)_{i=1,\ldots,p}$ is defined as follows:

$$X(t) - \sum_{i=1}^{p} a_i X(t-i) = Z(t),$$

where similar as in (1), the sequence $\{Z(t)\}$ is a white noise time series.

Amplitude response of an autoregressive model with coefficients $a = (a_i)_{i=1,\ldots,p}$ is defined as follows:

$$S(f) = \left| \frac{1}{\mathcal{FT}(a)} \right|,$$

where $\mathcal{FT}(a)$ is the discrete Fourier transform (DFT) of $a = (a_i)_{i=1,\ldots,p}$. The amplitude response of an autoregressive model is a tool used for interpretation of its coefficients. It illustrates how the model applied to an input signal increases amplitudes of input's spectral components. For an autoregressive model with time-varying coefficients, a natural extension of the amplitude response is a surface of amplitude response. Therefore, it depends not only on the frequency f, but also on the time instance t, as well. Thus, the surface of amplitude response is defined as

$$S(t, f) = \left| \frac{1}{\mathcal{FT}(a(t))} \right|. \tag{6}$$

Interpretation of the surface of amplitude response is similar to the classical, one-dimensional amplitude response, i.e. $S(t, f)$ describes how the model applied to an input signal increases amplitudes of input's spectral components at the time instance t.

In order to examine influence of the considered distribution to PAR model estimation, we calculate the mean square error (MSE) between estimated amplitude response and so-called "perfect surface", described in Sect. 4.2. To preserve fair results, every surface (estimated and "perfect") is normalized by its average value, i.e. arithmetic mean of the whole surface is subtracted from the surface. Higher MSE means that the procedure gives worse results in the considered case.

4.2 Simulation Results

The results are obtained using 50 simulated signals for one noise distribution each, i.e. 150 signals overall. Length of each signal is 4 s and sampling frequency is 8192 Hz. Each signal is consisted of 4 sine waves of frequencies 900, 1800, 2700 and 3600 Hz, respectively. Frequency modulation type is a saw-tooth, modulation depth—15 % and modulation frequency is 6 Hz. Figure 6 presents fragment of time series of the noiseless signal, i.e. noise consisted of a signal equal to 0 at every time t (top panel), fragment of the spectrogram of this signal containing only one period (bottom, left panel) and amplitude response surface of the fitted PAR model of order 15 (so-called "perfect surface"), see bottom right panel. The term "perfect surface" is related to the specific signal consisted of frequency modulated sine waves only, i.e. amplitude of the noise is 0 at every time t. Noiseless signal is treated as an ideal ("perfect") one.

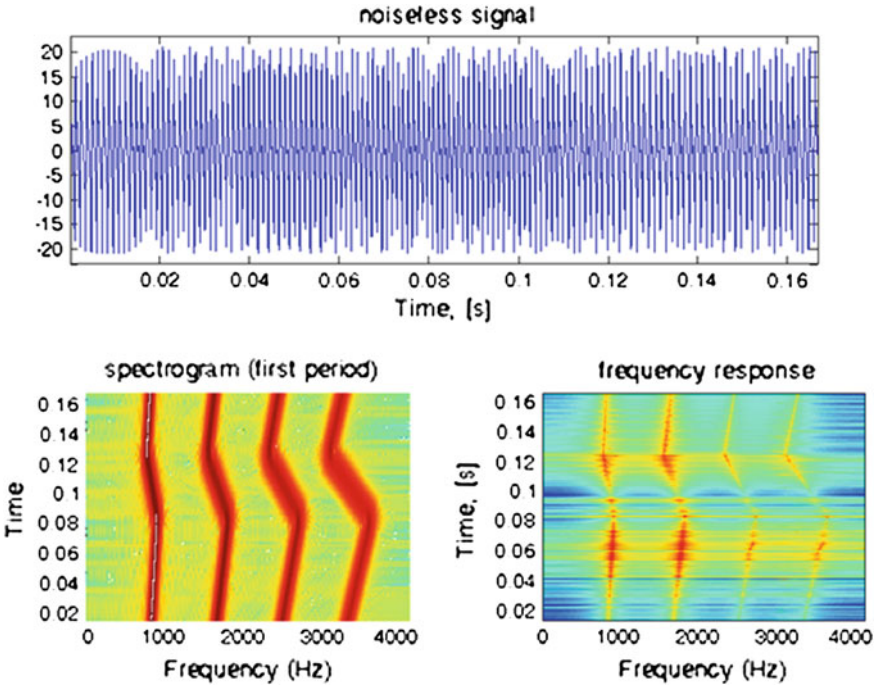


Fig. 6 Noiseless signal (time series, spectrogram, frequency response of fitted PAR filter)

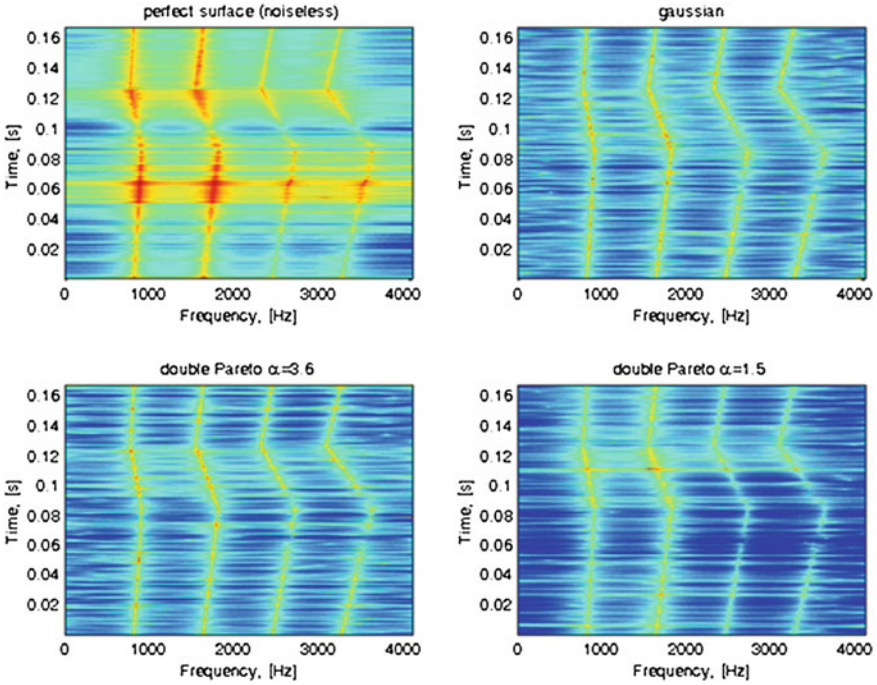


Fig. 7 Exemplary frequency responses (surfaces) for noiseless (*top left*), Gaussian (*top right*), Pareto 3.6 (*bottom left*), Pareto 1.5 (*bottom right*)

In Fig. 7 one can compare the "perfect surface" and exemplary estimated surfaces obtained for three considered distributions of noise (two Pareto with different shape parameters and Gaussian). It can be seen that the Gaussian noise seems to provide the surface which is the closest one to the "perfect" one. The property of heavy tails of the Pareto distribution might be seen in Fig. 8 (top panels). It is easy to verify that the Pareto distribution with the shape parameter 1.5 gives the highest amplitudes of noise, whereas energy of each signal is the same.

In Fig. 9 we present the boxplots of MSE for three considered distributions of noise. More precisely, in each case, we simulated 50 signals and for each of them we calculate the MSE of frequency responses. The boxplots are calculated on the basis of obtained measures. Boxplots in Fig. 9 clearly presents, that the MSE related to Gaussian noise is the lowest and MSE related to the Pareto distribution with $\alpha = 1.5$ is the highest. This result means that the heavier tail of the noise the worse results of PAR model estimation. Thus heavy tail of the noise distribution has a huge impact on the estimation procedure.

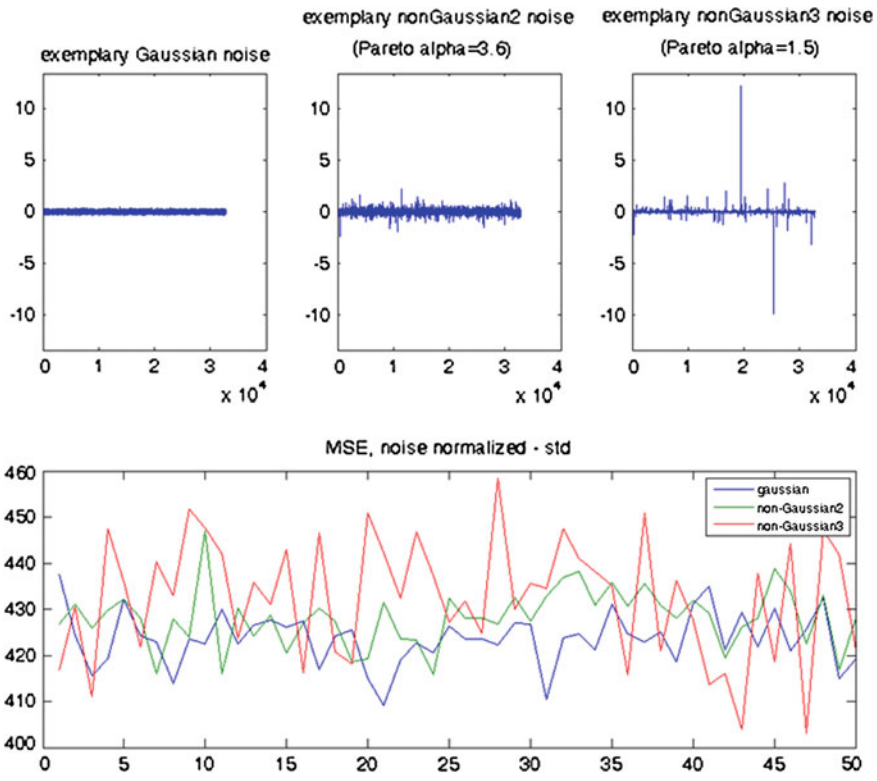


Fig. 8 Exemplary noise (3 cases), MSE of frequency responses (3 cases). Not so large differences, probably because of normalization by standard deviation (energy)

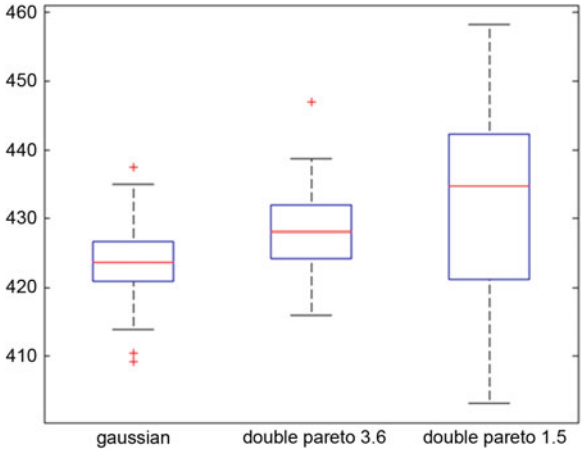


Fig. 9 Boxplots of MSE from bottom panel of Fig. 7. Even for noises normalized by energy there is a visible difference

5 Influence to PAR Estimation for Different Number of Period Repetitions

5.1 Methodology

In this section, we analyze influence of the number of period repetitions on PAR parameters estimation procedure. Such analysis is motivated by the question how long the data acquisition should be to ensure appropriate results using the PAR model. In order to examine the influence of different number of period replications, we analyze MSE between PAR parameters $\{a_i\ (t)\}_{i=1,\ldots,p,t=1,\ldots,T}$ estimated from the noiseless signals and corresponding noisy signals of different numbers of period repetitions. The minimum number of period repetitions is set to 3 and the maximum to 18. We compare boxplots and medians to clearly see how the estimation procedure is influenced by the number of period repetitions.

5.2 Simulations

In order to be consistent with Sect. 4, we analyze here the signal of the same kind, i.e. sum of 4 frequency modulated (saw tooth modulation type) sine waves with modulation frequency 6 Hz and modulation depth 15 % and noise. We consider two different kinds of noise—a white Gaussian noise and an amplitude modulated Gaussian noise, i.e. a white noise multiplied by a periodic impulsive function which reproduces impulses that might appear while the investigated machine is damaged.

Figure 10 presents boxplots of MSE mentioned in Sect. 5.1 in log scale. Each boxplot is obtained using 1000 signals of the given length. One can see, that results obtained for signals consisted of 3 whole periods are significantly worse than other. Boxplots for period repetitions from 4 to 16 are similar, thus it is circumstantiated to see median of these values separately. Signals consisted of 17 and 18 period repetitions lead to the best results, i.e. the smallest MSE. Amplitude modulation of the noise does not introduce significant gain in MSE.

Medians of MSE between PAR parameters for the noiseless signal and PAR parameters for noisy signals of different lengths are presented in Fig. 11. One can clearly see general decreasing trend of the medians while the number of period repetitions gains. This result is consistent with dispersion illustrated by boxplots in Fig. 10.

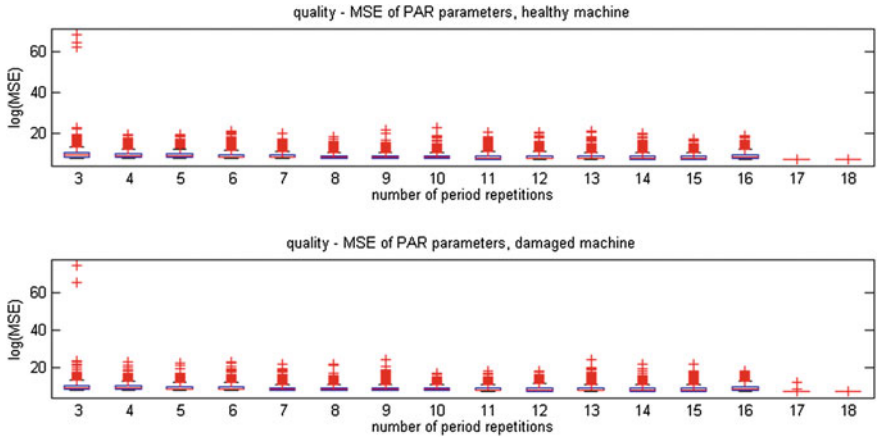


Fig. 10 Boxplots of MSE for PAR parameters in case of healthy machine (*top panel*) and damaged one (*bottom panel*)

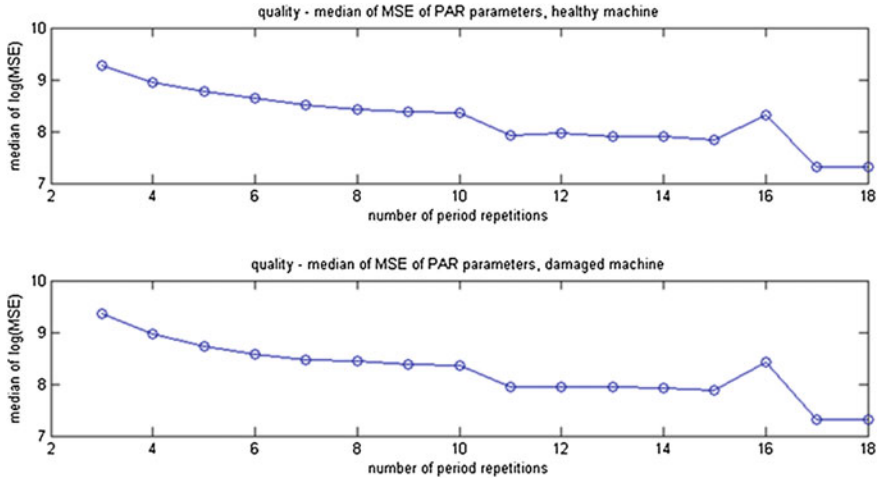


Fig. 11 Medians of MSE for PAR parameters in case of healthy machine (*top panel*) and damaged one (*bottom panel*)

6 Conclusions

In this paper, we have considered the problem of PAR model stability in case of the non-Gaussian noise and different number of period repetitions. We have introduced two methods that can be helpful in the considered problems. First of them, especially useful in case on non-Gaussian distribution of the noise, is based on the analysis of amplitude of frequency response of given signal by using fitted PAR model. The second method, applied in the problem of different number of period repetitions, is

concentrated on the analysis of fitted PAR model coefficients. The paper is motivated by analysis of the real signal from a belt conveyor driving system from an open-pit mine for which we observe non-Gaussian distribution. The results obtained in the paper allow to answer the question whether the PAR model might be applicable to industrial signals for which Gaussian distribution is not observable and the number of period replications is not very large.

References

1. Bartelmus, W., & Zimroz, R. (2009). Vibration condition monitoring of planetary gearbox under varying external load. *Mechanical Systems and Signal Processing*, *23*, 246–257.
2. Brockwell, P. J., & Davis, R. A. (2009). *Time series: Theory and methods*. New York: Springer.
3. Bloomfield, P., Hurd, H. L., & Lund, R. B. (1994). Periodic correlation in stratospheric ozone data. *Journal of Time Series Analysis*, *12*, 127–150.
4. Broszkiewicz-Suwaj, E., Makagon, A., Weron, R., & Wyłomańska, A. (2004). On detecting and modeling periodic correlation in financial data. *Physica A*, *336*, 196–205.
5. Cizek, P., Haerdle, W., & Weron, R. (2005). *Statistical tools for finance and insurance*. Berlin: Springer.
6. Chaari, F., Bartelmus, W., Zimroz, R., Fakhfakh, T., & Haddar, M. (2012). Gearbox vibration signal amplitude and frequency modulation. *Shock and Vibration*, *19*(4), 635–652.
7. Chambers, J., Cleveland, W., Kleiner, B., & Tukey, P. (1983). *Graphical methods for data analysis*. Belmont: Wadsworth.
8. Combet, F., & Zimroz, R. (2009). A new method for the estimation of instantenous speed relative fluctuation in a vibration signal based on the short time scale transform. *Mechanical Systems and Signal Processing*, *23*, 1382–1397.
9. Dybała, J., & Zimroz, R. (2014). Empirical mode decomposition of vibration signal for detection of local disturbances in planetary gearbox used in heavy machinery system. *Key Engineering Materials*, *588*, 109–116.
10. Gardner, W., & Franks, L. E. (1975). Characterisation of cyclostationary random signal processes. *IEEE Transactions of Information Theory*, *21*, 4–14.
11. Gladyshev, E. D. (1961). Periodically correlated random sequences. *Soviet Mathematics*, *2*, 385–388.
12. Hurd, H., & Miamee, A. (2007). *Periodically correlated random sequences. Spectral theory and practice*. New Jersey: Wiley.
13. Nowicka-Zagrajek, J., & Wyłomańska, A. (2006). The dependence structure for PARMA models with a-stable innovations. *Acta Physica Polonica B*, *37*(1), 3071–3081.
14. Parzen, E., & Pagano, M. (1979). An approach to modeling seasonally stationary time-series. *Journal of Econometrics*, *9*, 137–153.
15. Samuel, P. D., & Pines, D. J. (2005). A review of vibration-based techniques for helicopter transmission diagnostics. *Journal of Sound and Vibration*, *282*, 475–508.
16. Stefaniak, P., Wyłomańska, A., Obuchowski, J., & Zimroz, R. (2015). In C. Niemann-Delius (Ed.), *Procedures for decision thresholds finding in maintenance management of belt conveyor system—statistical modeling of diagnostic data*. Lecture Notes in Production Engineering (pp. 391–402).
17. Urbanek, J., Barszcz, T., & Antoni, J. (2013). Time-frequency approach to extraction of selected second-order cyclostationary vibration components for varying operational conditions. *Measurement*, *46*, 1454–1463.
18. Vecchia, A. V. (1985). Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Water Resources Bulletin*, *21*, 721–730.

19. Wyłomańska, A., Obuchowski, J., Zimroz, R., & Hurd, H. (2014). Periodic autoregressive modeling of vibration time series from planetary gearbox used in bucket wheel excavator. In F. Chaari, et al. (Eds.), *Cyclostationarity: Theory and methods*. Lecture Notes in Mechanical Engineering (pp. 171–186) Cham: Springer.
20. Wyłomańska, A. (2008). Spectral measures of PARMA sequences. *Journal of Time Series Analysis*, *29*(1), 1–13.
21. Yu, G., Li, Ch., & Zhang, J. (2013). A new statistical modeling and detection method for rolling element bearings faults based on alpha-stable distribution. *Mechanical Systems and Signal Processing*, *41*, 155–175.
22. Żak, G., Obuchowski, J., Wyłomańska, A., & Zimroz, R. (2014). Application of ARMA modelling and alpha-stable distribution for local damage detection in bearings. *Diagnostyka*, *15*(3), 3–11.

Limiting Distributions for Explosive PAR(1) Time Series with Strongly Mixing Innovation

Dominique Dehay

1 Introduction

Many man-made signals and data, even natural ones, exhibit periodicities. The non-stationary and seasonal behaviour is quite common for many random phenomena as rotating machinery in mechanics (see [8]), seasonal data in econometrics and climatology, but also signals in communication theory, biology to name a few (see e.g. [11, 15, 16, 19, 21, 22, 30] and references therein). Periodic autoregressive (PAR) models are one of the simplest linear models with a periodic structure. After more than 50 years of study, these models and their generalizations (periodic ARMA (PARMA), etc.) remain a subject of investigations of great interest as they can be applied in modeling periodic phenomena for which seasonal ARIMA models do not fit adequately (see e.g. [1, 10, 20, 26]).

It is well known that such linear models can be represented as vectorial autoregressive (VAR) models. However the general results known for VAR models do not take into account the whole periodic structure of the PARMA models [9, 32] in particular the fact that the innovation can be periodically distributed. Thus specific methods have been developed for PAR and PARMA models.

There is a large amount of publications on the estimation problem for the coefficients of PARMA models essentially whenever the model is stable, that is periodic stationary also called cyclostationary (see e.g. [1, 6, 9, 20, 27, 32–34]). The unstable case has been also studied when some autoregressive coefficients are in the boundary of the periodic stationary domain (see e.g. [2, 3, 5, 12, 13, 23] and references therein).

There are few results concerning explosive PAR model. Aknouche [4] studies the case of explosive PAR models driven by a periodically distributed independent

D. Dehay (✉)
Institut de Recherche Mathématique de Rennes, CNRS umr 6625,
Université de Rennes, Rennes, France
e-mail: dominique.dehay@uhb.fr

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods - II*,
Applied Condition Monitoring 3, DOI 10.1007/978-3-319-16330-7_6

innovation. However the independence of the innovation is too stringent in practice (see e.g. [6, 20] and references therein).

In this work, we relax the independence condition, and for simplicity of presentation, we consider periodic autoregressive of order 1 time series that is PAR(1) models. To state the convergence in distribution of the estimators (Theorems 1 and 2), we impose that the periodically distributed innovation is strongly mixing (see e.g. [14]). Thus it can be correlated and it satisfies some asymptotic independence between its past and its future (see condition (M) in Sect. 4). However, there is no condition on the rate of the asymptotic independence. This is similar to what [28] showed for the autoregressive time series with a unit root and constant coefficients.

Here we study the least squares estimators (LSE) of the PAR(1) coefficients, and the limiting distributions stated in Theorem 1 below generalize the results obtained in [25] for explosive AR models with independent Gaussian innovation (see also [7, 31]) and in [4] for explosive PAR models with independent innovation. The rate of convergence of the estimators depends on the product of the periodic coefficients of the PAR(1) model (Theorem 1). Actually this product determines whether the model is stable, unstable or explosive. Thus it is subject of great interest and we tackle the problem of its estimation. For this purpose, we consider two estimators : the product of the LSE of the PAR(1) coefficients (see [4] for independent innovation) and a least squares estimator (Theorem 2). By simulation, we detect no specific distinction between these estimators for the explosive PAR(1) models. The theoretical comparison of the limiting distributions of the two estimators is out of the scope of the paper and will be subject to another work.

The paper is organized as follows: In Sect. 2 the model under study is defined as well as the notations. Then the asymptotic behaviour as $n \to \infty$ of the scaled vector-valued time series $\left(\phi^{-n} X_{nP+r} : r = 1, \ldots, P\right)$ is stated in Sect. 3 where ϕ is the product of the periodic coefficients a_r of the explosive PAR(1) model, thus $|\phi| > 1$. The period is $P > 0$. Section 4 deals with the consistency of the least squares estimators $\widehat{a}_r$ of the PAR coefficients a_r, $r = 1, \ldots, P$, as well as the limiting distributions of the scaled errors $\phi^n(\widehat{a}_r - a_r)$, as $n \to \infty$. Next in Sect. 5 we consider the problem of estimation of the product ϕ. The asymptotic behaviour of the estimators introduced in this paper is illustrated by simulation in Sect. 6. For an easier reading and understanding of the statements of the paper, the proofs of the results are presented in Appendix.

2 Background : PAR(1) Time Series

Consider the following PAR(1) model

$$X_k = a_k X_{k-1} + u_k, \qquad k = 1, 2, \ldots, \tag{1}$$

where $\{a_k\}$ is a periodic sequence of real numbers and $\{u_k\}$ is a real-valued periodically distributed sequence of centred random variables defined on some underlying

probability space $(\Omega, \mathscr{F}, \mathrm{P})$. We assume that the periods of $\{a_k\}$ and of $\{u_k\}$ have the same value $P > 0$. Thus $a_{P+r} = a_r$ and $\mathscr{L}[u_{P+s}, \ldots, u_{P+r}] = \mathscr{L}[u_s, \ldots, u_r]$ for all integers s and r. To be short, in the sequel the sequence $\{u_k\}$ is called *innovation* of the model although it is not necessarily uncorrelated. Denote

$$A_s^{s-1} := 1, \qquad A_s^r := \prod_{j=s}^{r} a_j \quad \text{for } 1 \le s \le r, \qquad \text{and} \qquad A_s^r := 0 \quad \text{otherwise}$$

and let $\phi := A_1^P = \prod_{r=1}^{P} a_r$. Since $\{a_k\}$ is periodic with period $P > 0$, we have $A_1^{nP+r} = A_1^r \phi^n$ and we obtain the decomposition

$$X_{nP+r} = A_1^r X_{nP} + U_n^{(r)} = A_1^r \phi^n \big(X_0 + Z_{n-1}\big) + U_n^{(r)} \tag{2}$$

where

$$U_n^{(r)} := \sum_{s=1}^{r} A_{s+1}^r u_{nP+s},$$

$Z_{-1} := 0$ and for $n \ge 1$

$$Z_n := \sum_{l=0}^{n} \phi^{-l-1} \sum_{s=1}^{P} A_{s+1}^P u_{lP+s} = \sum_{l=0}^{n} \phi^{-l-1} U_l^{(P)}.$$

Note that the sequence $\big\{\big(U_n^{(1)}, \ldots, U_n^{(P)}\big) : n \in \mathbb{Z}\big\}$ is strictly stationary (stationarily distributed).

If $|\phi| < 1$, the model is stable and X_{nP+r} converges in distribution to some random variable $\zeta^{(r)}$ as $n \to \infty$. If $|\phi| = 1$, the model is unstable: its behaviour is similar to a random walk; indeed for each r, the time series $\{X_{nP+r}\}$ is a random walk when the innovation $\{u_k : k > nP + r\}$ is independent with respect to the random variable X_{nP+r}.

Henceforth, we assume that the time series $\{X_k\}$ satisfies the PAR(1) equation (1) with $|\phi| > 1$. We also assume that the initial random variable X_0 is square integrable. Moreover, the innovation $\{u_k\}$ is centred periodically distributed with second-order moments. Thus $\{u_k\}$ is periodically correlated (see [24]) and we have $\mathrm{E}[U_n^{(r)}] = 0$ as well as

$$\operatorname{cov}\Big[U_{n_1}^{(r_1)}, U_{n_2}^{(r_2)}\Big] = \sum_{s_1=1}^{r_1} \sum_{s_2=1}^{r_2} A_{s_1+1}^{r_1} A_{s_2+1}^{r_2} \operatorname{cov}\big[u_{s_1}\,,\, u_{(n_2-n_1)P+s_2}\big]$$

for all integers n, $n_1 \le n_2$, and $r, r_1, r_2 = 1, \ldots, P$. Denote $\sigma_r := \sqrt{\operatorname{var}[u_r]}$ and $K_n^{(r)} := \operatorname{cov}\Big[U_l^{(r)}, U_{l+n}^{(r)}\Big]$.

3 Explosive Asymptotic Behaviour of the Model

In the forthcoming proposition, we study the asymptotic behaviour of the time series $\{X_k\}$. Recall that we assume that $|\phi| > 1$.

Proposition 1 *For any* $r = 1, \ldots, P$

$$\lim_{n\to\infty} \phi^{-n} X_{nP+r} = A_1^r (X_0 + \zeta) \qquad \textit{a.s. and in q.m.}$$

where

$$\zeta := \lim_{n\to\infty} Z_n = \lim_{n\to\infty} \sum_{l=0}^{n} \phi^{-l-1} U_l^{(P)} \qquad \textit{a.s. and in q.m.}$$

Moreover

$$\lim_{n\to\infty} \mathscr{L}\left[X_{nP+r} - \phi^n A_1^r (X_0 + \zeta) : r = 1, \ldots, P\right] = \mathscr{L}\left[U_0^{(r)} - A_1^r \zeta : r = 1, \ldots, P\right].$$

Remark

(1) Assume that $\Pr[X_0 + \zeta \neq 0] \neq 0$, then conditionally that $X_0 + \zeta \neq 0$, the sequence $\{|X_k|\}$ converges to infinity almost surely as $k \to \infty$. Thus conditionally that $X_0 + \zeta \neq 0$, the paths of the time series $\{X_k\}$ are explosive.

(2) When $X_0 = -\zeta$ almost surely, the time series $\{X_k\}$ which follows the PAR(1) model with $|\phi| > 1$, is periodically distributed and satisfies the following stable PAR(1) equation

$$X_k = a_k^{-1} X_{k-1} + u_k^*$$

where $\{u_k^*\}$ is some periodically distributed time series. The estimation problem of the coefficients of such a PAR equation is now well known (see e.g. [6, 9, 20] and references therein).

(3) To state the convergence in quadratic mean in Proposition 1 we can easily replace the assumption that the innovation $\{u_k\}$ is periodically distributed by the less stringent one that the innovation is periodically correlated.

(4) When the innovation $\{u_k\}$ is uncorrelated and periodically distributed we obtain that

$$\mathrm{E}\left[(Z_n - \zeta)^2\right] = \frac{\phi^{-2n} K_0^{(P)}}{\phi^2 - 1} \qquad \text{and} \qquad \mathrm{var}\,[\zeta] = \frac{K_0^{(P)}}{\phi^2 - 1},$$

with $K_0^{(P)} = \sum_{s=1}^{P} (A_{s+1}^P)^2 \sigma_s$.

4 Least Squares Estimation of the Coefficients

Now we deal with the estimation problem of the coefficients $a_r, r = 1, \ldots, P$, from the observation $X_k, k = 0, \ldots, nP$, as $n \to \infty$. For that purpose, we determine the periodic sequence $\{b_k\}$ that minimizes the sum of the squared errors

$$\sum_{k=1}^{nP} \left(X_k - b_k X_{k-1}\right)^2 = \sum_{r=1}^{P} \sum_{j=0}^{n-1} \left(X_{jP+r} - b_r X_{jP+r-1}\right)^2.$$

Since $X_{jP+r} = a_r X_{jP+r-1} + u_{jP+r}$, for $j = 0, \ldots, n-1$, and $r = 1, \ldots, P$, the least squares estimator (LSE) of a_r is defined by

$$\widehat{a}_r := \frac{\sum_{j=0}^{n-1} X_{jP+r-1} X_{jP+r}}{\sum_{j=0}^{n-1} (X_{jP+r-1})^2} = a_r + \frac{\sum_{j=0}^{n-1} X_{jP+r-1}\, u_{jP+r}}{\sum_{j=0}^{n-1} (X_{jP+r-1})^2} = a_r + \frac{C_n^{(r)}}{B_n^{(r)}},$$

where from expression (2) we can write

$$B_n^{(r)} := \sum_{j=0}^{n-1} \left(X_{jP+r-1}\right)^2 = \sum_{j=0}^{n-1} \left(A_1^{r-1} \phi^j (X_0 + Z_{j-1}) + U_j^{(r-1)}\right)^2 \tag{3}$$

and

$$C_n^{(r)} := \sum_{j=0}^{n-1} X_{jP+r-1} u_{jP+r} = A_1^{r-1} \sum_{j=0}^{n-1} \phi^j (X_0 + Z_{j-1}) u_{jP+r} + \sum_{j=0}^{n-1} U_j^{(r-1)} u_{jP+r}. \tag{4}$$

Here $U_j^{(0)} := 0$.

Note that under Gaussian and independence assumptions on the periodically distributed innovation $\{u_k\}$, the LSE $\widehat{a}_r$ coincides with the maximum likelihood estimator of a_r.

To prove the convergence in distribution of the scaled errors in the following results, we use the next strong mixing condition. The notion of strong mixing, also called α-mixing, was introduced in [29] and it is largely used for modeling the asymptotic independence in time series. The condition could be weakened using the notion of weak dependence, but this is out of the scope of the paper. For more information about mixing time series and weak dependence (see e.g. [14, 17, 18] and references therein).

(M) $\lim_{n\to\infty} \alpha(n) = 0$ *where* $\alpha(n) = \sup \left|\mathrm{P}[A \cap B] - \mathrm{P}[A]\mathrm{P}[B]\right|$, *the supremum being taken over all* $k \in \mathbb{N}$ *and all sets* $A \in \mathscr{F}^k$, *and* $B \in \mathscr{F}_{k+n}$. *Here the* σ*-fields*

$\mathscr{F}^k$ *and* $\mathscr{F}_{k+n}$ *are defined by* $\mathscr{F}^k := \mathscr{F}(X_0, u_j : j \leq k)$ *and* $\mathscr{F}_{k+n} := \mathscr{F}(u_j : j \geq k+n)$.

Furthermore, in the following, we assume that the underlying probability space $(\Omega, \mathscr{F}, \mathrm{P})$ is sufficiently large so that there is a sequence of real-valued random variables $\{u_k^*\}$ which is independent with respect to X_0 and the innovation $\{u_k\}$, and such that $\mathscr{L}\left[u_0^*, \ldots, u_{nP-1}^*\right] = \mathscr{L}\left[u_{nP}, \ldots, u_1\right]$ for any integer $n > 1$. This is always possible at least by enlarging the probability space.

Now we state the strong consistency of the LSE $\widehat{a}_r$ of a_r, as well as the asymptotic limiting distribution of the scaled error $\phi^n\left(\widehat{a}_r - a_r\right)$ for the explosive PAR(1) model (1).

Theorem 1 *Conditionally that* $X_0 + \zeta \neq 0$, *the least squares estimator* $\widehat{a}_r$ *converges to* a_r *almost surely as* $n \to \infty$, *for* $r = 1, \ldots, P$. *Furthermore assume that* $\mathrm{P}[X_0 + \zeta = 0] = 0$ *and the mixing condition* (**M**) *is fulfilled, then the random vector of the scaled errors* $\left\{\phi^n\left(\widehat{a}_r - a_r\right) : r = 1, \ldots, P\right\}$ *converges in distribution to*

$$\left\{ \frac{(\phi^2 - 1)\zeta_r^*}{A_1^{r-1}(X_0 + \zeta)} : r = 1, \ldots, P \right\}$$

as $n \to \infty$. *The random variable* ζ *is defined in Proposition 1, the random vector* $\left(\zeta_1^*, \ldots, \zeta_P^*\right)$ *is independent with respect to* (X_0, ζ), *and its distribution is defined by*

$$\mathscr{L}\left[\zeta_r^* : r = 1 \ldots, P\right] = \mathscr{L}\left[\sum_{j=1}^{\infty} \phi^{-j} u_{jP-r}^* : r = 1, \ldots, P \right]$$

where the sequence $\{u_k^*\}$ *is independent with respect to* X_0 *and* $\{u_k\}$, *and such that* $\mathscr{L}\left[u_0^*, \ldots, u_{nP-1}^*\right] = \mathscr{L}\left[u_{nP}, \ldots, u_1\right]$ *for any integer* $n > 1$.

Note that in general the limiting distribution of $\phi^n(\widehat{a}_r - a_r)$ is not parameter free, that is, it depends on the parameters we are estimating. Under Gaussian assumption on the periodically distributed innovation $\{u_k\}$, the random variables ζ and ζ_r^* are Gaussian and independent, so the distribution of ratio ζ_r^*/ζ is a Cauchy distribution (see also [4]). In fact, the limiting distributions in Theorem 1 have heavy tails.

5 Estimation of ϕ

To estimate ϕ, the product of the coefficients a_r, $r = 1, \ldots, P$, we can consider the product of the estimators $\widehat{a}_r$:

$$\widetilde{\phi} = \prod_{r=1}^{P} \widehat{a}_r.$$

Then from Theorem 1, the estimator $\widetilde{\phi}$ converges almost surely to ϕ conditionally that $X_0 + \zeta \neq 0$. Thanks to the Delta method (Theorem 3.1 in [35]), we readily deduce the asymptotic law of the normalized error $\phi^n\big(\widetilde{\phi} - \phi\big)$.

$$\lim_{n\to\infty} \mathscr{L}\big[\phi^n\big(\widetilde{\phi} - \phi\big)\big] = \mathscr{L}\left[\frac{\phi^2 - 1}{X_0 + \zeta} \times \sum_{r=1}^{P} A_{r+1}^{P} \zeta_r^*\right]$$

when we assume that $\mathrm{P}[X_0 + \zeta = 0] = 0$ and the mixing condition (**M**) is fulfilled. See [4] for independent innovation.

Besides, we can define a least squares estimator of ϕ. For that purpose, note that from relation (1) and the periodicity of the coefficients, we obtain for all $j \in \mathbb{N}^*$ and $r = 1, \ldots, P$, that $X_{jP+r} = \phi X_{(j-1)P+r} + V_j^{(r)}$ where

$$V_j^{(r)} = \sum_{k=0}^{P-1} A_{r-k+1}^{r} u_{jP+r-k}.$$

Since the innovation $\{u_k\}$ is periodically distributed with the same period P, the sequence $\big\{\big(V_j^{(1)}, \ldots, V_j^{(P)}\big) : j \in \mathbb{Z}\big\}$ is stationarily distributed (strictly stationary). Then minimizing the sum of the squared errors

$$\sum_{k=P+1}^{nP} \big(X_k - bX_{k-P}\big)^2 = \sum_{j=1}^{n-1} \sum_{r=1}^{P} \big(X_{jP+r} - bX_{(j-1)P+r}\big)^2,$$

we define the least squares estimator $\widehat{\phi}$ as

$$\widehat{\phi} = \phi + \frac{C_n}{B_n}$$

where

$$C_n = \sum_{j=1}^{n-1} \sum_{r=1}^{P} X_{(j-1)P+r} V_j^{(r)} \quad \text{and} \quad B_n = \sum_{k=P+1}^{nP} (X_{k-P})^2 = \sum_{j=1}^{n-1} \sum_{r=1}^{P} \big(X_{(j-1)P+r}\big)^2.$$

Then following the same arguments as for the LSE $\widehat{a}_r$, we can state the forthcoming result.

Theorem 2 *Conditionally that $X_0 + \zeta \neq 0$, the least squares estimator $\widehat{\phi}$ converges to ϕ almost surely. Assume that $\mathrm{P}[X_0 + \zeta = 0] = 0$ and the mixing condition* (**M**) *is fulfilled, then the scaled error $\phi^n\big(\widehat{\phi} - \phi\big)$ converges in distribution*

$$\lim_{n\to\infty} \mathscr{L}\big[\phi^n\big(\widehat{\phi} - \phi\big)\big] = \mathscr{L}\left[\frac{(\phi^2 - 1)\zeta^*}{\sum_{r=1}^{P} \big(A_{r+1}^{P}\big)^{-2} (X_0 + \zeta)}\right] \tag{5}$$

conditionally that $X_0+\zeta \neq 0$, *as* $n \to \infty$. *Here* ζ^* *is a random variable independent with respect to* X_0 *and* ζ. *The distribution of* ζ^* *coincides with the distribution of*

$$\sum_{r=1}^{P} \sum_{k=0}^{P-1} A_{r+1}^{P} A_{r-k+1}^{r} \sum_{j=1}^{\infty} \phi^{-j} u^*_{jP-r+k}.$$

In the next section, we see by simulation that the distributions of $\tilde{\phi}$ and of $\widehat{\phi}$ seem to be similar when $|\phi| > 1$. The theoretical comparison of these distributions is out of the scope of the paper.

6 Simulation

Here we present the simulations of some explosive PAR(1) time series, and we illustrate the behaviour of the LSE $\widehat{a}_r$, $\widehat{\phi}$ and of $\tilde{\phi}$ for different values of the coefficients $a_r, r = 1, \ldots, P$ and for different types of innovation. For that purpose, we consider the PAR(1) model (1) with period $P = 6$. The periodic coefficients a_rs are given in Table 1.
Thus we simulate the cases $|\phi| > 1$, $|\phi|$ close to 1, $\phi = 1$ and $|\phi| < 1$. The innovation is defined by

$$u_k = \cos\left(\frac{\pi k}{3}\right) v_k \qquad \text{where} \qquad v_k = \frac{1}{\sqrt{m+1}} \sum_{i=0}^{m} \varepsilon_{k+i},$$

the random variables ε_k, $k \in \mathbb{N}$, are independent and identically distributed, and $m \in \{0\,,\ 2000\}$. When $m = 0$, we have $v_k = \varepsilon_k$, $k \in \mathbb{N}$, and the random variables u_k are periodically distributed and independent. When $m = 2000$, the random variables u_k are periodically distributed and correlated. Actually they are m-dependent, thus strongly mixing. We consider two distributions for the ε_ks : the standard normal distribution $\mathscr{N}(0, 1)$, and the uniform distribution $\mathscr{U}[-1000, 1000]$. Hence in the last case the distribution of the ε_ks is spread out.

To sum up, for each family of coefficients, we obtain four PAR(1) time series that we simulate with different lengths $T = nP = 6n$. In each case we perform 100

Table 1 PAR(1) coefficients

	a_1	a_2	a_3	a_4	a_5	a_6	ϕ
Family 1	0.8	1.2	1	1.5	1.1	0.9	1.4256
Family 2	0.8	1.1	1	1.5	1.1	0.7	1.0164
Family 3	0.5	1	1	2.5	1.6	0.5	1
Family 4	0.5	1	1.5	1.62	1.6	0.5	0.972

replications. The algorithm of the simulation is implemented in 'R' software code. Below we present some of the tables and histograms that we obtain to compare the results.

In each table, we write down the mean and the median of the values of each estimator that we have obtained from the 100 replications, as well as some box plot characteristics of the errors : extreme of upper whishers, upper hinge (3rd quarter), lower hinge (1rt quarter) and extreme of the lower whishers. We also give the 95 % percentiles of the absolute values of the errors.

First we note that the rates of convergence of the estimates decrease with $|\phi|$ (see Tables 2, 3, 4, 7 and 8). Actually, from the theoretical point of view, the rate of convergence is $|\phi|^n$ when $|\phi| > 1$ (Theorems 1 and 2 above). It is n when $|\phi| = 1$ (see [12]) and $n^{1/2}$ when $|\phi| < 1$ (see [9]). Thus we produce the histograms of scaled errors, the scale factor being ϕ^{-n} when $|\phi| > 1$, n^{-1} when $\phi = 1$ and $n^{-1/2}$ when $|\phi| < 1$.

We observe when $\phi = 1.4256$ and when $\phi = 1.0164$ in Figs. 1, 2, 3, 4 and 5, that the histograms of the scaled errors have long tails. In Tables 2, 3, 4, 5 and 6, we note that the ratios of the hinges (upper hinge, lower hinge) to the sigmas of the errors are of order of magnitude 10^{-1} or less. It is the same for the ratios of the hinges to the whiskers. These phenomena correspond to the fact that the limiting distributions have heavy tails. See [4] for independent innovation.

When $\phi = 1$, in Fig. 6 and in Table 7 we observe the distributions for some estimates $(\widehat{a}_1, \widehat{a}_3, \widehat{a}_6, \widehat{\phi})$ have also relatively long tails. But the phenomenon is very less apparent than previously. See [28] for the autoregressive model with a unit root.

When $\phi = 0.972$, in Fig. 7 the tails of the histograms are shorter than for the others values of ϕ. In Table 8 the hinges and the whishers are often with the same

Table 2 $|\phi| = 1.4256$ and uncorrelated Gaussian innovation: $\mathscr{L}[\varepsilon_k] = \mathscr{N}(0, 1)$, $m = 0$, $n = 20$

Parameter	0.8	1.2	1	1.5	1.1	0.9	1.4256	
Estimate	$\widehat{a}_1$	$\widehat{a}_2$	$\widehat{a}_3$	$\widehat{a}_4$	$\widehat{a}_5$	$\widehat{a}_6$	$\widehat{\phi}$	$\widetilde{\phi}$
Mean	0.8001	1.1999	0.9999	1.4999	1.0999	0.9000	1.4257	1.4256
Median	0.8000	1.1999	1.0000	1.4999	1.0999	0.9000	1.4256	1.4255
Error								
Mean	2e-04	-4e-05	-2e-06	-8e-05	-1e-05	7e-06	2e-04	9e-05
Sigma	2e-03	4e-04	8e-06	8e-04	9e-05	9e-05	3e-03	2e-03
Boxplot								
u. whisker	8e-04	2e-04	8e-07	4e-04	6e-05	6e-05	9e-04	8e-04
u. hinge	3e-04	5e-05	3e-07	9e-05	2e-05	2e-05	3e-04	3e-04
l. hinge	-2e-04	-7e-05	-4e-07	-2e-04	-2e-05	-2e-05	-2e-04	-2e-04
l. whisker	-8e-04	-2e-04	-9e-07	-5e-04	-6e-05	-5e-05	-7e-04	-8e-04
Percentile								
abs 0.95	4e-03	1e-03	7e-06	2e-03	3e-04	3e-04	4e-03	4e-03

Table 3 $|\phi| = 1.0164$ and uncorrelated Gaussian innovation: $\mathscr{L}[\varepsilon_k] = \mathscr{N}(0, 1), m = 0, n = 200$

Parameter	0.8	1.1	1	1.5	1.1	0.7	1.0164	
Estimate	$\widehat{a}_1$	$\widehat{a}_2$	$\widehat{a}_3$	$\widehat{a}_4$	$\widehat{a}_5$	$\widehat{a}_6$	$\widehat{\phi}$	$\widetilde{\phi}$
Mean	0.7965	1.0996	0.9999	1.4986	1.0999	0.7000	1.0131	1.0107
Median	0.7994	1.0999	1.0000	1.4999	1.1000	0.6999	1.0158	1.0154
Error								
Mean	-4e-03	-4e-04	-3e-05	-2e-03	-4e-05	4e-05	-4e-03	-6e-03
Sigma	1e-02	2e-03	2e-04	7e-03	5e-04	1e-03	9e-03	3e-02
Boxplot								
u. whisker	4e-03	4e-04	2e-05	2e-03	3e-04	9e-04	4e-03	3e-03
u. hinge	2e-04	5e-05	5e-06	4e-04	8e-05	3e-04	2e-04	-5e-05
l. hinge	-3e-03	-3e-04	-4e-06	-1e-03	-9e-05	-3e-04	-3e-03	-2e-03
l. whisker	-5e-03	-6e-04	-2e-05	-3e-03	-4e-04	-8e-04	-5e-04	-7e-03
Percentile								
abs 0.95	3e-02	3e-03	4e-04	9e-03	7e-04	2e-03	3e-02	5e-02

Table 4 $|\phi| = 1.0164$ and uncorrelated Gaussian innovation: $\mathscr{L}[\varepsilon_k] = \mathscr{N}(0, 1), m = 0, n = 400$

Parameter	0.8	1.1	1	1.5	1.1	0.7	1.0164	
Estimate	$\widehat{a}_1$	$\widehat{a}_2$	$\widehat{a}_3$	$\widehat{a}_4$	$\widehat{a}_5$	$\widehat{a}_6$	$\widehat{\phi}$	$\widetilde{\phi}$
Mean	0.7999	1.1000	1.0000	1.5000	1.1000	0.6999	1.0163	1.0163
Median	0.8000	1.0999	1.0000	1.4999	1.0999	0.7000	1.0164	1.0164
Error								
Mean	-8e-05	4e-07	4e-07	5e-06	2e-06	-8e-06	-7e-05	-1e-04
Sigma	7e-04	5e-05	3e-06	3e-04	4e-05	1e-04	7e-04	9e-04
Boxplot								
u. whisker	2e-04	2e-05	2e-08	8e-05	2e-05	3e-05	2e-04	2e-04
u. hinge	5e-05	3e-06	5e-09	2e-05	2e-06	2e-05	5e-05	5e-05
l. hinge	-4e-05	-8e-06	-6e-09	-4e-05	-4e-06	-6e-06	-4e-05	-5e-05
l. whisker	-2e-04	-2e-05	-2e-08	-7e-05	-2e-05	-3e-05	-2e-04	-2e-04
Percentile								
abs 0.95	7e-04	8e-05	2e-06	4e-04	5e-05	2e-04	7e-04	8e-04

order of magnitude or larger than the sigmas. This fits to the theoretical result, the limiting distribution being Gaussian (see [9]).

When the innovation $\{u_k\}$ is correlated with $m = 2000$ for $\phi = 1.0162$, in Table 5 we observe few change in the performances of the estimates with respect to the case when the innovation is independent (Table 4). The confidence intervals are smaller for $a_3 = 1$, and larger for $a_4 = 1.5$. However in Fig. 4 with $\phi = 1.0164$ and $m = 2000$, the tails of the histograms are farther from 0 than in Fig. 3 when $m = 0$.

Table 5 $|\phi| = 1.0164$ and correlated Gaussian innovation: $\mathscr{L}[\varepsilon_k] = \mathscr{N}(0, 1), m = 2000, n = 400$

Parameter	0.8	1.1	1	1.5	1.1	0.7	1.0164	
Estimate	$\widehat{a}_1$	$\widehat{a}_2$	$\widehat{a}_3$	$\widehat{a}_4$	$\widehat{a}_5$	$\widehat{a}_6$	$\widehat{\phi}$	$\widetilde{\phi}$
Mean	0.7999	1.1000	1.0000	1.5000	1.1000	0.6999	1.0163	1.0163
Median	0.7999	1.1000	1.0000	1.5000	1.1000	0.6999	1.0163	1.0163
Error								
Mean	-8e-05	9e-06	-3e-10	5e-05	6e-06	-2e-05	-8e-05	-8e-05
Sigma	7e-04	9e-05	2e-09	4e-04	5e-05	2e-04	7e-04	7e-04
Boxplot								
u. whisker	2e-04	3e-05	5e-12	2e-04	2e-05	5e-05	2e-04	2e-04
u. hinge	5e-05	9e-06	-2e-12	4e-05	6e-06	2e-05	5e-05	5e-05
l. hinge	-7e-05	-7e-06	-2e-11	-3e-05	-4e-06	-2e-05	-7e-05	-7e-05
l. whisker	-2e-04	-3e-05	-4e-11	-2e-04	-2e-05	-5e-05	-3e-04	-3e-04
Percentile								
abs 0.95	9e-04	2e-04	5e-10	5e-04	7e-05	2e-04	9e-04	9e-04

Table 6 $|\phi| = 1.0164$ and uncorrelated uniformly distributed innovation: $\mathscr{L}[\varepsilon_r] = \mathscr{U}[-1000, 1000], m = 0, n = 400$

Parameter	0.8	1.1	1	1.5	1.1	0.7	1.0164	
Estimate	$\widehat{a}_1$	$\widehat{a}_2$	$\widehat{a}_3$	$\widehat{a}_4$	$\widehat{a}_5$	$\widehat{a}_6$	$\widehat{\phi}$	$\widetilde{\phi}$
Mean	0.7991	1.0998	1.0000	1.4996	1.0999	0.7000	1.0156	1.0150
Median	0.7999	1.1000	1.0000	1.5000	1.1000	0.7000	1.0163	1.0163
Error								
Mean	-9e-04	-2e-04	1e-06	-4e-04	-2e-05	2e-05	-8e-04	-2e-03
Sigma	6e-03	8e-04	2e-05	3e-03	1e-04	2e-04	6e-03	1e-02
Boxplot								
u. whisker	2e-04	3e-05	4e-08	2e-04	2e-05	4e-05	2e-04	2e-04
u. hinge	4e-05	8e-06	1e-08	4e-05	5e-06	9e-06	4e-05	4e-05
l. hinge	-8e-05	-6e-06	-4e-10	-3e-05	-4e-06	-2e-05	-8e-05	-8e-05
l. whisker	-3e-04	-3e-05	-4e-05	-2e-04	-2e-05	-4e-05	-3e-04	-3e-04
Percentile								
abs 0.95	6e-04	1e-04	7e-06	4e-04	6e-05	2e-04	6e-04	7e-04

Finally, comparing the statistics of the errors of the two estimates $\widehat{\phi}$ and $\widetilde{\phi}$ in each table, we find out that they have globally the same order of magnitude whenever $|\phi| > 1$ and the histograms are quite similar. However when $|\phi| \leq 1$ it seems that $\widehat{\phi}$ gives better results than $\tilde{\phi}$. The comparison of these two estimators needs more investigation to determine whether one of them is better than the other.

Table 7 $|\phi| = 1$ and uncorrelated Gaussian innovation: $\mathscr{L}[\varepsilon_r] = \mathscr{N}(0, 1)$, $m = 0$, $n = 400$

Parameter	0.5	1	1	2.5	1.6	0.5	1	
Estimates	$\widehat{a}_1$	$\widehat{a}_2$	$\widehat{a}_3$	$\widehat{a}_4$	$\widehat{a}_5$	$\widehat{a}_6$	$\widehat{\phi}$	$\widetilde{\phi}$
Mean	0.4948	1.0000	1.0000	2.4372	1.5972	0.5004	0.9944	0.9644
Median	0.4973	1.0000	1.0000	2.4566	1.5978	0.5002	0.9972	0.9765
Errors								
Mean	-5e-03	-2e-06	1e-05	-6e-02	-2e-03	4e-03	-6e-03	4e-02
Sigma	9e-03	2e-04	4e-04	6e-02	3e-03	6e-04	9e-03	4e-02
Boxplot								
u. whisker	4e-03	8e-05	4e-04	-8e-03	-4e-04	2e-03	5e-03	2e-03
u. hinge	-4e-04	2e-05	8e-05	-3e-02	-2e-03	8e-04	-5e-04	-2e-02
l. hinge	-9e-03	-2e-05	-9e-05	-8e-02	-4e-03	4e-05	-1e-02	-5e-02
l. whisker	-2e-02	-8e-05	-4e-04	-2e-01	-7e-03	-1e-03	-3e-02	-1e-01
Percentiles								
abs 0.95	3e-03	5e-04	9e-04	3e-01	8e-03	2e-03	3e-02	2e-01

Table 8 $|\phi| = 0.972$ and uncorrelated Gaussian innovation: $\mathscr{L}[\varepsilon_r] = \mathscr{N}(0, 1)$, $m = 0$, $n = 400$

Parameter	0.5	1	1.5	1.62	1.6	0.5	0.972	
Estimate	$\widehat{a}_1$	$\widehat{a}_2$	$\widehat{a}_3$	$\widehat{a}_4$	$\widehat{a}_5$	$\widehat{a}_6$	$\widehat{\phi}$	$\widetilde{\phi}$
Mean	0.4938	1.0000	1.3983	1.5579	1.5765	0.5071	0.9662	0.8612
Median	0.4971	1.0000	1.4055	1.5624	1.5777	0.5072	0.9693	0.8745
Error								
Mean	-7e-03	5e-06	-2e-01	-7e-02	-3e-02	8e-03	-6e-03	-2e-01
Sigma	2e-02	5e-04	3e-02	2e-02	4e-03	1e-03	2e-02	6e-02
Boxplot								
u. whisker	2e-02	5e-04	-6e-02	-4e-02	-2e-02	9e-03	2e-02	-3e-02
u. hinge	3e-03	2e-04	-9e-02	-6e-02	-3e-02	8e-03	4e-03	-8e-02
l. hinge	-2e-02	-2e-04	-2e-01	-8e-02	-3e-02	7e-03	-2e-02	-2e-01
l. whisker	-4e-02	-5e-04	-2e-01	-1e-01	-4e-02	7e-03	-3e-02	-2e-01
Percentile								
abs 0.95	3e-02	2e-03	2e-01	1e-01	4e-02	9e-03	3e-02	2e-01

Conclusion

In this paper we have studied the least squares estimators of the coefficients of explosive PAR(1) time series under relatively weak dependence assumptions. It is quite interesting to see how heavy-tailed distributions enter in this context. We have also constructed two estimators of the product of these coefficients, which characterizes the explosive behaviour of the model. It would be worth to investigate the comparison of these estimators and also to consider more general PAR and PARMA models.

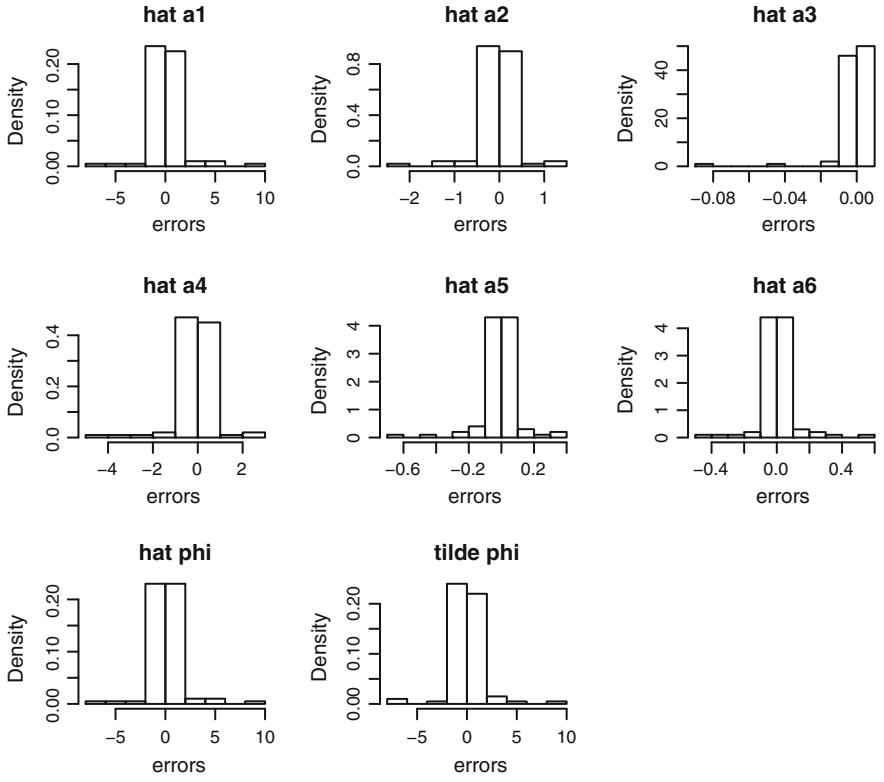


Fig. 1 Histograms of the scaled errors: $\phi = 1.4256$, $\mathscr{L}[\varepsilon_r] = \mathscr{N}(0, 1)$, $m = 0$, $n = 20$

Appendix

Proof of Proposition 1

Let $r = 1, \ldots, P$ be fixed. We know that $\phi^{-n} X_{nP+r} = A_1^r (X_0 + Z_{n-1}) + \phi^{-n} U_n^{(r)}$. Since the sequence $\{U_n^{(r)}\}$ is stationary with finite second-order moments and $|\phi| > 1$, we can readily establish that

$$\lim_{n \to \infty} \phi^{-n} U_n^{(r)} = 0 \qquad \text{a.e and in q.m..}$$

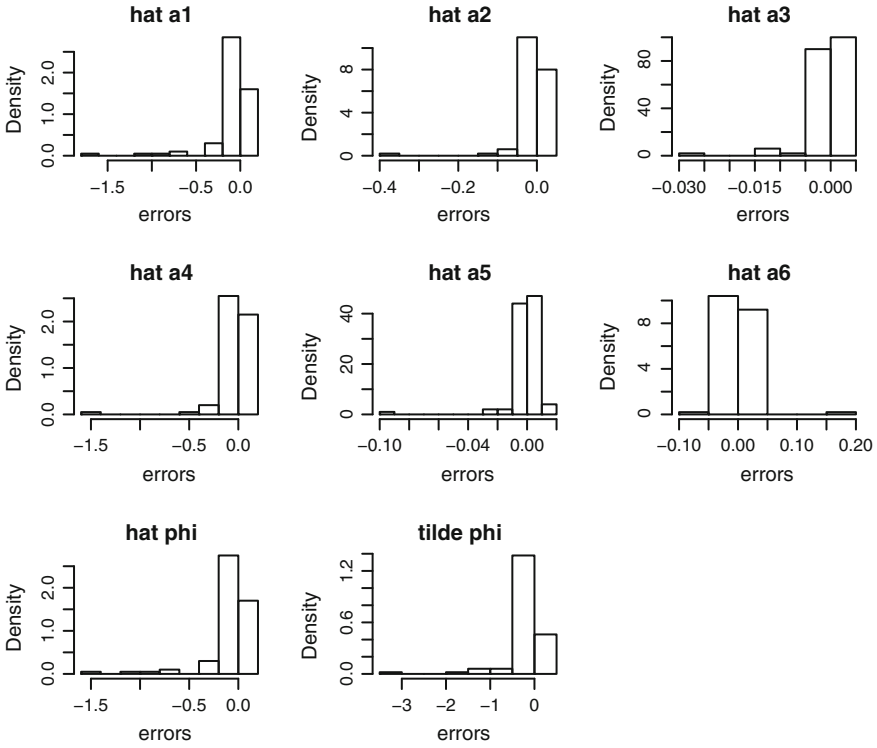


Fig. 2 Histograms of the scaled errors: $\phi = 1.0164$, $\mathscr{L}[\varepsilon_r] = \mathscr{N}(0,1)$, $m = 0$, $n = 200$

Now we show that Z_n converges almost surely and in quadratique mean. First we have

$$\begin{aligned}
&\operatorname{var}\left[\sum_{l=j}^{k} \phi^{-l-1} U_l^{(P)}\right] \\
&\quad = \sum_{l=j}^{k} \phi^{-2l-2} \operatorname{var}\left[U_l^{(P)}\right] + 2\sum_{l_1=j}^{k-1} \sum_{l_2=l_1+1}^{k} \phi^{-l_1-l_2-2} \operatorname{cov}\left[U_{l_1}^{(P)}, U_{l_2}^{(P)}\right] \\
&\quad \le \frac{\phi^{-2j} K_0^{(P)}}{\phi^2 - 1} + 2\frac{\phi^{-2j} K_0^{(P)}}{(\phi^2 - 1)(|\phi| - 1)} = \frac{\phi^{-2j} K_0^{(P)}}{(|\phi| - 1)^2}. \qquad (6)
\end{aligned}$$

Then the sequence $\{Z_n\}$ is a Cauchy sequence in the Hilbert space $L^2(\mathrm{P})$, thus this sequence converges to some random variable ζ in quadratic mean. Moreover

$$\mathrm{E}\left[(Z_n - \zeta)^2\right] \le \frac{\phi^{-2n} K_0^{(P)}}{(|\phi| - 1)^2}.$$

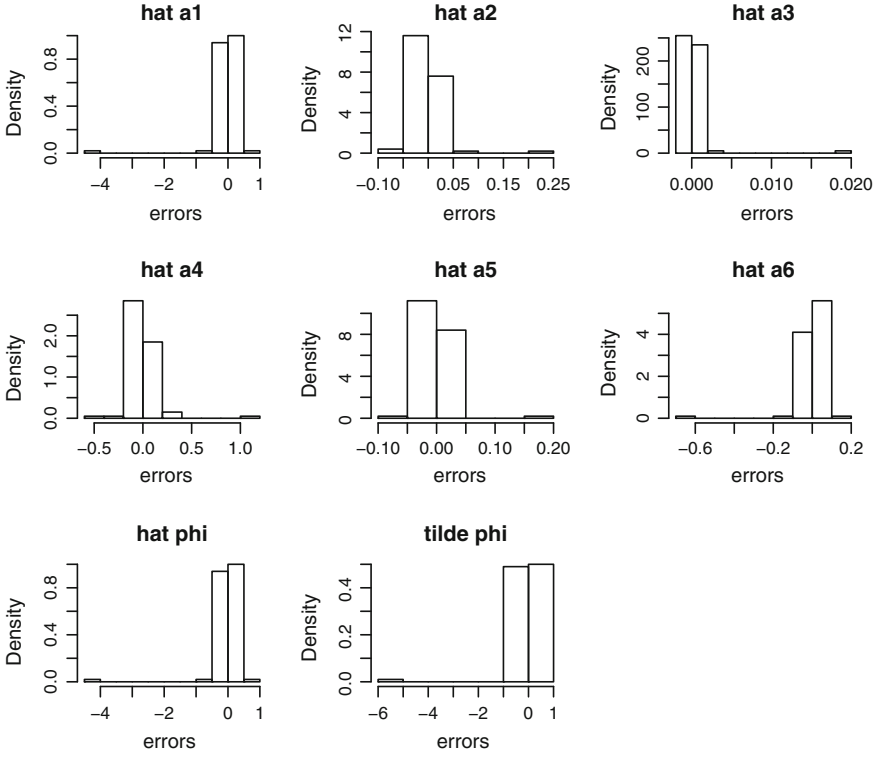


Fig. 3 Histograms of the scaled errors: $\phi = 1.0164$, $\mathscr{L}[\varepsilon_r] = \mathscr{N}(0, 1)$, $m = 0$, $n = 400$

From the exponential decreasing to 0 of the right-hand side of the last inequality, we can readily deduce the convergence almost sure following the usual method applying Borel Cantelli lemma.

As for the second part of the proposition, note that

$$X_{nP+r} - \phi^n A_1^r (X_0 + \zeta) = U_n^{(r)} + \phi^n A_1^r (Z_{n-1} - \zeta)$$

and

$$\phi^n (Z_{n-1} - \zeta) = -\sum_{l=n}^{\infty} \phi^{n-l-1} U_l^{(P)} = -\sum_{l=0}^{\infty} \phi^{-l-1} U_{n+l}^{(P)}.$$

Since the sequence of random vectors $\{(U_n^1, \ldots, U_n^{(P)}) : n \in \mathbb{Z}\}$ is stationarily distributed, we deduce that

$$\mathscr{L}\left[\left(U_n^{(r)}, \sum_{l=0}^{\infty} \phi^{-l-1} U_{n+l}^{(r)}\right) : r = 1, \ldots, P\right] = \mathscr{L}\left[\left(U_0^{(r)}, \sum_{l=0}^{\infty} \phi^{-l-1} U_l^{(r)}\right) : r = 1, \ldots, P\right].$$

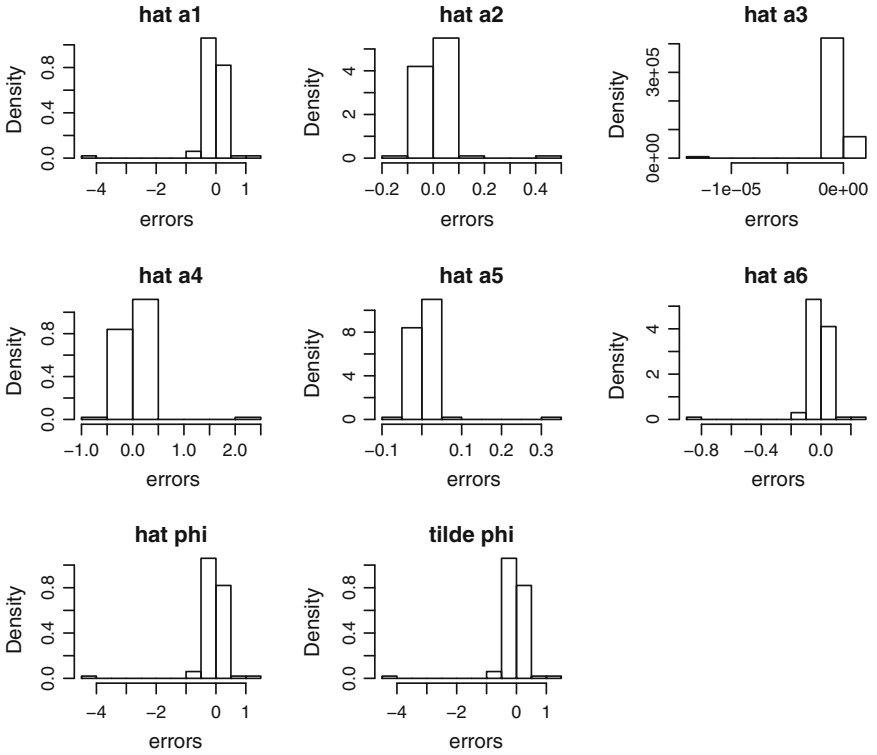


Fig. 4 Histograms of the scaled errors: $\phi = 1.0164$, $\mathscr{L}[\varepsilon_r] = \mathscr{N}(0, 1)$, $m = 2000$, $n = 400$

Then from the definition of ζ, we readily achieve the proof of Proposition 1. □

Proof of Theorem 1

First in the two following lemmas, we study the asymptotic behaviours of $B_n^{(r)}$ and $C_n^{(r)}$.

Lemma 1

$$\lim_{n\to\infty} \phi^{-2n} B_n^{(r)} = \frac{\left(A_1^{r-1}\right)^2}{\phi^2 - 1}(X_0 + \zeta)^2 \qquad \textit{a.s. and in } L^1(\mathrm{P}).$$

Proof We have seen that that $\phi^{-n} X_{nP+r-1}$ converges to $A_1^{r-1}(X_0 + \zeta)$ almost surely and in quadratic mean. Then Toeplitz lemma on series convergence gives the result. □

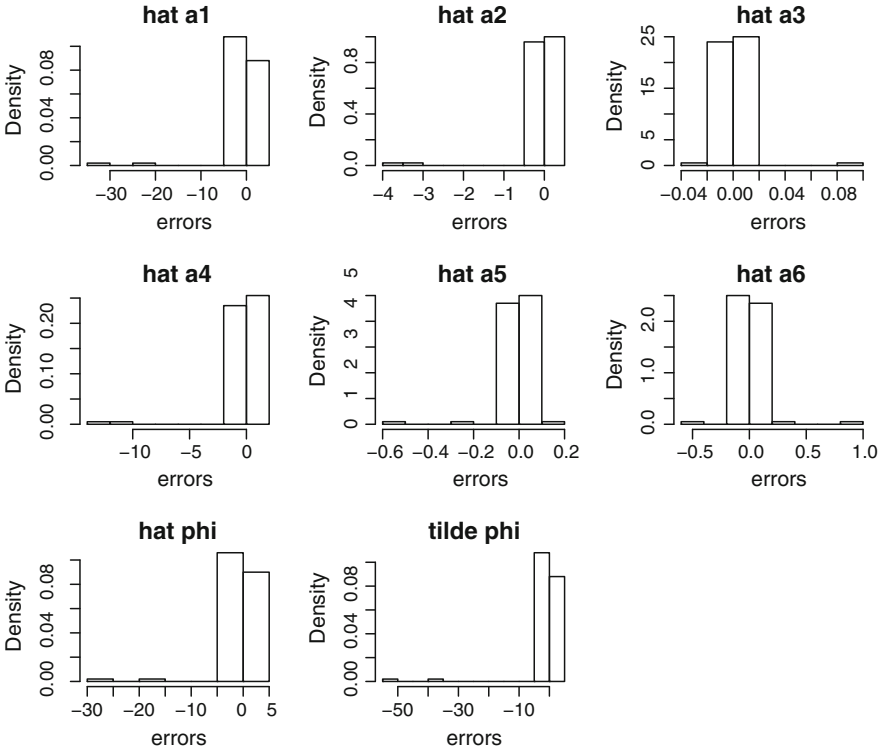


Fig. 5 Histograms of the scaled errors: $\phi = 1.0164$, $\mathscr{L}[\varepsilon_r] = \mathscr{U}[-1000, 1000]$, $m = 0$, $n = 400$

Lemma 2

$$\lim_{n\to\infty} \phi^{-2n} C_n^{(r)} = 0 \qquad \textit{a.s. and in } L^1(\mathrm{P})$$

for any $r = 1, \ldots, P$. *Moreover, under the mixing hypothesis* (**M**) *we have*

$$\lim_{n\to\infty} \mathscr{L}\left[\phi^{-n} C_n^{(r)} : r = 1, \ldots, P\right] = \mathscr{L}\left[A_1^{r-1}(X_0 + \zeta)\zeta_r^* : r = 1, \ldots, P\right].$$

Proof To prove the lemma, we study the left-hand side of equality (4).

(1) For the last term of equality (4), Cauchy Schwarz inequality entails

$$\mathrm{E}\left[\left|\sum_{j=0}^{n-1} U_j^{(r-1)} u_{jP+r}\right|\right] \le \sum_{j=0}^{n-1} \mathrm{E}\left[\left|U_j^{(r-1)}\right|^2\right]^{1/2} \mathrm{E}\left[\left|u_{jP+r}\right|^2\right]^{1/2}.$$

As the sequences $\{u_{jP+r}\}$ and $\{U_j^{(r-1)}\}$ are stationary, we have

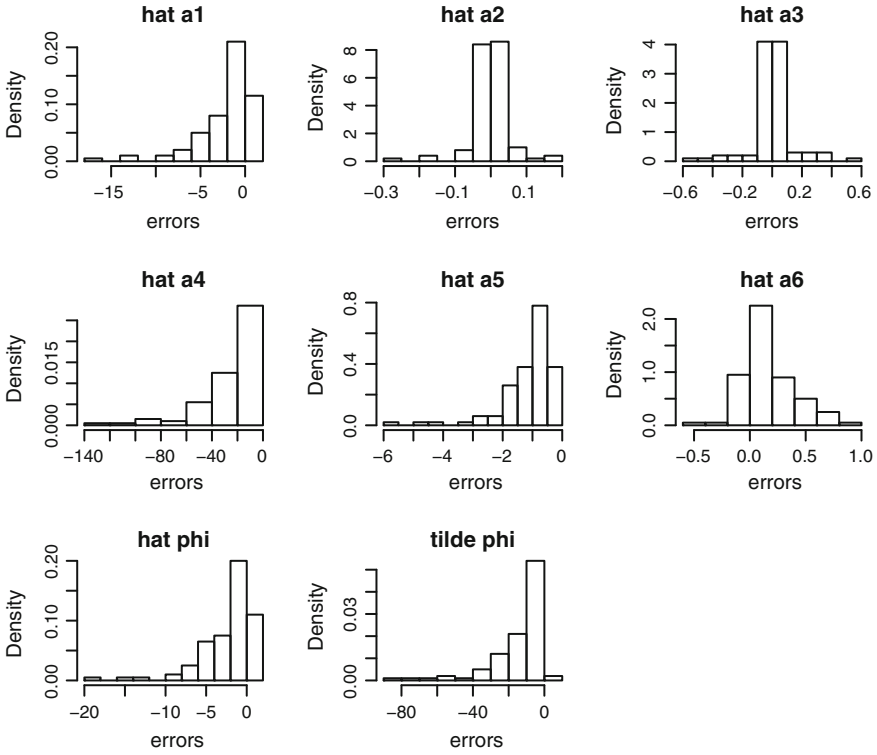


Fig. 6 Histograms of the scaled errors: $\phi = 1$, $\mathscr{L}[\varepsilon_r] = \mathscr{N}(0,1)$, $m = 0$, $n = 400$

$$\mathrm{E}\left[\left|\sum_{j=0}^{n-1} U_j^{(r-1)} u_{jP+r}\right|\right] = n\sqrt{K_0^{(r-1)}}\sigma_r.$$

Thus

$$\lim_{n\to\infty} \phi^{-n} \sum_{j=0}^{n-1} U_j^{(r-1)} u_{jP+r} = 0 \qquad \text{in } L^1(\mathrm{P}).$$

Furthermore, thanks to the exponential decreasing rate of convergence to 0 in $L^1(\mathrm{P})$, applying Borel Cantelli lemma, we easily establish the almost sure convergence

$$\lim_{n\to\infty} \phi^{-n} \sum_{j=0}^{n-1} U_j^{(r-1)} u_{jP+r} = 0 \qquad \text{a.s.}$$

(2) To study the first term of left-hand side of equality (4), the idea is first to isolate the sums of Z_{j-1} and of u_{jP+r}. Then define blocks that separate the first and the last

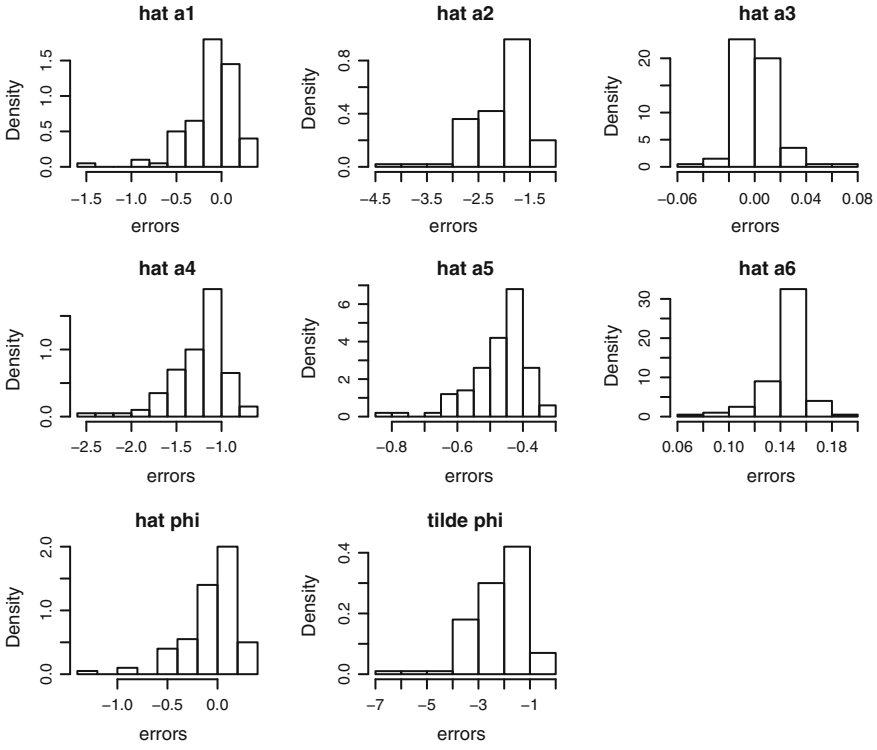


Fig. 7 Histograms of the scaled errors: $\phi = 0.972$, $\mathscr{L}[\varepsilon_r] = \mathscr{N}(0, 1)$, $m = 0$, $n = 400$

terms of the time series in order to be able to use the asymptotic independence which is given by the strong mixing condition on the innovation. Thus assume without lost of generality that n is a multiple of 4 and let $n_1 = n/4$, and $n_2 = n/2$. Then we can write

$$\sum_{j=0}^{n-1} \phi^j (X_0 + Z_{j-1}) u_{jP+r} = \sum_{j=0}^{n-1} \phi^j (Z_{j-1} - Z_{n_1}) u_{jP+r} + \sum_{j=0}^{n_2-1} \phi^j (X_0 + Z_{n_1}) u_{jP+r} + \sum_{j=n_2}^{n-1} \phi^j (X_0 + Z_{n_1}) u_{jP+r}. \quad (7)$$

(i) Thanks to inequality (6), we have

$$\mathrm{E}\left[\left|\sum_{j=0}^{n-1}\phi^j(Z_{j-1}-Z_{n_1})u_{jP+r}\right|\right] \le \sum_{j=0}^{n-1}|\phi|^j\mathrm{E}\big[|Z_{j-1}-Z_{n_1}|^2\big]^{1/2}\mathrm{E}\big[|u_{jP+r}|^2\big]^{1/2}$$
$$\le \sum_{j=0}^{n_1}|\phi|^j\mathrm{E}\big[|Z_{j-1}-Z_{n_1}|^2\big]^{1/2}\sigma_r + \sum_{j=n_1+1}^{n-1}|\phi|^j\mathrm{E}\big[|Z_{j-1}-Z_{n_1}|^2\big]^{1/2}\sigma_r$$
$$\le \frac{n_1+1+|\phi|^{n-n_1}}{|\phi|-1}\sqrt{K_0^{(P)}}\sigma_r.$$

Hence

$$\lim_{n\to\infty}\phi^{-n}\sum_{j=0}^{n-1}\phi^j(Z_{j-1}-Z_{n_1})u_{jP+r}=0 \qquad \text{in } L^1(\mathrm{P}).$$

Using Borel Cantelli lemma, the exponential decreasing rate of convergence permits to prove the almost sure convergence.
(ii) Besides, the second term of the right-hand side of equality (7) can be estimated as follows

$$\mathrm{E}\left[\left|(X_0+Z_{n_1})\sum_{j=0}^{n_2-1}\phi^j u_{jP+r}\right|\right] \le \mathrm{E}\big[|X_0+Z_{n_1}|^2\big]^{1/2}\sum_{j=0}^{n_2-1}|\phi|^j\mathrm{E}\big[|u_{jP+r}|^2\big]^{1/2}$$
$$\le \left(\mathrm{E}\big[|X_0|^2\big]^{1/2}+\sum_{j=0}^{n_1-1}|\phi|^{-l-1}\mathrm{E}\big[|U_l^{(P)}|^2\big]^{1/2}\right)\left(\sum_{j=0}^{n_2-1}|\phi|^j\mathrm{E}\big[(u_{jP+r})^2\big]^{1/2}\right)$$
$$\le \frac{\left(\mathrm{E}\big[|X_0|^2\big]^{1/2}(|\phi|-1)+\sqrt{K_0^{(P)}}\right)\sigma_r|\phi|^{n_2}}{(|\phi|-1)^2}.$$

Thus

$$\lim_{n\to\infty}\phi^{-n}(X_0+Z_{n_1})\sum_{j=0}^{n_2-1}\phi^j u_{jP+r}=0 \qquad \text{in } L^1(\mathrm{P}).$$

As in part (i), we obtain the almost sure convergence.
(iii) It remains to study the asymptotic behaviour of $(X_0+Z_{n_1})\Psi_{n_2}^{n,r}$ where

$$\Psi_{n_2}^{n,r} := \phi^{-n}\sum_{j=n_2}^{n-1}\phi^j u_{jP+r} = \sum_{j=n_2}^{n-1}\phi^{j-n}u_{jP+r} = \sum_{j=1}^{n-n_2}\phi^{-j}u_{(n-j)P+r}. \qquad (8)$$

We know that $X_0+Z_{n_1}$ converges to $X_0+\zeta$ almost surely and in quadratic mean. (Proposition 1). Since

$$\mathrm{E}\big[|\Psi_{n_2}^{n,r}|^2\big]^{1/2} \le \sum_{j=1}^{n-n_2}|\phi|^{-j}\mathrm{E}\big[|u_{(n-j)P+r}|^2\big]^{1/2} \le \frac{\sigma_r}{|\phi|-1},$$

$\phi^{-n}\Psi_{n_2}^{n,r}$ converges to 0 in quadratic mean and also almost surely. Hence $\phi^{-n}(X_0 + Z_{n_1})\Psi_{n_2}^{n,r}$ converges to 0 in $L^1(\mathrm{P})$ and almost surely.
(iv) Now we establish the convergence in distribution of $(X_0 + Z_{n_1})\Psi_{n_2}^{n,r}$. Note that $(X_0 + Z_{n_1})$ can be expressed with $X_0, u_1, u_2, \ldots, u_{(n_1+1)P}$ while $\Psi_{n_2}^{n,r}$ can be expressed with $u_{n_2 P}, \ldots, u_{nP}$. Hence, as $n_1 = n_2/2 = n/4$, the mixing property entails that

$$\left|\mathrm{P}\left[X_0 + Z_{n_1} \in A, \Psi_{n_2}^{n,r} \in B\right] - \mathrm{P}\left[X_0 + Z_{n_1} \in A\right]\mathrm{P}\left[\Psi_{n_2}^{n,r} \in B\right]\right| \leq \alpha\big((n/4 - 1)P\big)$$

for all Borel subsets A and B of $\mathbb{R}$, where $\alpha(\cdot)$ is the strong mixing coefficient. The mixing hypothesis entails that $\alpha\big((n/4 - 1)P\big)\big)$ tends to 0 as n goes to ∞, thus

$$\lim_{n\to\infty}\left(\mathrm{P}\left[X_0 + Z_{n_1} \in A, \Psi_{n_2}^{n,r} \in B\right] - \mathrm{P}\left[X_0 + Z_{n_1} \in A\right] \times \mathrm{P}\left[\Psi_{n_2}^{n,r} \in B\right]\right) = 0.$$

for all Borel subsets A and B. So $X_0 + Z_{n_1}$ and $\Psi_{n_2}^{n,r}$ are asymptotically independent. We know that $X_0 + Z_{n_1}$ converges in quadratic mean so in distribution to $X_0 + \zeta$ (Proposition 1). Now it remains to study the behaviour of $\Psi_{n_2}^{n,r}$.
(v) Since the time series $\{u_k\}$ is periodically distributed, the time series $\{u_k^*\}$ is also periodically distributed. Denoting

$$\Psi_n^{*(r)} := \sum_{j=1}^{n} \phi^{-j} u_{jP-r}^*,$$

we have $\mathscr{L}[\Psi_{n-n_2}^{*(r)}] = \mathscr{L}[\Psi_{n_2}^{n,r}]$, and the sequence $\{\Psi_n^{*(r)}\}$ converges almost surely and in quadratic mean to some random variable ζ_r^*. Then $\Psi_{n_2}^{n,r}$ converge in distribution to ζ_r^* as $n - n_2 = n/2 \to \infty$. Consequently, $(X_0 + Z_{n_1})\Psi_{n_2}^{n,r}$ converges in distribution to $(X_0 + \zeta)\zeta^{*(r)}$ where $X_0 + \zeta$ and $\zeta^{*(r)}$ are independent random variables. Furthermore from definition (8), we easily deduce the distribution of the ζ_r^*s.

Following the same lines with Cramér device, we can establish the multidimensional convergence. □

Proof of Theorem 1 The almost sure convergence is a direct consequence of Lemmas 1 and 2. To prove the convergence in distribution, we first apply Cramér device to prove the convergence in distribution of $\big(\phi^{-2n}B_n^{(1)}, \ldots, \phi^{-2n}B_n^{(P)}, \phi^{-n}C_n^{(1)}, \ldots, \phi^{-n}C_n^{(P)}\big)$. For this purpose, let $\alpha_1, \ldots, \alpha_P, \beta_1, \ldots, \beta_P \in \mathbb{R}$ and establish the convergence in distribution of

$$S_n := \sum_{r=1}^{P} \alpha_r \phi^{-2n} B_n^{(r)} + \sum_{r=1}^{P} \beta_r \phi^{-n} C_n^{(r)}.$$

Following the same method as in the proof of Lemma 2 we define blocks to separate the terms $B_n^{(r)}$ and $C_n^{(r)}$, as well as to apply the asymptotic independence given by the

strong mixing condition. Denote $n_1 = n_2/2 = n/4$. Then S_n can be decomposed as

$$S_n = \sum_{r=1}^{P} \alpha_r \left(\phi^{-2n} B_n^{(r)} - \phi^{-2n_1} B_{n_1}^{(r)}\right) + \phi^{-2n_1} \sum_{r=1}^{P} \alpha_r B_{n_1}^{(r)} + \phi^{-n} \sum_{r=1}^{P} \beta_r C_n^{(r)}.$$

Thanks to Lemma 1, $\phi^{-2n} B_n^{(r)} - \phi^{-2n_1} B_{n_1}^{(r)}$ converges to 0 almost surely. Moreover thanks to the parts 1, 2(i) and 2(ii) of the proof of Lemma 2, it remains to study

$$\sum_{r=1}^{n} \alpha_r \phi^{-2n_1} B_{n_1}^{(r)} + \sum_{r=1}^{n} \beta_r (X_0 + Z_{n_1}) \Psi_{n_2}^{n,r}.$$

Then from the strong mixing condition, $\left(B_{n_1}^{(1)}, \ldots, B_{n_1}^{(P)}, X_0 + Z_{n_1}\right)$ is asymptotically independent with respect to $\left(\Psi_{n_2}^{n,1}, \ldots, \Psi_{n_2}^{n,P}\right)$, and following the same lines as in the part 2(v) of the proof of Lemma 2 we deduce the convergence in distribution of S_n as $n \to \infty$.

Finally, the application of the continuous mapping theorem for convergence in distribution completes the proof of the theorem. □

Proof of Theorem 2

Theorem 2 is a direct consequence of the following lemma about the asymptotic behaviours of B_n and C_n.

Lemma 3

$$\lim_{n\to\infty} \phi^{-2n} B_n = \sum_{r=1}^{P} \left(A_{r+1}^{P}\right)^{-2} \frac{(X_0 + \zeta)^2}{\phi^2 - 1} \quad \textit{a.s. and in } L^1(\mathrm{P}); \tag{9}$$

$$\lim_{n\to\infty} \phi^{-2n} C_n = 0 \quad \textit{a.s. and in } L^1(\mathrm{P});$$

and

$$\lim_{n\to\infty} \mathscr{L}\left[\phi^{-n} C_n\right] = \mathscr{L}\left[(X_0 + \zeta)\zeta^*\right]. \tag{10}$$

Proof (1) First note that B_n can be expressed as follows

$$B_n = \sum_{r=2}^{P} B_{n-1}^{(r)} + B_n^{(P)} - (X_0)^2.$$

Then thanks to Lemma 1

$$\lim_{n\to\infty} \phi^{-2n} B_n = \left(1 + \phi^{-2} \sum_{r=2}^{P} \left(A_1^{r-1}\right)^2\right) \frac{(X_0+\zeta)^2}{\phi^2 - 1} \qquad \text{a.s. and in } L^1(\mathrm{P}).$$

Using the fact that $\phi^{-1} A_1^{r-1} = \left(A_r^P\right)^{-1}$, we deduce limit (9).

(2) The convergence almost sure and in $L^1(\mathrm{P})$ of $\phi^{-2n} C_n$ to 0, can be easily obtained following the lines of the proof of the convergence almost sure and in $L^1(\mathrm{P})$ of $\phi^{-2n} C_n^{(r)}$ in Lemma 2. Thus the proof is left to the reader.

(3) From its definition, C_n can be expressed by

$$C_n = \sum_{r=1}^{P} A_1^r \sum_{j=0}^{n-2} \phi^j \left(X_0 + Z_{j-1}\right) V_{j+1}^{(r)} + \sum_{r=1}^{P} \sum_{j=0}^{n-2} U_j^{(r)} V_{j+1}^{(r)}. \tag{11}$$

(i) Thanks to the stationarity of the sequences $\left\{\left(U_j^{(1)}, \ldots, U_j^{(P)}\right) : j \in \mathbb{Z}\right\}$ and $\left\{\left(V_j^{(1)}, \ldots, V_j^{(P)}\right) : j \in \mathbb{Z}\right\}$, the second term of expression (11) is of order of magnitude n in probability. Indeed

$$\mathrm{E}\left[\left|\sum_{r=1}^{P} \sum_{j=0}^{n-2} U_j^{(r)} V_{j+1}^{(r)}\right|\right] \le (n-1) \sum_{r=1}^{P} \mathrm{E}\left[\left(U_0^{(r)}\right)^2\right]^{1/2} \mathrm{E}\left[\left(V_0^{(r)}\right)^2\right]^{1/2}.$$

(ii) Following the same lines as in the proof of Lemma 2, and using the stationarity of $\left\{\left(V_j^{(1)}, \ldots, V_j^{(P)}\right) : j \in \mathbb{Z}\right\}$ we can readily prove that

$$\phi^{-n} \sum_{r=1}^{P} A_1^r \sum_{j=0}^{n-2} \phi^j \left(X_0 + Z_{j-1}\right) V_{j+1}^{(r)} = \left(X_0 + Z_{n_1}\right) \Psi_{n_2}^n + o_{\mathrm{P}}(1)$$

where $n_1 = n_2/2 = n/4$ and

$$\Psi_{n_2}^n := \sum_{r=1}^{P} A_1^r \sum_{j=n_2}^{n-2} \phi^{j-n} V_{j+1}^{(r)} = \sum_{r=1}^{P} A_1^r \sum_{j=2}^{n-n_2} \phi^{-j} V_{n-j+1}^{(r)}. \tag{12}$$

The mixing hypothesis (**M**) entails that $X_0 + Z_{n_1}$ and $\Psi_{n_2}^n$ are asymptotically independent. Besides the distribution of

$$\Psi_{n_2}^n = \sum_{r=1}^{P} \sum_{k=0}^{P-1} A_1^r A_{r-k+1}^r \sum_{j=2}^{n-n_2} \phi^{-j} u_{(n-j+1)P+r-k}.$$

coincides with the distribution of $\Psi^*_{n-n_2}$ where

$$\Psi^*_n := \sum_{r=1}^{P} \sum_{k=0}^{P-1} A_1^r A_{r-k+1}^r \sum_{j=2}^{n} \phi^{-j} u^*_{jP-r+k}$$

and the sequence $\{u_k^*\}$ is defined in part 2(v) of the proof of Lemma 2. The sequence $\{u_k^*\}$ is independent with respect to X_0 and $\{u_k\}$, thus the sequence $\{\Psi_n\}$ is also independent with respect to X_0 and $\{u_k\}$. Since the sequence $\{u_k^*\}$ is centred periodically distributed with second-order moments, the sequence $\{\Psi_n^*\}$ converges almost surely and in quadratic mean to some random variable ζ^*. Thanks to the definition (12) of $\Psi_{n_2}^n$ and the periodicity of the distribution of the innovation $\{u_k\}$, we deduce the distribution of ζ^*. Then limit (10) is proved. □

References

1. Adams, G. J., & Goodwin, G. C. (1995). Parameter estimation for periodic ARMA models. *Journal of Time Series Analysis, 16*(2), 127–147.
2. Aknouche, A. (2012). Implication of instability on econometric and financial time series modeling. In S. A. Mendez & A. M. Vega (Eds.), *Econometrics: New research*. New York: Nova Publishers.
3. Aknouche, A. (2012). Multi-stage weighted least squares estimation of ARCH processes in the stable and unstable cases. *Statistical Inference for Stochastic Processes, 15*, 241–256.
4. Aknouche, A. (2015). Explosive strong autoregression with multiplicity one. *Journal of Statistical Planning and Inference*, *161*, 50–72.
5. Aknouche, A., & Al-Eid, E. (2012). Asymptotic inference of unstable periodic ARCH processes. *Statistical Inference for Stochastic Processes, 15*, 61–79.
6. Aknouche, A., & Bibi, A. (2009). Quasi-maximum likelihood estimation of periodic GARCH and periodic ARMA-GARCH processes. *Journal of Time Series Analysis, 30*, 19–46.
7. Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics, 30*, 676–687.
8. Antoni, J. (2009). Cyclostationarity by examples. *Mechanical Systems and Signal Processing, 23*, 987–1036.
9. Basawa, I. V., & Lund, R. (2001). Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis, 22*(6), 651–663.
10. Bittanti, S., & Colaneri, P. (2009). *Periodic systems: Filtering and control*. New York: Springer.
11. Bloomfield, P., Hurd, H. L., & Lund, R. B. (1994). Periodic correlation in stratospheric ozone data. *Journal of Time Series Analysis, 15*(2), 127–150.
12. Boswijk, H. P., & Franses, P. H. (1995). Testing for periodic integration. *Economics Letters, 48*, 241–248.
13. Boswijk, H. P., & Franses, P. H. (1996). Unit roots for periodic integration. *Journal of Time Series Analysis, 17*, 221–245.
14. Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys, 2*, 107–144.
15. Chaari, F., Leśkow, J., Napolitano, A., & Sanchez-Ramirez, A. (Eds.). (2014). *Cyclostationarity: Theory and methods*. Lecture notes in mechanical engineering. Cham: Springer.
16. Collet, P., & Martinez, S. (2008). Asymptotic velocity of one dimensional diffusions with periodic drift. *Journal of Mathematical Biology, 56*, 765–792.

17. Dedecker, J., Doukhan, P., Lang, G., León, R. J. R., Louhichi, S., & Prieur, C. (2007). *Weak dependence: With examples and applications*. Lecture Notes in Statistics (Vol. 190). New York: Springer.
18. Doukhan, P. (1994). *Mixing: Properties and examples*. Lecture Notes in Statistics (Vol. 85). New York: Springer.
19. Dragan, Ya., & Javors'kyj, I. (1982). *Rhythmics of sea waving and underwater acoustic signals*. Kiev (Kijev) (in Russian): Naukova dumka.
20. Francq, C., Roy, R., & Saidi, A. (2011). Asymptotic properties of weighted least squares estimation in weak PARMA models. *Journal of Time Series Analysis, 32*, 699–723.
21. Franses, P., & Paap, R. (2004). *Periodic time series*. Oxford: Oxford University Press.
22. Gardner, W. A., Napolitano, A., & Paura, L. (2006). Cyclostationarity: Half a century of research. *Signal Processing, 86*, 639–697.
23. Ghysels, E., Hall, A., & Lee, H. S. (1996). On periodic structures and testing for seasonal unit root. *Journal of the American Statistical Association, 91*, 1551–1559.
24. Hurd, H. L., Makagon, A., & Miamee, A. G. (2002). On AR(1) models with periodic and almost periodic coefficients. *Stochastic Processes and Their Applications, 100*, 167–185.
25. Monsour, J. M., & Mikulski, P. W. (1998). On limiting distributions in explosive autoregressive processess. *Statistics & Probability Letters, 37*, 141–147.
26. Osborn, D. R., Chui, A. P. L., Smith, J. P., & Birchenhall, C. R. (1988). Seasonality and the order of integration for consumption. *Oxford Bulletin of Economic Statistics, 50*, 361–377.
27. Pagano, M. (1978). On periodic and multiple autoregression. *Annals of Statistics, 6*, 1310–1317.
28. Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica, 55*(2), 277–301.
29. Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences USA, 27*, 832–837.
30. Serpedin, E., Pandura, F., Sari, I., & Giannakis, G. B. (2005). Bibliography on cyclostationarity. *Signal Processing, 85*, 2233–2303.
31. Stigum, B. P. (1974). Asymptotic properties of dynamic stochastic parameter estimates (III). *Journal of Multivariate Analysis, 4*, 351–381.
32. Tiao, G. C., & Grupe, M. R. (1980). Hidden periodic autoregressive-moving average models in time series data. *Biometrika, 67*, 365–373.
33. Troutman, B. M. (1979). Some results in periodic autoregressions. *Biometrika, 66*, 219–228.
34. Vecchia, A. (1985). Maximum likelihood estimation for periodic autoregressive moving average models. *Technometrics, 27*, 375–384.
35. van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge: Cambridge University Press.

PARMA Models with Applications in R

Anna E. Dudek, Harry Hurd and Wioletta Wójtowicz

1 Introduction

Any second-order random process that is generated by the mixing (in the workings of a system) of randomness and periodicity will likely have the structure of periodic correlation. Precisely, a random sequence $\{X_t, t \in \mathscr{Z}\}$ with finite second moments is called periodically correlated with period T (PC-T) if it has periodic mean and covariance functions, e.g.

$$E(X_t) = E(X_{t+T}) \quad \text{and} \quad Cov(X_t, X_s) = Cov(X_{t+T}, X_{s+T}) \tag{1}$$

for each $t, s \in \mathscr{Z}$. To avoid ambiguity, the period T is taken as the smallest positive integer such that (1) holds.

The studies of PC sequences, which were initiated by [15], result in the appreciable theory and some practical approaches as well. Many real data have periodic structure, so they can be described by periodically correlated sequences. The cyclic nature of environmental and social phenomena impart a seasonal mean and correlation structure into many climatological [21, 26] as well as economical [11, 19, 20] time series. Other examples of PC data could be found in e.g. [9, 16, 22]. Gardner investigated perceptively the nature of cyclostationarity [12, 13]. Hurd and Miamee [16] provide substantial study of PC issues with many motivating and illustrative examples.

A.E. Dudek (✉)
AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Krakow, Poland
e-mail: aedudek@agh.edu.pl

H. Hurd
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill 27599-3260, USA

W. Wójtowicz
Institute of Applied Computer Science, Cracow University of Technology, Al. Jana Pawla II 37, 31-864 Cracow, Poland

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods - II*,
Applied Condition Monitoring 3, DOI 10.1007/978-3-319-16330-7_7

In this paper we focus on identification and estimation of PC structure and on PARMA (periodic autoregressive–moving-average) modelling introduced in [27, 28]. PARMA sequences are formed by introducing time-varying coefficients into the ARMA set-up; under a mild restriction on the coefficients, the resulting PARMA sequence is periodically correlated (PC) [28]. The periodically correlated processes and PARMA modelling are studied in [4, 5, 9] and lately also in [1]. The methods proposed by those authors, although based on statistical principles, are not well known or available to a wide audience. Currently, difficulties connected with analysis of periodic sequences arise from deficient knowledge about mathematical achievement in this field. The intent of this work is to provide an accessible package for periodic time series analysis in R. In this paper, we present an R package called `perARMA` that is available from the Comprehensive R Archive Network (CRAN) (see online supplementary material and [10]). The package implements non-parametric analysis, characterization and identification, PARMA model fitting and prediction. Missing observations are allowed in some characterization procedures. The implemented procedures are loosely based on the Hurd's Matlab functions available from his Web page and introduced in [16]. As a result, the applied researcher obtains quite easy and very intuitive tool that can be easily used in many applications. To our knowledge there is no R package (under CRAN) for PARMA time series analysis although the `pear` [3] and `partsm` [18] packages provide for PAR analysis.

The paper is organized in the following way. In Sect. 2 we provide some theoretical background of PARMA time series. Section 3 presents analysis of real dataset from energy market. In Sect. 4 the estimation of full PARMA model for simulated dataset is performed. Finally, Sect. 5 concludes our study.

2 PARMA Time Series Analysis

PARMA modelling arises from the introduction of periodic correlation of PC sequences into a stationary ARMA model when the coefficients of the model are allowed to vary periodically in time. Precisely, the random sequence X_t is called PARMA(p,q) with period T if it satisfies

$$X_t = \sum_{j=1}^{p} \phi_j(t) X_{t-j} + \sum_{k=1}^{q} \theta_k(t) \xi_{t-k} + \sigma(t) \xi_t, \tag{2}$$

where $\phi_j(t) = \phi_j(t+T), \theta_k(t) = \theta_k(t+T), \sigma(t) = \sigma(t+T)$ for all $j = 1, \ldots, p$, $k = 1, \ldots, q$ are periodic coefficients, and ξ_t is mean zero white noise with variance equal to one.

The PARMA systems are widely applied in modelling climatology [4, 5], meteorology [24], hydrology [2, 21, 26, 28] and economics data [7, 11].

PARMA time series analysis is performed in three main processing steps: (1) identification, (2) parameter estimation and (3) diagnostic checking. The same general

process is used in the stationary case, where the tools are simpler. In the following sections, we describe the tools that are essential to achieve each of these three steps. But first we focus on the ARMA(p,q) fitting process to remind the reader of the general idea.

Identification in the stationary case refers to the determination of the model order parameters p and q, which provide an adequate fit to the data. Initial guesses of p and q are usually suggested from the identification tools, which are the sample covariance, sample autocorrelation (ACF) and sample partial autocorrelation (PACF). Parameter estimation refers to the process of estimating the values of the parameters in the chosen representation. For AR models we can use the Yule–Walker equations. For general ARMA we use maximum likelihood. Diagnostic checking in the stationary case consists of determining if the residuals (based on some parameter estimates) are consistent with white noise. If not, then modifications to p and q are made based essentially on the application of the identification step to the residuals (determine what structure is not yet explained) and estimation is rerun.

Below we present how those ideas can be transformed to the periodic nonstationary case. In the subsequent we assume that $\{X_1, X_2, \ldots, X_N\}$ is a mean zero PC time series with period T. Moreover, we assume without loss of generality that the data record contains d full cycles, e.g. N is an integer multiple of period T ($N = dT$).

2.1 Identification of PC-T Structure

There are two processes that we include under the heading of identification: (1) the determination of period T when it is unknown and some basic characterizing quantities such as sample periodic mean $\widehat{m}_{t,d}$, sample periodic deviation $\widehat{\sigma}_d\,(t)$ and sample periodic covariance $\widehat{R}_d\,(t+\tau, t)$; (2) the determination of p and q, the orders of the PAR and PMA parts of a PARMA model.

Preliminary Identification

An important preliminary step in the identification process is the determination of period T when it is not known. In this case, the *periodogram* and *squared coherence statistic* can be used. The usual periodogram can detect additive frequency components in the time series and this includes frequencies belonging to the additive periodic mean. So if a periodic mean is present, the periodogram can illuminate its frequencies and help in the determination of T. The value of T may also be inferred from spacing of the support lines in the (harmonizable) spectral measure of a PC process. These support lines are seen empirically in the images produced by the magnitude-squared coherence statistic. Using both simultaneously gives a complete picture.

To clarify the idea of the magnitude-squared coherence, we need to use some features of the spectral measure of a PC sequence. The spectral measure of periodically correlated sequence is determined on the two-dimensional set $[0, 2\pi) \times [0, 2\pi)$, so we always deal with the pairs of frequencies $(\lambda_p, \lambda_q) \in [0, 2\pi)^2$, and the support

of the measure is contained in the subset of parallel lines $\lambda_q = \lambda_p + 2j\pi/T$ for $j = -(T-1), \ldots, -1, 0, 1, \ldots, (T-1)$. For more details see [14, 16].

The concept of determining the period length using the squared coherence statistic directly corresponds to its features. We begin with Discrete Fourier Transform (DFT), $\tilde{X}_N(\lambda_j)$, of the series $\{X_1, X_2, \ldots, X_N\}$, defined for the Fourier frequencies $\lambda_j = 2\pi j/N$, j = 0, 1, ..., N−1. The squared coherence statistic,

$$\left|\hat{\gamma}(\lambda_p, \lambda_q, M)\right|^2 = \frac{\left|\sum_{m=1}^{M-1} \tilde{X}_N(\lambda_{p-M/2+m})\overline{\tilde{X}_N(\lambda_{q-M/2+m})}\right|^2}{\sum_{m=1}^{M-1}\left|\tilde{X}_N(\lambda_{p-M/2+m})\right|^2 \sum_{m=1}^{M-1}\left|\tilde{X}_N(\lambda_{q-M/2+m})\right|^2} \tag{3}$$

is computed for a specified collection of pairs $(\lambda_p, \lambda_q) \in [0, 2\pi)^2$; note it is the sample correlation between two M-dimensional complex-valued vectors and satisfies $0 \le \left|\hat{\gamma}(\lambda_p, \lambda_q, M)\right|^2 \le 1$. Having computed (3) for (λ_p, λ_q) in some subset of $[0, 2\pi)^2$, one may determine frequency pairs indexed by (p, q) for which the sample correlation is significant in the sense that threshold determined by the distribution of $\left|\hat{\gamma}(\lambda_p, \lambda_q, M)\right|^2$, under the null hypothesis of stationary white noise, is exceeded [16, p. 310]. Plotting those points on the square $[0, 2\pi)^2$, we can say something about the nature of the analyzed time series, according to some general rules:

- if in the square only the main diagonal appears, then X_t is a stationary time series;
- if there are some significant values of statistic and they seem to lie along the parallel equally spaced diagonal lines, then X_t is likely PC-T, where T is the "fundamental" line spacing. Algebraically, T would be the gcd of the line spacings from the diagonal; for a sequence to be PC-T, not all lines are required to be present;
- if there are some significant values of statistic but they occur in some non-regular places, then X_t is a nonstationary time series in other than periodic sense; but note there are many hypotheses being tested, so some threshold exceedances are to be expected.

We need to comment also the choice of the parameter M that controls the length of the smoothing window appearing in (3). Too small or too large values of M can affect the results [16, p. 311]. Hurd and Miamee suggest to observe the results for a collection of smoothing parameters; a suggested beginning choice is $M = 8, 16, 32$. Under the null case the sample-squared coherence statistics has probability density

$$p\left(|\gamma|^2\right) = (M-1)\left(1 - |\gamma|^2\right)^{M-2}, \quad 0 \le |\gamma|^2 \le 1, \tag{4}$$

because, for X_t a Gaussian white noise, $\widetilde{X}_N(\lambda_j)$ are complex Gaussian with uncorrelated real and imaginary parts for each j. As a result, to determine the squared coherence α-threshold we use

$$|\gamma|_\alpha^2 = 1 - \exp\left(\log(\alpha)/(M-1)\right). \tag{5}$$

For more details we refer the reader to [14, 16].

When the period length is known, the estimation of basic sequence statistics: periodic mean, periodic deviation and periodic covariance is possible. Detailed description of all estimators presented below can be found in [16].

The sample periodic mean function is given by

$$\widehat{m}_{t,d} = \frac{1}{d}\sum_{k=0}^{d-1} X_{t+kT}. \tag{6}$$

The sample periodic covariance function is defined for any lag τ as

$$\widehat{R}_d\,(t+\tau,t) = \frac{1}{d}\sum_{k=0}^{d-1}\left(X_{t+kT+\tau} - \widehat{m}_{t+\tau,d}\right)\left(X_{t+kT} - \widehat{m}_{t,d}\right) \tag{7}$$

for $t = 0, 1, \ldots, T-1$. The number of τ for which (7) is evaluated can be determined (or estimated) from the usual (stationary) ACF. Then it is quite straightforward to get the periodic sample deviation by putting

$$\widehat{\sigma}_d\,(t) = \widehat{R}_d\,(t,t)\,. \tag{8}$$

The Fourier representation of the covariance function [15] for PC-T processes is based on the periodicity $R_d(t+\tau,t) = R_d(t+\tau+T, t+T)$ and the Fourier series representation

$$R_d(t+\tau,t) = \sum_{k=0}^{T-1} B_k(\tau)\exp(i2kt\pi/T), \tag{9}$$

where $B_k(\tau)$ ($\tau \in \mathscr{Z}$ and $k = 0, 1, \ldots, T-1$) are the Fourier coefficients given by

$$B_k(\tau) = \frac{1}{T}\sum_{t=0}^{T-1}\exp(-i2kt\pi/T)R_d(t+\tau,t). \tag{10}$$

The problem of computing the Fourier representation is thus reduced to the problem of determining the coefficients $B_k(\tau)$. Moreover, $B_0(\tau)$ is always non-negative definite and hence is the covariance of stationary sequence. The Fourier coefficients of the sample covariance, for $k = 0, 1, \ldots, T-1$ and $\tau \in$ set of lags, can be estimated

$$\widehat{B}_{k,dT}(\tau) = \frac{1}{dT}\sum_{t\in I_{dT,\tau}}\left(X_{t+\tau} - \widehat{m}_{t+\tau,d}\right)\left(X_t - \widehat{m}_{t,d}\right)e^{-i2\pi kt/T}, \tag{11}$$

where

$$I_{dT,\tau} = \begin{cases} [0, dT-\tau-1] & \text{for } \tau \geq 0 \\ [-\tau, dT-1] & \text{for } \tau < 0. \end{cases}$$

Hurd and Miamee proposed a test with the null hypothesis $B_k(\tau) = 0$ for fixed τ and k, which is based on the properties of the sample Fourier transform of $Y_{t,\tau} = \left(X_{t+\tau} - \widehat{m}_{t+\tau,d}\right)\left(X_t - \widehat{m}_{t,d}\right)$. The p values for the test of $|B_k(\tau)|^2 = 0$ are based on the ratio of magnitude squares of amplitudes of a high resolution Fourier decomposition. Magnitudes for the frequency corresponding to index k are compared to the magnitudes of neighbouring frequencies (via the F distribution) [16, pp. 272–282, 288–292].

Determination of p and q

Before fitting a periodic model (PARMA(p,q)) to data, the orders of maximum lag p and/or q are required. Similarly to the stationary ARMA, we need to look at the sample autocovariance (ACF) and the sample partial autocovariance (PACF). Of course, their periodic versions are essential in that case. Below we present formulas for both of them.

Sample periodic ACF for $t = 0, 1, \ldots, T-1$ is defined as

$$\widehat{\rho}_d\,(t+\tau, t) = \frac{\widehat{R}_d\,(t+\tau, t)}{\widehat{\sigma}_d\,(t+\tau)\,\widehat{\sigma}_d\,(t)}, \tag{12}$$

where τ is a lag. To calculate confidence bands for $\widehat{\rho}_d\,(t+\tau, t)$ the Fisher transformation

$$Z = \frac{1}{2}\log\frac{1+\widehat{\rho}}{1-\widehat{\rho}} \tag{13}$$

is used, producing Z that are approximately normal $N\left(\mu_Z, \sigma_Z^2\right)$, where $\mu_Z \approx \zeta + (1/2d)\rho$, ζ is the Fisher transformation of ρ and $\sigma_Z^2 = 1/(d-3)$. The confidence limits for ρ can be calculated using those for Z with the assumption that the term $(1/2d)\rho$ can be ignored.

Two useful tests may be additionally constructed. First, a test for null hypothesis $\rho(t+\tau, t) \equiv \rho(\tau)$, where $\rho(\tau)$ is some unknown constant, (no time variation at lag τ) is based on statistic

$$S_\rho^2\,(\tau) = \frac{\left(Z_t - \widehat{Z}_t\right)^2}{1/\,(d-3)}, \tag{14}$$

which under the null hypothesis is approximately $\chi^2(T-1)$. The second, a test for null hypothesis $\rho(t+\tau, t) \equiv 0$ for some specific τ, (no correlation at this lag τ) is equivalent to the test for $\mu_Z = 0$. For more details about those tests we refer the reader to [8, p. 399] and [16, p. 285].

The sample periodic PACF, for $t = 0, 1, \ldots, T-1$ and $n \in$ set of orders is defined as

$$\hat{\pi}_{n+1}(t) = \frac{\hat{R}_d(t+1, t-n) - \hat{\mathbf{r}}'_{t-n,t:t-n+1}\hat{\boldsymbol{\alpha}}_n^{(t+1)}}{\hat{\sigma}_n(t+1)\hat{\sigma}_n(t-n)}, \tag{15}$$

where

$$\hat{\sigma}_n^2(t+1) = \hat{R}_d(t+1, t+1) - \hat{\mathbf{r}}'_{t+1,t:t-n+1}\hat{\boldsymbol{\alpha}}_n^{(t+1)},$$

$$\hat{\sigma}_n^2(t-n) = \hat{R}_d(t-n, t-n) - \hat{\mathbf{r}}'_{t-n,t:t-n+1}\hat{\boldsymbol{\beta}}_n^{(t-n)}$$

and

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_n^{(t+1)} &= \hat{\mathbf{r}}^{-1}(t,n)\hat{\mathbf{r}}_{t+1,t:t-n+1},\\ \hat{\boldsymbol{\beta}}_n^{(t-n)} &= \hat{\mathbf{r}}^{-1}(t,n)\hat{\mathbf{r}}_{t-n,t:t-n+1}.\end{aligned}$$

The confidence bands are based on the asymptotic normality assumption [6, Chap. 3].

2.2 Estimation of PARMA Parameters

Parameter estimation refers to the process of estimating the values of the parameters in the chosen representation. The full PARMA model has T parameters for each autoregressive order, for each moving average order and for $\sigma(t) = \theta_0(t), t = 0, 1, \ldots, T-1$, and thus a total of $T(p+q+1)$ parameters, so the cost of computing all of them could be quite high. Therefore, it is recommended first trying to fit simple models, which explain the data with the fewest possible parameters.

Two useful methods for estimation of periodic models coefficients are implemented in `perARMA` package: the Yule–Walker moment estimators for the general PAR model and approximate maximum likelihood estimation for the general PARMA. It is easy to show that for PAR models both techniques are equivalent, e.g. for PAR(1) model provide following estimators of coefficients:

$$\hat{\phi}(t) = \frac{\hat{\gamma}_t(1)}{\hat{\gamma}_{t-1}(0)}, \qquad \hat{\sigma}^2(t) = \hat{\gamma}_t(0) - \hat{\phi}(t)\hat{\gamma}_t(1), \qquad t = 0, 1, \ldots, T-1, \quad (16)$$

where $\hat{\gamma}_t(j) = \frac{1}{d}\sum_{n=0}^{d-1} X_{nT+t-j}X_{nT+t}$ and $d = N/T$ is number of cycles.

The Yule–Walker estimation is a simple method, but whenever PMA terms are present, it will be inadequate. In these cases maximum likelihood estimation is used. But whenever there are a large number of parameters, the optimization has difficulty due to presence of local maxima and due to slowness caused by the approximation of derivatives.

Rather than searching over the entire Φ, Θ parameter space, the dimension of the search space can often be substantially reduced by transforming Φ, Θ to the Fourier coefficients $\mathbf{A}, \mathbf{B}$ appearing in the DFT representation introduced in [17]:

$$\phi_j(t) = a_{j,1} + \sum_{n=1}^{\lfloor T/2\rfloor} a_{j,2n}\cos(2\pi nt/T) + \sum_{n=1}^{\lfloor T/2\rfloor} a_{j,2n+1}\sin(2\pi nt/T), \quad (17)$$

$$\theta_k(t) = b_{k,1} + \sum_{n=1}^{\lfloor T/2\rfloor} b_{k,2n}\cos(2\pi nt/T) + \sum_{n=1}^{\lfloor T/2\rfloor} b_{k,2n+1}\sin(2\pi nt/T) \quad (18)$$

for $j = 1, \ldots, p$, $k = 0, \ldots, q$ and $t = 0, 1, \ldots, T-1$. Reduction of the parameter search to subspaces of $\mathbf{A}, \mathbf{B}$ can be easily accomplished by constraining some (or even most) frequencies to have zero amplitude. Often this can be justified by physical considerations that correspond to smoothness in the time dependence. Since the mapping (DFT) from Φ, Θ to $\mathbf{A}, \mathbf{B}$ is linear and invertible, a linear subspace of $\mathbf{A}, \mathbf{B}$ transforms back to a linear subspace of Φ, Θ, although our methods make no explicit use of this fact. In our experience, reduction of the search space in this manner not only reduces computation time, but tends to produce unique solutions, whereas the likelihood often has many local extremes in the entire space Φ, Θ.

This parameterization was first used in [28] in the context of maximum likelihood parameter estimation and in [23] where are developed expressions for the harmonizable spectral density in terms of the parameters $\mathbf{A}, \mathbf{B}$.

Computation speed of PARMA maximum likelihood estimation can be further improved by the use of Ansley's transformation, first applied to PARMA(p,q) sequences in [27, 28]. This method focuses on the conditional version which ignores the first $m = \max(p; q)$ samples in order to avoid the cumbersome calculation of the full covariance.

2.3 Goodness of Fit and Model Selection

To confirm that fitted model is appropriate the residuals should be examined to ensure independence (whiteness) and normality, but this does not differ from stationary case. However, in the periodic case we do permit the residuals to be periodic white noise, this can be easily converted to white noise by scaling out the empirical periodic variance.

Sometimes there are several sets of model parameters that give reasonable fits. Then to choose the simplest model that explains data the best the penalized likelihoods could be computed. The AIC and BIC methods compute penalties for the number of parameters used and thus encourage the simplicity (or parsimony) of the selected fit. The parameter set that minimizes the penalized likelihoods is considered to the best fit. To calculate AIC and BIC values we use following (k is the total number of parameters in the parameter set Φ, Θ or $\mathbf{A}, \mathbf{B}$ in the Fourier parametrization and N is the number of linearly independent samples):

$$AIC^{(k)} = -2 \ln L\left(\widehat{\Phi}, \widehat{\Theta}\right) + 2k,$$
$$BIC^{(k)} = -2 \ln L\left(\widehat{\Phi}, \widehat{\Theta}\right) + 2k \log N.$$

3 Real Data Example

In this section identification of PC-T structure, PAR modelling and forecasting procedure is developed for a non-volatile segment of real time series contained hourly observations from volumes of energy traded on the Nord Pool Spot exchange. The Nord Pool Spot runs the largest market for electrical energy in the world, measured in volume traded (TWh) and in market share. The data were found on the Nord Pool Spot Exchange Web page http://www.npspot.com.

3.1 Data

Analyzed data aggregates volume in MWh bought and sold by participants in each bidding area of this market. Thus, this data reflects the demand on the energy on the daily basis. In Fig. 1 there is presented `volumes_2010` time series that contains all hourly observations within 2010 year with non-volatile segment identified. This segment (after removing weekends records) as `volumes` time series is analyzed in our study. It contains hourly observations of the volumes of energy traded from 6 July to 31 August of 2010 ($N = 984$ records). We decided to omit the weekends, as including them would increase the complexity of the model with unknown benefit. So for simplicity and clarity, we focus on understanding the dynamics within the trading part of the week and leave the weekend effects for a future study.

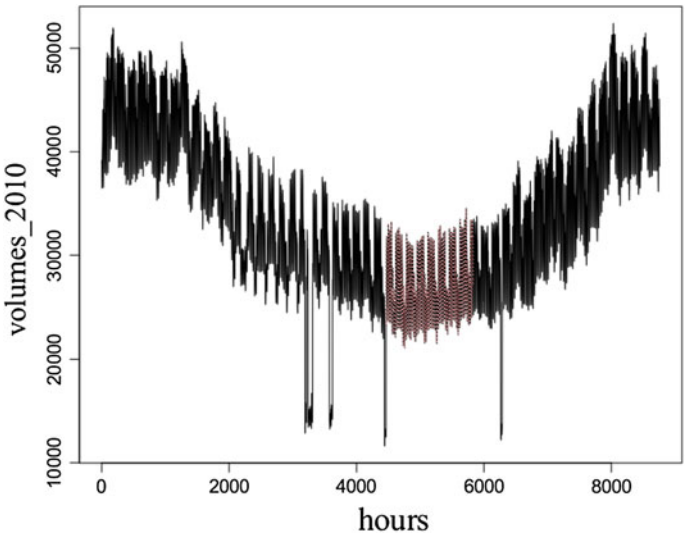


Fig. 1 Volumes of energy traded hourly on the Nord Pool Spot Exchange in 2010 with the non-volatile segment of volumes data (from 6 July 2010 to 31 August 2010)

3.2 Non-parametric Spectral Analysis

As one may expect for `volumes` time series presented in Fig. 1, the daily periodic structure is observed. Even though the period length $T = 24$ is rather obvious in this case, in some cases it may need to be estimated, as the dominant frequencies of the process may not be known at the beginning of analysis. Therefore in our example, we confirm that the period length was properly chosen.

In the `perARMA` package two procedures, based on the spectral analysis, are useful for estimating the period: `pgram()` and `scoh()`. Function `pgram()` plots periodogram of the series and provides test to find candidate for the period length assuming that the period of the second-order structure is the same as the period of the first order structure (i.e. in the instantaneous mean $E\{X(t)\}$ of the series itself). Recall that the FFT index j (where a big peak occurs) corresponds to the number of cycles in the FFT window, so the period corresponding to the index $j = 41$ where the first highly significant peak occurs can be easily computed as $T = 984/41 = 24$ (see Fig. 2). The function `scoh()`, which computes and plots the magnitude-squared coherence, can be used to confirm the presence of the PC structure, to estimate the period of the PC structure and to determine the strength of the PC structure on the different support lines. Specifically, the magnitude-squared coherence $|\hat{\gamma}(\lambda_p, \lambda_q, M)|^2$ is computed in a specified square set of (λ_p, λ_q) and using a specified smoothing window, M. The perception of this empirical spectral coherence is aided by plotting the coherence values only at points where threshold is exceeded. To ensure that periodic structure seen in the spectral coherence image is not a consequence of an additive periodic mean, it is recommended that the periodic mean should be first removed from the series. In Fig. 2 there are presented periodogram plot and magnitude-squared coherence values. In the right, the first significant

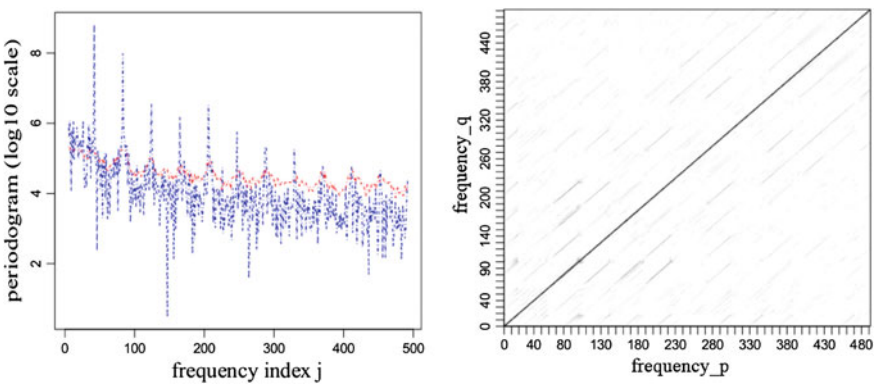


Fig. 2 Periodogram of 'volumes' in the logarithmic scale (*solid line*) with threshold line (*dashed*) for $\alpha = 0.05$ computed from $F(2, 2 * M)$ statistic with $M = 4$. Peaks at $j = 41, 82, \ldots$ (*left-hand side*). Squared coherence statistic, values exceeding an $\alpha = 0.05$ threshold, applied to the 'volumes' series after removing periodic mean with $M = 16$ (*right-hand side*)

off-diagonal is at $|p-q| = 41$ which verifies the first significant peak at $j = 41$ in the periodogram plot on the left. This shows that there is a second-order PC structure with period $T = 494/41 = 24$ in the data.

3.3 Preliminary Identification and Conditioning

Knowing the period length enables one to compute the basic periodic characteristics of the series: periodic mean and periodic standard deviations, which are computed and plotted by procedures `permest()` and `persigest()`, respectively. Both are plotted as a function of season with $1-\alpha$ confidence intervals based on all non-missing values present for each particular season. For periodic mean confidence intervals are based on the t distribution, whereas for periodic standard deviations the chi-square distribution was used. Figure 3 presents plots of the periodic sample mean and periodic sample deviation, along with 95 % confidence intervals, for the `volumes` data with $T = 24$; it is clearly indicated that the sample mean and periodic sample deviation are not constant and thus the data are non-stationary. The p value for a one-way ANOVA test for equality of seasonal means and p value for Barttlet's test for homogeneity of variance $\sigma(t) \equiv \sigma$ are also computed; by $\sigma(t) \equiv \sigma$ we mean $\sigma(t) = \sigma, t = 0, 1, \ldots, T-1$. These tests resulted in p value $= 9.74 * 10^{-60}$ for the ANOVA test and p value $= 0$ for Barttlet's test. Rejection of homogeneity based on the p value indicates a properly periodic variance, but leaves open whether or not series is simply the result of a stationary process subjected to amplitude-scale modulation. To resolve this, $R(t+\tau, t)$ for some set of τ values with $\tau \neq 0$ needs to be calculated. If we cannot reject '$R(t, t+\tau)$ is properly periodic' for some $\tau \neq 0$ then the series is an amplitude-modulated stationary sequence.

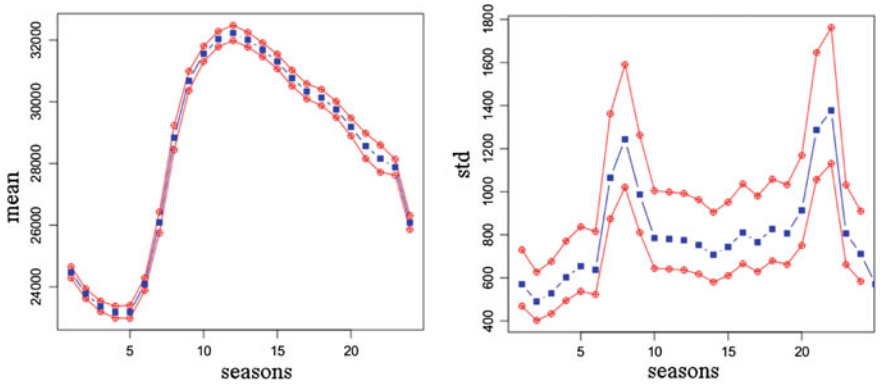


Fig. 3 Estimated periodic mean (*left-hand side*) and periodic standard deviations (*right-hand side*), with $1-\alpha$ confidence intervals for $\alpha = 0.05$, for 'volumes' series. Number of samples per season = number of periods = 41

3.4 Identification

In this section the determination of p and/or q orders will be carried out. Therefore, periodic version of ACF and PACF functions for `volumes` data will be computed.

3.4.1 Autocovariance Function of PC Processes

At the identification step the covariance function $R(t+\tau, t)$ (see (7)) and/or correlation coefficients $\rho(t+\tau, t)$ for each particular season $t = 0, 1, \ldots, T-1$ and lags τ should be computed. The essential calculation could be performed in two different ways: direct and indirect.

- Direct method: tests directly on $\rho(t+\tau, t)$
  We first compute correlation coefficients $\rho(t+\tau, t)$ (formula (12)), where $t = 0, 1, \ldots, T-1$ are seasons and τ is lag (procedure `peracf()`). For each possible pair of t and τ confidence limits for these coefficients are also computed using Fisher transformation, see Fig. 4. Two important tests are also performed:

  (a) $\rho(t+\tau, t) \equiv 0$ meaning $\rho(t+\tau, t) = 0, t = 0, 1, \ldots, T-1$ the rejection for any $\tau \neq 0$ indicates that the sequence is not PC white noise.
  (b) $\rho(t+\tau, t) \equiv \rho(\tau)$ meaning $\rho(t+\tau, t) = \rho(\tau), t = 0, 1, \ldots, T-1$, rejection for any $\tau \neq 0$ indicates that series is properly PC and is not just an amplitude-modulated stationary sequence. That is, there exist lags τ for which $\rho(t, \tau)$ is properly periodic in variable t.

  An issue is the number of lags that need to be tested. A quick estimate can be obtained from examining the usual ACF applied to the data and using the largest lag producing significant non-zero correlation.

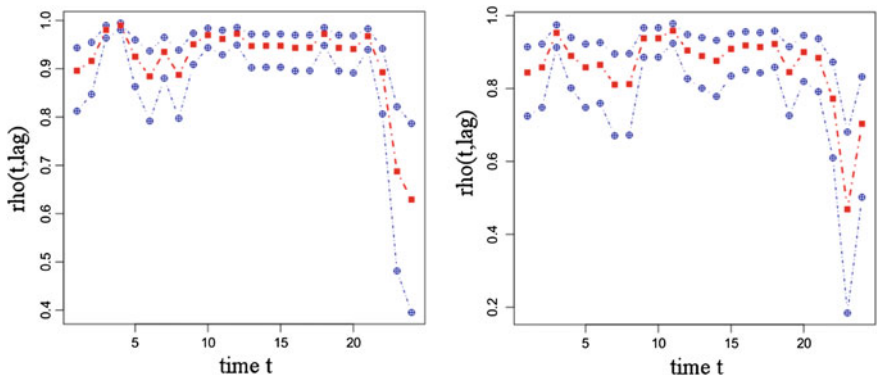


Fig. 4 Correlation coefficients of 'volumes' series: $\hat{\rho}(t, 1)$ versus t and $\hat{\rho}(t, 2)$ versus t with $1-\alpha$ confidence intervals for $\alpha = 0.05$

Table 1 Constancy of correlation coefficients for 'volumes' data: for each fixed τ we give the p values for tests of constancy in correlations for $\tau = 1, 2, \ldots, 12$

τ	$(a) : \rho_t(\tau) \equiv 0$	$(b) : \rho_t(\tau) = \text{const}$
1	0.00	0.00
2	0.00	4.86623×10^{-9}
3	0.00	2.30767×10^{-4}
4	0.00	0.02262
5	0.00	0.15584
6	0.00	0.41075
7	0.00	0.48365
8	0.00	0.25407
9	0.00	0.038110
10	0.00	0.00770
11	0.00	0.00846
12	0.00	0.00202

Tests are (a) if the coefficients are zero for all t: $\rho_t(\tau) \equiv 0$ and (b) if they are constant: $\rho_t(\tau) \equiv \rho(\tau)$

We note that if $\rho(t + \tau, t) \equiv 0$ is rejected for some $\tau \neq 0$ then also $\rho(t + \tau, t) \equiv \rho(\tau)$ is rejected for that τ. This follows also because if a process is PC white noise, then it is also an amplitude-modulated stationary sequence. Table 1 presents results for both tests when lags $\tau = 1, 2, \ldots, T$ were considered.
Conclusions from direct tests:

1. test (a): hypothesis $\rho_t(\tau) \equiv 0$ for all analyzed lags is rejected. It means that the process exhibits non-zero correlation at non-zero lags, meaning it is not PC neither stationary white noise;
2. test (b): hypothesis $\rho_t(\tau) \equiv \rho(\tau)$ is rejected for lags $\tau = 1, 2, 3$, weakly rejected for $\tau = 10, 11, 12$ and not rejected for other lags. Significant values at lag $\tau = 1, 2, 3$ show that there is periodicity in correlations even when periodic variance is scaled out.

- Indirect method: tests on Fourier coefficients $B_k(\tau)$
  In the second approach computation of the complex estimators $\hat{B}_k(\tau)$ (see (10)) for particular lags τ and $k = 0, 1, \ldots, T - 1$ is performed (procedure `Bcoeff()`). Moreover, p values for the test of $B_k(\tau) = 0$ for all τ and k are returned. Additionally, note that testing if $R_d(t, t) = \sigma_t^2 \equiv \sigma^2$ is equivalent to testing if $B_k(0) = 0$ for $k = 1, \ldots, T - 1$.
  Computations are made for each specified τ, the values of $\hat{B}_k(\tau)$ only for $k = 0, 1, \ldots, \lfloor (T - 1)/2 \rfloor$ because $B_k(\tau) = \overline{B_{T-k}(\tau)}$. In addition, the p values for the test if $B_k(\tau) = 0$ are presented in Table 2. These p values should be treated with caution because the requisite assumptions may not be met (see [16]). In the first two columns we have results for $\tau = 0, 1$ for the original series `volumes`, whereas in the next columns there are results for $\tau = 1, 2$ for the normalized

Table 2 The results of testing the nullity of Fourier coefficients $B_k(\tau)$ and $\rho_k(\tau)$

k	$B_k(0)$	$B_k(1)$	$\rho_k(1)$	$\rho_k(2)$
0	0.00	0.00	0.00	0.00
1	1.48×10^{-1}	8.68×10^{-2}	7.09×10^{-1}	5.36×10^{-1}
2	1.11×10^{-16}	2.30×10^{-12}	8.25×10^{-1}	8.04×10^{-1}
3	8.02×10^{-5}	4.35×10^{-3}	5.51×10^{-1}	4.13×10^{-1}
4	1.43×10^{-3}	3.76×10^{-3}	7.62×10^{-1}	7.43×10^{-1}
5	4.98×10^{-5}	1.16×10^{-3}	8.17×10^{-1}	7.54×10^{-1}
6	5.06×10^{-1}	6.46×10^{-1}	9.43×10^{-1}	8.39×10^{-1}
7	3.46×10^{-3}	4.25×10^{-2}	9.14×10^{-1}	9.23×10^{-1}
8	4.66×10^{-1}	6.54×10^{-1}	8.99×10^{-1}	8.54×10^{-1}
9	3.05×10^{-1}	6.12×10^{-1}	9.57×10^{-1}	9.37×10^{-1}
10	5.75×10^{-1}	7.44×10^{-1}	1.00	9.63×10^{-1}
11	7.54×10^{-1}	9.65×10^{-1}	9.76×10^{-1}	9.49×10^{-1}
12	7.55×10^{-1}	7.49×10^{-1}	9.84×10^{-1}	9.99×10^{-1}

Significant values of $B_k(\tau)$ occurs for $k = 0$, 1 and of $\rho_k(\tau)$ for $k = 0$

series (`volumes` normalized by the sample variance $\hat{\sigma}^2(t)$). If the series is the result of an amplitude-scale modulation of a stationary series, then we expect that $\rho_k(\tau) = 0$ will be rejected only for $k = 0$ and $\tau = 0$ and possibly some other τ and it will never be rejected for any other $k \geq 1$ and lag τ.
Conclusions from indirect tests:

1. Rejection for $\tau = 0$ of hypothesis $B_0(0) = 0$ was expected because $B_0(0)$ is the average variance of the sequence and therefore is non-zero for nontrivial sequences;
2. The strong rejection of that $B_2(\tau) = 0$ for $\tau = 0$ indicates the periodicity in the variance (result is consistent with the `persigest()` function output);
3. Hypothesis that $B_0(1) = 0$ and $B_2(1) = 0$ are strongly rejected, showing very significant periodic variation in correlation. It also indicates that covariance function $R(t + \tau, t)$ is periodic for lags τ with frequency $\lambda = 4\pi/T$;
4. Hypothesis $\rho_k(\tau) = 0$ is strongly rejected for $k = 0$ and $\tau = 1, 2$, which means that there exist high correlation coefficients in the normalized `volumes` series, so it is not white noise;
5. Hypothesis that $\rho_k(\tau) = 0$ is never rejected for $k > 0$ and $\tau = 1, 2$, so using this method we cannot reject the hypothesis that the correlation coefficients for lag τ are constant (again consistent with normalized series being stationary, not necessarily white noise).

Note that indirect method of covariance function estimation provides the opposite conclusion in comparison to the direct method. In this case we cannot reject the possibility that analyzed series is a result of amplitude modulation of stationary series (the hypothesis about correlation coefficients equal to zero was not rejected). Surprisingly, it can happen because direct method examines the sample time-dependent

correlations in the period, whereas indirect method returns estimators for Fourier coefficients. It is argued that usually we can reject $\rho(t+1,t) \equiv \rho_1$ better in time domain than in the frequency domain [16, pp. 228–292].

3.4.2 Partial Autocovariance of PC-T Processes

Computation of periodic sample correlation coefficients $\hat{\pi}_{n+1}(t) = \hat{\pi}(n+1,t)$ (see (15)) is provided by the `perpacf()` function, see Fig. 5. Also `ppfcoeffab()` procedure is applied to represent $\pi(n,t)$ by its Fourier coefficients. If the variation in time of $\pi(n,t)$ is sufficiently smooth over the period T, then looking at these Fourier coefficients may be a more sensitive detector of linear dependence of x_{t+1} on the preceding n samples (n is fixed here) than looking at $\pi(n,t)$ for individual times. Two additional hypothesis tests are also provided:

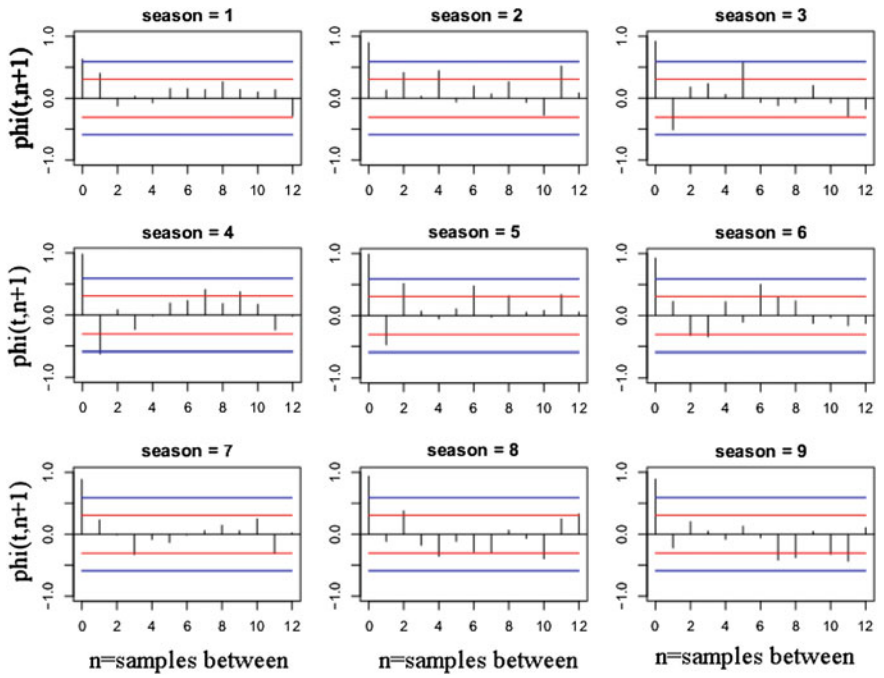


Fig. 5 Partial correlation coefficients $\hat{\pi}(n+1,t)$ of the 'volumes' series for specified number of samples $n = 1,\ldots,12$ and seasons $t = 1,\ldots,9$ with $1-\alpha$ confidence intervals, $\alpha = 0.05$; inner lines are confidence intervals for the null hypothesis of $\pi(n+1;t) = 0$ and are based on the asymptotic null distribution; the outer lines are based on the same test but with Bonferroni correction for the number of n; coefficients values are presented on separate plots for particular $t = 1,\ldots,T$

(a) if $\pi(n_0 + 1, t) = 0$ for $t = 0, 1, \ldots, T - 1$ and each fixed n_0 for $n_0 = 1, 2, \ldots, nmax$,
(b) if $\pi(n + 1, t) = 0$ for $t = 0, 1, \ldots, T - 1$ and $n_0 \leq n \leq nmax$.

For `volumes` data only for $n = 0, 2$ null hypothesis are rejected. Thus, only $\pi(1, t) \neq 0$ and $\pi(3, t) \neq 0$ for some values of t. For $n = 1$ and $n > 2$ and all t we can assume that $\pi(n + 1, t) = 0$. The obtained results suggest fitting a PAR(3) model to the data.

3.4.3 Usual ACF and PACF

Finally, to check the strength of dependence between the variables usual (meaning not periodic) ACF and PACF functions are plotted. It happens that for PC time series the usual ACF and PACF are still useful in the identification of model orders and in some cases can be more sensitive than the periodic versions. When subjected to a truly PC sequence, the usual ACF and PACF produce an average (of sorts; the exact relationship is an open question) of the instantaneous (time indexed) values produced by periodic ACF and periodic PACF. Depending therefore on correlations between these averaged quantities, the usual ACF and PACF may be more or less sensitive than the instantaneous ones. Function `acfpacf()` plots values of usual ACF and PACF functions with confidence intervals. It is possible to run this procedure on original data which include the periodic mean as a deterministic component. But typically the periodic mean can distort our understanding (or view) of the random fluctuations, thus using data after removing periodic mean is recommended as well (see Fig. 6). As a result of identification stage the orders of lags p and/or q for model of PARMA type: PARMA(p,q), PAR(p) or PMA(q), should be determined. For `volumes` data, periodic ACF values point to significant periodic correlations to many lags, which is consistent with PAR model with large coefficients. One can observe the significant periodic PACF values at $n = 0$ and maybe $n = 1$, which suggests starting with

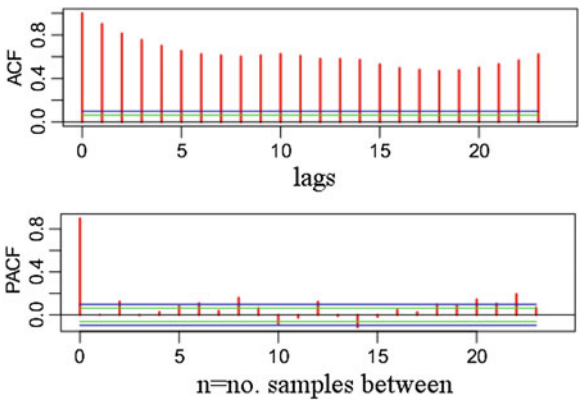


Fig. 6 Usual ACF and PACF values of 'volumes' series together with 95 % confidence intervals (*top* and *bottom* plot, respectively); think inner dotted line is 95 % CI for null hypothesis of $\hat{\pi}(n + 1) = 0$ and outer is same but Bonferonni corrected for the number of n plotted (i.e. 24)

PAR(1) model and increasing the order of lags if it will be necessary. Significant values of the usual ACF and PACF are consistent with the periodic ones (strong autocorrelation values on ACF plot for all specified lags and significant values of PACF for some n, i.e. $n = 0, 2$, mainly for $n = 0$, p value equal to 0). This indicates a strong average lag 1 correlation and a weaker lag 3 ($lag = n + 1$) correlation suggesting a PAR(3) with a high average $\phi_1(t)$ and a low average $\phi_3(t)$.

3.5 Model Fitting

The PARMA(p,q) model has in total $T(p + q + 1)$ coefficients, thus, especially for long period length, the problem of estimation is computationally burdensome. Therefore, at least at the beginning, to reduce the number of parameters, it is recommended to fit models which explain that data with the fewest possible parameters. Then, only if proposed model is not sufficient, the order of lags should be increased. In the `perARMA` package, the Yule–Walker moment estimates for the general PAR model (procedure `perYW()`) and the approximate maximum likelihood parameter estimates for the general PARMA model (procedure `parmaf()`) are implemented. To illustrate the functionality of these procedures we apply them to the `volumes` data, after removing periodic mean, elaborating identification clues. First, we fit the simplest possible model, i.e. PAR(1)

$$X_t - \phi(t)X_{t-1} = \sigma(t)\xi_t,$$

where ξ_t is white noise with zero mean and variance equal to 1. The total number of parameters in this model is equal to $2T$. As it is shown in Figs. 7 and 8 the

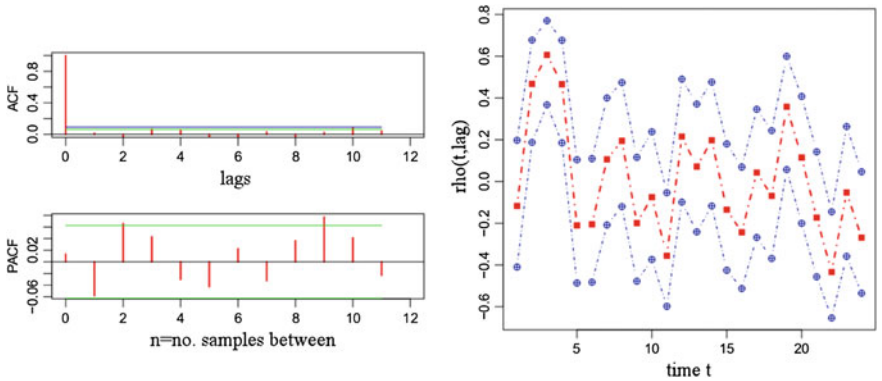


Fig. 7 PAR(1) model residuals evaluation: Usual ACF and PACF values together with 95 % confidence intervals (*top* and *bottom* plot, respectively); think inner dotted line is 95 % CI for null hypothesis of $\hat{\pi}(n + 1) = 0$ and outer is same but Bonferonni corrected for the number of n plotted (i.e. 12) (*left-hand side*). Correlation coefficients $\hat{\rho}(t, 1)$ versus t with confidence intervals for $\alpha = 0.05$ (*right-hand side*)

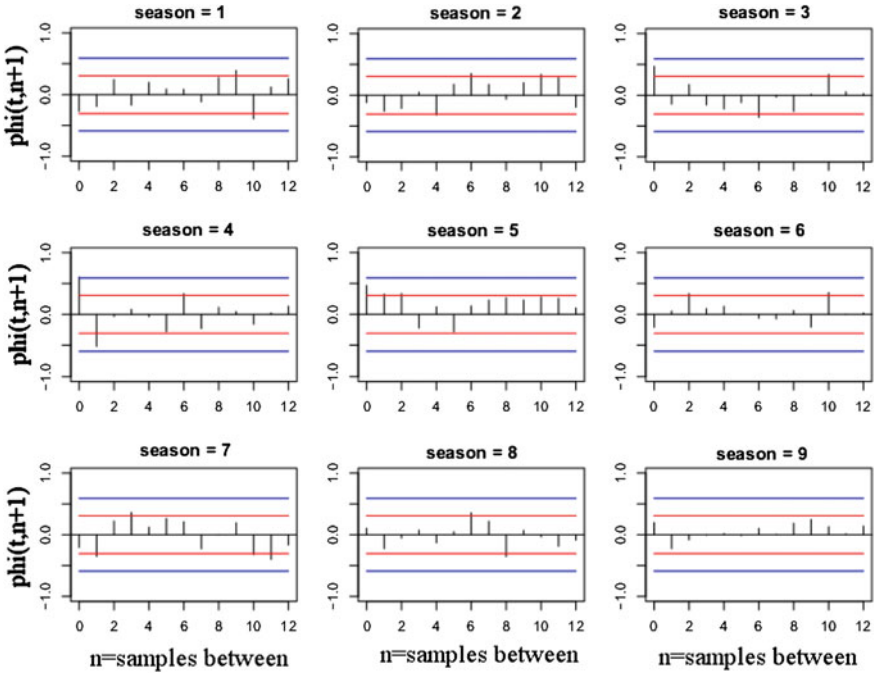


Fig. 8 Partial correlation coefficients $\hat{\pi}(n+1,t)$ of PAR(1) model residuals for specified number of samples $n = 1,\ldots,12$ and seasons $t = 1,\ldots,9$ with $1-\alpha$ confidence intervals, $\alpha = 0.05$; inner lines are confidence intervals for the null hypothesis of $\pi(n+1,t) = 0$ and are based on the asymptotic null distribution; the outer lines are based on the same test but with Bonferroni correction for the number of n; coefficients values are presented on separate plots for particular $t = 1,\ldots,12$

residuals are not consistent with periodic white noise ones. Even if correlations significantly decreased, for $\tau = 1$ test for equality still rejects the null hypothesis with p value equal to $1.54E-7$. Then we fit another model with the increased order of autoregression p, i.e. PAR(2)

$$X_t - \hat{\phi}_1(t)X_{t-1} - \hat{\phi}_2(t)X_{t-2}(t) = \hat{\sigma}(t)\xi_t,$$

where ξ_t is white noise with zero mean and variance equal to 1. The total number of parameters is equal to 3T. We repeat the whole procedure of examining residuals and checking if residuals are consistent with periodic white noise. In Fig. 9 one can observe that significant periodic correlations in the residuals are completely absent in residuals of PAR(2) model. For lag $\tau = 1$ both null hypothesis (provided by `peracf()` procedure) were not rejected with p values equal to 0.99 and 0.81. Also usual ACF and PACF functions for residuals are consistent with periodic white noise ones. In Fig. 10 there are presented coefficients $\hat{\pi}(n+1,t)$ for PAR(2) residuals ($n = 1,\ldots,8$ and $t = 1,\ldots,12$) together with confidence intervals for $\alpha = 0.05$: inner for $\pi(n+1,t) = 0$ and outer for Bonferroni correction. There are finally

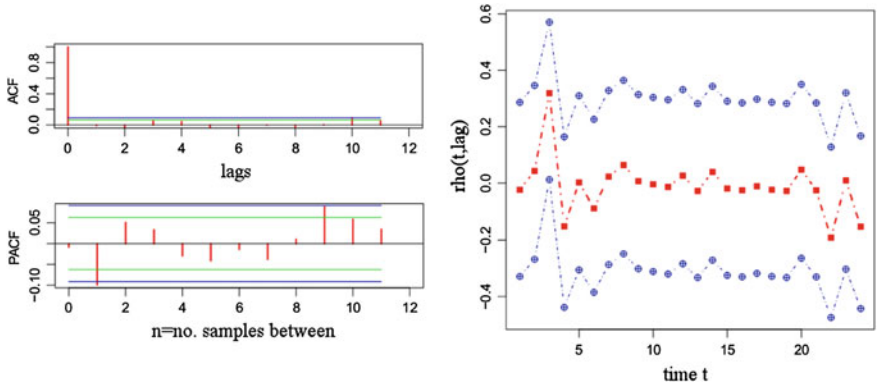


Fig. 9 PAR(2) model residuals evaluation: Usual ACF and PACF values together with 95 % confidence intervals (*top* and *bottom* plot, respectively); think inner dotted line is 95 % CI for null hypothesis of $\hat{\pi}(n+1) = 0$ and outer is same but Bonferonni corrected for the number of n plotted (i.e. 12) (*left-hand side*). Correlation coefficients $\hat{\rho}(t, 1)$ versus t with confidence intervals for $\alpha = 0.05$ (*right-hand side*)

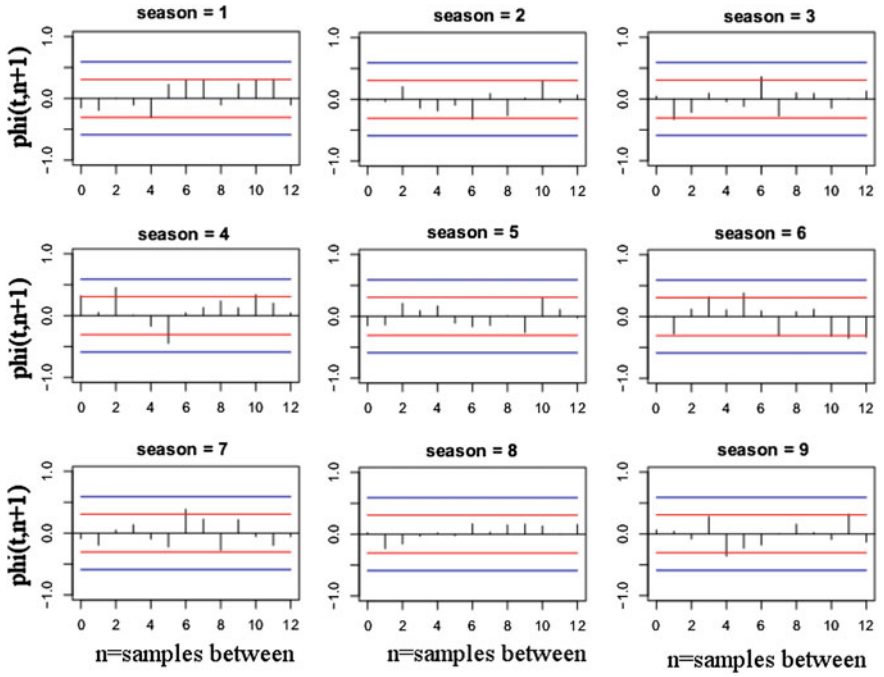


Fig. 10 Partial correlation coefficients $\hat{\pi}(n+1, t)$ of PAR(2) model residuals for specified number of samples $n = 1, \ldots, 12$ and seasons $t = 1, \ldots, 9$ with $1 - \alpha$ confidence intervals, $\alpha = 0.05$; inner lines are confidence intervals for the null hypothesis of $\pi(n + 1, t) = 0$ and are based on the asymptotic null distribution; the outer lines are based on the same test but with Bonferroni correction for the number of n; coefficients values are presented on separate plots for particular $t = 1, \ldots, 12$)

no values which exceed the outer threshold. For periodic white noise we expect $\pi(n+1,t) = 0$, for all t and n and in this case this hypothesis is definitely not rejected (p value $= 0.27$).

Additionally, the penalized likelihoods for both PAR models were computed using `loglikec()` procedure. AIC values were equal to 6873.231 and 6821.9 for PAR(1) and PAR(2) models, respectively. BIC values were equal to 6883.012 and 6836.569 for PAR(1) and PAR(2) models, respectively. Both criteria prefers PAR(2) model and analysis of residuals confirms PAR(2) model as a better fit. Furthermore, as residuals of PAR(2) model are consistent with periodic white noise, it seems that `volumes` data did not require a full PARMA model.

3.6 Short Time Prediction for PAR Models

In this section a diagnostic checking is carried out by comparing predicted values of `volumes` series (based on PAR models coefficients) with the corresponding real values of this series on 1 and 2 September 2010. First procedure `predictperYW()`,

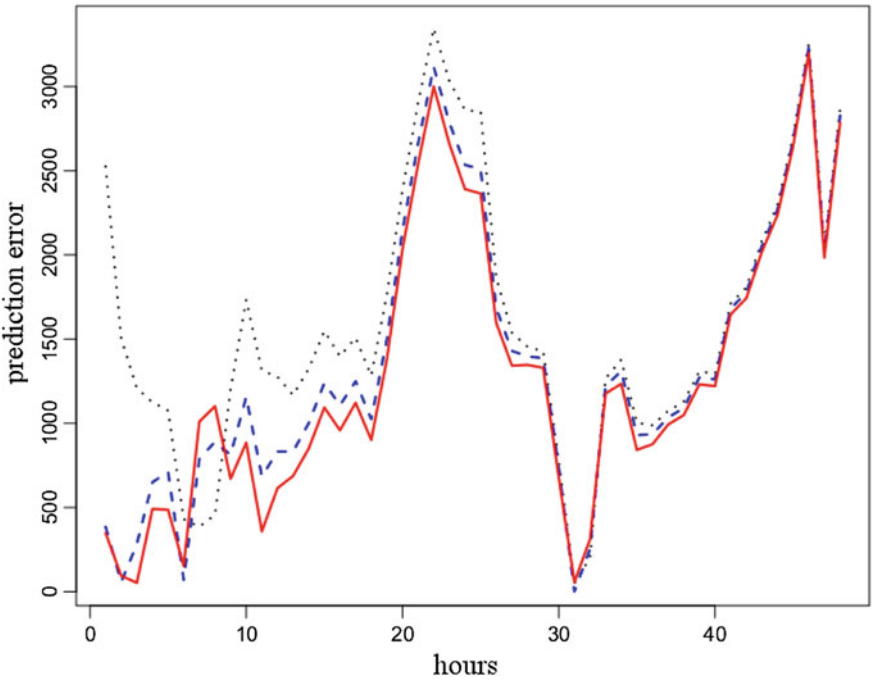


Fig. 11 Prediction errors (i.e. difference between predicted and true values of the series) on 1 and 2 September 2010 for various ways of predicting: PAR(1) (*dashed line*), PAR(2) (*solid line*) and sample periodic mean (*dotted line*)

based on the Yule–Walker estimators, is applied to demeaned `volumes` series to forecast values for the next 2 days (48 observations). Then sample periodic mean is added to obtained predictors and results are compared with the real (observed) values of the series. Instead of analyzing particular observations values the prediction error, i.e. observed values minus predicted values is considered. In Fig. 11 there are presented prediction errors for both PAR(1) and PAR(2) models and just sample periodic mean used as a predictor as well. Taking these results into account, it seems that the forecast based on PAR(2) model approximates the real values the best.

4 Simulated Data Example

In the previous section PAR(2) model was the optimal choice and Yule–Walker estimators were sufficient but not always this needs to be the case. Therefore, in this section, to complete our considerations about PARMA modelling, we illustrate full PARMA model estimation method, based on maximization of log-likelihood function. We use simulated dataset to test the performance of the estimation procedures for the PARMA(p,q) model.

We decided to consider the simulated datasets to check the performance of our procedures for full PARMA model estimation. In the example below, a PARMA(2,1) sequence is generated with period length $T = 12$ and series length equal to $N = 480$. Knowing the orders of the original PARMA sequence we can compare them with obtained output from estimation procedures presented in the next section.

For general PARMA we use non-linear optimization methods to obtain maximum logarithm of likelihood function. In `perARMA` package procedure `parmaf()` enables to estimate the values of the PARMA(p,q) model. This method of computation of log-likelihood function is based on the representation of the periodically varying parameters by Fourier series (see (17)). Inside the procedure the negative logarithm of likelihood function from the PARMA sequence for matrices of coefficients in their Fourier representation is computed. This alternative parametrization of PARMA system, introduced by [17], can sometimes substantially reduce the number of parameters required to represent a PARMA system. This permits estimation of the values of the chosen representation of PARMA(p,q) model using non-linear optimization methods. Initial values of parameters are computed using Yule–Walker equations.

To illustrate functionality of these procedure we try to fit PARMA-type model to the series generated as PARMA(2,1). We consider three PARMA-type models: PARMA(0,1), PARMA(2,0) and PARMA(2,1) and compare the result with reality. According to log-likelihood, AIC and BIC values presented in Table 3, PARMA(2,1) is considered to be the best fitted. It is consistent with our expectations as analyzed series y was generated using orders $p = 2, q = 1$.

Table 3 The information criteria for validation of PARMA models fits when true model was PARMA(2,1)

model	negloglik	AIC	BIC
PARMA(0,1)	969.05	977.05	993.74
PARMA(2,0)	748.98	762.98	792.17
PARMA(2,1)	653.38	673.38	715.07

5 Conclusions

In the present study we showed that PARMA model approach works for real data with periodic structure. We follow all steps of standard model fitting procedure in regard to PARMA models through non-parametric spectral analysis, model identification, parameter estimation, diagnostic checking (model verification). Additionally, we perform test for period length detection, prediction for PAR models and estimation of full PARMA for simulated data. The results presented here illustrate how PARMA approach can be applied to model periodic structure of energy market data and confirm the considerable value of this method in forecasting short-term energy demand.

To our knowledge `perARMA` package is the only existing tool in statistical software programs that provide PARMA time series analysis and allow to follow all steps of standard model fitting procedure. Additionally, it deals with missing data, provides period length detection tools and prediction for PAR models. Moreover, `perARMA` package permits simulation of PARMA time series.

Acknowledgments Research of Anna E. Dudek was partially supported by the Polish Ministry of Science and Higher Education and AGH local grant.

References

1. Anderson, P. L., Meerschaert, M. M., & Zhang, K. (2013). Forecasting with prediction intervals for periodic autoregressive moving average models. *Journal of Time Series Analysis, 34*(2), 187–193.
2. Anderson, P. L., Tesfaye, Y. G., & Meerschaert, M. M. (2007). Fourier-PARMA models and their application to river flows. *Journal of Hydrologic Engineering, 12*(5), 462–472.
3. Balcilar, M., & McLeod, A. I. (2011). pear: Package for periodic autoregression analysis. http://cran.r-project.org/web/packages/pear/.
4. Bloomfield, P., Hurd, H. L., & Lund, R. (1994). Periodic correlation in stratospheric ozone data. *Journal of Time Series Analysis, 15*, 127–150.
5. Bloomfield, P., Hurd, H. L., Lund, R. B., & Smith, R. (1995). Climatological time series with periodic correlation. *Journal of Climate, 8*, 2787–2809.
6. Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting*. New York: Springer.
7. Broszkiewicz-Suwaj, E., Makagon, A., Weron, R., & Wyłomańska, A. (2004). On detecting and modeling periodic correlation in financial data. *Physica A, 336*, 196–205.
8. Cramér, H. (1961). *Methods of mathematical statistics*. New York: Princeton University Press.

9. Dehay, D., & Hurd, H. L. (1994). Representation and estimation for periodically and almost periodically correlated random processes. In W. A. Gardner (Ed.), *Cyclostationarity in communications and signal processing*. New York: IEEE Press.
10. Dudek, A. E., Hurd, H., & Wójtowicz, W. perARMA: Package for periodic time series analysis. R package version 1.5, http://cran.r-project.org/web/packages/perARMA/.
11. Franses, P. H. (1996). *Periodicity and stochastic trends in economic time series*. Oxford: Oxford Press.
12. Gardner, W. A. (1994). *Cyclostationarity in communications and and signal processing*. New York: IEEE Press.
13. Gardner, W. A., Napolitano, A., & Paura, L. (2006). Cyclostationarity: Half a century of research. *Signal Processing, 86*, 639–697.
14. Gerr, N. L., & Hurd, H. L. (1991). Graphical methods for determining the presence of periodic correlation in time series. *Journal of Time Series Analysis, 12*, 337–350.
15. Gladyshev, E. G. (1961). Periodically correlated random sequences. *Soviet Mathematics, 2*, 385–388.
16. Hurd, H. L., & Miamee, A. G. (2007). *Periodically correlated random sequences: Spectral theory and practice*. Hoboken: Wiley.
17. Jones, R., & Brelsford, W. (1967). Time series with periodic structure. *Biometrika, 54*, 403–408.
18. López-de-Lacalle, J. (2012). partsm: Periodic autoregressive time series models. R package version 1.1, http://cran.r-project.org/web/packages/partsm/.
19. Pagano, M. (1978). On periodic and multiple autoregressions. *Annals of Statistics, 6*(6), 1310–1317.
20. Parzen, E., & Pagano, M. (1979). An approach to modelling seasonally stationary time series. *Journal of Econometrics, 9*(1–2), 137–153.
21. Rasmussen, P. F., Salas, J., Fagherazzi, L., Rassam, J. C., & Bobée, B. (1996). Estimation and validation of contemporaneous PARMA models for streamflow simulation. *Water Resources Research, 32*(10), 3151–3160.
22. Sabri, K., Badaoui, M., Guillet, F., Belli, A., Millet, G., & Morin, J. B. (2010). Cyclostationary modelling of ground reaction force signals. *Signal Processing*, *90*, 1146–1152.
23. Sakai, H. (1991). On the spectral density matrix of a periodic ARMA process. *Journal of Time Series Analysis, 12*(2), 73–82.
24. Smadi, A. A. (2009). Periodic auto-regression modeling of the temperature data of Jordan. *Pakistan Journal of Statistics, 25*(3), 323–332.
25. Swider, D. J., & Weber, C. (2007). Extended ARMA models for estimating price developments on day-ahead electricity markets. *Electric Power Research, 77*, 583–593.
26. Tesfaye, Y., Meerschaert, M. M., & Anderson, P. L. (2006). Identification of periodic autoregressive moving average models and their application to the modeling of river flows. *Water Resources Research, 42*, W01419.
27. Vecchia, A. V. (1985). Maximum likelihood estimation for periodic autoregressive moving average models. *Technometrics, 27*, 375–384.
28. Vecchia, A. V. (1985). Periodic autoregressive moving average (PARMA) modeling with applications to water resources. *Water Resources Bulletin, 21*, 721–730.

Multidimensional Analysis of New Zealand Electricity Prices

Matylda Jabłońska-Sabuka and Agnieszka Wylomanska

1 Introduction

Electricity prices have become some of the most challenging time series to be modelled mathematically. This is due to electricity market deregulation, which has by now been carried out in many countries across the world. Its aim was to allow fair competition in the market and thus open it to more participants. Indeed, the goal of reducing average price level through competition has been achieved, however, electricity prices have also become more volatile and less predictable than ever before.

An important factor which contributes to high volatility in most of the markets is the large variations in demand and supply of electricity, both of which are very uncertain in deregulated markets [5]. For instance, temperature strongly affects the demand; in total, the demand varies between 50 and 100 %. Thus, as some say, forecasting demand is almost equivalent to forecasting weather [21]. Next to any climatic factors, hydrological balance, demand and base load supply [31] can be considered with equal importance as the key spot price drivers.

Electricity prices can be studied both in their inter- and intraday format (hourly or half-hourly, depending on the market). Substantial research effort has so far been focused on investigating spot price interdependencies. For instance, the New Zealand spot prices can be divided into five intraday groups: overnight off-peak, morning peak, day-time off-peak, evening peak and evening off-peak. Then it appears that prices within these groups are a lot more correlated than between these groups along

M. Jabłońska-Sabuka
Department of Mathematics and Physics, Lappeenranta University of Technology, Lappeenranta, Finland
e-mail: matylda.jablonska-sabuka@lut.fi

A. Wylomanska (✉)
Hugo Steinhaus Center, Department of Mathematics, Wrocław University of Technology, Wrocław, Poland
e-mail: agnieszka.wylomanska@pwr.edu.pl

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods - II*,
Applied Condition Monitoring 3, DOI 10.1007/978-3-319-16330-7_8

different trading periods [10]. Another work analysed a group of models classified as Markov regime-switching (MRS) [13]. There, the focus was on the performance of different models in terms of statistical goodness-of-fit and the results showed that the best one was an independent spike 3-regime model with time-varying transition probabilities, heteroscedastic diffusion-type base regime dynamics and shifted spike regime distributions.

Thorough understanding of intraday price structure makes day-ahead forecasting these days possible up to a significant level of accuracy. However, also long-term predictions are equally important for risk management. It is known that electricity price trend and cyclic structure comes in big part from the influence of some well-recognized factors. Among those one should list hydrological storage (for heavily hydro-dependent markets, like Nord Pool ad New Zealand) and thus rainfall, weather and demand [11, 17, 24, 32]. A stochastic model using a number of explanatory variables has been proposed, for instance, for the California electricity market [14], among others. For the case of New Zealand prices, there is lack of modelling approaches that would be able to thoroughly explain the behaviour of prices. Only some works have just identified that the nature of prices has changed after market deregulation in various ways [34]. Some other researches have used discrete wavelet transforms to investigate the demand-price relationship [22] and market volatility in New Zealand with the use of GARCH models [26].

The main contribution of this study is twofold. First, it exploits the significant interdependencies between the prices in different trading nodes, as well as external factors influencing them, like hydro storage, rainfall, air temperature and demand. This is done through multivariate autoregressive models which provide a lot of advantages over classical regression analysis. Secondly, both proposed models consider non-Gaussian noise with α-stable structure, which allows to capture statistically the occurrence of price spikes. This is a novelty when compared to the classical ARMA-GARCH methodology. Results show that both models provide accurate day-ahead forecasts for the period of a year ahead. Therefore, we demonstrate that a novel combination of VAR models with non-Gaussian noise provides a powerful tool for analysis and day-ahead forecasting of electricity spot prices, including a new way of coping with simulation of price spikes.

This article is structured as follows. Section 2 describes the data set used in the study together with the main statistical features of the data. In Sect. 3, the multivariate autoregressive model with non-Gaussian structure is introduced. Section 4 applies the model to the real data and collects fit and forecast results. Section 5 concludes the results.

2 Main Features of New Zealand Electricity Prices

2.1 New Zealand Electricity Market

The New Zealand electricity market is a very interesting case study for spot prices analysis from many points of view. Its electricity sector is principally (70 %) based on renewable energy sources such as hydropower, geothermal power and a steadily increasing wind energy, which makes New Zealand one of the most sustainable countries in the world when it comes to energy generation. On the other hand, its electricity demand has been significantly growing in the past by an average of 2.4 % per year since 1974 and by 1.7 % over 1997–2007.

New Zealand is characterized by a geographically unbalanced demand–supply relation. The highest electric power production takes place on the South Island, whereas the highest demand comes from the more populated and industrialized North Island. Moreover, the electricity market in New Zealand is not pooled. The main

Fig. 1 Location of the 11 nodes in the New Zealand grid used in the analysis

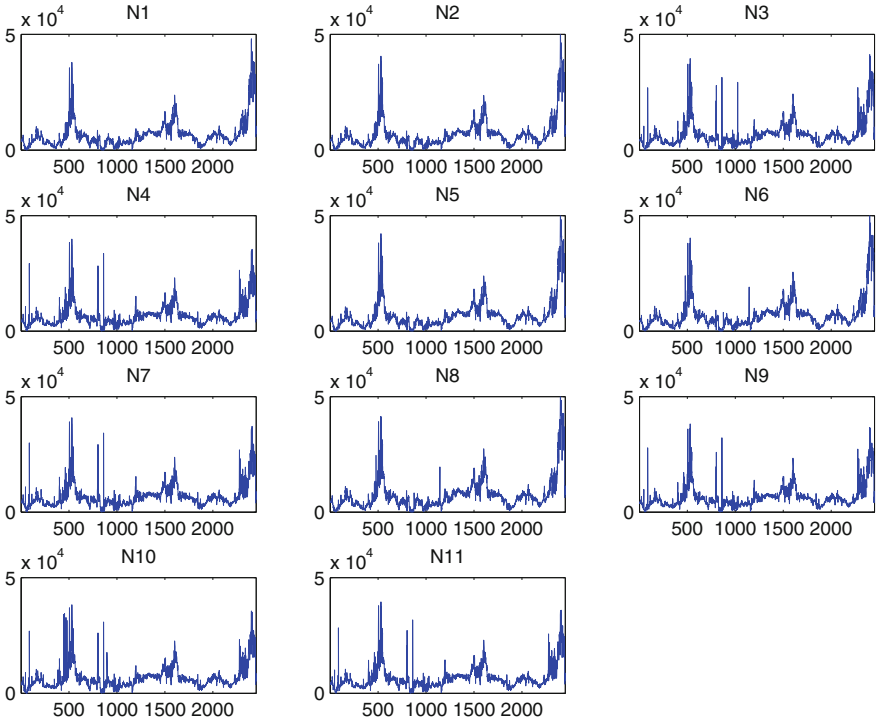


Fig. 2 The examined 11 data sets corresponding to electricity prices in New Zealand in the examined period

participants are seven generators/retailers who trade at over 200 nodes across the transmission grid.

New Zealand market data is available free of charge from the Electricity Authority (former Electricity Commission) in the form of Centralized Dataset. The data consists of half-hourly information on nodal prices, bids and offers, metre data and binding constraints, as well as additional daily hydrology and network configuration data. The authors utilized price data from a number of nodes, as well as the centralized demand and hydrological storage values. Moreover, additional data on rainfall and temperatures in each of the considered nodes was used.

The analysis presented in this paper considers 11 trading nodes spread across both South and North Island, as presented in Fig. 1. Six of them are located near power plants. In particular, Benmore, Tuai and Whakamaru that represent hydropower generation, while Huntly, Otahuhu and Stratford are based on geothermal generation. The remaining five nodes are only splitting substations. The data covers a period from 1 January 1999 to 31 July 2009. Data are quoted daily.

The choice of the trading nodes was deliberate to properly represent the most important power generators as well as the key splitting substations. For instance,

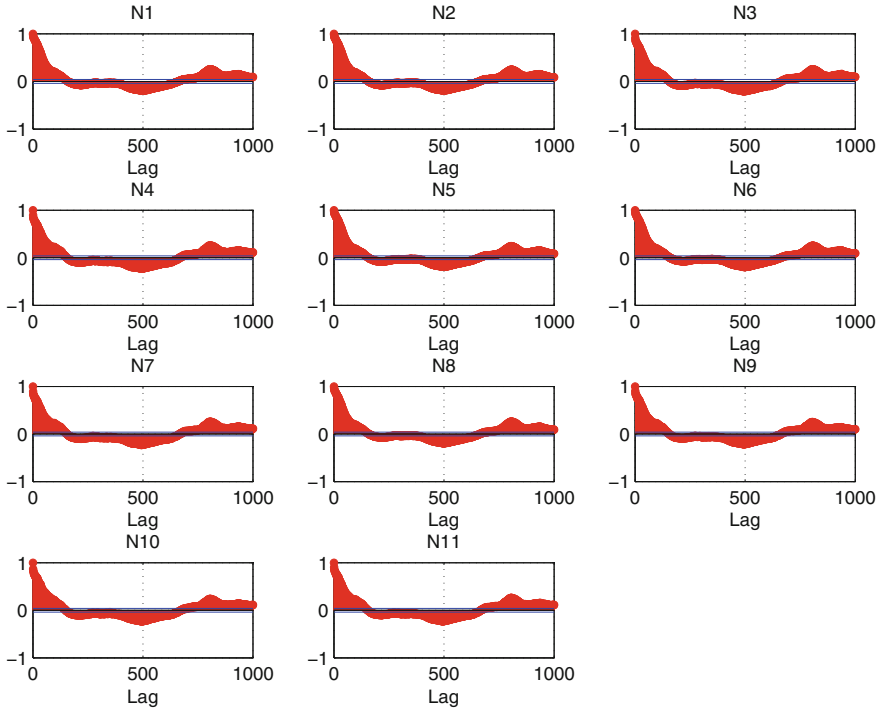


Fig. 3 The ACFs of electricity prices in New Zealand in the examined period

Haywards is the key node splitting power delivered through High Voltage Direct Current (HVDC) connection between the South Island (originating in Benmore) and North Island. Numerous times, price spikes can occur in that node due to electricity transmission constraints. Also, fair balance between the South and North Island was needed as, for instance, prices in the South used to be on average lower than in the North until an additional fee was imposed on the generators for using the HVDC connector to transmit power.

Let us note that the 11 nodes create just a small representative subset of the total of over 200 nodes across the country. Theoretically, it is of course possible to work with all the grid nodes in one model. However, this would, in our opinion, make the system over-represented and, certainly, computationally heavy when it comes to parameter estimation (there would be many more parameters to be estimated that data points available).

2.2 Relationship Between Respective Price Data Corresponding to Analysed Nodes

The first approach proposed in this study is based on relationships between price data for the 11 considered nodes. The analysed data sets, which correspond to 11 vectors of electricity prices in chosen nodes in New Zealand during the examined period, exhibit similar behaviour. In Fig. 2, we present the analysed prices. Moreover, the data sets have also similar statistical properties. The first property that we observe here is the non-stationarity of each vector. This non-stationarity can be easily proven by observing the autocorrelation functions (ACFs) of appropriate data sets. Moreover, in the ACFs we can also see the seasonal behaviour of all the considered vectors of observations. The corresponding ACFs are presented in Fig. 3.

The second statistical property that is observable for all considered data sets is the weekly seasonality. This kind of seasonality can be seen in ACFs of differenced series. In Fig. 4 we present the autocorrelation functions for 11 electricity price vectors, where the weekly seasonality is easily observable for lags that are multiples of seven.

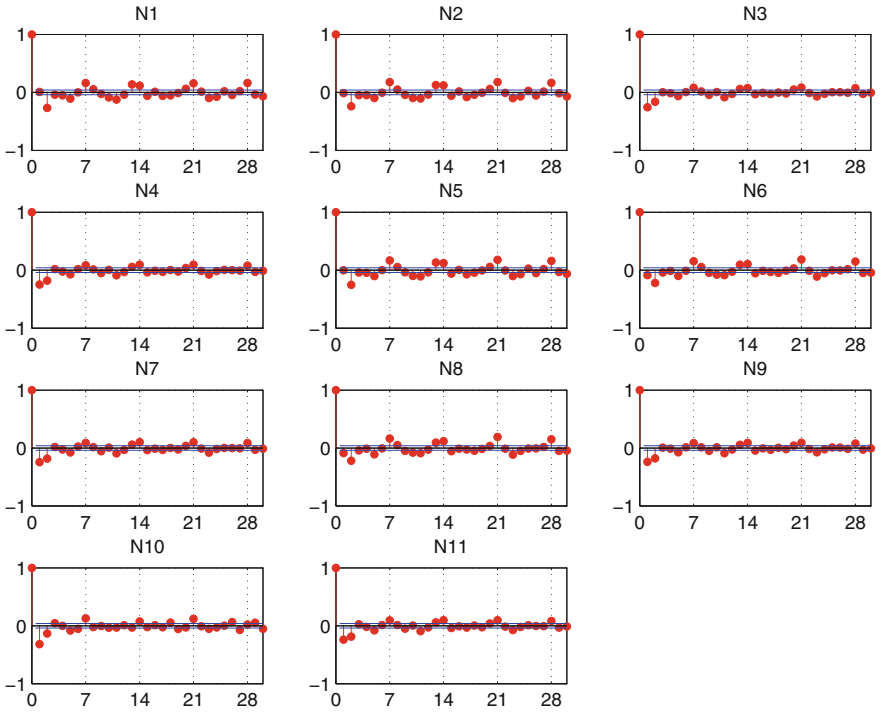


Fig. 4 The ACFs of differenced series. The larger values of autocorrelations at lags being multiples of seven indicate at week seasonality

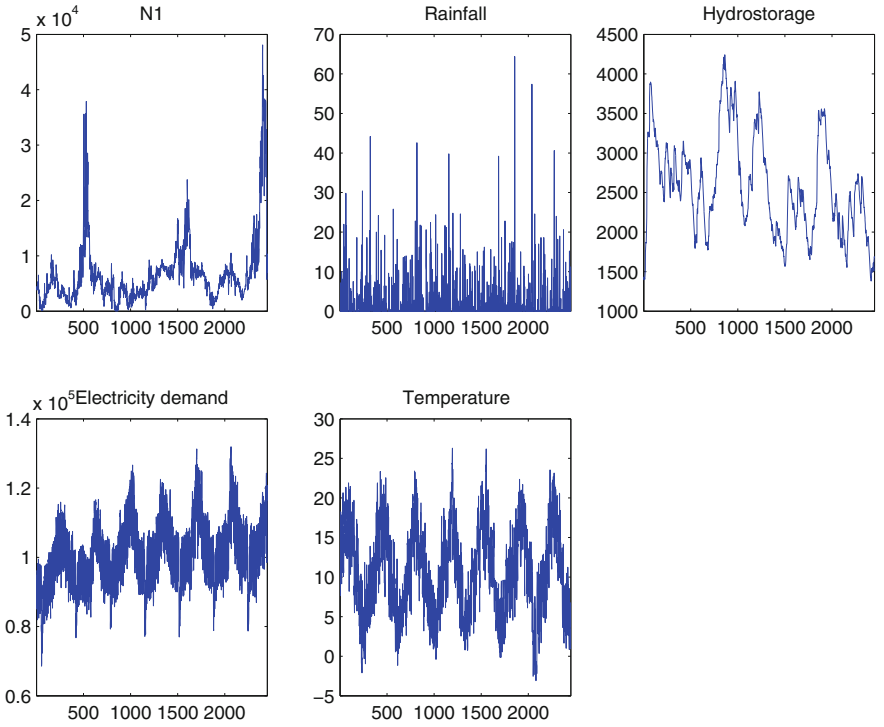


Fig. 5 The five examined data sets corresponding to electricity price $N1$, rainfall, hydro storage, electricity demand and temperature. The rainfall and temperature are measured in the region corresponding to node 1

The last property that should be emphasized here is a strong relation between the analysed data sets. The correlation coefficients are large and the test for its significance indicates at non-zero correlation between analysed prices. That is, all the p-values (for testing the hypothesis of no correlation against the alternative that there is a non-zero correlation) are very close to zero. In particular, all correlation coefficients among the 11 price series themselves exceed 0.9. High correlations could raise a multicollinearity concern, however, this is problematic with classical regression models, whereas it has not been proven such in the case of autoregressive models. The idea of multivariate autoregressive models used in this study is that all the variables are equally important and we do not distinguish between dependent and independent variables. Therefore, there is no need to remove any of the highly correlated components.

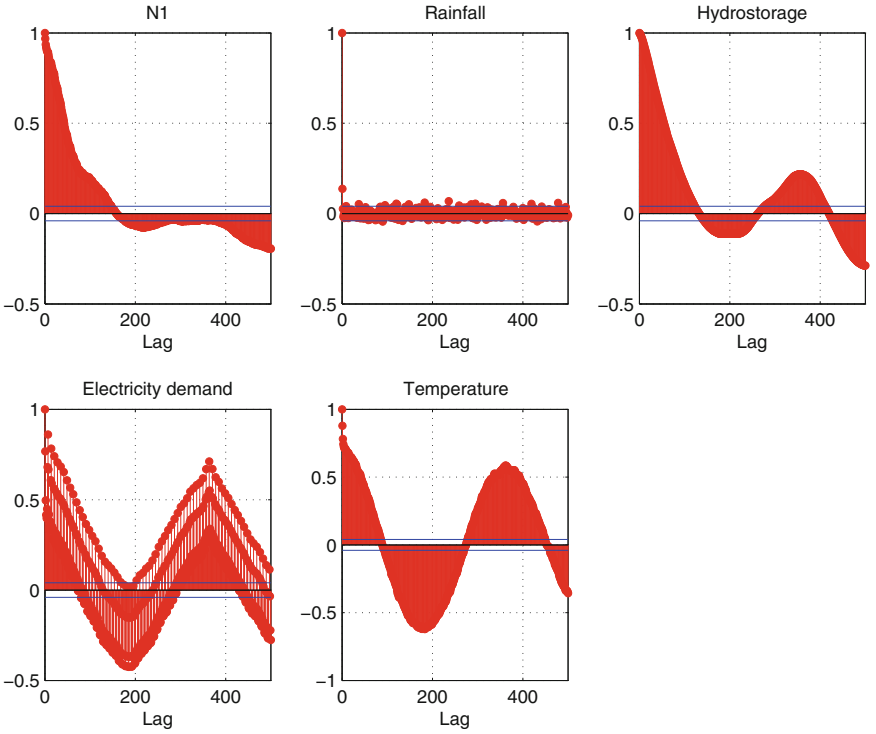


Fig. 6 The ACFs of examined 5 data sets corresponding to electricity price $N1$, rainfall, hydro storage, electricity demand and temperature

2.3 Relationship Between Price Data and Deterministic Factors

The second proposed approach is based on the relationship between price in one node and other variables that can have influence on the price. Along with the nodal spot price, the deterministic factors considered here are: rainfall in the analysed node, country's hydro storage, electricity demand and temperature in the region.

The choice of the variables was deliberate with respect to their level of correlation with the prices, as well as the ability to represent both local and country-level situation. Therefore, some of them are considered as measurements from the particular node location, and some are aggregated for the whole country. First, as we mentioned before, the New Zealand electricity market is heavily hydro-dependent and the information about country's hydrological information is of key importance to price modelling [30]. Therefore, we include two variables which can account for this feature. The amount of rainfall is measured regionally in millimetres per square metre. This variable tells a lot about whether a given season/year is dry or not in the region. On the other hand, we know that prices in all nodes depend on the entire

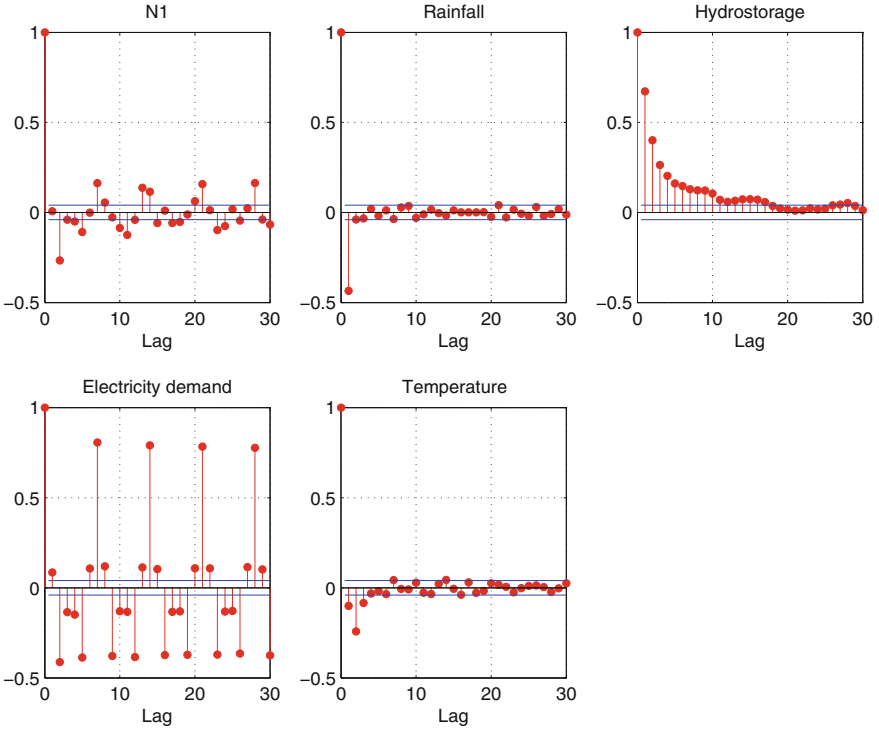


Fig. 7 ACFs of the five differenced data sets corresponding to electricity price $N1$, rainfall, hydro storage, electricity demand and temperature

country's hydro reservoir levels, with correlations between the spot price and the hydro information ranging between -0.39 and -0.49 in various nodes. Namely, when the overall reserves drop, this triggers the prices to rise. Therefore, we add another variable which is the New Zealand's total hydro storage, since the study considers trading nodes that representing not only hydropower plants but also other plants and purely splitting substations. In the same manner we include two demand-related variables, as it is known that when the consumption significantly increases, it may cause the prices to go the same way. One of them is regionally measured air temperature, and another is the total demand calculated for the entire country.

In Fig. 5 we present the five analysed data sets. As we observe, there is a strong relation between the studied variables, which is especially visible in the seasonal behaviour of all data series.

Similarly, as in the first approach, here we should also emphasize the non-stationarity of the data. This non-stationarity can be proved by using ACFs, where we can observe a strong relation between the analysed processes, see Fig. 6. The only exception is rainfall, not showing non-stationarity.

It is worth mentioning that some of the analysed vectors of observations exhibit also weekly seasonal behaviour. We observe this especially for electricity price data and electricity demand. In Fig. 7, we present ACFs of the differenced series, where weekly seasonality is visible for price and demand data.

The last property is a strong relation between the analysed data sets that can be proved not only by visualization but also a strong statistical test. The correlation coefficients are large and the test for its significance indicates at non-zero correlation between analysed variables. All p-values for testing the hypothesis of no correlation against the alternative that there is a non-zero correlation are very close to zero, which indicates the correlations are significantly different from zero. In particular, the correlation coefficients between the price and other variables vary between -0.54 and 0.2. Only one exception is the correlation between rainfall and temperature. For those variables the p-value of the test for no-correlation is equal to 0.6954, which means the correlation is at zero level.

3 Multivariate Autoregressive Model with Non-Gaussian Structure

An m-dimensional process $\{\mathbf{X}_t\} = \{X_t^1, \ldots X_t^m\}$ is an m-variate autoregressive moving average (ARMA) of order (p, q) if it is stationary and if for every t it satisfies the following equation [4]:

$$\mathbf{X}_t - \Phi_1 \mathbf{X}_{t-1} - \cdots - \Phi_p \mathbf{X}_{t-p} = \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \cdots + \Theta_q \varepsilon_{t-q}, \tag{1}$$

where $\{\varepsilon_t\}$ is an m-variate white noise of mean vector $\mathbf{0}$ and covariance matrix $\Gamma(t, t+h)$ that is independent of t and has the following form:

$$\Gamma(t, t+h) = \Gamma(h) \begin{cases} \Sigma \text{ for } h = 0 \\ \\ 0 \quad \text{otherwise} \end{cases} \tag{2}$$

Let us recall that the covariance matrix for m-dimensional vector is the matrix $\Gamma(t, t+h)$ defined as follows:

$$\Gamma(t, t+h) = \begin{bmatrix} \gamma_{11}(t, t+h) & \ldots \gamma_{1m}(t, t+h) \\ \ldots & \ldots \ldots \\ \gamma_{m1}(t, t+h) & \ldots \gamma_{mm}(t, t+h) \end{bmatrix} \tag{3}$$

In the above definition $\gamma_{ij}(t, t+h) = Cov(X_t^i, X_{t+h}^j)$. Let us recall that the Φ_i $i = 1, 2, \ldots, p$ and Θ_i, $i = 1, 2, \ldots, q$ parameters in Eq. (1) are matrices. In further analysis we denote $\{\varepsilon_t\} \sim WN(0, \Sigma)$ (white noise). The process $\{\mathbf{X}_t\}$ is called ARMA(p, q) with mean μ if $\{\mathbf{X}_t\} - \mu$ is an ARMA(p, q) system.

The multivariate ARMA processes are very useful in practice. They found many applications, especially, because of their simple form. They are extensions of known one-dimensional ARMA models extremely popular in different fields. Some interesting applications of multivariate ARMA models can be found in [8, 23, 27, 29].

In the classical approach, the residual series $\{\varepsilon_t\}$ is assumed to be an m-variate Gaussian distributed random variable, i.e. random variable of the following probability density function:

$$f(\mathbf{x}) = (2\pi)^{-m/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \tag{4}$$

where $\mathbf{x} = (x^1, \ldots, x^m)$ and $\mu = (\mu^1, \ldots, \mu^m)$ is the mean vector. Let us mention that if the random variable has m-dimensional Gaussian distribution, each component has one-dimensional Gaussian distribution of appropriate parameters.

For simplicity, in further analysis we concentrate only on multivariate autoregressive (AR) models that we denote as VAR (vector AR). In this case, the procedure of estimating the parameters is based on the method of moments and is an extension of the popular Yule-Walker method applied to one-dimensional AR systems [4]. The method applied to multidimensional AR systems is called the Whittle algorithm [33]. If we multiply the causal VAR process:

$$\mathbf{X}_t = \Phi_1\mathbf{X}_{t-1} + \ldots \Phi_p\mathbf{X}_{t-p} + \varepsilon_t \tag{5}$$

by $\mathbf{X}'_{t-j}$ for $j = 0, 1, \ldots, p$ and take expectations of both sides, then we obtain the following equations:

$$\Sigma = \Gamma(0) - \sum_{j=1}^{p}\Phi_j\Gamma(j),\ \ \Gamma(i) = \sum_{j=1}^{p}\Gamma(i-j),\ \ j = 1, 2, \ldots p$$

Now, taking the empirical equivalences of matrices $\Gamma(0), \ldots, \Gamma(p)$, we can estimate the matrices $\Phi_1, \ldots, \Phi_p$ and Σ. Let us recall that a natural estimator of covariance matrix is the sample covariance, which for the vector $\mathbf{X} = (X^1, \ldots, X^m)$ (each component of n elements) is given by:

$$\hat{\Gamma}(h) = \begin{cases} n^{-1}\sum_{j=1}^{n-h}(\mathbf{X}_{t+h} - \hat{\mu})(\mathbf{X}_t - \hat{\mu}) & \text{for } 0 \le h \le n-1 \\ \hat{\Gamma}'(-h) & \text{for } -n+1 \le h < 0 \end{cases} \tag{6}$$

In the above formula, $\hat{\mu}$ denotes the sample mean vector.

3.1 The VAR Model with Non-Gaussian Structure

Because many real data series exhibit behaviour not adequate for Gaussian systems, in this paper we extend the classical VAR model presented above by replacing the m-dimensional Gaussian distribution with more general class of distributions, namely α-stable (called also stable), which belong to the family of the so-called heavy-tailed distributions. The α-stable distribution is flexible for data modelling and includes Gaussian distribution as a special case. The importance of this class of distributions is strongly supported by the limit theorems which indicate that the stable distribution is the only possible limiting distribution for the normed sum of independent and identically distributed random variables. The interesting applications of a class of α-stable distributions can be found, for instance, in [9, 18, 28]. See also [19, 20].

We start by introducing a one-dimensional α-stable random variable. A random variable S has one-dimensional stable distribution if, for any numbers $A, B > 0$, there exist numbers $C > 0$ and $D \in R$ such that

$$AS_1 + BS_2 \stackrel{d}{=} CS + D, \tag{7}$$

where S_1 and S_2 are two independent copies of S [25]. For each stable random variable S there exists a number $\alpha \in (0, 2]$ such that the constant C in Eq. (7) satisfies the following relation:

$$C^\alpha = A^\alpha + B^\alpha.$$

The second equivalent definition is based on the characteristic function of random variable S.

The random variable S has α-stable distribution if there exist parameters $\alpha \in (0, 2]$, $\sigma > 0$, $-1 \leq \beta \leq 1$ and $\mu \in R$ such that the characteristic function of S takes the form:

$$\mathbb{E}e^{ixS} = \begin{cases} \exp\{-\sigma^\alpha |x|^\alpha (1 - i\beta sign(x) tan(\pi\alpha/2)) + i\mu x\} & \text{for } \alpha \neq 1 \\ \exp\{-\sigma|x|(1 + i\beta(2/\pi) sign(x) \log(|x|)) + i\mu x\} & \text{for } \alpha = 1. \end{cases} \tag{8}$$

Stability index α, scale parameter σ, skewness parameter β and shift parameter μ in a unique way define the distribution of a random variable S. In further analysis, we denote $S \sim S(\alpha, \sigma, \beta, \mu)$. The probability density function for most α-stable distributions has no explicit form. However, there are three exceptions, namely Gaussian (for $\alpha = 2$), Cauchy (for $\alpha = 1$ and $\beta = 1$) and Lévy (for $\alpha = 0.5$ and $\beta = 1$).

The extension of a univariate stable distribution is the multivariate one. A characteristic function of an m-dimensional random vector $\mathbf{S} = (S^1, \ldots, S^m)$ is defined as follows:

$$\Phi(\mathbf{x}) = Ee^{i\sum_{k=1}^m x_k S_k}.$$

The m-dimensional vector $\mathbf{S}$ is α-stable in R^m if and only if there exists a finite-dimensional measure G on the unit sphere S_m and the vector μ such that:

$$\Phi(\mathbf{x}) = e^{-I(\mathbf{x})+i\sum_{k=1}^{m} x_k \mu^k},$$

where $I(\mathbf{x}) = \int_{S_m} \psi(\sum_{k=1}^{m} s_i x_i) G(ds_1, \ldots, ds_m)$ and

$$\psi(u) = \begin{cases} |u|^{\alpha}\,(1 - isign(u)tan(\pi\alpha/2))\,, & \text{for } \alpha \neq 1 \\ |u|\left(1 + i\frac{2}{\pi}sign(u)\log(|u|)\right), & \text{otherwise} \end{cases}$$

Moreover, (G, μ) is unique. It is worth mentioning that the stability index α, measure G (called spectral measure), and the shift vector μ in a unique way define the m-dimensional stable distribution denoted as $S^m(\alpha, G, \mu)$. If $\mathbf{S}$ has m-dimensional stable distribution, then each component has one-dimensional stable distribution with appropriate parameters.

The multidimensional AR model with α-stable structure is the system defined in (1) for which the residual series $\{\varepsilon_t\}$ comes from an m-dimensional α-stable distribution.

4 Real Data Analysis

4.1 Model 1

In the first proposed model, we use the relationship between prices corresponding to 11 nodes. In the first step of our analysis we remove the seasonal behaviour from the original data sets, which is related to the annual seasonality observable in electricity prices. In order to do this, we fit to all 11 vectors the sum of sinusoidal functions by using the least squares method. After removing the sinusoidal functions, we differentiate the data. For each vector a different function is fitted (we choose the best fitting) but results after the transformations are similar. This is especially observable in the autocorrelation functions where only weekly seasonality is visible. As an example, we present the vector $N1$ after mentioned transformation (see Fig. 8 top panel) and its autocorrelation function (see Fig. 8 bottom panel).

Because the autoregressive models can be used only on stationary series, before we fit the VAR model, we have to remove weekly seasonality from the data. The simplest method is to remove the seasonal mean which is calculated by using the data corresponding to each season. More precisely, from the analysed data set $X_1, \ldots, X_{nT}$ (T is the observed season) we remove function w which is calculated as follows:

$$w(t) = \frac{1}{n}\sum_{k=0}^{n-1} X_{kT+t},\ t = 0, 1, 2, \ldots \tag{9}$$

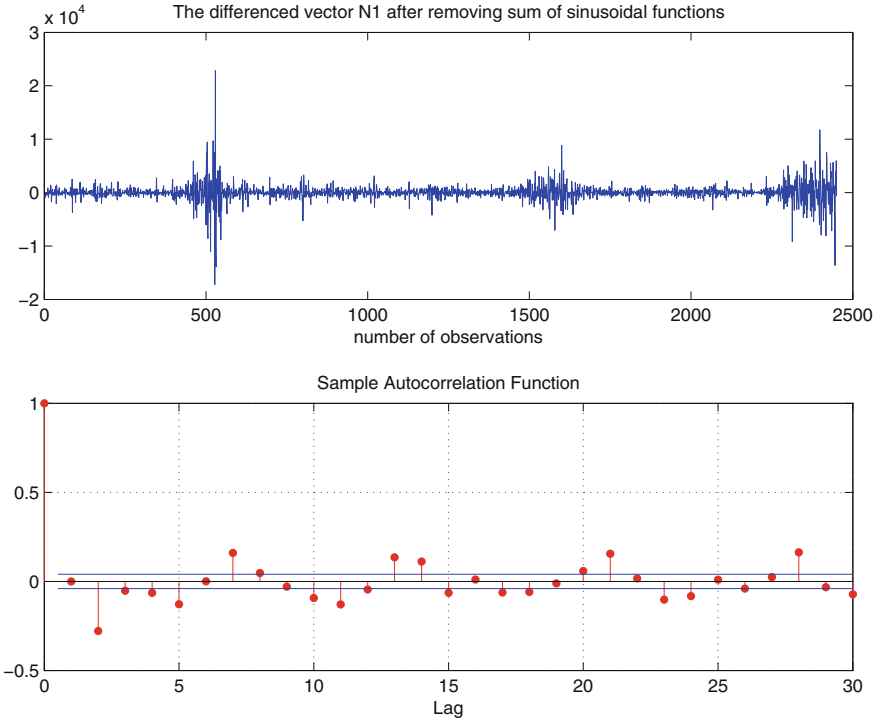


Fig. 8 The exemplary differenced vector $N1$ after removing the sum of sinusoidal functions (*top panel*) and its autocorrelation function (*bottom panel*). The ACF indicates the existence of weekly seasonality in the data set

After removing the seasonal behaviour from the original data sets, we can fit the best VAR model. In order to find the best order p we use the Schwartz-Bayesian criterion (BIC) which, for the vector $X^1, \ldots, X^m$ (each of n observations), is defined as follows:

$$BIC(p) = \log(|\Sigma(p)|) + \frac{\log(m)}{m} pn^2, \tag{10}$$

where $\Sigma(p) = \frac{1}{m} \sum_{t=1}^{m} \varepsilon_t \varepsilon_t'$ is the residual covariance matrix from a VAR(p) model. The general approach is to fit a VAR(p) model with order $p = 0, \ldots, p_{max}$ and choose the value of p which minimizes the selection criterion. In Fig. 9, we present the BIC statistic for order p taking values between 1 and 15. The plot clearly indicates that the best model is VAR(4).

After selecting an appropriate order p, we can estimate the parameters. As mentioned earlier, here we use the Whittle algorithm which is based on the method of moments [4]. The model contains $11 \times 11 \times 4 = 5324$ parameters, therefore, we do not present the estimated values. In the next step of the analysis, we examine

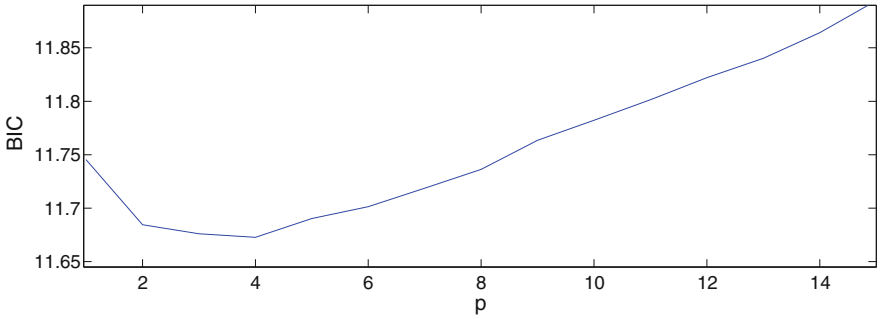


Fig. 9 The BIC statistic for selection of the best model that takes into consideration electricity prices relation. For considered data the best model is VAR(4)

the residual series. Since we consider the multidimensional model, the series is also multidimensional. We test the distribution of the residual series for each vector separately. For all vectors the residuals exhibit non-Gaussian behaviour. This can be proved by using the Jacque-Bera test for normality. The statistic of JB test for the vector $\varepsilon_1, \ldots, \varepsilon_n$ is defined as follows [6, 12]:

$$JB = \frac{n}{6}\left(S^2 + \frac{K-3)^2}{4}\right), \tag{11}$$

where S and K are the sample skewness and kurtosis, respectively, namely:

$$S = \frac{1/n \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})^3}{\left(\sqrt{1/n \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})}\right)^2},$$

$$K = \frac{1/n \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})^4}{\left(\sqrt{1/n \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})}\right)^2}.$$

The value of the JB statistic given by (11) forms a random variable which converges to zero if the underlying distribution has skewness zero and kurtosis 3 (e.g., Gaussian). Any deviation of skewness from zero and deviation of kurtosis from 3 increases the JB statistic. For distributions with infinite kurtosis (like α-stable with $\alpha < 2$) it diverges to infinity. The test is quite standard and implemented in various numerical packages, for example, R or MATLAB. Similar as for all statistical tests, the calculated p-value indicates whether the zero hypothesis can be accepted at the given significance level. If p-value is small, the hypothesis (here of Gaussian distribution) should be rejected. For the analysed residuals from VAR(4) model the obtained p-values are at the level of 0.001 which indicates non-Gaussian distribution. In order to show the analysed residual series comes from an α-stable distribution, we use two goodness-of-fit-test statistics, namely Kolmogorov-Smirnov and Anderson-Darling.

The most well-known supremum statistic is the Kolmogorov-Smirnov (KS) statistic. It is just the supremum of the set of distances:

$$KS = sup_x \left| ECDF(x) - F(x) \right|, \tag{12}$$

where F is the cumulative distribution function of the testing distribution and $ECDF(x)$ is the empirical cumulative distribution function, which for the series $\varepsilon_1, \ldots, \varepsilon_n$ is calculated as follows:

$$ECDF(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\varepsilon_i \leq x\} \tag{13}$$

In the above definition $\mathbf{1}\{A\}$ denotes the indicator of a set A.

The Anderson-Darling statistic belongs to the Cramer-von-Misses family of statistics, which incorporate the idea of quadratic norm. The Cramer-von-Misses statistic for vector $\varepsilon_1, \ldots, \varepsilon_n$ is defined by [6]

$$Q = n \int_{-\infty}^{\infty} \left(ECDF\,(x) - F\,(x) \right)^2 \phi\,(x)\ \mathrm{d}x \tag{14}$$

where $\phi\,(x)$ is a suitable function which puts weights to the squared difference $(ECDF\,(x) - F\,(x))^2$. When $\phi(x) = 1$, Q is called the Cramer-von-Misses statistic. If $\phi\,(x) = [F\,(x)\,(1 - F\,(x))]^{-1}$, the above definition yields the Anderson-Darling (AD) statistic. Similar as for the JB test, there exists a statistical test that allows to test the proper distribution of examined data. More details can be found in [1, 2, 7].

In Table 1 we present the values of the KS and AD statistics and corresponding p-values of tests for α-stable distribution for the 11 analysed data sets. For the α-stable distribution the cumulative distribution function is not given explicitly,

Table 1 The values of KS and AD statistics and corresponding p-values of goodness-of-fit tests

Vector	KS	p-value (KS)	AD	p-value (AD)
$N1$	0.55	0.60	0.29	0.68
$N2$	0.43	0.95	0.16	0.95
$N3$	0.54	0.60	0.61	0.48
$N4$	0.72	0.27	0.46	0.54
$N5$	0.52	0.65	0.23	0.86
$N6$	0.37	0.97	0.18	0.89
$N7$	0.53	0.67	0.35	0.75
$N8$	0.52	0.63	0.24	0.73
$N9$	0.74	0.18	0.61	0.40
$N10$	0.44	0.88	0.29	0.74
$N11$	0.72	0.23	0.54	0.48

thus to obtain the corresponding p-values we use the Monte Carlo method with 1000 repetitions. As we observe, the obtained p-values significantly exceed the significance level 0.05, therefore we cannot reject the hypothesis of α-stable distribution of residual series.

Next we estimate the parameters of α-stable distribution for all residual series. We use here the regression method [15, 16]. The method is based on the form of the characteristic function of an α-stable random variable $S(\alpha, \sigma, \beta, \mu)$ given in (8). For a random sample $\varepsilon_1, \ldots, \varepsilon_n$ from the α-stable distribution, we calculate the empirical characteristic function:

$$ECHF(x) = \frac{1}{n} \sum_{j=1}^{n} e^{ix\varepsilon_j}. \tag{15}$$

In the regression method, the empirical characteristic function is compared to the theoretical one and by using least squares method one can estimate the parameters. For details see [3]. In Table 2, we present the estimated parameters for all 11 vectors of residuals. As we observe, the estimated α parameters of the residual series are on the level $1.3 - 1.4$, so tails of distributions of data related to the considered nodes are very similar. Moreover, the σ parameter in all considered cases is approximately 500, so the scaling in all cases is the same. The β parameters indicate the data are symmetric (β close to zero). Only the shift parameter μ indicates differences between the distribution of residuals.

By using the fitted model, we can simulate the VAR(4) system with α-stable residuals. In Fig. 10 we present the exemplary vector $N1$ together with the quantile lines on the $10, \ldots, 90\,\%$ levels from the simulated samples. The quantile lines are constructed using Monte Carlo methods with 1000 repetitions. Note that the observed price does not exceed the bounds given by the 10 and 90 % quantile lines.

Table 2 The estimated parameters of α-stable distribution for the residual series from Model 1

Vector	α	σ	β	μ
$N1$	1.39	463.73	−0.02	−9.42
$N2$	1.42	483.86	−0.04	−9.23
$N3$	1.28	523.45	0.03	21.50
$N4$	1.28	536.31	0.04	17.81
$N5$	1.40	488.23	−0.03	−12.53
$N6$	1.38	513.43	−0.04	−21.46
$N7$	1.28	559.92	0.02	−3.09
$N8$	1.38	540.24	−0.03	−13.56
$N9$	1.27	513.06	0.05	37.71
$N10$	1.28	553.97	−0.03	−42.38
$N11$	1.28	530.19	0.06	36.20

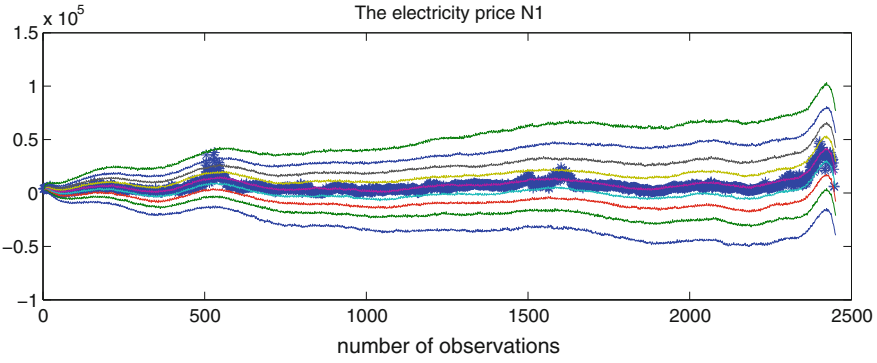


Fig. 10 Quantile lines of level 10, 20, . . . , 90 % and the measured price $N1$ (*thick blue line*)

4.2 Model 2

In the second approach we propose to use the relationship between the electricity price corresponding to node 1 and other variables such as rainfall, hydro storage, electricity demand and temperature. Let us recall that the rainfall and temperature were measured for the region related to node 1, whereas the storage and demand have been aggregated for the entire country (that is the original format in which the data have been provided by Electricity Commission).

Similar to the first approach, in the first step of the analysis we remove the non-stationarity of the data by fitting the sum of sinusoidal functions for all considered vectors of observations. Only one exception is the rainfall where the non-stationarity is not visible, see Fig. 6. Next, we differentiate the data after removing the fitted functions. In the second step, for the data with weekly seasonal behaviour (price and electricity demand), we calculate the periodic mean and remove it. The formula for periodic mean is presented in (9).

Next, a proper order p should be calculated. Similar to the previous approach, here we use the BIC criterion defined in (10). In Fig. 11 we present the plot of BIC statistic for order p varied between 1 and 15.

After the proper model has been fitted, we can analyse the residuals. First, we check if they constitute samples from Gaussian distribution. Similar to Model 1, we use the JB statistic defined in (11) and test based on it. In Table 3 we present the p-values of the test for gaussianity. Let us recall, the small p-value (less than significance level 0.05) indicates non-Gaussian behaviour of underlying random sample. As we observe, only the residuals related to temperature can be treated as a sample from Gaussian distribution. The estimated parameters of Gaussian distribution for temperature residuals are $\mu = 0.0034$ and $\sigma = 2.45$.

Next, we check if the residuals related to price, rainfall, hydro storage and demand constitute samples from α-stable distribution. In order to do this, we use the KS and AD goodness-of-fit tests presented above. In Table 4 we present the values of the

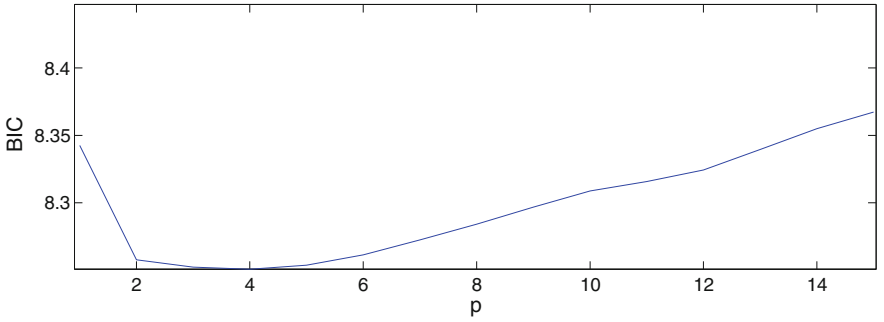


Fig. 11 The BIC statistic for selection of the best model that takes into consideration the relationship between electricity price $N1$ and rainfall, hydro storage, electricity demand and temperature. For considered data the best model is VAR(4)

Table 3 The p-values of the JB test for gaussianity

Price	Rainfall	Hydro storage	Demand	Temperature
0.001	0.001	0.001	0.001	0.1786

Table 4 The values of KS and AD statistics and corresponding p-values of goodness-of-fit tests

Vector	KS	p-value (KS)	AD	p-value (AD)
Price	0.64	0.39	0.31	0.72
Rainfall	49.37	0	01.26	0.56
Hydro storage	49.55	0	1.71	0.82
Demand	1.02	0	2.5	0.13

statistics and the corresponding p-values calculated on the basis of 1000 Monte Carlo simulations.

As we observe, the KS test indicates that only the price comes from an α-stable distribution. This can be seen from the prominent values of the statistic. But the AD test indicates the residuals of mentioned in Table 4 variables can be treated as from the α-stable family. Therefore, we estimate the parameters of this distribution. In Table 5, we present the estimated parameters by using the regression method. Here the situation is different from that in Model 1. There is a difference in α parameters so the tail behavior is different for the considered variables. The other parameters are also different for different variables.

By using the fitted model we can simulate the VAR(4) system with residuals which come from the fitted distributions. In Fig. 12 we present the price vector $N1$ together with the quantile lines on the level $10, \ldots, 90\,\%$ from the simulated samples. The quantile lines are constructed using Monte Carlo methods with 1000 repetitions. Note that the observed price does not exceed the bounds given by the 10 and 90 % quantile lines.

Table 5 The estimated parameters of α-stable distribution for the residual series from Model 2

Vector	α	σ	β	μ
Price	1.4	457.49	0.07	26.60
Rainfall	1.10	1.11	−0.18	108.96
Hydro storage	1.19	7.46	0.56	751.87
Demand	1.58	1.390	0.19	237.75

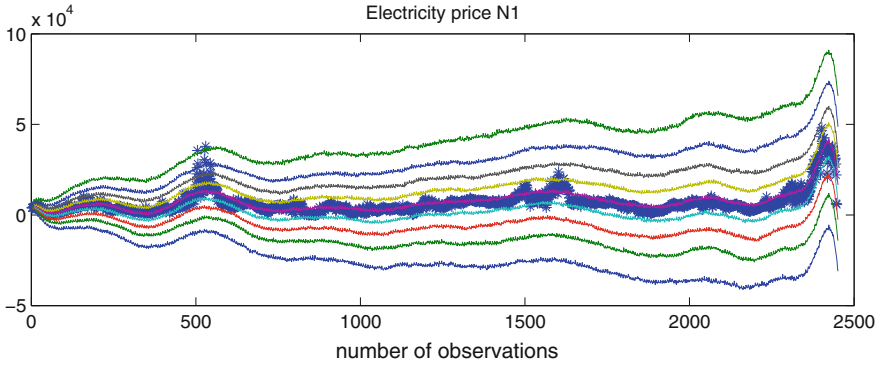


Fig. 12 Quantile lines of level 10, 20, . . . , 90 % and the electricity price $N1$ (*blue thick line*)

Finally, in order to illustrate how beneficial the obtained models might be for the problem of electricity price description, we calculate the price prediction for the next year. The obtained predicted values are then validated by comparing them to the actual values. To this end, we fit the model to the first 2091 observations and then based on the obtained estimates we calculate the prediction for the next 360. The obtained values are plotted in Fig. 13. As can be observed in the figure, the predicted values visually resemble the actual values for both considered models.

In order to check which model gives better prediction results, we calculate three measures: mean squared error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) of the forecast for the next year. The measures are defined as:

$$MSE = \frac{1}{k}\sum_{t=1}^{k}(X_t - \hat{X}_t)^2,\ MAE = \frac{1}{k}\sum_{t=1}^{k}|X_t - \hat{X}_t|,\ MAPE = \frac{1}{k}\sum_{t=1}^{k}\frac{|X_t - \hat{X}_t|}{X_t},$$

where k is the number of predicted values, X_t is the measured observation at point t and $\hat{X}_t$ its prediction. The obtained values of mentioned measures are given in Table 6.

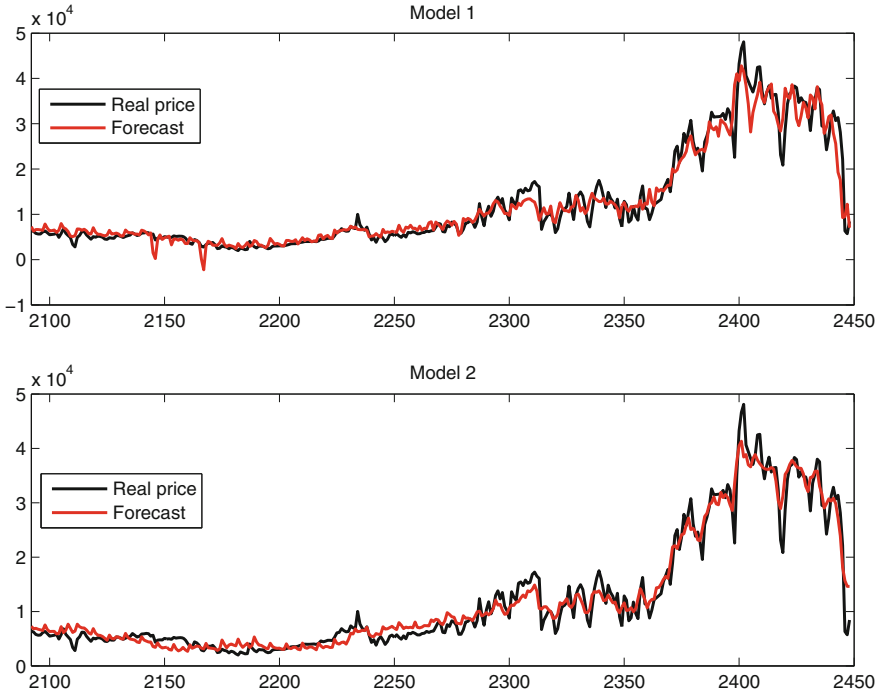


Fig. 13 Values of electricity prices $N1$ together with the prediction for the next year. The predicted values were calculated using Model 1 (*top panel*) and Model 2 (*bottom panel*)

5 Conclusions

In this work two multivariate autoregressive models with non-Gaussian noise structure for forecasting day-ahead electricity prices in New Zealand have been proposed. The approaches have clear advantages over classical ARMA-GARCH type modelling. First, they do not rely only on the historical values of the price itself, but account for other data series as well. In one of the models we have referred one of the nodal prices to ten other nodes. This is explained by the fact that prices throughout the country are interconnected and strongly correlated. The choice of nodes (11 out of over 200) has been carefully designed to pick representative nodes, but not to make the model too big.

Table 6 Mean square error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) of the prediction for the next year

	MSE	MAE	MAPE (%)
Model 1	7.48×10^6	1.73×10^3	17
Model 2	4.99×10^6	1.59×10^3	19

The second approach modelled one of the nodal prices with the use of the regressed price itself together with some deterministic factors having significant influence on price dynamics. These were country's hydrological storage, centralized demand, node's rainfall and air temperature from the studied region. This model proved better for forecasting due to two main reasons. First, the forecasting skill was higher than for Model 1. Secondly, the model had less parameters than the first one, which made it more appropriate from modelling and parameter estimation point of view. Finally, the second model has better results in comparison to Model 1 showing that incorporation of deterministic factors in price modelling is of key importance for high forecasting accuracy.

The robustness of our results was confirmed by the fact that both applied models were optimal with the same order, namely VAR(4). Moreover, the models had non-Gaussian noise included which allowed us to capture statically the non-Gaussian distribution of prices themselves. One could argue that more trading nodes should be included in the analysis. However, we argue that this would unnecessarily complicate parameter estimation without much added value.

Our work provides a sound basis for risk analysis for electricity market traders in New Zealand. Further steps should include the use of Markov Chain Monte Carlo Methods in parameter estimation, to allow proper understanding of price noise structure.

References

1. Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain 'goodness-of-fit' criteria based on stochastic processes. *Annals of Mathematical Statistics, 23*, 193–212.
2. Anderson, T. W., & Darling, D. A. (1954). A test of goodness-of-fit. *Journal of the American Statistical Association, 49,* 765–769.
3. Borak, S., Haerdle, W., & Weron, R. (2005). *Stable distributions in statistical tools for finance and insurance*. Berlin: Springer.
4. Brockwell, P. J., & Davis, R. A. (1996). *Introduction to time series and forecasting*. New York: Springer.
5. Burger, M., Klar, B., Müller, A., & Schindlmayr, G. (2004). A spot market model for pricing derivatives in electricity markets. *Quantitative Finance, 4,* 109–122.
6. Burnecki, K., Janczura, J., & Weron, R. (2011). Building loss models. In P., Cizek, W. H., Hardle, & R. Weron (Eds.), *Statistical tools for finance and insurance*. Berlin: Springer.
7. Burnecki, K., Wyłomańska, A., Beletskii, A., Gonchar, V., & Chechkin, A. (2012). Recognition of stable distribution with Levy index alpha close to 2. *Physical Review E, 85*, 056711.
8. du Preeza, J., & Witt, S. F. (2003). Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting, 19,* 435–451.
9. Fichea, A., Cexusa, J.-C., Martin, A., & Khenchaf, A. (2013). Features modeling with an α-stable distribution: Application to pattern recognition based on continuous belief functions. *Information Fusion, 14,* 504–520.
10. Guthriea, G., & Videbeck, S. (2007). Electricity spot price dynamics: Beyond financial models. *Energy Policy, 35,* 5614–5621.
11. Jabłońska, M., Nampala, H., & Kauranne, T. (2011). Multiple mean reversion jump diffusion model for Nordic electricity spot prices. *The Journal of Energy Markets, 4,* 2.

12. Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters, 6*(3), 255–259.
13. Janczura, J., & Weron, A. (2010). An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy Economics, 32,* 1059–1073.
14. Kian, A., and Keyhani, A. (2001) Stochastic price modeling of electricity in deregulated energy markets. In: *IEEE Proceedings of the 34th Hawaii International Conference on System Sciences*.
15. Kogon, S. M., & Williams, D. B. (1998). In R., Adler, R., Feldman, & M. Taqqu (Eds.), *A Practical Guide to Heavy Tails*. Boston: Birkhauser.
16. Koutrouvelis, I. A. (1980). Regression-type Estimation of the Parameters of Stable Laws. *Journal of the American Statistical Association, 75,* 918.
17. Laitinen, K., Hovila, J., Mannila, T., & Korpinen, L. (2000). The influences of climatic factors on electricity prices in liberalized market in Finland. In: *International Conference on Electric Utility Deregulation and Restructuring and Power Technologies,* pp. 544–548.
18. Mittnik, S., & Ratchev, S. T. (2000). *Stable paretian models in finance*. New York: Wiley.
19. Nowicka-Zagrajek, J., & Wyłomańska, A. (2006). The dependence structure for PARMA models with a—Stable innovations. *Acta Physica Polonica B, 37,* 3071–3081.
20. Nowicka-Zagrajek, J., & Wyłomańska, A. (2008). Measures of dependence for stable AR(1) models with time-varying coefficients. *Stochastic Models, 24,* 58–70.
21. Podraza, E. (2006). Challenges in forecasting electric load in deregulated markets. *The Journal of Business Forecasting, 25,* 31–35.
22. Qureshi, W. A., Nair, N.-K. C., & Farid, M. M. (2009). Demand-price analysis of electricity market using Discrete Wavelet Transform, Power Engineering Conference, AUPEC 2009. Australasian Universities.
23. Ranam, H., & Sunilkumar, N. (1995). Multivariate modeling of water resources time series using artificial neural networks, *40*, 145–163.
24. Ruibal, C. M., & Mazumdar, M. (2008). Forecasting the mean and the variance of electricity prices in deregulated markets. *IEEE Transactions on Power Systems, 23,* 25–32.
25. Samorodnitsky, G., & Taqqu, M. S. (1994). *Stable non-gaussian random processes*. New York: Chapman and Hall.
26. Situ, R. F., & Nair, N.-K. C. (2007). Applying Wavelet Transform Techniques For New Zealand Electricity Market Volatility Analysis, Power Engineering Conference, AUPEC 2007. Australasian Universities.
27. Stathopoulos, A., & Karlaftis, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies, 11,* 121–135.
28. Stuck, B. W., & Kleiner, B. (1974). A statistical analysis of telephone noise. *Bell System Technical Journal, 53,* 1263–1320.
29. Tiao, G. C., & Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society Series B, 51,* 157–213.
30. Tipping, J. P. & Grant Read, E. & Mcnickle, D. C. (2004). The incorporation of hydro storage into a spot price model for the New Zealand electricity market.
31. Vehviläinen, I., & Pyykkönen, T. (2005). Stochastic factor model for electricity spot price—The case of the Nordic market. *Energy Economics, 27,* 351–367.
32. Vucetic, S., Tomsovic, K., & Obradovic, Z. (2001). Discovering price-load relationships in California's electricity market. *IEEE Transactions on Power Systems, 2,* 280–286.
33. Whittle, P. (1963). On the fitting of multivariate autoregressions and the approximation of canonical factorization of a spectral density matrix. *Biometrika, 40,* 129–134.
34. Ying, Y., & Flynn, P. C. (2003). Power price in deregulated markets. *IEEE Power Engineering Society General Meeting*, *2*.

Imputation of Missing Observations for Heavy Tailed Cyclostationary Time Series

Christiana Drake, Jacek Leskow and Aldo M. Garay

1 Introduction

Periodically correlated (or cyclostationary) time series have many applications in engineering, environmental sciences, and economics (see [4]). However, in many practical situations data that can be modelled with such time series is incomplete. An algorithm addressing imputation when some observations are missing at random has been developed in [2], where a version of the EM method had been employed. The mentioned algorithm was developed for AR(p) times series with Gaussian errors. In this paper we propose four algorithms for a K-dependent time series that come from the multivariate t-distribution.

We assume that a zero-mean second-order cyclostationary time series $\{Y_t, t \in Z\}$ is observed. This means that if $B_Y(t, \tau) = Cov(Y_t, Y_{t+\tau})$ denotes the autocovariance function of $\{Y_t\}$, then this function is periodic in t for each τ. Our sample size is T and we assume that T is an integer multiple of the period H and also T is an integer

C. Drake
University of Davis, Davis, CA, USA
e-mail: cmdrake@ucdavis.edu

J. Leskow (✉)
Institute of Mathematics, Cracow Technical University, Cracow, Poland
e-mail: jleskow@pk.edu.pl

A.M. Garay
Departamento de Estatística, Universidade Estadual de Campinas, Campinas, Brazil
e-mail: medina_garay@yahoo.com

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods - II*,
Applied Condition Monitoring 3, DOI 10.1007/978-3-319-16330-7_9

multiple of K, the size of the dependence. We assume that the period H is known. Following [2], the model is defined as

$$Y_t = X_t \cdot c_t, \tag{1}$$

where we make the following assumptions:

(AS1) The deterministic sequence $c_t > 0$, $\forall t \in \{1, \ldots, T\}$ is periodic with period H.

(AS2) We assume that the stationary time series $\{X_t; 1 \leq t \leq T\}$ comes from multivariate t-distribution $t_{K+\nu}(\mathbf{0}, \Sigma, \nu)$. In the sequel, we will assume that $V(X_t) = \frac{\nu}{\nu-2}$. Therefore, the matrix Σ can be considered as a correlation matrix for $\{X_t; 1 \leq t \leq T\}$.

The assumption of K-dependence implies the following factorization of Σ:

$$\Sigma = \begin{pmatrix} \mathbf{A} & \mathbf{C} & \mathbf{0} & \ldots\ldots & \mathbf{0} \\ \mathbf{B} & \mathbf{A} & \mathbf{C} & \ldots\ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{A} & \mathbf{C} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \ldots\ldots & \mathbf{B} & \mathbf{A} & \mathbf{C} \\ \mathbf{0} & \ldots\ldots & \mathbf{0} & \mathbf{B} & \mathbf{A} \end{pmatrix} \tag{2}$$

where

$$\mathbf{A} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \ldots\ldots & \rho_{K-1} \\ \rho_1 & 1 & \rho_1 & \ldots\ldots & \rho_{K-2} \\ \rho_2 & \rho_1 & 1 & \rho_1 & \ldots & \rho_{K-3} \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \rho_{K-2} & \ddots & \ldots & \rho_1 & 1 & \rho_1 \\ \rho_{K-1} & \ldots\ldots & \rho_2 & \rho_1 & 1 \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} 0 & \rho_{K-1} & \rho_{K-2} & \ldots & \rho_1 \\ 0 & 0 & \rho_{K-1} & \ldots & \rho_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \rho_{K-1} \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix}$$

and

$$\mathbf{C} = \mathbf{B}^\top.$$

Assumption AS1 assures that the distribution of Y_t is not degenerate at zero.

In our study, we assume that the times series X_t (thus Y_t) is K dependent. This means that X_t, X_{t+s} are dependent only when $-(K-1) \leq s \leq (K-1)$.

The periodic behavior of the observed process Y_t is entirely determined by the c_t and the correlation structure of X_t. We denote the period by H. The dimension of the parameter vector $(c_1, \ldots, c_H)$ is H and the variance of Y_t is given by

$$V(Y_t) = c_t^2 \times V(X_t) = c_t^2 \times \frac{\nu}{\nu - 2} \quad \text{for } t = 1, \ldots, H. \tag{3}$$

We will need to estimate $c_1, \ldots, c_H$, and the parameters of the correlation matrix $\rho_1, \ldots, \rho_{K-1}$ for the K-dependent process X_t. Let $\theta = (\rho_1, \ldots \rho_{K-1}, c_1, \ldots c_H)^T$ be the vector of all unknown parameters and suppose that we have T observations from this model. We will make use of the structure of our process in developing an iterative process for estimation of the model parameters.

We assume that K is much smaller than H. Furthermore, we will also assume that the length of the time series $T = L \times H$ is a multiple of the period H and therefore is much larger than K.

Note that the correlation matrix of the X_t process, which is unobserved, equals the correlation matrix of the Y_t process, which is observed.

2 Estimation of the Model Parameters

Assume that the period H is known. Then $c_t = c_{t+H} = \cdots = c_{t+lH}$ for $t = 1, \ldots, H$ and $l = 1, \ldots, L-1$. The parameters $c_1, \ldots, c_H$ determine the variance of the process Y_t. Since we assume that X_t is K-dependent X_t, X_u are independent if $|t - u| \geq K$ for each t and u. Therefore, the sequence $Y_t, Y_{t+H}, Y_{t+2H}, \ldots, Y_{t+(L-1)H}$ consists of independent and identically distributed random variables for $t = 1, \ldots, H$. These sequences provide us with a simple method to estimate c_t. Furthermore, when data are missing completely at random we can use the remaining data which constitute a random subsample of the complete data to estimate c_t without bias. We assume ν is fixed and known ($2 < \nu \leq 6$). Its value determines the kurtosis of the X_t. We will treat this parameter as known and investigate the dependence of our estimates on the value we choose for ν.

We also need to determine K. For this we will use the multivariate asymptotic normality of the sample autocovariance function [7]. We propose to proceed as follows:

1. Extract the sequence of iid random variables $\mathbf{Y}_h = (Y_h, Y_{h+H}, \ldots, Y_{h+LH})$ for $h = 1, \ldots, H$.
2. Estimate the values of $\hat{c}_h$ for $h = 1, \ldots, H$.
3. Calculate $\tilde{\mathbf{X}}_h = (\tilde{X}_h, \tilde{X}_{h+H}, \ldots, \tilde{X}_{h+LH}) = \left(\frac{Y_h}{\hat{c}_h}, \frac{Y_{h+H}}{\hat{c}_h}, \ldots, \frac{Y_{h+LH}}{\hat{c}_h}\right)$.

4. Compute the sample correlation function from the data

$$\hat{\rho}_\tau(T) = T^{-1} \sum_{t=1}^{T-\tau} \left(\tilde{X}_{t+\tau} - \bar{\tilde{X}}\right)\left(\tilde{X}_t - \bar{\tilde{X}}\right).$$

5. Based on the asymptotic normality of $\hat{\rho}_\tau(T)$ for $\tau_1, \ldots, \tau_H$ we will determine $K \ll H$ to establish the value K of the dependence structure. For details, see Shumway and Stoffer [7].
6. After the value of K has been established we will estimate $\hat{R}$, the correlation matrix by methods of moments. For this we will use the correlations $\hat{\rho}_\tau(T)$ for $\tau = 0, \ldots, K-1$ calculated in Step 4.
7. Once we have determined the correlation structure and estimated the correlation matrix we will perform multiple imputation to fill in any gaps in our data as follows. For example if Y_t is missing we will impute this value based on

$$E[X_t \mid X_{t-K+1}, \ldots, X_{t-1}, X_{t+1}, \ldots X_{t+K-1}],$$
$$E[X_t^2 \mid X_{t-K+1}, \ldots, X_{t-1}, X_{t+1}, \ldots X_{t+K-1}],$$

and obtain the variance

$$V(X_t \mid X_{t-K+1}, \ldots, X_{t-1}, X_{t+1}, \ldots X_{t+K-1}).$$

3 Simulation Study

In order to study the performance of our method, we present this simulation study for the cyclostationary time series described in Sect. 1. In this study we will benefit from the main properties of the Multivariate Student t distribution developed by Arellano-Valle and Bolfarine [1], Matos et al. [5] and more recently by Garay et al. [3]. Thus, for this purpose, we proceeded as follows:

Step 1: *Data Generation*

- Simulate $T+2$ different dependent processes of order $k = 2p+1$, using moving averages of independent and identically distributed random variables from a Student t distribution, denoted by u_i for $i = 1, \ldots, T+2$.
- Generate the process X_t as follows:

$$X_t = \frac{1}{k} \sum_{i=t-p}^{t+p} u_i, \quad \text{where} \quad u_i \overset{\text{iid}}{\sim} t\,(0, 1; \nu) \quad \text{and} \quad t = 1, \ldots, T.$$

- We furthermore generate a set of periodic functions c_t with period $H \gg K$ as follows:

$$c_j = \sin\left(\frac{2\pi j}{H}\right) + 1.1, \quad \text{for} \quad j = 1, \ldots, H.$$

The constant 1.1 is added to obtain $c_j > 0$.

- Generate the cyclostationary time series y_t, as defined in (1), for $t = 1, \ldots, T$.

Step 2: *Estimation of* c_j

- The estimator $\widehat{c}_j$ of c_j is obtained by

$$\widehat{c}_j = S_j\sqrt{3}\left(\frac{\nu}{\nu-2}\right)^{-\frac{1}{2}}, \quad \text{for} \quad j = 1, \ldots, H. \tag{4}$$

where

$$S_j = \frac{1}{H-1}\sqrt{\left(\sum[y^{\tilde{h}} - \bar{y}^{\tilde{h}}]^2\right)} \text{ and } \overline{y}_{j.} = \frac{1}{H}\sum_{i=1}^{H} y_{(i-1)H+j}.$$

The estimator (4) is obtained by the method of moments.

Step 3: *Imputation of* Y_t

- After the data have been generated (see step 1), we will estimate the parameters c_j, as defined before, from the complete data. Then we proceed to generate a times series Y_t with the number I of missing observations. We will assume that observations are missing completely at random MCAR [6].
- We know that every linear combination of independent univariate Student t distribution is again Student t distribution, i.e., the t-distribution is closed under addition. We will also use the marginal-conditional decomposition of a Student-t random vector developed in the Proposition in the Appendix. We propose the estimation and four schemes of imputation of process. The impute values will be denoted by $\widetilde{Y}_i$, for $i = 1, \ldots, I$

– *Backward*

$$\widetilde{Y}_{Fi} = E\left[Y_i | Y_{i+1}, Y_{i+2}, \ldots, min(Y_{i+2(k-1)}, Y_T)\right],$$

where $min(a, b)$ represents the minimum value between a and b.

– *Forward*

$$\widetilde{Y}_{Bi} = E\left[Y_i | Y_{i-1}, Y_{i-2}, \ldots, max(Y_{i-2(k-1)}, Y_1)\right],$$

where $max(a, b)$ represents the maximum value between a and b.

– *Average*

$$\widetilde{Y}_{Mi} = \frac{1}{2}\left(\widetilde{Y}_{Fi} + \widetilde{Y}_{Bi}\right).$$

– *Generalized* In this case $\widetilde{Y}_{Gi}$ is obtained via sampling

$$\pi(Y_i|Y_{i-1}, Y_{i-2}, \ldots, max(Y_{i-2(k-1)}, Y_1), Y_{i+1}, Y_{i+2}, \ldots, min(Y_{i+2(k-1)})),$$

which is a Multivariate-t Distribution.

Step 4: *Evaluation*

- In order to compare the performance of four different strategies defined before, we use two empirical discrepancy measures called the mean absolute error (MAE) and mean square error (MSE), see Wang et al. [9], Wang [8] and Garay et al. [3]. They are defined as

$$\text{MAE} = \frac{1}{I}\sum_{i=1}^{I}\left|Y_i - \widetilde{Y}_i\right| \quad \text{and} \quad \text{MSE} = \frac{1}{I}\sum_{i=1}^{I}\left(Y_i - \widetilde{Y}_i\right)^2, \tag{5}$$

where Y_i is the original simulated value and $\widetilde{Y}_i$ is the impute value, for $i = 1, \ldots, I$.

Step 5: *Results of simulation study*.

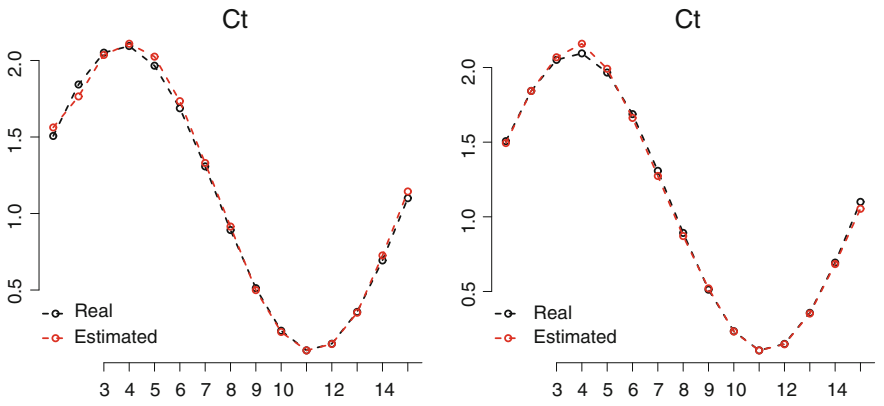


Fig. 1 *Simulated data*. Comparison between the real and estimated values of c_t for two different samples

Table 1 *Simulated data*

Schemes of imputation process				
Imput %	Backward		Forward	
	MAE	MSE	MAE	MSE
1	0.42962	0.43228	0.34861	0.41287
3	0.45266	0.45154	0.37053	0.42639
5	0.47523	0.47637	0.39064	0.44320
10	0.53847	0.53928	0.44673	0.47981
Imput %	Average		Generalized	
	MAE	MSE	MAE	MSE
1	0.50903	0.48155	0.30626	0.41415
3	0.53261	0.53681	0.33438	0.44637
5	0.60392	0.61190	0.37528	0.48658
10	0.82201	0.81727	0.49780	0.57421

Arithmetics means of the MAE and MSE over $M = 100$ datasets

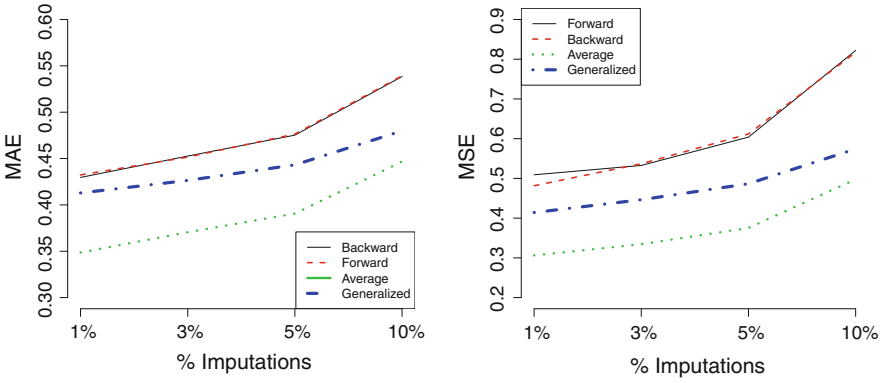


Fig. 2 *Simulated data*. Arithmetic means of the MAE and MSE over $M = 100$ datasets

We generated $M = 100$ different dependent processes, each of them with length $T = 15{,}000$ and order of dependence $k = 3$ and $H = 15$. Each process is considered with proportions of imputation, $\gamma = 1, 3, 5$ and $10\,\%$. It is important to note that the goal here is to compare the four schemes of imputation process.

Figure 1 shows the comparison between the real and estimated value of the c_j for two random samples, considering the methodology described in Step 2. We observe that these values are very similar.

Arithmetic means of MSE and MAE over the 100 datasets are displayed in Table 1 and Fig. 2. We can see that in all cases, the scheme *average* has the smallest MSE and MAE.

4 Conclusions

In our paper, we have provided the methodology and a working algorithm to perform data imputation for cyclostationary times series. The times series under study was heavy tailed and this property was reflected by choosing the multivariate t-student distribution, with a small number of degree of freedom. Four different imputation algorithms were considered. The algorithm calculating the average of the backward and forward turned out to be the best in the MSE sense.

The future work of the authors will be focused on other heavy tailed models and the corresponding imputation algorithms.

Acknowledgments Jacek Leskow would like to acknowledge the support of the grant by the Polish National Center for Science, grant number UMO-2013/10/M/ST1/00096. Moreover, while working on this paper Jacek Leskow was also supported by the Grant 2014/11831-3 from FAPESP-Brazil. Aldo M. Garay would like to acknowledge the support of the Fundação de Amparo à Pesquisa do Estado de São Paulo (Grant 2014/13994-7 from FAPESP-Brazil).

Appendix

The following Proposition provided by Arellano-Valle and Bolfarine [1] and worked by Garay et al. [3] is used in this paper for implementation of the simulation study and represents the marginal-conditional decomposition of a Student-t random vector.

Proposition *Let* $\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *and* $\mathbf{Y}$ *be partitioned as* $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$, *with* $dim(\mathbf{Y}_1) = p_1$, $dim(\mathbf{Y}_2) = p_2$, $p_1 + p_2 = p$, *and where* $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ *and* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$, *are the corresponding partitions of* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\mu}$. *Then, we have*

(i) $\mathbf{Y}_1 \sim t_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)$*; and*
(ii) the conditional cdf of $\mathbf{Y}_2|\mathbf{Y}_1 = y_1$ *is given by*

$$P(\mathbf{Y}_2 \leq y_2|\mathbf{Y}_1 = y_1) = T_{p_2}\left(y_2|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1\right),$$

where $\widetilde{\boldsymbol{\Sigma}}_{22.1} = \left(\frac{\nu + \delta_1}{\nu + p_1}\right)\boldsymbol{\Sigma}_{22.1}$, $\delta_1 = (y_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^{-1}(y_1 - \boldsymbol{\mu}_1)$, $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\mu_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(y_1 - \boldsymbol{\mu}_1)$, *and* $T_r(.|\ldots)$ *represents a cdf of the Student-t random vector of order r.*

References

1. Arellano-Valle, R., & Bolfarine, H. (1995). On some characterizations of the t-distribution. *Statistics & Probability Letters*, *25*, 79–85.
2. Drake, C., Knapik, O., & Leśkow, J. (2014). EM-based inference for cyclostanioraty times series with missing observations. In Chaari, et al. (Eds.), *cyclostationarity: Theory and Methods*. Lecture Notes in Mechanical Engineering Springer.
3. Garay, A. M., Castro, L. M., Leskow, J., & Lachos, V. H. (2014). Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. *Statistical Methods in Medical Research*. doi:10.1177/0962280214551191.
4. Gardner, W. A., Napolitano, A., & Paura, L. (2006). Cyclostationarity: Half a century of research. *Signal Processing*, *86*, 639–697.
5. Matos, L. A., Prates, M. O., Chen, M. H., & Lachos, V. H. (2013). Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica*, *23*, 1323–1342.
6. Schaffer, J. L. (1997). *Analysis of incomplete data*. Chapman and Hall/CRC.
7. Shumway, R. H., & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications*. New York: Springer—Texts in Statistics.
8. Wang, W.-L. (2013). Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biometrical Journal*, *55*(4), 554–571.
9. Wang, W.-L., & Fan, T.-H. (2010). ECM based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. *Computational Statistics & Data Analysis*, *54*(5), 1328–1341.

The Dependence Structure for Symmetric α-stable CARMA(p,q) Processes

Agnieszka Wylomanska

1 Introduction

In modern mathematical finance, continuous time models play a crucial role because they allow handling unequally spaced data and even high frequency data, which are realistic for many real time series. Probably, the most famous example is the Black-Scholes model, which is build out of Brownian motion and models the logarithm of an asset price by the solution to the arithmetic Brownian motion, see [7]. The asset pricing model implies that the aggregate returns are normal and independly distributed. But the assumption is unsatisfactory for many observed data. One approach is to replace the Brownian motion in Black-Scholes model by a heavier tailed Lévy process. This will allow to model returns by heavy-tailed and skewed distribution and take into account jumps. However, the returns will be independent and stationary, since every Lévy process has stationary independent increments. This approach was proposed by Brockwell and Marquardt in [3], where second-order Lévy-driven CARMA (continuous time ARMA) processes are reviewed. Gaussian CARMA processes are special cases in which the driving Lévy process is Brownian motion. The use of more general Lévy processes permits the specification of CARMA processes with a wide variety of marginal distributions which may be asymmetric and heavier tailed than Gaussian. Non-negative CARMA processes are of special interest, partly because of the introduction by Barndorff-Nielsen and Shepardt [1] of non-negative Ornstein-Uhlenbeck process as model for stochastic volatility [3].

Because many studies have shown that heavy-tailed distributions allow for modelling different kinds of phenomena when the assumption of normality for the observations seems not to be reasonable [9, 24, 35]; we extend the definition of the second-order Lévy-driven CARMA processes considered in [3]. We propose to replace the second-order process with the symmetric α-stable Lévy motion. The role that α-stable Lévy motion plays among stable processes is similar to the role that Brownian motion plays among Gaussian processes. Moreover, the α-stable (stable) distributions have found many practical applications, for instance in finance [18], physics

A. Wylomanska (✉)
Hugo Steinhaus Center, Department of Mathematics,
Wrocław University of Technology, Wrocław, Poland
e-mail: agnieszka.wylomanska@pwr.edu.pl

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods - II*,
Applied Condition Monitoring 3, DOI 10.1007/978-3-319-16330-7_10

[10], electrical engineering [27] and in the vibro-diagnostics [35]. The importance of this class of distributions is strongly supported by the limit theorems which indicate that the stable distribution is the only possible limiting distribution for the normed sum of independent and identically distributed random variables [21].

On one hand, the α-stable CARMA processes are extension of second-order Lévy-driven CARMA models; on the other hand, they are extension of the ARMA time series models with α-stable innovations described in [19]. The mentioned time series models found practical applications, see [34, 35]. The discrete ARMA models with innovations from the stable distributions are a special case of considered in [20] PARMA models with α-stable innovations as well as ARMA models with time-varying coefficients and α-stable innovations presented in [21]. We should mention, the PARMA models are main discrete time series used to description of cyclostationary processes and appear to be useful in applications to vibration time series for machines that operate in time-varying conditions [33] as well as for electricity data [4]. The CARMA processes with periodic coefficients (which are extension of presented CARMA models with fixed coefficients) can be also examined in the context of cyclostationary processes. Therefore, the presented results, can be a starting point to analysis of a continuous time model adequate for cyclostationary processes.

The structure of dependence is one of the most important characteristic of the process. One of the most common function that allows to analyze the dependence inside the examined process is the covariance function. Unfortunately, for the considered stable models, the covariance function is not defined and therefore other measures of dependence have to be used. In this paper, we consider the codifference and the covariation—the most popular measures of dependence defined for symmetric α-stable random variables [26, 31] . We prove the codifference and the covariation are asymptotically proportional with the coefficient of proportionality equal to α. Let us mention, the similar asymptotic behaviour we observe for corresponding discrete models. We also consider the alternative measure defined for infinitely divisible stochastic processes called the Lévy correlation cascade.

The rest of the paper is organized as follows: In Sect. 2, we introduce the general class of continuous-time autoregressive moving average time series (CARMA) driven by symmetric α-stable Lévy process. Next, in Sect. 3, we define the measures of dependence adequate for processes with infinite variance for which the covariance can be a tool describing the structure of dependence. In Sect. 4, we study the asymptotic behaviour of mentioned measures of dependence for symmetric α-stable CARMA processes and indicate the asymptotic behaviour of covariation (CV) and codifference (CD)—the most popular measures adequate for stable processes. The main result, namely the asymptotic behaviour of the ratio CD/CV obtained for examined processes is similar to that obtained for discrete version of CARMA process, namely ARMA one. In Sect. 5, we concentrate on the special case of CARMA process, namely CAR(1), known in the literature as the Ornstein-Uhlenbeck process. We show the explicit form of the solution of such system and formulas for considered measures of dependence. Similarly, as in general case, we indicate the asymptotic behaviour of the ratio CD/CV. In order to illustrate the

theoretical results, in Sect. 6, we examine the real financial dataset by using examined processes. Last section contains conclusions.

2 CARMA Processes

In this paper, we use the definition proposed in [2] that is an extension of second-order Lévy-driven continuous ARMA(p,q) process given in [3] and study the CARMA process with symmetric α-stable Lévy motion indexed by $\mathbb{R}$, i.e. the processes satisfying following equation:

$$a(D)Y(t) = b(D)DL^*(t),\ t \in \mathbb{R}, \tag{1}$$

in which D denotes differentiation with respect to t, $\{L^*(t),\ t \in \mathbb{R}\}$ is a two-sided symmetric α-stable Lévy process defined as:

$$L^*(t) = \begin{cases} L(t), & \text{when } t \geq 0, \\ L(-t), & \text{otherwise} \end{cases} \tag{2}$$

and $\{L(t)\ t \geq 0\}$ is an α-stable Lévy process called also α-stable Lévy motion [32]. Let us mention, the α-stable Lévy motion is the process $\{L(t),\ t \geq 0\}$, with independent stationary increments possessing α-stable distribution, i.e. in general case for each t the random variable $L(t)$ has the stable distribution with index of stability $0 < \alpha \leq 2$, scale parameter $t^{1/\alpha}\sigma > 0$, skewness $-1 \leq \beta \leq 1$ and shift parameter $\mu = 0$, i.e. with the characteristic function defined as:

$$\begin{aligned} \Phi_{L(t)}(k) &= \mathbb{E}\exp(ikL(t)) \\ &= \begin{cases} \exp\left\{-t\sigma^\alpha |k|^\alpha \left(1 - i\beta sgn(k)\tan\frac{\pi\alpha}{2}\right)\right\} & \text{if } \alpha \neq 1, \\ \exp\left\{-t\sigma|k|\left(1 - i\beta\frac{2}{\pi}sgn(k)\ln|k|\right)\right\} & \text{if } \alpha = 1. \end{cases} \end{aligned} \tag{3}$$

The α-stable Lévy motion is called symmetric α-stable if $\beta = 0$. In the further analysis, we consider only the symmetric case with $\sigma = 1$. Moreover, the polynomials in (1) are defined by

$$\begin{aligned} a(z) &= z^p + a_1 z^{p-1} + \cdots + a_p, \\ b(z) &= b_0 + b_1 z + \cdots b_{p-1} z^{p-1}. \end{aligned}$$

In the further analysis, we assume $p > q$ and take the $b_q \neq 0$ and $b_j = 0$ for $q < j < p$. Equation (1) can be written in the equivalent form [2]:

$$Y(t) = \mathbf{b}'\mathbf{X(t)}, \tag{4}$$

where the process $\{\mathbf{X(t)},\ t \in \mathbb{R}\}$ satisfies the following stochastic differential equation with respect to the process $\{L^*(t),\ t \in \mathbb{R}\}$:

$$d\mathbf{X(t)} - A\mathbf{X(t)}dt = \mathbf{e}dL^*(t), \tag{5}$$

for

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \dots & -a_1 \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-2} \\ b_{p-1} \end{bmatrix}. \tag{6}$$

The solution of (5) is given by:

$$\mathbf{X(t)} = e^{A(t-s)}\mathbf{X(s)} + \int_s^t e^{A(t-u)}\mathbf{e}dL^*(u), \tag{7}$$

for $s < t$ and the process $\{Y(t),\ t \in \mathbb{R}\}$ defined as

$$Y(t) = \int_{-\infty}^t g(t-u)dL^*(u), \tag{8}$$

where

$$g(x) = \mathbf{b}^{'} e^{Ax}\mathbf{e}\mathbb{I}_{\{(0,\infty)\}}(x) \tag{9}$$

satisfies Eq. (1). Let us mention, process given in (8) is called moving-average symmtric α-stable process [26].

Lemma 1 *If the following condition*

$$\int_0^\infty |\boldsymbol{b}^{'} e^{Ax}\boldsymbol{e}|^\alpha dx < \infty \tag{10}$$

is satisfied, then the stochastic process $\{Y(t),\ t \in R\}$ *given in (8) is stationary.*

Proof The proof follows directly from the form of the solution of Eq. (1), namely (8) and Example 3.6.2 in [26]. □

Remark 1 By the assumption all eigenvalues of A have negative real parts, condition (10) is satisfied, see [2].

Remark 2 Using Proposition 1.11 in [15] we obtain that there exist $x_0 \geq 0$ and constants $c, C > 0$ and $\tilde{c}, \tilde{C} > 0$ such that

$$|g(x)| \leq Ce^{-cx},\ x \in \mathbb{R}$$

$$|g(x)| \geq \tilde{C}e^{-\tilde{c}x},\ x \geq x_0,$$

where $g(x) : \mathbb{R} \to \mathbb{R}$ is given in (9). Moreover the above conditions give the following limit

$$\lim_{x \to \infty} g(x) = 0.$$

3 Measures of Dependence of α-stable Random Processes

Let X and Y be jointly symmetric α-stable random variables (SαS for short) and let Γ be the spectral measure of the random vector (X, Y) (see for instance [26]). We say that random variables X and Y are jointly SαS if the vector (X, Y) is SαS in $\mathbb{R}^2$. If $\alpha < 2$ then the covariance is not defined and thus other measures of dependence have to be used. The most popular measures are the covariation $CV(X, Y)$ and the codifference $CD(X, Y)$ given in Definitions 1 and 2, respectively.

Definition 1 Let X and Y be jointly SαS. The covariation $CV(X, Y)$ of X on Y defined for $1 < \alpha \leq 2$ is the real number [26]:

$$CV(X, Y) = \int_{S_2} s_1 s_2^{<\alpha-1>} \Gamma(d\mathbf{s}), \tag{11}$$

where Γ is the spectral measure of the random vector (X, Y), $\mathbf{s} = (s_1, s_2)$ and the signed power $z^{<p>}$ is given by $z^{<p>} = |z|^{p-1}z$.

Definition 2 Let X and Y be jointly SαS. The codifference $CD(X, Y)$ of X on Y defined for $0 < \alpha \leq 2$ is defined as [26, 32]:

$$CD(X, Y) = \ln \mathbb{E} \exp\{i(X - Y)\} - \ln \mathbb{E} \exp\{iX\} - \ln \mathbb{E} \exp\{-iY\}. \tag{12}$$

Properties of the considered measures of dependence one can find in [26, 32]. Let us only mention here that, in contrast to the codifference, the covariation is not symmetric in its arguments. Moreover, when $\alpha = 2$ both measures reduce to the covariance, namely

$$Cov(X, Y) = 2CV(X, Y) = CD(X, Y). \tag{13}$$

If $\alpha > 1$, then the covariation induces a norm $||.||_\alpha$ on the linear space S_α of jointly SαS random variables.

Definition 3 The covariation norm of $X \in S_\alpha, \alpha > 1$, is defined as

$$||X||_\alpha = (CV(X, X))^{1/\alpha}. \tag{14}$$

The covariation norm of a SαS random variable X is equal to the scale parameter of this variable, see [26]. For $1 < \alpha \leq 2$ the codifference of SαS random variables X and Y can be rewritten in the form:

$$CD(X,Y) = ||X||_\alpha^\alpha + ||Y||_\alpha^\alpha - ||X-Y||_\alpha^\alpha. \tag{15}$$

We should mention here the properties of codifference confirm that the measure is an appropriate mathematical tool for measuring the dependence between α-stable random variables as well as random variables from more general class of distributions (e.g., infinitely divisible). Moreover in contrast to covariation, which is of limited practical importance, according to the definition of codifference (12), it is easy to evaluate the empirical codifference which is based on the empirical characteristic function of the analyzed data [22, 23, 32].

The alternative measure of dependence is defined for independently divisible stochastic process $\{Y(t),\ t \in \mathbb{R}\}$ with the following integral representation

$$Y(t) = \int_X K(t,x)M(dx),$$

where M is an independly scattered infinitely divisible random measure on some measurable space S with control measure m, such that for every m-finite set $A \subseteq S$ (Lévy-Khinchin formula for infinitely divisible random measure):

$$\mathbb{E}\exp[i\theta M(A)] = \exp\left[m(A)\left\{i\theta\mu - \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}}(e^{i\theta x} - 1 - i\theta x I(|x|<1))M(dx)\right\}\right].$$

The measure was introduced in [6] as a new concept of correlation cascades, which is a promising tool for exploiting the properties of the Poissonian part of $Y(t)$ and the dependence structure of this stochastic process. Eliazar and Klafter in [6] proceed in the following way—first, define a Poissonian tail-rate function Λ of the Lévy measure Q as:

$$\Lambda(l) = \int_{|x|>l} Q(dx),\ l > 0,$$

next, for $t_1, t_2, \ldots, t_n \in \mathbb{R}$ and $l > 0$ introduce a family of functions

$$C_l(t_1, t_2, \ldots, t_n) = \int_X \Lambda\left(\frac{l}{\min\{K(t_1,x), \ldots, K(t_n,x)\}}\right) m(dx),$$

called the Lévy correlation cascade, see also [14]. In the special symmetric α-stable case the tail function takes the simple form:

$$\Lambda(l) = \frac{2}{\alpha}\cdot l^{-\alpha} \tag{16}$$

and for the symmetric α-stable moving-average process $\{Y(t),\ t \in \mathbb{R}\}$ with the integral representation

$$Y(t) = \int_{-\infty}^{t} f(t-x)M(dx)$$

the function $C_l(t,s),\ s,t \in \mathbb{R}$ has the form:

$$C_l(t,s) = \frac{2}{\alpha} \cdot l^{-\alpha} \int\limits_{|t-s|} |f(x)|^{\alpha} m(dx). \tag{17}$$

Many significant properties and results related to the Lévy correlation cascade for infinitely divisible processes are presented in [6, 14]. We only mention here the function $C_l(t_1, t_2, \ldots, t_n)$ tells us, how dependant the coordinates of the vector $(Y(t_1), Y(t_2), \ldots, Y(t_n))$ are. Therefore, $C_l(t_1, t_2, \ldots, t_n)$ can be considered as an appropriate measure of dependence for the Poissonian part of the infinitely divisible process. In particular, the function $C_l(t,s)$ can serve as an analogue of the covariance and the function

$$r_l(t,s) \frac{C_l(t,s)}{\sqrt{C_l(t)C_l(s)}} \tag{18}$$

can play a role of the correlation coefficient. Note that in the α-stable case the right-hand side of (18) does not depend on the l parameter, [14, 31].

4 Asymptotic Behaviour of the Measures of Dependence for Symmetric α-stable CARMA(p,q) Processes

Proposition 1 *Let $\{Y(t),\ t \in \mathbb{R}\}$ be the stationary solution of (1) given by (8). Moreover, we assume the condition (10) is satisfied. Then for $1 < \alpha \leq 2$ the covariation of $Y(t+h)$ on $Y(t)$ for $t \in \mathbb{R}$ and $h > 0$ has the following form*

$$CV(h) = CV(Y(t+h), Y(t)) = \int_0^{\infty} g(h+x)g(x)^{<\alpha-1>}dx,$$

where the function $g(x)$ is defined in (9).

Proof The result follows from the form of the stationary solution of Eq. (1) and Proposition 3.5.2 in [26].

□

Proposition 2 *If $\{Y(t),\ t \in \mathbb{R}\}$ is the stationary solution of (1) given in (8) and condition (10) is satisfied, then for $1 < \alpha \leq 2$ the codifference of $Y(t+h)$ on $Y(t)$ ($t \in \mathbb{R}$, $h > 0$) has the following form:*

$$CD(h) = CD(Y(t+h), Y(t)) = \int_0^\infty \left(|g(h+x)|^\alpha + |g(x)|^\alpha - |g(h+x) - g(x)|^\alpha\right) dx.$$

Proof The proof follows from the relation between the covariation and codifference given in Eq. (15) as well as from Proposition 1. □

Lemma 2 *Let $\{Y(t),\ t \in \mathbb{R}\}$ be the stationary solution of (1) given in (8). Moreover, let us assume when condition (10) is satisfied, then there exist constants $c, D_1 > 0$ such that for each $h \in \mathbb{R}$ we have*

$$|CV(h)| \leq e^{-ch} D_1.$$

Proof Using Proposition 1 and Remark 2 we obtain

$$\begin{aligned}
|CV(h)| &= \left| \int_0^\infty g(x+h) g(x)^{<\alpha-1>} dx \right| \\
&\leq \int_0^\infty \left| g(x+h) g(x)^{<\alpha-1>} \right| dx \\
&\leq Ce^{-ch} \int_0^\infty e^{-cx} \left| g(x)^{<\alpha-1>} \right| dx \\
&= Ce^{-ch} \int_0^\infty e^{-cx} |g(x)|^{\alpha-2} \left| \overline{g(x)} \right| dx \\
&= Ce^{-ch} \int_0^\infty e^{-cx} |g(x)|^{\alpha-1} dx \\
&\leq C^2 \int_0^\infty e^{-c\alpha x} = D_1 e^{-ch},
\end{aligned}$$

where $D_1 = \frac{C^2}{c\alpha} > 0$. □

Lemma 3 *Let $\{Y(t),\ t \in \mathbb{R}\}$ be the stationary solution of (1) given in (8). Moreover let us assume condition (10) is satisfied. Then the following holds:*

$$CD(h) \longrightarrow \alpha \int_0^\infty \frac{g(h+x)}{g(x)} |g(x)|^\alpha dx \ \text{ as } \ h \longrightarrow \infty. \tag{19}$$

Proof Using Proposition 2, we have

$$
\begin{aligned}
CD(h) &= \int_0^\infty \left(|g(h+x)|^\alpha + |g(x)|^\alpha - |g(h+x) - g(x)|^\alpha\right) dx \\
&= \int_0^\infty |g(x)|^\alpha \left(\left| \frac{g(h+x)}{g(x)} \right|^\alpha + 1 - \left| \frac{g(h+x)}{g(x)} - 1 \right|^\alpha \right) dx \\
&= \int_0^\infty \frac{g(h+x)}{g(x)} |g(x)|^\alpha \frac{\left| \frac{g(h+x)}{g(x)} \right|^\alpha + 1 - \left| \frac{g(h+x)}{g(x)} - 1 \right|^\alpha}{\frac{g(h+x)}{g(x)}} dx.
\end{aligned}
$$

Taking into consideration Proposition 1.11 in [15], we obtain there exist constants $C, c > 0$ such that

$$
\left| \frac{g(h+x)}{g(x)} \right| \leq \frac{Ce^{-c(h-x)}}{g(x)},
$$

which for each $x \in \mathbb{R}$ gives:

$$
\frac{g(h+x)}{g(x)} \longrightarrow 0, \text{ as } h \longrightarrow \infty.
$$

Since for $1 < \alpha \leq 2$ we have $\lim_{x \to 0} \frac{1-|1-x|^\alpha}{x} = \alpha$, $\lim_{x \to 0} \frac{|x|^\alpha}{x} = 0$, then we obtain the result. □

Theorem 1 *If $\{Y(t),\ t \in \mathbb{R}\}$ is the stationary solution of Eq. (1) given in (8) and condition (10) is satisfied, then for $1 < \alpha \leq 2$ the following holds*

$$
\lim_{h \to \infty} \frac{CD(h)}{CV(h)} = \alpha. \tag{20}
$$

Proof Applying Proposition 1 and Lemma 3 for $1 < \alpha \leq 2$ we get formulas for the covariation and the asymptotic behaviour of codifference:

$$
CV(h) = \int_0^\infty \frac{g(h+x)}{g(x)} |g(x)|^\alpha,\ h > 0
$$

$$
CD(h) \longrightarrow \alpha \int_0^\infty \frac{g(h+x)}{g(x)} |g(x)|^\alpha \text{ as } h \longrightarrow \infty.
$$

Therefore for $1 < \alpha \leq 2$ we obtain

$$
\lim_{h \to \infty} \frac{CD(h)}{CV(h)} = \alpha.
$$

□

Result similar to this obtained in (20), we also obtain for discrete ARMA models with symmetric α-stable innovations, see [19] as well as for periodic ARMA models

(PARMA) with stable structure reviewed in [20]. The special cases of considered stable CARMA(p,q) processes and the asymptotic behaviour of their measures of dependence are presented in [32].

Lemma 4 *Let* $\{Y(t),\ t \in \mathbb{R}\}$ *be the stationary solution of (1) given in (8). Moreover, let us assume condition (10) is satisfied. Then there exist constants* c, $D_2 > 0$ *such that for each* $h \in \mathbb{R}$ *we have*

$$|C_l(h,0)| \leq l^{-\alpha} e^{-ch} D_2.$$

Proof Using the form of the solution of Eq. (1) given in (8) and formula (17) we obtain

$$C_l(h,0) = \frac{2}{\alpha} l^{-\alpha} \int_h^\infty |g(x)|^\alpha dx,$$

where the function $g(x)$ is defined in (9). Moreover using Remark 2 we have

$$|C_l(h,0)| \leq \frac{2C}{\alpha} l^{-\alpha} \int_h^\infty e^{-cx\alpha} dx = l^{-\alpha} e^{-ch} D_2,$$

where $D_2 = \frac{2C}{c\alpha}$. □

The results presented in this section can be useful in analysis of real time series. As it is known, the measures of dependence are main tools useful in the problem of finding of the proper model for real data. By comparing the theoretical measure of dependence and its empirical equivalent, we can easily conclude the model that describes the data as well as estimate the parameters of the model. This issue is also discussed in [32] where the codifference is examined as a practical tool to measure interdependence of given process.

5 Symmetric α-stable CAR(1) Process

As an example, we consider the symmetric α-stable CAR(1) (CARMA(1,0)) process given by the following formula:

$$DY(t) + \lambda Y(t) = DL^*(t), \tag{21}$$

where $\{L^*(t),\ t \in \mathbb{R}\}$ is a two-sided symmetric α-stable Lévy motion.
The stationary solution of Eq. (1) is therefore given by

$$Y(t) = \int_{-\infty}^t e^{-\lambda(t-u)} dL^*(u), \tag{22}$$

Let us mention, the general CAR(1) process is known in the literature as the Ornstein-Uhlenbeck process [29]. When $\{L^*(t)\ t \in \mathbb{R}\}$ in Eq. (21) is Gaussian, then the system is called the classical Ornstein-Uhlenbeck process. It is one of the famous examples of continuous time models and was originally introduced by Uhlenbeck and Ornstein [29] as a suitable model for the velocity process in the Brownian diffusion. The Ornstein-Uhlenbeck process has been of fundamental importance for theoretical studies in physics and mathematics, but it has also been used in many applications including financial data such as interest rates, currency exchange rates and commodity prices. In finance, it is best known in connection with the Vasiček interest rate model [30], which was one of the earliest stochastic models of the term structure. The model exhibits mean reversion, which means that if the interest rate is above the long run mean, then the drift becomes negative so that the rate will be pushed down to be closer to the mean level. Likewise, if the rate is below the long run mean, then the drift remains positive so that the rate will be pushed up to the mean level. Such mean reversion feature complies with the economic phenomenon that in the long-time period interest rates appear to be pulled back to some average value [9]. There are many extensions of the classical Ornstein-Uhlenbeck process. One of them is to replace the Gaussian process by other Lévy processes, for example α-stable one. The subordinated Ornstein-Uhlenbeck process with symmetric α-stable structure was used in [9] to description of interest rates data. Moreover, the Ornstein-Uhlenbeck process with other than α-stable distributions was considered in [24, 31]. We should also mention, the discrete version of symmetric α-stable CAR(1) process is an AR(1) time series with α-stable innovations considered in [21]. In the next propositions, we give the form of considered measures of dependence for symmetric α-stable CAR(1) defined by the Eq. (21).

Proposition 3 *Let $\{Y(t),\ t \in \mathbb{R}\}$ be the stationary solution of (21) and $\lambda > 0$, then for $1 < \alpha \leq 2$ the covariation of $Y(t)$ on $Y(s)$ for $s, t \in \mathbb{R}$ has the following form:*

$$CV(Y(t), Y(s)) = \begin{cases} \dfrac{e^{-\lambda(t-s)}}{\lambda\alpha} & \text{for } s < t, \\ \dfrac{e^{\lambda(\alpha-1)(t-s)}}{\lambda\alpha}, & \text{for } s \geq t. \end{cases}$$

Proof The process $\{Y(t),\ t \in \mathbb{R}\}$ is given by (22), thus to obtain the covariation of $Y(t)$ on $Y(s)$ $(t, s \in \mathbb{R})$ we use Proposition 3.5.2 in [26] and obtain:

$$CV(Y(t), Y(s)) = \int_{-\infty}^{\infty} f(t-x) f(s-x)^{<\alpha-1>} dx,$$

where

$$f(x) = e^{-\lambda x} I_{[0,\infty)}(x) \tag{23}$$

Therefore we have

$$\begin{aligned} CV(Y(t), Y(s)) &= \int_{-\infty}^{\min(s,t)} e^{-\lambda(t-x)} e^{-\lambda(\alpha-1)(s-x)} dx \\ &= e^{-\lambda t - \lambda(\alpha-1)s} \int_{-\infty}^{\min(s,t)} e^{\lambda\alpha x} dx \\ &= \frac{\exp\{-\lambda t - \lambda(\alpha-1)s + \lambda\alpha \min(s,t)\}}{\lambda\alpha}, \end{aligned}$$

which for $s < t$ gives

$$CV(Y(t), Y(s)) = \frac{e^{-\lambda(t-s)}}{\lambda\alpha}.$$

For $s \geq t$ we have

$$CV(Y(t), Y(s)) = \frac{e^{\lambda(\alpha-1)(t-s)}}{\lambda\alpha}.$$

□

Proposition 4 *If* $\{Y(t),\ t \in \mathbb{R}\}$ *is the stationary solution of equation (21) and* $\lambda > 0$, *then for* $1 < \alpha \leq 2$ *the codifference of* $Y(t)$ *on* $Y(s)$ $(s, t \in \mathbb{R})$ *has the following form:*

$$CD(Y(t), Y(s)) = \frac{1 + e^{-\lambda\alpha|t-s|} - |1 - e^{-\lambda|t-s|}|^{\alpha}}{\lambda\alpha}. \tag{24}$$

Proof The calculations one can find in [32], see also [26]. □

Theorem 2 *If* $\{Y(t),\ t \in \mathbb{R}\}$ *is the stationary solution of Eq. (21) and* $\lambda > 0$, *then:*

(a) for $1 < \alpha \leq 2$ *the following formula holds:*

$$\lim_{h\to\infty} \frac{CD(h)}{CV(h)} = \alpha, \tag{25}$$

(b) for $1 < \alpha < 2$ *the following formula holds:*

$$\lim_{h\to\infty} \frac{CD(-h)}{CV(-h)} = 0. \tag{26}$$

Proof (a) The proof of this part follows directly from Theorem 1).
(b) Applying Propositions 1 and 2 for $1 < \alpha \leq 2$, $t \in \mathbb{R}$ and $h > 0$ we have the following formulas:

$$CV(Y(t-h),Y(t)) = CV(Y(t),Y(t+h)) = \frac{e^{-\lambda(\alpha-1)h}}{\lambda\alpha},$$

$$CD(Y(t-h),Y(t)) = CD(Y(t),Y(t+h)) = \frac{1+e^{-\lambda\alpha h}-|1-e^{-\lambda h}|^{\alpha}}{\lambda\alpha}.$$

Therefore for $1<\alpha\leq 2$, $t\in\mathbb{R}$ and $h>0$ one has:

$$\frac{CD(Y(t-h),Y(t))}{CV(Y(t-h),Y(t))} = \frac{1+e^{-\lambda\alpha h}-|1-e^{-\lambda h}|^{\alpha}}{e^{-\lambda(\alpha-1)h}}.$$

Since for $1<\alpha<2$ we have:

$$\lim_{x\to 0}\frac{1+|x|^{\alpha}-|1-x|^{\alpha}}{x^{\alpha-1}} = \lim_{x\to 0} x^{2-\alpha}\left(\frac{1+|x|^{\alpha}-|1-x|^{\alpha}}{x}\right) = 0,$$

then for $1<\alpha<2$ and $t\in\mathbb{R}$ we obtain:

$$\lim_{h\to\infty}\frac{CD(-h)}{CV(-h)} = \lim_{h\to\infty}\frac{CD(Y(t-h),Y(t))}{CV(Y(t-h),Y(t))} = 0.$$

□

Proposition 5 *If $\{Y(t),\ t\in\mathbb{R}\}$ is the stationary solution of symmetric α-stable CAR(1) process given in (22) and $\lambda>0$, then for $0<\alpha\leq 2$ the Lévy correlation cascade of $Y(t)$ on $Y(s)$ $(s,t\in\mathbb{R})$ is given by:*

$$C_l(t,s) = \frac{2}{\lambda\alpha^2 l^{\alpha}}e^{-\lambda\alpha|t-s|},\ l>0. \tag{27}$$

The correlation coefficient $r_l(t,s)$ defined in (18) has the following form:

$$r_l(t,s) = e^{-\lambda\alpha|t-s|}. \tag{28}$$

Proof The proof follows directly from the form of the solution of equation (21) as well as from (17) and the fact that for symmetric α-stable processes $\Lambda(l) = \frac{2}{\alpha}l^{-\alpha}$. The form of the $r_l(t,s)$ function follows directly from relation (18). □

In Fig. 1 we present exemplary trajectories of symmetric α-stable CAR(1) process with $\lambda = 0.0018$ and three values of α parameter $\alpha = 1.3$, $\alpha = 1.46$ and $\alpha = 1.9$. Moreover in Fig. 2 we show the ratio $CV(h)/CD(h)$ for three considered cases of α parameter and $h\in[2000, 12000]$.

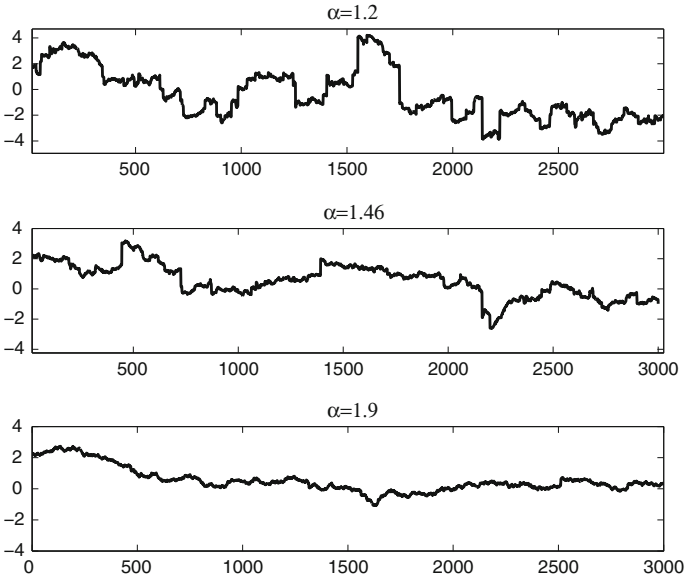


Fig. 1 The exemplary trajectories of symmetric α-stable Ornstein-Uhlenbeck process with $\alpha = 1.2$, $\alpha = 1.46$ and $\alpha = 1.9$

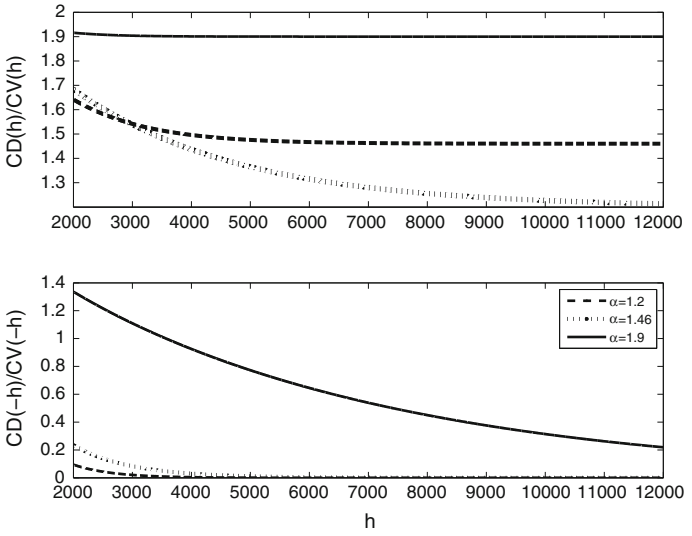


Fig. 2 The ratio $CV(h)/CD(h)$ (*top panel*) and $CV(-h)/CD(-h)$ (*bottom panel*) for symmetric α-stable CAR(1) process (21) for $\lambda = 0.0018$ and three considered cases of α parameter

6 Applications

In this section, we use the considered models to describe the real financial data. In the analysis, we use the estimation procedure consisted in the estimation of embedding in CARMA systems the discrete ARMA models. In our calculations we consider the estimator for parameters of ARMA models with infinite variance innovations proposed in [17], which is based on the Whittle method. For the simplicity we consider only the autoregressive models (i.e. without the MA part).

As an example, we consider the United Kingdom interbank rates LIBID/LIBOR for 12 months quoted daily from the period 02.01.1995–03.07.2006. In Fig. 3 we present the examined dataset after removing the sample mean. Because the Gaussian CAR(1) was originally used by Vasiček for the analysis of interest rates, we propose to describe the examined vector of observations by using this process. However, in the dataset we observe significant jumps that may suggest the α-stable distribution. As it was mentioned, first we estimate the parameters corresponding to considered model by using the Whittle estimation method based on the sample periodogram. We apply the method to the original dataset after removing its sample mean. As a result, we obtain the following model:

$$DY(t) + 0.0018Y(t) = DL^*(t),$$

where $\{L^*(t),\ t \in \mathbb{R}\}$ is a two-sided symmetric α-stable Lévy motion. As it was mention, the discrete version of symmetric α-stable CAR(1) process is an AR(1) time series with symmetric α-stable innovations, [9]. After fitting the appropriate model, we calculate the model residuals, see Fig. 4. Our assumption od α-stable distribution we confirm by using Jacque-Bera test for Gaussian distribution [5, 8, 25] and the Anderson-Darling test for stable behaviour of examined dataset [5]. Namely, we test if the residuals can be considered as Gaussian sample (by Jacque-Bera test). As a result, we obtain p-value equal to 0.001 which means, on the confidence level 0.05 we can not assume the Gaussian distribution. Then we test the α-stable distribution by Anderson-Darling test for stable behaviour. As a result, we obtain p-value equal

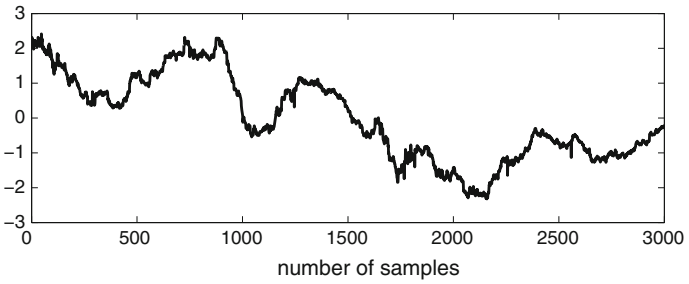


Fig. 3 The real dataset (after removing the sample mean) corresponding to the United Kingdom interbank rates LIBID/LIBOR for 12 months quoted daily from the period 02.01.1995–03.07.2006

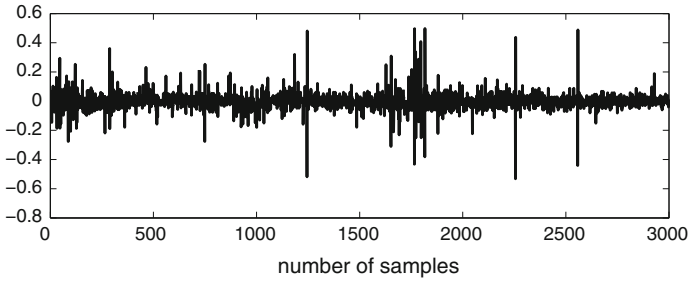


Fig. 4 The residuals of fitted symmetric α-stable CAR(1) models to dataset corresponding to the United Kingdom interbank rates LIBID/LIBOR from the period 02.01.1995–03.07.2006

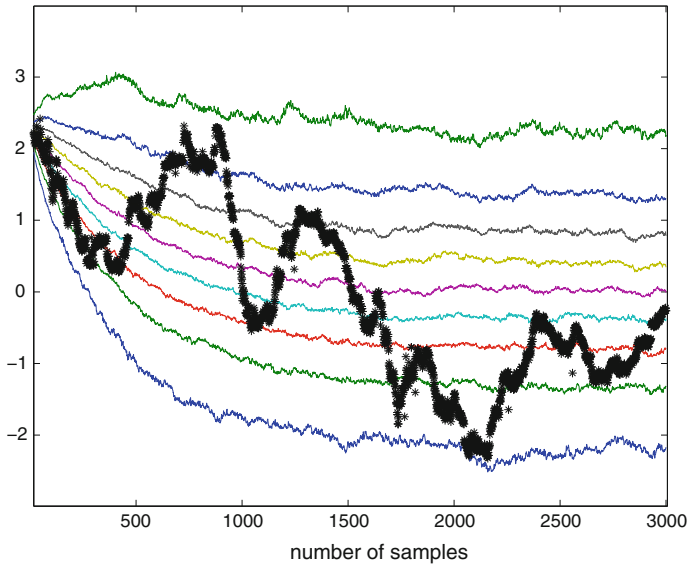


Fig. 5 The quantile lines of levels 10, 20…, 90 % obtained using 1000 simulated trajectories of fitted symmetric α-stable CAR(1) process and the analyzed dataset (*black thick lines*)

to 0.35 which means, on the 0.05 confidence level we can not reject the hypothesis of the stable distribution. Finally, we estimate the stability parameter α by using the McCulloch method [12]. We obtain $\hat{\alpha} = 1.46$. Next, we calculate the quantile lines, i.e. the curves describing the values that with a given probability will not be exceeded by the corresponding process [11]. The quantile lines of levels 10, 20, 90 % obtained using 1000 simulated trajectories are presented in Fig. 5. Additionally, in the same figure we plot the analyzed dataset. Observe that the observed interest rates doe not exceed the bounds given by the 10 and 90 % quantile lines [13].

7 Conclusions

In this paper, we have analyzed the symmetric α-stable CARMA processes and examined their structure of dependence. Because for the considered models the covariance function is not defined, then the dependence is expressed by using the alternative measures. We have studied the codifference and the covariation—the most popular measures of dependence defined for symmetric α-stable random variables. We have proved, similar as for the discrete counterparts of examined processes, the codifference and the covariation are asymptotically proportional with the coefficient of proportionality equal to the stability parameter of the model. We have also considered the alternative measure defined for infinitely divisible stochastic processes, namely the Lévy correlation cascade. As a special case, we have considered the symmetric α-stable CAR(1) process, known in the literature as the Ornstein-Uhlenbeck process. The theoretical results are illustrated by real dataset that describes the interest rates from the United Kingdom market. We believe the presented results can be a base to obtain the alternative methods of estimation of unknown parameters of symmetric α-stable CARMA processes.

References

1. Barndorff-Nielsen, O., & Shepardt, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B, 63*, 1–42.
2. Brockwell, P. J. (2001). Lévy-driven CARMA processes. *Annals of the Institute of Statistical Mathematics, 53*(1), 113–124.
3. Brockwell, P. J., & Marquardt, T. (2005). Lévy-driven and fractionally integrated ARMA processes with continuous time parameter. *Statistica Sinica, 15*, 477–494.
4. Broszkiewicz-Suwaj, E., Makagon, A., Weron, R., & Wyłomańska Agnieszka, A. (2004). On detecting and modeling periodic correlation in financial data. *Physica A, 336*, 196–205.
5. Burnecki, K., Wyłomańska, A., Beletskii, A., Gonchar, V., & Chechkin, A. (2012). Recognition of stable distribution with Levy index alpha close to 2. *Physical Review E, 85*, 056711.
6. Eliazar, I., & Klafter, J. (2007). Correlation cascades of Lévy-driven random processes. *Physica A, 376*, 1–26.
7. Gajda, J., & Wyłomańska, A. (2012). Geometric Brownian motion with tempered stable waiting times. *Journal of Statistical Physics, 148*, 296–305.
8. Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters, 6*(3), 255–259.
9. Janczura, J., Orzeł, S., & Wyłomańska, A. (2011). Subordinated alpha-stable Ornstein-Uhlenbeck process as a tool of financial data description. *Physica A, 390*, 4379–4387.
10. Janicki, A., & Weron, A. (1994). Can one see α-stable variables and processes? *Statistical Science, 9*, 109–126.
11. Janicki, A., & Weron, A. (1994). *Simulation and chaotic behaviour of α-stable stochastic processes*. New York: Marcel Dekker.
12. MacCulloch, J. H. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics-Simulation and Computation, 15*, 1109–1136.
13. Maciejewska, M., Szczurek, A., Janczura, J., & Wyłomańska, A. (2013). Stochastic modeling of indoor air temperature. *Journal of Statistical Physics, 152*, 979–994.

14. Magdziarz, M. (2009). Correlation cascades, Ergodic properties and long memory of infinitely divisible processes. *Stochastic Processes and Their Applications, 119*, 3416–3434.
15. Marquardt, T. (2006). Fractional Lévy processes, CARMA processes and related topics. *Doctoral Thesis*, TU Munchen.
16. Marquardt, T., & Stelzer, R. (2007). Multivariate CARMA processes. *Stochastic Processes and Their Applications, 117*, 96–120.
17. Mikosch, T., Gadrich, T., Klueppelberg, C., & Adler, R. J. (1995). Parameter estimation form ARMA models with infinite variance innovations. *Annals of Statistics, 23*(1), 305–326.
18. Mittnik, S., & Rachev, S. T. (2000). *Stable paretian models in finance*. New York: Wiley.
19. Nowicka, J. (1997). Asymptotic behavior of the covariation and the codifference for ARMA models with stable innovations. *Stochastic Models, 13*, 673–685.
20. Nowicka-Zagrajek, J., & Wyłomańska, A. (2006). The dependence structure for PARMA models with $\alpha-$stable innovations. *Acta Physica Polonica, 37*(11), 3071–3082.
21. Nowicka-Zagrajek, J., & Wyłomańska, A. (2008). Measures of dependence for stable AR(1) models with time-varying coefficients. *Stochastic Models, 24*(1), 58–70.
22. Rosadi, D. (2005). Asymptotic behavior of the codifference and the dynamical function for ARMA models with infinite variance. *Journal of Indonesian Mathematical Society (MIHMI), 11*(1), 59–69.
23. Rosadi, D., & Deistler, M. (2011). Estimating the codifference function of linear time series models with infinite variance. *Metrika, 73*(3), 395–429.
24. Obuchowski, J., & Wyłomańska, A. (2013). The Ornstein-Uhlenbeck process with non-Gaussian structure. *Acta Physica Polonica A, 44*(5), 1123–1136.
25. Obuchowski, J., Wyłomańska, A., & Zimroz, R. (2014). Selection of informative frequency band in local damage detection in rotating machinery. *Mechanical Systems and Signal Processing, 48*, 138–152.
26. Samorodnitsky, G., & Taqqu, M. S. (1994). *Stable non-Gaussian random processes*. New York: Chapman & Hall.
27. Stuck, B. W., & Kleiner, B. (1974). A statistical analysis of telephone noise. *Bell System Technical Journal, 53*, 1263–1320.
28. Uchaikin, V. V., & Zolotarev, V. M. (1999). *Chance and stability, stable distributions and their applications*. Utrecht: VSP.
29. Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical Review, 36*, 823–841.
30. Vasiček, O. (1977). An equilibrium characterisation of the term structure. *Journal of Financial Economics, 5*(2), 177–188.
31. Wyłomańska, A. (2011). Measures of dependence for Ornstein-Uhlenbeck process with tempered stable distribution. *Acta Physica Polonica, 42*(10), 2049–2062.
32. Wyłomańska, A., Chechkin, A., Sokolov, I., & Gajda, J. (2015). Codifference as a practical tool to measure interdependence. *Physica A*, *421*, 412–429.
33. Wyłomańska, A., Obuchowski, J., Zimroz, R., Hurd, H. (2014). Periodic autoregressive modeling of vibration time series from planetary gearbox used in bucket wheel excavator, In Fakher Chaari et al. (Eds.), *Cyclostationarity: Theory and methods*. Lecture Notes in Mechanical Engineering, (pp. 171–186).
34. Yu, G., & Li, Ch. (2013). A new statistical modeling and detection method for rolling element bearings faults based on alpha-stable distribution. *Mechanical Systems and Signal Processing, 41*, 155–175.
35. Zak, G., Obuchowski, J., Wyłomańska, A., & Zimroz, R. (2014). Application of ARMA modelling and alpha-stable distribution for local damage detection in bearings. *Diagnostyka,15*(3), 3–11.