



HAL
open science

Spectral density estimation for nonstationary data with nonzero mean function

Anna E. Dudek, Lukasz Lenart

► **To cite this version:**

Anna E. Dudek, Lukasz Lenart. Spectral density estimation for nonstationary data with nonzero mean function. 2020. hal-02442913v1

HAL Id: hal-02442913

<https://hal.science/hal-02442913v1>

Preprint submitted on 16 Jan 2020 (v1), last revised 1 Mar 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Spectral density estimation for nonstationary data with nonzero mean function

Anna E. Dudek*

Department of Applied Mathematics, AGH University of Science and Technology,
al. Mickiewicza 30, 30-059 Krakow, Poland

and

Łukasz Lenart †

Department of Mathematics, Cracow University of Economics,
ul. Rakowicka 27, 31-510 Cracow, Poland

January 16, 2020

Abstract

We introduce a new approach for nonparametric spectral density estimation based on the subsampling technique, which we apply to the important class of nonstationary time series. These are almost periodically correlated sequences. In contrary to existing methods our technique does not require demeaning of the data. On the simulated data examples we compare our estimator of spectral density function with the classical one. Additionally, we propose a modified estimator, which allows to reduce the leakage effect. Moreover, we provide two real data economic applications.

Keywords: subsampling, almost periodically correlated, covariance matrix estimation, mixing.

*Anna Dudek was partially supported by the Faculty of Applied Mathematics AGH UST statutory tasks within subsidy of Ministry of Science and Higher Education.

†Łukasz Lenart was financed from the funds granted to the College of Economics, Finance and Law at CUE, within the framework of the subsidy for the maintenance of research potential.

1 Introduction

Spectral theory for stationary time series is well developed (see Priestley (1981), Žurbenko (1986), Stoica and Moses (2005), and many others). However, in the real data applications one deals often with nonstationary phenomena. In such a case statistical tools designed for stationary processes should be appropriately modified or generalized. If this is not possible a new approach needs to be proposed.

In this paper we focus on the spectral density estimation problem in the nonstationary setting. Classical approach for spectral density estimation requires removing the mean function from the data. However, as we discuss below, this may be very challenging in the nonstationary case. Thus, it is crucial to propose a new method that provides consistent estimates and does not assume the nullity of the mean function.

In the sequel we introduce and develop a novel methodology for spectral densities estimation for nonstationary mixing sequences with nonzero mean function. Our main results cover the class of almost periodically correlated (APC) time series, which is an important class of nonstationary time series containing among others periodically correlated (PC) and covariance stationary sequences. APC and PC processes are broadly used to model data with cyclic features. Many interesting examples concerning e.g., telecommunication, vibroacoustics, mechanics, economics and climatology can be found in Antoni (2009), Gardner et al. (2006), Hurd and Miamee (2007), Napolitano (2012, 2016). To be precise, a time series $\{X_t, t \in \mathbb{Z}\}$ with finite second moments is called PC if it has periodic mean and covariance functions, while in APC case the mean and covariance functions are almost periodic functions of time argument. A function $f(t) : \mathbb{Z} \rightarrow \mathbb{R}$ is called almost periodic if for any $\epsilon > 0$ there exists an integer $L_\epsilon > 0$ such that among any L_ϵ consecutive integers there is an integer p_ϵ such that $\sup_{t \in \mathbb{Z}} |f(t + p_\epsilon) - f(t)| < \epsilon$ (see Corduneanu (1989)).

Problem of spectral density estimation for PC and APC sequences is not new (see Nematollahi and Rao (2005), Lii and Rosenblatt (2006), Hurd and Miamee (2007), Napolitano (2012, 2016)), but as we mentioned before all methods assume that $EX_t \equiv 0$. In the case of stationary and PC time series this is not a restrictive condition, because the data can be demeaned (see e.g., Brillinger (2001) and Hurd and Miamee (2007)). However, for APC time series this problem is much more difficult. In general case it is not easy to estimate an almost periodic mean function. Moreover, in the presence of an almost periodic mean, the spectral densities and cyclic spectra contain Dirac impulses as a consequence of the presence of an almost periodic component in the cyclic autocovariance (see Napolitano (2016)). Estimation of the spectral densities or the cyclic spectrum without removing the mean, results in spikes in correspondence of the ideal Dirac deltas. In such a case, Napolitano and Spooner (2000) proposed a median filtering technique to remove the spikes. However, for now the theoretical properties of this approach were not studied in the literature.

Our idea for spectral densities estimation is based on the subsampling method, which allows to approximate the asymptotic variance of the rescaled statistics of interest. Let us indicate that the general consistency result concerning subsampling estimator of the asymptotic variance that we present does not require periodicity or almost periodicity of the data. However, we apply it to an APC process to provide at the same time an application that is interesting and important from practical point of view. For that purpose we use results from Lenart (2013). Lenart considered estimation problem for the Fourier coefficients of the mean function of an APC time series. He showed that their rescaled estimators are asymptotically normal and that elements of the asymptotic covariance matrix are linear functions of the spectral densities of the considered process. Applying subsampling to estimate elements of this matrix we get consistent estimators of the spectral densities.

Paper is organized as follows. In Section 2 the known methods for spectral density estimation in stationary and almost periodic case are summarized. The main results are provided in Section 3. In particular the subsampling estimator of the spectral density functions of an APC time series with the nonzero mean function is constructed and its consistency is shown. Moreover, to reduce the leakage effect the modification of this estimator is discussed. In Section 4 real data example is provided. Finally, all proofs are presented in the supplementary materials.

2 Spectral density estimation for APC time series

Let $\{X_t, t \in \mathbb{Z}\}$ be an APC time series. Its mean function $\mu(t) = E(X_t)$ and the autocovariance function $B(t, \tau) = \text{cov}(X_t, X_{t+\tau})$ (for any fixed $\tau \in \mathbb{Z}$) can be represented as Fourier series

$$\mu(t) \sim \sum_{\psi \in \Psi} m(\psi) \exp(i\psi t), \quad B(t, \tau) \sim \sum_{\lambda \in \Lambda_\tau} a(\lambda, \tau) \exp(i\lambda t), \quad (1)$$

where the sets $\Psi = \{\psi \in [0, 2\pi) : m(\psi) \neq 0\}$, $\Lambda_\tau = \{\lambda \in [0, 2\pi) : a(\lambda, \tau) \neq 0\}$ are countable, and

$$m(\psi) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mu(j) \exp(-i\psi j), \quad a(\lambda, \tau) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n B(j, \tau) \exp(-i\lambda j).$$

Symbol \sim indicates some association between the Fourier series and the almost periodic function. The sup-norm used in the definition of an almost periodic function (see Section 1) results in the uniform convergence of the Fourier series (1) to $\mu(t)$ and $B(t, \tau)$, respectively. For more details we refer the reader to Napolitano (2012).

If the sets of frequencies Ψ and $\Lambda = \bigcup_{\tau \in \mathbb{Z}} \Lambda_\tau$ are finite, then in the expressions (1) we have equalities. This condition is always fulfilled for PC time series, with $\Psi, \Lambda \subset \{2k\pi/d, k = 0, 1, \dots, d-1\}$ (see e.g., Hurd and Miamee (2007)).

Under some regularity conditions (see e.g., Lenart (2011)) the spectral density function $P(\cdot, \cdot)$ of an APC time series is defined on the bifrequency square $[0, 2\pi)^2$ and has the form

$$P(\nu, \omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} a(\nu - \omega, \tau) \exp(-i\nu\tau), \quad (\nu, \omega) \in [0, 2\pi)^2. \quad (2)$$

Below we shortly summarize the known results concerning estimation of $P(\nu, \omega)$.

In the stationary case i.e., when $\Lambda = \{0\}$, the consistent estimator of the spectral density function is obtained by tapering the covariance estimator or equivalently by smoothing the periodogram (see e.g., Priestley (1981), Brillinger (2001)). Let us recall that when a PC process with known period d is considered, then the spectral density function $P(\nu, \omega)$ can be nonzero only on the $2d - 1$ diagonal lines $\{(\nu, \omega) \in [0, 2\pi)^2 : \omega = \nu - 2k\pi/d, k = -d + 1, \dots, d - 1\}$ (see Hurd (1991), Dehay and Hurd (1994)). As a result the consistent estimator of the spectral density functions can be obtained by smoothing the shifted periodogram along the support lines or equivalently by weighting the estimators of the Fourier coefficients of the autocovariance function (ch. 10.2 in Hurd and Miamee (2007)). On the other hand, in Nematollahi and Rao (2005) to construct an estimator of the spectral density matrix, the relation between the PC time series and the multivariate stationary time series is used. Nematollahi and Rao proposed to estimate the spectral matrix of the stationary vector time series and then to use it to construct the estimator of the spectral density matrix of the considered PC time series. Finally, in Lii and Rosenblatt (2006) the estimation of the frequencies in almost periodic autocovariance structure of a zero mean Gaussian harmonizable APC time series is considered. In Napolitano (2016) the review of known

results for the APC processes is provided. Two estimators of the spectral density functions are discussed: the frequency-smoothed cyclic periodogram and the time-smoothed cyclic periodogram (see Section 2 in Napolitano (2016)).

Independently on the type of the process (stationary, PC, APC) to construct the aforementioned estimators of the spectral density function, the nullity of the mean function is assumed. If the considered process is not zero mean, before one estimates the spectral density, the data should be demeaned. This can be easily achieved when the mean function is constant or periodic with the known period length. It is also possible when one deals with an APC time series and the set Ψ is finite and known. Then, having sample (X_1, \dots, X_n) , one can apply the estimator of the mean function $\hat{\mu}_n(t) = \sum_{\psi \in \Psi} \hat{m}_n(\psi) \exp(i\psi t)$, where $\hat{m}_n(\psi) = \frac{1}{n} \sum_{j=1}^n X_j \exp(-i\psi j)$ is the estimator of the Fourier coefficient $m(\psi)$, $\psi \in [0, 2\pi)$. Finally, the estimator of the spectral density $P(\nu, \omega)$ can be obtained by replacing $\mu(t)$ by its estimate $\hat{\mu}_n(t)$ (see Lenart (2011, 2016)). According to our knowledge such estimator was not considered in the literature in APC case when the set Ψ is unknown and needs to be estimated. It is worth to note that in the general APC case, when $\mu(t)$ is an almost periodic function with the unknown set Ψ , it is very difficult to remove entirely the mean function from the data. Moreover, even in purely periodic case one may deal with so called hidden periodicities (Yavorskyj et al. (2012)) and in consequence estimation of the mean function can be challenging. In Napolitano (2016) (see (2.9)) it was shown that if the almost periodic mean is not removed from the data, the classical estimates of the spectral densities for zero mean case contain Dirac impulses, which can be observed as additional spikes. The shape of the spikes is related to the length of the frequency-smoothing window if one uses the frequency-smoothing cyclic periodogram method for the estimation. In such a case the estimator of spectral density function is not consistent. To solve this problem

in Napolitano and Spooner (2000) a median filtering technique is proposed, but so far the theoretical properties of this approach were not studied in the literature.

In the next section we introduce a new method for the spectral density function $P(\nu, \omega)$ (see (2)) estimation on the bifrequency square $(\nu, \omega) \in [0, 2\pi)^2$ in a case when the almost periodic mean function is unknown, i.e., when set Ψ is unknown. Our approach does not require estimation of the mean function and can be applied directly to the raw data.

3 Main results

In the first part of this section we propose the subsampling covariance matrix estimator for general weakly dependent time series and show its mean square consistency. In the next step we use this general result in the construction of the spectral density function estimator for an APC time series in the case when the set Ψ is unknown.

From now on by $\text{Re}(z)$ and $\text{Im}(z)$ we denote the real and the imaginary part of a complex number z . Symbol $(\cdot)'$ denotes transpose of a vector. By $\mathcal{N}_r(\mathbf{a}, \mathbf{\Sigma})$ we denote the r -variate Gaussian distribution with mean \mathbf{a} and covariance matrix $\mathbf{\Sigma}$. Finally, for any random variable Z and real number $p > 0$, $\|Z\|_p = (\mathbb{E}|Z|^p)^{1/p}$.

3.1 General subsampling consistency result

The idea of subsampling (see Politis et al. (1999)) is to approximate the sampling distribution of the statistics using subsamples of the data. Let $\{X_t, t \in \mathbb{Z}\}$ be a considered time series and (X_1, \dots, X_n) be an observed sample. Moreover, let r be any positive integer. By $\boldsymbol{\theta} = (\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(r)})' \in \mathbb{R}^r$ and $\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_n^{(1)}, \widehat{\theta}_n^{(2)}, \dots, \widehat{\theta}_n^{(r)})'$ we denote an unknown parameter and its estimator based on the sample (X_1, X_2, \dots, X_n) . We assume

that there exists a non-degenerate limiting distribution of $\tau_n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$, where τ_n is a normalizing sequence. Finally, $\widehat{\boldsymbol{\theta}}_{n,b,t}$ is a subsampling version of $\widehat{\boldsymbol{\theta}}_n$ based on the subsample (X_t, \dots, X_{t+b-1}) , where $b = b(n) \rightarrow \infty$, $1 \leq t \leq n - b + 1$ and

$$L_{n,b}(A) = \frac{1}{n - b + 1} \sum_{t=1}^{n-b+1} \mathbf{1} \left\{ \tau_b \left(\widehat{\boldsymbol{\theta}}_{n,b,t} - \widehat{\boldsymbol{\theta}}_n \right) \in A \right\} \quad (3)$$

is an r -variate empirical distribution which approximates the distribution of $\tau_n \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \right)$, A is any Borel set in \mathbb{R}^r and $\mathbf{1}\{C\}$ is an indicator function of the event C .

The important feature of subsampling (and its recent generalizations) is that it provides valid quantiles for confidence intervals and hypothesis tests in stationary and nonstationary models under very mild regularity conditions (see e.g., Politis et al. (1999), Dehay et al. (2014), Dudek and Lenart (2017), Lenart (2011, 2013, 2018), Tewes et al. (2019)). Moreover, to obtain subsampling consistency one does not need to know the exact form of the limiting distribution of $\tau_n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$. However, in this paper we are only interested in subsampling approximation of the elements of the covariance matrix of this limiting distribution. By $\boldsymbol{\Sigma} = [\sigma_{ij}]_{r \times r}$ we denote this limiting covariance matrix.

Recall that for fixed $1 \leq i, j \leq r$ and any sequence \tilde{b} such that $\tilde{b} \rightarrow \infty$, $\tilde{b}/n \rightarrow 0$ the estimator

$$\widehat{\sigma}_{n,\tilde{b}}^{(i,j)} = \frac{\tau_{\tilde{b}}^2}{n - \tilde{b} + 1} \sum_{t=1}^{n-\tilde{b}+1} \left(\widehat{\theta}_{n,\tilde{b},t}^{(i)} - \widehat{\theta}_{n,\tilde{b},\cdot}^{(i)} \right) \left(\widehat{\theta}_{n,\tilde{b},t}^{(j)} - \widehat{\theta}_{n,\tilde{b},\cdot}^{(j)} \right) \quad (4)$$

is a classical subsampling estimator of σ_{ij} based on block length \tilde{b} , where $\widehat{\theta}_{n,\tilde{b},\cdot}^{(i)} = 1/(n - \tilde{b} + 1) \sum_{t=1}^{n-\tilde{b}+1} \widehat{\theta}_{n,\tilde{b},t}^{(i)}$ and $\widehat{\theta}_{n,\tilde{b},t}^{(i)}$ is a subsampling estimator of $\theta^{(i)}$ (see Section 3.8.1 and 4.6 in Politis et al. (1999)). However, in this paper we define some more general subsampling estimator of σ_{ij} . For this purpose let us define two sequences of positive integers b_1, b_2

such that $b_1, b_2 \rightarrow \infty$ as $n \rightarrow \infty$, $b_1 \leq b_2$ and $b_2/n \rightarrow 0$ as $n \rightarrow \infty$. Next, denote $w = w(b_1, b_2) = b_2 - b_1 + 1$ and $q = q(n, b) = n - b + 1$ for $b_1 \leq b \leq b_2$. Finally, by

$$\widehat{\sigma}_{n, b_1, b_2}^{(i, j)} = \frac{1}{w} \sum_{b=b_1}^{b_2} \frac{\tau_b^2}{q} \sum_{t=1}^q \left(\widehat{\theta}_{n, b, t}^{(i)} - \widehat{\theta}_{n, b, \cdot}^{(i)} \right) \left(\widehat{\theta}_{n, b, t}^{(j)} - \widehat{\theta}_{n, b, \cdot}^{(j)} \right), \quad (5)$$

we define the generalized subsampling estimator of the σ_{ij} , which is based on block lengths belonging to the set $\{b_1, b_1 + 1, \dots, b_2\}$.

The idea behind the construction of the estimator (5) is as follows. Take any sequence \tilde{b} of integers such that $b_1 \leq \tilde{b} \leq b_2$. Let t^* be a random variable with uniform distribution on the set $\{1, 2, \dots, n - \tilde{b} + 1\}$. Note that the elements of the covariance matrix of the distribution $\tau_{\tilde{b}}(\widehat{\theta}_{n, \tilde{b}, t^*} - \widehat{\theta}_n)$ (in the resampling world) are of the form (4) i.e., $\text{cov}^*(\tau_{\tilde{b}}(\widehat{\theta}_{n, \tilde{b}, t^*} - \widehat{\theta}_n)) = \widehat{\Sigma}_{n, \tilde{b}} = [\widehat{\sigma}_{n, \tilde{b}}^{(i, j)}]_{i, j=1, 2, \dots, r}$. Thus, $\sigma_{n, b_1, b_2}^{(i, j)} = \frac{1}{w} \sum_{b=b_1}^{b_2} \widehat{\sigma}_{n, b}^{(i, j)}$ is a mean of the subsampling estimators (4) based on different block lengths (similar idea of generalized subsampling procedure can be found in Lenart (2018)).

Finally, let $\mathbf{A} = [a_{ij}]_{r \times r}$ be a known real valued matrix and let

$$\sigma = \sum_{i=1}^r \sum_{j=1}^r a_{ij} \sigma_{ij} \quad (6)$$

be a parameter of interests, which is a linear combination of elements of the matrix Σ . Based on (5) we propose the following subsampling estimator of σ

$$\widehat{\sigma}_n = \sum_{i=1}^r \sum_{j=1}^r a_{ij} \left[\frac{1}{w} \sum_{b=b_1}^{b_2} \frac{\tau_b^2}{q} \sum_{t=1}^q \left(\widehat{\theta}_{n, b, t}^{(i)} - \widehat{\theta}_{n, b, \cdot}^{(i)} \right) \left(\widehat{\theta}_{n, b, t}^{(j)} - \widehat{\theta}_{n, b, \cdot}^{(j)} \right) \right]. \quad (7)$$

Below we formulate conditions under which the subsampling estimator $\widehat{\sigma}_n$ is mean square consistent.

Theorem 3.1. *Assume that $\{X_t, t \in \mathbb{Z}\}$ is an α -mixing time series and*

(i) there exists $\tilde{\sigma}$ such that for any sequence of integers $\tilde{b} = \tilde{b}(n)$ such that $0 < b_1 \leq \tilde{b} \leq b_2 < n$, $b_1, b_2 \rightarrow \infty$ as $n \rightarrow \infty$, but $b_2/n \rightarrow 0$ as $n \rightarrow \infty$ and any sequence of integers t_n such that $1 \leq t_n \leq n - \tilde{b} + 1$ we have that

$$E \left[\sum_{i=1}^r \sum_{j=1}^r a_{ij} \tau_{\tilde{b}}^2 \left(\hat{\theta}_{n, \tilde{b}, t_n}^{(i)} - E \left(\hat{\theta}_{n, \tilde{b}, \cdot}^{(i)} \right) \right) \left(\hat{\theta}_{n, \tilde{b}, t_n}^{(j)} - E \left(\hat{\theta}_{n, \tilde{b}, \cdot}^{(j)} \right) \right) \right] \rightarrow \tilde{\sigma};$$

(i') OR there exists $\tilde{\sigma}$ such that for any sequence of integers \tilde{b} such that $0 < b_1 \leq \tilde{b} \leq b_2 < n$, $b_1, b_2 \rightarrow \infty$ as $n \rightarrow \infty$, but $b_2/n \rightarrow 0$ as $n \rightarrow \infty$ we have that

$$\frac{1}{n - \tilde{b} + 1} \sum_{t=1}^{n-\tilde{b}+1} E \left[\sum_{i=1}^r \sum_{j=1}^r a_{ij} \tau_{\tilde{b}}^2 \left(\hat{\theta}_{n, \tilde{b}, t}^{(i)} - E \left(\hat{\theta}_{n, \tilde{b}, \cdot}^{(i)} \right) \right) \left(\hat{\theta}_{n, \tilde{b}, t}^{(j)} - E \left(\hat{\theta}_{n, \tilde{b}, \cdot}^{(j)} \right) \right) \right] \rightarrow \tilde{\sigma};$$

(ii) there exist $\xi > 0$ and $\Delta > 0$ such that

$$\sup_{1 \leq i \leq r, b_1 \leq b \leq b_2, 1 \leq t \leq n-b+1, n \in \mathbb{Z}} \tau_b \left\| \hat{\theta}_{n, b, t}^{(i)} - E \left(\hat{\theta}_{n, b, \cdot}^{(i)} \right) \right\|_{4+2\xi} < \Delta \quad \text{and} \quad \sum_{k=1}^{\infty} \alpha^{\frac{\xi}{2+\xi}}(k) < \infty;$$

(iii) for any $1 \leq i \leq r$ and any sequence \tilde{b} such that $b_1 \leq \tilde{b} \leq b_2$ we have convergence

$$\left\| \frac{\tau_{\tilde{b}}}{n - \tilde{b} + 1} \sum_{1 \leq t \leq n - \tilde{b} + 1} \left(\hat{\theta}_{n, \tilde{b}, t}^{(i)} - E \left(\hat{\theta}_{n, \tilde{b}, \cdot}^{(i)} \right) \right) \right\|_2 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Then the estimator $\hat{\sigma}_n$ (given by (7)) is mean square consistent i.e.,

$$\hat{\sigma}_n \xrightarrow{L_2} \tilde{\sigma} \quad \text{as } n \rightarrow \infty.$$

3.2 Spectral density estimation for general APC time series

In this section we show how to apply results from the previous section in the problem of spectral density estimation of an APC time series with an almost periodic mean function,

under general assumption that the set Ψ is unknown and finite. At first we formulate the assumptions and introduce some additional notation. Next, we provide an auxiliary central limit theorem for the Fourier coefficients of the mean function (Section 3.2.1). Finally, we present the main results concerning consistency of our subsampling estimators (Sections 3.2.2 and 3.2.3).

Let $(X_{c_n}, \dots, X_{c_n+d_n-1})$ be a sample of the size d_n from an APC time series $\{X_t, t \in \mathbb{Z}\}$ with the mean function $\mu(t)$, where $\{c_n\}_{n \in \mathbb{Z}}$ and $\{d_n\}_{n \in \mathbb{Z}}$ are arbitrary sequences of positive integers such that $d_n \rightarrow \infty$ as $n \rightarrow \infty$. Let us recall that when the set Ψ is finite, the mean function $\mu(t)$ can be expressed as follows $\mu(t) = \sum_{\psi \in \Psi} m(\psi) \exp(i\psi t)$. Moreover, its estimator based on the sample $(X_{c_n}, \dots, X_{c_n+d_n-1})$ (see also Lenart (2011)) is of the form $\hat{\mu}_{d,c}(t) = \sum_{\psi \in \Psi} \hat{m}_{d,c}(\psi) \exp(i\psi t)$, where

$$\hat{m}_{d,c}(\psi) = \frac{1}{d_n} \sum_{j=c_n}^{c_n+d_n-1} X_j \exp(-i\psi j). \quad (8)$$

In the sequel we assume the following conditions

A1 $\{X_t, t \in \mathbb{Z}\}$ is an APC time series such that the sets Ψ and Λ are finite;

A2 $\{X_t, t \in \mathbb{Z}\}$ is α -mixing (for definition see e.g., Doukhan (1994));

A3 the complex measure r_0 , where $a(0, \tau) = \int_0^{2\pi} e^{i\xi\tau} r_0(d\xi)$ is absolutely continuous with respect to the Lebesgue measure.

Under the condition **A1** the considered APC time series is of the form $X_t = \mu(t) + Y_t$, where Y_t is a zero mean harmonizable sequence (in the sense of Loève (1963); see Definition 5.4 in Hurd and Miamee (2007)). Harmonizability of Y_t can be obtained using the same arguments as presented in Hurd and Miamee (2007) for PC sequences (see also Gladyshev

(1961)). The condition **A3** ensures that for any $\lambda \in \Lambda$ there exists the spectral density function $g_\lambda(\nu)$, $\nu \in [0, 2\pi)$. The set of spectral densities can be simply extended to the bifrequency square $[0, 2\pi)^2$ (see (2)) taking $P(\nu, \omega) = g_\lambda(\nu)$ for $|\nu - \omega| = \lambda$, $(\nu, \omega) \in [0, 2\pi)^2$ and $P(\nu, \omega) = 0$ otherwise (see Hurd (1991), Dehay and Hurd (1994), Hurd and Miamee (2007), Napolitano (2012), Lenart (2016)). Finally, the existence of complex measure r_0 follows from Lemma 7.3 in Lenart (2016).

3.2.1 Central limit theorem for the Fourier coefficients

Take $\nu, \omega \in [0, 2\pi)$. Below we state the asymptotic distribution of

$$\widehat{\mathbf{m}}_{d,c}(\nu, \omega) = (\text{Re}(\widehat{m}_{d,c}(\nu)), \text{Im}(\widehat{m}_{d,c}(\nu)), \text{Re}(\widehat{m}_{d,c}(\omega)), \text{Im}(\widehat{m}_{d,c}(\omega)))', \quad (9)$$

i.e., the estimator of the parameter

$$\mathbf{m}(\nu, \omega) = (\text{Re}(m(\nu)), \text{Im}(m(\nu)), \text{Re}(m(\omega)), \text{Im}(m(\omega)))'. \quad (10)$$

However, our main goal is to provide the exact form of the asymptotic covariance matrix. It turns out that its elements are linear combinations of the spectral density function. In a consecutive subsection we use this feature to construct the estimator of $P(\nu, \omega)$.

Theorem 3.2. *Assume that conditions **A1** – **A3** hold and that $\sup_t \|X_t\|_{2+\delta} < \infty$ and $\sum_{k=0}^{\infty} \alpha^{\frac{\delta}{2+\delta}}(k) < \infty$, for some $\delta > 0$. Then $\sqrt{d_n}(\widehat{\mathbf{m}}_{d,c}(\nu, \omega) - \mathbf{m}(\nu, \omega)) \xrightarrow{d} \mathcal{N}_4(\mathbf{0}, \mathbf{\Gamma}(\nu, \omega))$, where $\mathbf{\Gamma}(\nu, \omega) = [\gamma_{ij}(\nu, \omega)]_{4 \times 4}$ is of the form*

$$\pi \begin{pmatrix} g_0(\nu) + P_R(\nu, -\nu) & P_I(\nu, -\nu) & P_R(\nu, \omega) + P_R(\nu, -\omega) & -P_I(\nu, \omega) + P_I(\nu, -\omega) \\ P_I(\nu, -\nu) & g_0(\nu) - P_R(\nu, -\nu) & -P_I(\omega, \nu) + P_I(\omega, -\nu) & P_R(\nu, \omega) - P_R(\nu, -\omega) \\ P_R(\nu, \omega) + P_R(\nu, -\omega) & -P_I(\omega, \nu) + P_I(\omega, -\nu) & g_0(\omega) + P_R(\omega, -\omega) & P_I(\omega, -\omega) \\ -P_I(\nu, \omega) + P_I(\nu, -\omega) & P_R(\nu, \omega) - P_R(\nu, -\omega) & P_I(\omega, -\omega) & g_0(\omega) - P_R(\omega, -\omega) \end{pmatrix}$$

and $P_R(\cdot, \cdot) = \text{Re}[P(\cdot, \cdot)]$, $P_I(\cdot, \cdot) = \text{Im}[P(\cdot, \cdot)]$.

Note that the elements of the matrix $\mathbf{\Gamma}(\nu, \omega)$ are functions of $P(\cdot, \cdot)$ i.e., they are functions of the extension of the spectral density matrix function to the bifrequency square $[0, 2\pi)^2$. Moreover, taking linear combinations of elements of $\mathbf{\Gamma}(\nu, \omega)$, one can get $P(\nu, \omega)$. In particular, we have that

$$\gamma_{13}(\nu, \omega) + \gamma_{24}(\nu, \omega) = 2\pi P_R(\nu, \omega), \quad \gamma_{23}(\nu, \omega) - \gamma_{14}(\nu, \omega) = 2\pi P_I(\nu, \omega), \quad (11)$$

because $P_I(\omega, \nu) = -P_I(\nu, \omega)$ and $P_I(\omega, -\nu) = P_I(\nu, -\omega)$.

Our idea to obtain an estimate of $P(\nu, \omega)$ is to apply the subsampling method in order to approximate the covariance matrix of the distribution $\mathcal{N}_4(\mathbf{0}, \mathbf{\Gamma}(\nu, \omega))$. Then, having estimates of elements of the matrix $\mathbf{\Gamma}(\nu, \omega)$ and using the formulas (11), we can easily get the estimate of $P(\nu, \omega)$. In this way, in our approach we do not need to know the form of the set Ψ i.e., we do not need to know the Fourier frequencies of the mean function.

It is worth to indicate that in the literature dedicated to PC and APC processes, asymptotic normality results corresponding to Theorem 3.2, are treated as the preliminary step for construction of confidence intervals for parameters of interest. Since the asymptotic covariance matrix usually is of complicated form and depends on unknown parameters, in practice it is very difficult to estimate. Thus, in such situations to approximate quantiles of the asymptotic distribution, resampling methods are often applied (see e.g., Dudek et al. (2014); Lenart (2016); Dudek (2018) and references therein). Although, in our problem we need to estimate the asymptotic covariance matrix $\mathbf{\Gamma}(\nu, \omega)$ and not $\mathbf{m}(\cdot, \cdot)$, we also use a resampling method. More precisely, we adopt the idea of the subsampling estimator of the limiting variance examined in the stationary case in Carlstein (1986) and generalized to the nonstationary cases in Fukuchi (1999) and Politis et al. (1999). It is necessary to indicate that the consistency of subsampling is not a sufficient condition to show that the subsampling estimator of the limiting variance (or covariance matrix) converges to the

asymptotic variance (or the covariance matrix) and hence the mentioned consistency results are not sufficient for our purpose. Thus, in the next subsection we introduce the consistent subsampling estimators of the linear combinations of the elements of the matrix $\mathbf{\Gamma}(\nu, \omega)$ (see Theorem 3.2). To show their consistency we adopt ideas presented in Fukuchi (1999), Carlstein (1986) and Politis et al. (1999).

3.2.2 Subsampling consistency results

Let $(\nu, \omega) \in [0, 2\pi)^2$, $r = 4$ and (X_1, \dots, X_n) be an observed sample. Following the notation introduced in Section 3.1 we set $\boldsymbol{\theta} = \boldsymbol{\theta}(\nu, \omega) = \mathbf{m}(\nu, \omega)$, $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_n(\nu, \omega) = \hat{\mathbf{m}}_{n,1}(\nu, \omega)$ and $\hat{\boldsymbol{\theta}}_{n,b,t} = \hat{\boldsymbol{\theta}}_{n,b,t}(\nu, \omega) = \hat{\mathbf{m}}_{b,t}(\nu, \omega)$. In such a case the estimator (7) is of the form

$$\hat{\sigma}_n(\nu, \omega) = \sum_{i,j=1}^4 a_{ij} \left[\frac{1}{w} \sum_{b=b_1}^{b_2} \frac{\tau_b^2}{q} \sum_{t=1}^q \left(\hat{\theta}_{n,b,t}^{(i)}(\nu, \omega) - \hat{\theta}_{n,b,\cdot}^{(i)}(\nu, \omega) \right) \left(\hat{\theta}_{n,b,t}^{(j)}(\nu, \omega) - \hat{\theta}_{n,b,\cdot}^{(j)}(\nu, \omega) \right) \right]. \quad (12)$$

Theorem below states the mean square consistency of $\hat{\sigma}_n(\nu, \omega)$.

Theorem 3.3. *Assume that the conditions of Theorem 3.2 hold. Additionally, assume that there exists $\xi > 0$ such that $\sup_t \|X_t\|_{4+3\xi} < \infty$ and $\sum_{k=1}^{\infty} (k+1)^{c-2} \alpha^{\frac{\xi}{c+\xi}}(k) < \infty$, where c is the smallest even integer such that $c \geq 4 + 2\xi$. Then the conclusion of Theorem 3.1 holds for any real valued matrix $\mathbf{A} = [a_{ij}]_{i,j=1,\dots,4}$ and $\tilde{\sigma} = \sigma(\nu, \omega)$ given by (6) with $\sigma_{ij}(\nu, \omega) = \gamma_{ij}(\nu, \omega)$, i.e.,*

$$\hat{\sigma}_n(\nu, \omega) \xrightarrow{L_2} \sigma(\nu, \omega) = \sum_{i=1}^4 \sum_{j=1}^4 a_{ij} \gamma_{ij}(\nu, \omega).$$

Setting $a_{13} = 1$, $a_{24} = 1$ and $a_{ij} = 0$ otherwise, we get the estimate of $2\pi P_R(\nu, \omega)$, while for $a_{14} = -1$, $a_{23} = 1$ and $a_{ij} = 0$ otherwise the estimate of $2\pi P_I(\nu, \omega)$ (see 11). Example below illustrates performance of our estimator $\hat{\sigma}_n(\nu, \omega)$ and at the same time problem of leakage, which inspired us to propose in the next subsection a modification of $\hat{\sigma}_n(\nu, \omega)$.

Example 3.1. We consider a PC series $X_t = 5 \sin(2\pi t/30) + 2(\sin(2\pi t/12) + 21/20)\epsilon_t$, where ϵ_t is a normalized Gaussian white noise. Thus, X_t is composition of the PC white noise $2(\sin(\frac{2\pi t}{12}) + \frac{21}{20})\epsilon_t$ and the periodic deterministic component $5 \sin(2\pi t/30)$, with $\Lambda \subset \{\lambda_k = 2k\pi/12 : k = 0, \dots, 11\}$. It is easy to show that the spectral densities (denoted here by $g_k(\cdot)$) are constant on the interval $[0, 2\pi)$ and $g_0 = 6.41$, $g_1 = g_{11} = -4.2i$, $g_2 = g_{10} = -1$, $g_3 = \dots = g_9 = 0$ (see p. 170 in Hurd and Miamee (2007)). To construct our estimator of spectral density function we generated two samples of sizes $n = 1000$ and $n = 10000$, respectively. We considered the subsample length $b_1 = b_2 = b$, with $b \in \{\lfloor \sqrt{n} \rfloor, \lfloor 2\sqrt{n} \rfloor, \lfloor 3\sqrt{n} \rfloor\}$. The simulation results are presented in Figures 1-2. Since the main conclusions do not depend of the chosen values of parameters, we decided to restrict the number of figures only to those that are necessary to illustrate performance of our estimator.

In Figure 1 one may observe the estimated rescaled spectral density $2\pi g_0$ for different values of b . Let us recall that the main advantage of our approach is the ability to estimate the spectral densities without need of removing the mean function from the data. In this example the mean function is periodic and its true significant frequency ψ_0 is equal to $2\pi/30$. First of all, one may notice that independently of the subsample length b our estimator provides an accurate estimate of the rescaled spectral density value at ψ_0 . However, in the neighborhood of ψ_0 there are always two side-lobes representing the spectral leakage in the frequency domain. The size of lobes depends on the value of b i.e., on the subsample length. In Figure 2 we present the estimate of the spectral density functions on the bifrequency square. One may note that the real part of the estimate is concentrated on the main diagonal, while the imaginary part on two lines that are parallel to diagonal. Moreover, for the real and the imaginary parts, we observe leakage effect around vertical lines $\nu = \psi_0$, $\nu = 2\pi - \psi_0$ and

horizontal lines $\omega = \psi_0$, $\omega = 2\pi - \psi_0$.

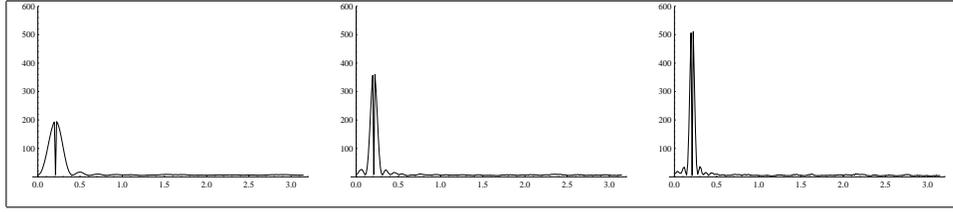


Figure 1: Estimated spectral density $2\pi g_0$ on interval $[0, \pi)$. Sample size $n = 1000$. From left results for subsample length b equal to $\lfloor \sqrt{n} \rfloor$, $\lfloor 2\sqrt{n} \rfloor$, $\lfloor 3\sqrt{n} \rfloor$, respectively.

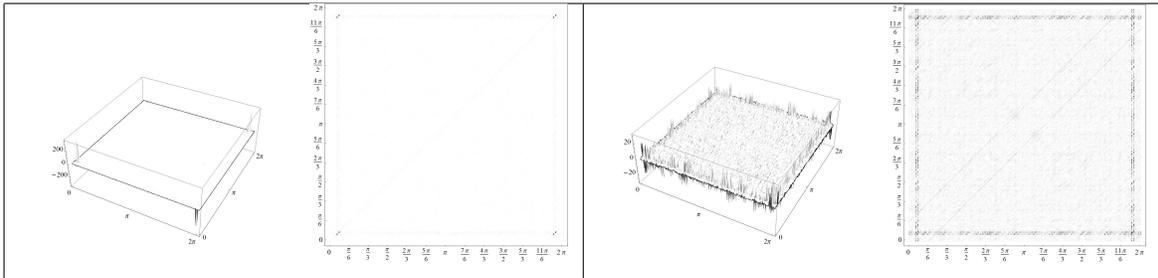


Figure 2: Estimates of $2\pi \text{Re}(P(\nu, \omega))$ (left-hand side) and $2\pi \text{Im}(P(\nu, \omega))$ (right-hand side) for $(\nu, \omega) \in [0, 2\pi)^2$ together with their projections on the xy -plane. Sample size $n = 1000$, $b = \lfloor 3\sqrt{n} \rfloor$.

As we mentioned above the observed leakage is a consequence of applying the rectangular window of the length b to the data. In the result we do not calculate just the Discrete Fourier Transform (DFT) of X_t but of the product of our signal X_t and the rectangular window function. The Fourier transform of the product is equal to the convolution of the Fourier transforms. Moreover, the Fourier transform of the rectangular window function is a sinc function. Finally, the convolution of the DFT of the signal with a sinc produces the observed side-lobes.

The leakage phenomenon is a typical problem in signal processing. Taking sample of the

finite size for Fourier transform is equivalent to application of a rectangular window on a data stream. In the literature one may find many methods designed to reduce the leakage effect. The most popular are based on non-rectangular windows. However, reducing the spectral leakage results in decreasing the spectral resolution. More details on that topic can be found for instance in the book Proakis and Manolakis (2009).

Example 3.2. This example shows some advantages of our estimator over the classical estimators of the spectral density function. In contrary to Example 3.1 we considered time series with nonconstant spectral density functions. These are two AR(2) signals of the following form:

$$\mathbf{M1} \quad X_t = 0.85X_{t-1} - 0.7X_{t-2} + \epsilon_t,$$

$$\mathbf{M2} \quad X_t - \mu(t) = 0.85(X_{t-1} - \mu(t-1)) - 0.7(X_{t-2} - \mu(t-2)) + \epsilon_t,$$

where ϵ_t is normalized Gaussian white noise and $\mu(t) = 1.6 \sin(\pi t/3)$. Model **M1** is the example of an AR(2) type stationary time series. Model **M2** is having periodic mean function. Both models have the same spectral density function which concentrates the spectral mass around frequency $\pi/3$. We chose those two simple series to illustrate problems of classical estimator of spectral density function when the mean function is nonzero and to show that our method provides comparable results to the classical approach in the standard stationary series.

We estimated the spectral density function using our estimator $\hat{\sigma}_n(\nu, \omega)$ (see Section 3.2.2) and the classical estimator based on smoothing the covariance function using the Bartlett window. We considered two sample sizes $n = 500$ and $n = 1000$ and block lengths $b_1 = \lfloor 5\sqrt[3]{n} \rfloor$, $b_2 = \lfloor 3\sqrt{n} \rfloor$. The length of smoothing window c was set $c = \lfloor (b_1 + b_2)/2 \rfloor$. Since the conclusions do not depend of the sample size we decided to present results only

for $n = 1000$ (see Figure 3). In panels on the left hand-side one may observe the generated realizations for models **M1** and **M2**. Panels on the right-hand side contain the theoretical spectral density function (dashed line) together with its estimates obtained using our non-modified subsampling estimator (solid line) and the classical estimator with Bartlett window of length c (shadowed area).

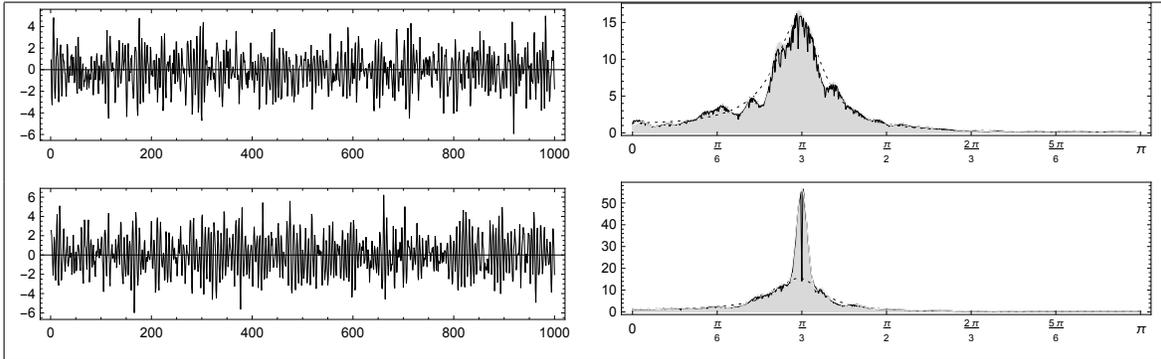


Figure 3: From top results for models **M1-M2**, respectively. Left panels: generated realizations; right panels: classical spectral density estimates obtained using Bartlett window of length c (shadowed area), subsampling estimates (solid line), theoretical spectral density functions (dashed line).

In the case of **M1** the mean function is equal to zero and hence assumption of the classical approach is fulfilled. As a result performances of the classical estimator and the subsampling estimator are comparable and obtained estimates are close to the true values. Model **M2** is having periodic mean function with the true significant frequency equal to $\pi/3$. Note that the subsampling estimator provides estimate close to the true value of the spectral density function at the frequency $\pi/3$. At the same time the classical estimator fails at this frequency. Moreover, for **M2** we observe the spectral leakage around $\pi/3$.

Examples above show that our estimator performs correctly and sometimes outperforms

the classical one. In contrary to the classical approach we do not need to remove the mean function from the data to be able to estimate the spectral density functions. That is very important in the real data applications since the period length of the mean function may not be known prior to the performed analysis on the data or the mean function may be almost periodic. In such cases removing the first-order component from the data may be very challenging.

It should be emphasized that the leakage phenomenon that we observed appears naturally when DFT is applied to the data. We are aware that there exist variety of approaches trying to deal with the leakage problem. The aim of this paper is not to provide the new solution for the leakage problem, however in the next section we propose the modified subsampling estimator that allows for significant reduction of the leakage effect.

3.2.3 Modified subsampling estimator

Below we introduce the modified subsampling estimator that significantly reduces the leakage problem. Instead of using $\hat{\sigma}_n(\nu, \omega)$ we apply $\hat{\sigma}_n(\nu + v_n, \omega + v_n)$, where v_n is a deterministic sequence of real numbers such that $v_n \rightarrow 0$ as $n \rightarrow \infty$. In other words, we consider the small neighborhood of (ν, ω) . At first we show mean square convergence of $\hat{\sigma}_n(\nu + v_n, \omega + v_n)$ to $\hat{\sigma}_n(\nu, \omega)$. In the second step to reduce leakage effect we use a finite number of estimators $\hat{\sigma}_n(\nu + v_n, \omega + v_n)$ (with different sequences v_n) to define the new estimator as their median.

Theorem 3.4. *Assume that the conditions of Theorem 3.3 hold. Additionally, assume that $v_n = Db_1^{-\rho}$, where $\rho < 1/4$ and $D \in \mathbb{R}$. Then, as $n \rightarrow \infty$ we have*

$$a) \hat{\sigma}_n(\nu + v_n, \omega + v_n) \xrightarrow{L_2} \sigma(\nu, \omega) = 2\pi P_R(\nu, \omega) \text{ for } A = [a_{ij}]_{4 \times 4} \text{ with } a_{13} = 1, a_{24} = 1 \\ \text{and } a_{ij} = 0 \text{ otherwise;}$$

b) $\widehat{\sigma}_n(\nu + v_n, \omega + v_n) \xrightarrow{L_2} \sigma(\nu, \omega) = 2\pi P_I(\nu, \omega)$ for $A = [a_{ij}]_{4 \times 4}$ with $a_{14} = -1$, $a_{23} = 1$ and $a_{ij} = 0$ otherwise.

Let $k > 0$ be a fixed odd integer and $\nu, \omega \in [0, 2\pi)$ be fixed frequencies. Moreover, let sequences $v_{n,1}, \dots, v_{n,k}$ be of the form $v_{n,i} = D_i b_1^{-\rho_i}$, where $0 < \rho_i < 1/4$ and $D_i \in \mathbb{R}$, for $i = 1, 2, \dots, k$. To reduce the leakage effect we will use the following estimator

$$\widehat{\sigma}_n^{med}(\nu, \omega) = \text{med} \{ \widehat{\sigma}_n(\nu + v_{n,1}, \omega + v_{n,1}), \dots, \widehat{\sigma}_n(\nu + v_{n,k}, \omega + v_{n,k}) \}, \quad (13)$$

where the symbol 'med' denotes the median. Below we state consistency of $\widehat{\sigma}_n^{med}(\nu, \omega)$ and we discuss the possible choice of the sequences $v_{n,i}$, $i = 1, 2, \dots, k$.

Theorem 3.5. *Under assumptions of Theorem 3.4 the conclusions of Theorem 3.4 hold for $\widehat{\sigma}_n^{med}(\nu, \omega)$ with any $v_{n,i} = D_i b_1^{-\rho_i}$, where $0 < \rho_i < 1/4$ and $D_i \in \mathbb{R}$, $i = 1, 2, \dots, k$.*

Remark 3.1. *Let $v_n = D b_1^{-\rho}$, where $0 < \rho < 1/4$ and $D \in \mathbb{R}$ and take any $\tilde{k} \in \mathbb{N}$. Then each sequence of the form $v_{n,i} = (i - \tilde{k} - 1)v_n/\tilde{k}$, where $i = 1, \dots, 2\tilde{k} + 1$ can be equivalently expressed as $v_{n,i} = D_i b_1^{-\rho}$ with $D_i = D(i - \tilde{k} - 1)/\tilde{k}$, $i = 1, 2, \dots, 2\tilde{k} + 1$ and hence Theorem 3.5 holds for the estimator*

$$\widehat{\sigma}_{n,\tilde{k}}^{med}(\nu, \omega) = \text{med} \left\{ \widehat{\sigma}_n(\nu + (i - \tilde{k} - 1)v_n/\tilde{k}, \omega + (i - \tilde{k} - 1)v_n/\tilde{k}) : i = 1, \dots, 2\tilde{k} + 1 \right\}.$$

Example 3.3. *(continuation of Example 3.1) We calculated the value of the estimator $\widehat{\sigma}_n^{med}$ using the same model and the same values of the parameters as were set in Section 3.2.2. To get $\widehat{\sigma}_n^{med}$ we took sequences $v_{n,1} = b^{-0.19}$, $v_{n,i} = b^{-0.19+0.01(i-1)}$, $i = 2, \dots, 6$ and $v_{n,i} = -v_{n,i-5}$, $i = 7, \dots, 11$. Results are presented in Figures 4-5. As before the obtained conclusions are independent of the values of the parameters and hence we do not provide all figures.*

In Figure 4 one may observe the estimated rescaled spectral density $2\pi g_0$ for different values

of b . Definitely, using the estimator $\widehat{\sigma}_n^{med}$ allowed for significant reduction of the leakage effect. The side-lobes are no longer present around $\lambda_0 = 2\pi/30$. The same conclusion can be drawn looking on the estimates on the bifrequency square (see Figure 5).

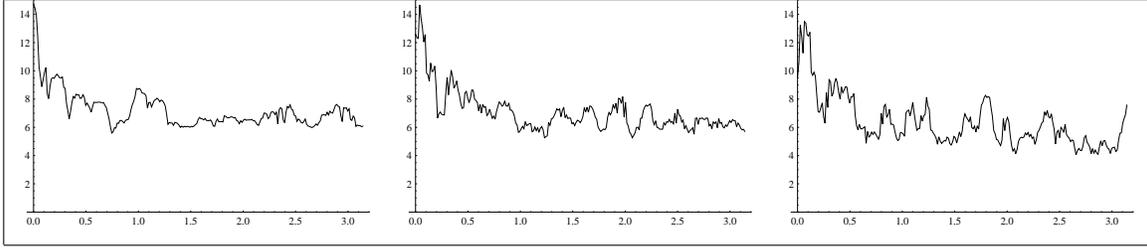


Figure 4: Estimated spectral density $2\pi g_0$ on interval $[0, 2\pi)$. Sample size $n = 1000$. From left results for subsample length b equal to $\lfloor \sqrt{n} \rfloor$, $\lfloor 2\sqrt{n} \rfloor$, $\lfloor 3\sqrt{n} \rfloor$, respectively.

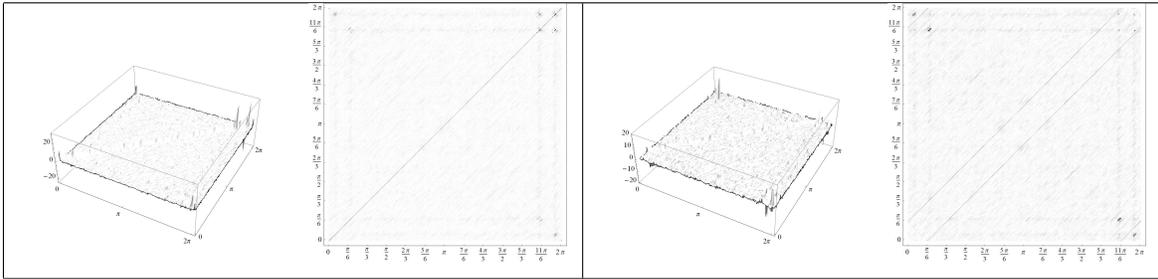


Figure 5: Estimates of $2\pi \text{Re}(P(\nu, \omega))$ (left-hand side) and $2\pi \text{Im}(P(\nu, \omega))$ (right-hand side) for $(\nu, \omega) \in [0, 2\pi)^2$ together with their projections on the xy -plane. Sample size $n = 1000$, $b = \lfloor 3\sqrt{n} \rfloor$.

4 Real data applications

In this section we present two motivating examples. We focus on economic datasets, however we would like to indicate that our methodology can be applied to any data containing

periodic or almost periodic structure. Thus, it can be also particularly useful for instance in telecommunication, mechanics, vibroacoustics, climatology and many others.

Below we show two very different applications of our results. At first, we consider an important problem of distinction between types of cyclical fluctuations, i.e., we discuss the type of cyclical fluctuations (deterministic or stochastic). For that purpose we consider data characterized by different types of cyclical fluctuations, i.e., business fluctuations and fluctuations related to trading-day effect. In the second part we show how to detect the second-order periodic structure in the considered data without need of estimation the almost periodic mean function.

4.1 Cyclicity of industrial production

In this subsection we consider the growth cycle for monthly industrial production (percentage change compared to same period in previous year) for three very well developed European economies: Germany, France and Italy. We use unadjusted data (i.e. neither seasonally adjusted nor calendar adjusted data) available on Eurostat webpage. The analyzed sequences cover period of 18 years in the case of France and Italy (January 1991 - December 2018, number of observations $n = 336$) and 17 years in the case of Germany (January 1992 - December 2018, $n = 324$). For Germany the year 1991 is not available on Eurostat.

Usually growth cycles for industrial production are considered to be covariance stationary and hence we assume that the spectral density is 0 beyond the main diagonal line on the bifrequency square, i.e., $\Lambda = \{0\}$. Moreover, we assume that their mean functions are almost periodic and the corresponding sets Ψ are unknown. Our main aim in this subsection is to recognize which of the observed cyclical fluctuations can be classified as deterministic

i.e., related to some frequency in the set Ψ and which can be classified rather as stochastic. Let us recall that the set Ψ contains all frequencies for which the corresponding Fourier coefficients of the mean function are nonzero. In the considered real data example these frequencies can be related to well known business cycle or so called calendar effects, which are caused by time-varying numbers of working days in consecutive months. Note that the typical (examined in the literature) business fluctuations are having length from 1.5 to 12 years and are related (in the case of monthly data) to frequencies from interval $[\frac{2\pi}{12 \times 12}, \frac{2\pi}{1.5 \times 12}] \approx [0.044, 0.349]$. On the other hand, the calendar effects are related to trading-day frequencies. The first ten predominant theoretical trading-day frequencies are 2.18733, 2.71093, 2.52506, 1.66374, 1.47786, 3.04865, 2.31823, 1.90852, 1.38492, 2.18603 (see Table 11.3 in Ladiray (2012)). Note that the tenth predominant frequency is almost the same as the first one and hence they are indistinguishable. Therefore in our analysis we consider only the first nine frequencies.

In Figure 6 the periodogram for each of considered sequences is presented. In all cases on the upper part of the frame the first nine theoretical trading-day frequencies are marked. Independently of the chosen country one may observe periodogram peaks at some of those theoretical trading-day frequencies and some peaks in the interval $[0.044, 0.349]$, which contains frequencies related to the business fluctuations. To investigate which of those peaks may correspond to frequencies in the set Ψ or equivalently which fluctuations we recognize as deterministic, in Figure 6 we present spectral density estimates obtained using our subsampling estimator (12) and the classical spectral density estimator with Bartlett window. The subsampling estimate is less smooth than the classical one. Moreover, it has local minima with recognizable leakage effect in their neighbourhood at the trading-day frequencies for which we observed before the periodogram peaks. This suggests that fluctuations

related to those frequencies can be considered as deterministic. Moreover, as we discussed in Section 3, the classical estimator fails in that case, i.e., it provides incorrect estimates at deterministic frequencies (see Figure 6). Note also that in the interval $[0.044, 0.349]$ (corresponding to business fluctuations) we do not observe leakage effect. For each considered country both estimates of the spectral density function (classical and subsampling one) have similar shapes in that frequency interval, which suggests that the business fluctuations should be rather considered as stochastic. This fact confirms the classical paradigm that time-varying amplitude and phase of business cycle fluctuations cannot be described just by an almost periodic mean function and that their presence is reflected in the higher order moments.

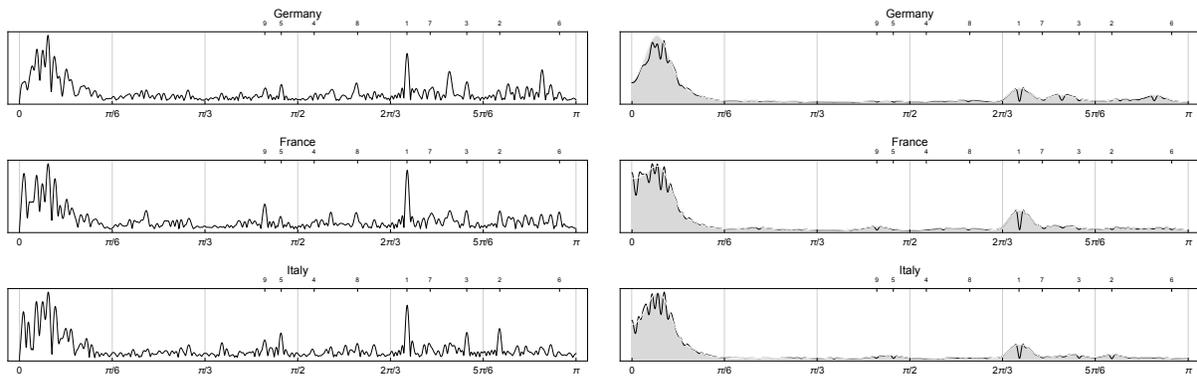


Figure 6: Left-hand side: periodograms. Right-hand side: subsampling estimator of spectral density (12) with $b_1 = \lfloor 5\sqrt[3]{n} \rfloor$ and $b_2 = \lfloor 3\sqrt{n} \rfloor$ (solid line) together with classical estimator with Bartlett window and $c = \lfloor (b_1 + b_2)/2 \rfloor$ (shaded area).

4.2 Seasonal volatility of electricity spot prices

Energy market data are recognized to have periodic first and second moments (Broszkiewicz-Suwaj et al. (2004), Dudek et al. (2015)). However, so far detection of periodicity in the

second moments was possible only when the mean function was removed from the data. In this subsection we study a typical energy market dataset. These are hourly observations from electricity spot prices in EUR/MWh of energy for Lithuania traded on the Nord Pool Spot exchange from 2013 to 2018 (52 584 observations). Considered data are having two types of periodicity in the mean function: daily and weakly (see e.g., Weron (2006), Nowotarski and Weron (2015)). Thus, for the mean function it is natural to set the period length $24 \times 7 = 168$ (hours). In consequence to estimate the periodic mean function we need to deal with 168 unknown parameters (Fourier coefficients). However, for our subsampling approach demeaning data is no longer necessary. We apply it to show that we are able to detect periodicity of the autocovariance function (with period length equal to 24 hours).

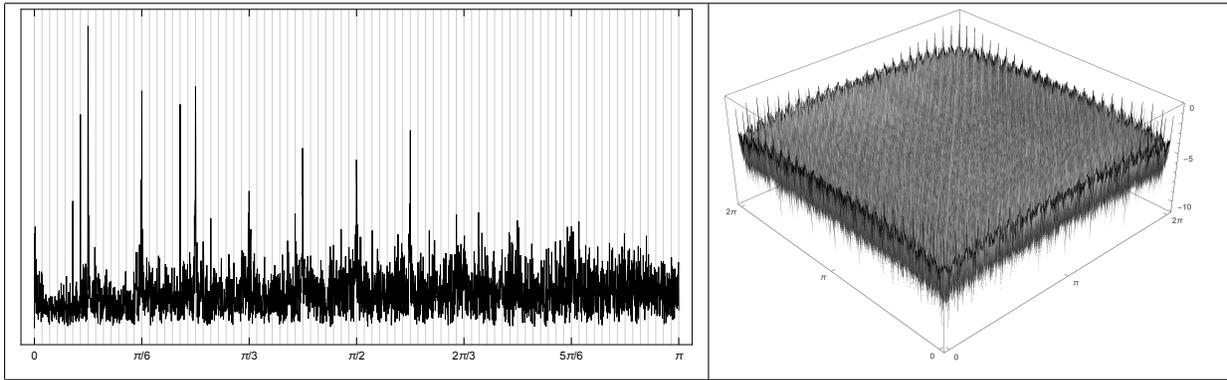


Figure 7: Left-hand side: periodogram for electricity spot prices. Right-hand side: subsampling estimator of spectral density in the log scale with $b_1 = b_2 = \lfloor 20\sqrt[3]{n} \rfloor$.

In Figure 7 we present the periodogram and the subsampling spectral density estimator (12) with $b_1 = b_2 = \lfloor 20\sqrt[3]{n} \rfloor$. To improve the clarity in the latter case we decided to use the log scale. However, due to the large number of peaks (on the bifrequency square) it is still difficult to decide where the spectral density is nonzero. Thus, we use the smoothed

estimator i.e., our modified subsampling estimator $\widehat{\sigma}_{n,\tilde{k}}^{med}$ (see Remark 3.1 in Section 3.2.3). We consider two cases (see Figure 8) by setting $v_n = 0.4/\sqrt[15]{n}$, $\tilde{k} = 30$ and $v_n = 0.3/\sqrt[15]{n}$, $\tilde{k} = 20$. One may notice that independently of the chosen value of parameters, the spectral density is nonzero on lines that are parallel to the main diagonal. Moreover, the distance between those lines allows us to state that the autocovariance function is periodic with period length equal to 24 hours.

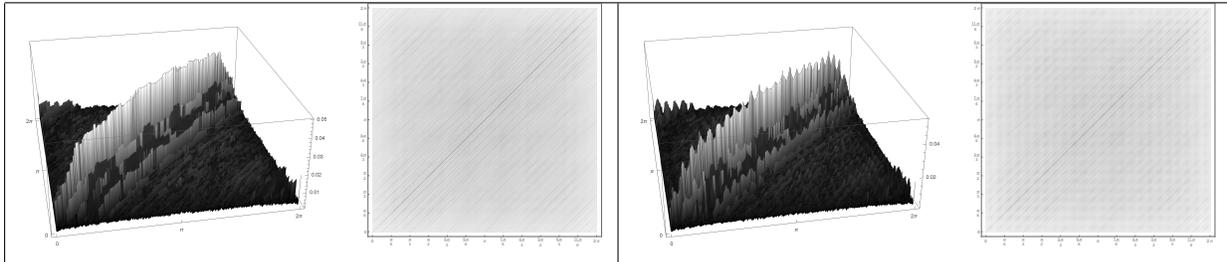


Figure 8: Modified subsampling estimators defined in Remark 3.1 with $\tilde{k} = 30$, $v_n = 0.4/\sqrt[15]{n}$ (left-hand side) and with $\tilde{k} = 20$, $v_n = 0.3/\sqrt[15]{n}$ (right-hand side) and their projections on the xy-plane.

5 Conclusions

In this paper we proposed a new nonparametric approach for spectral density estimation, which does not require demeaning of the data and is based on the subsampling method. Our estimator was constructed for almost periodically correlated sequences, which are an important subclass of nonstationary time series. Additionally, we introduced the modified estimator, which allows to reduce the leakage effect. We showed the ability of our approach to analyze the cyclical fluctuations and to determine the period length on the examples of typical economic datasets. Our method may be used to construct testing tools to recognize

the type of cyclical fluctuations i.e., to decide if they are described by the periodic mean function or by the the higher-order moments.

Presented results are providing new open problems concerning subsampling estimation for the higher-order moments. Since the classical approach for the higher-order spectral estimation also requires removing the mean function from the data, we believe that ideas presented in this paper can be adapted in higher-order spectral estimation problem under nonzero mean assumption. This and other related issues are the subject of the current research of the authors.

SUPPLEMENTARY MATERIAL

Companion Document: contains all proofs of the paper. (pdf file)

References

- Antoni, J. (2009). Cyclostationarity by examples. *Mech. Syst Sig Process.* 23(4), 987–1036.
- Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*. SIAM.
- Broszkiewicz-Suwaj, E., A. Makagon, R. Weron, and A. Wyłomańska (2004). On detecting and modeling periodic correlation in financial data. *Physica A* 336, 196–205.
- Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Annals of Statist.* 14, 1171–1179.
- Corduneanu, C. (1989). *Almost Periodic Functions*. Chelsea, New York.
- Dehay, D., A. Dudek, and J. Leśkow (2014). Subsampling for continuous-time nonstationary stochastic processes. *J. Stat. Plan. Inf.* 150, 142–158.

- Dehay, D. and H. Hurd (1994). Representation and estimation for periodically and almost periodically correlated random processes. In W. Gardner (Ed.), *Cyclostationarity in Communications and Signal Processing*, pp. 295–329. IEEE Press.
- Doukhan, P. (1994). *Mixing. Properties and Examples*. Springer-Verlag, New York.
- Dudek, A. (2018). Block bootstrap for periodic characteristics of periodically correlated time series. *Journal of Nonparametric Statistics* 30(1), 87–124.
- Dudek, A., H. Hurd, and W. Wójtowicz (2015). Parma models with applications in r. In F. e. a. Chaari (Ed.), *Cyclostationarity: Theory and Methods - II*, pp. 131–153. Springer International Publishing Switzerland.
- Dudek, A. and L. Lenart (2017). Subsampling for nonstationary time series with non-zero mean function. *Statistics and Probability Letters* 129, 252–259.
- Dudek, A., J. Leśkow, E. Paparoditis, and D. Politis (2014). A generalized block bootstrap for seasonal time series. *J. Time Ser. Anal.* 35, 89–114.
- Fukuchi, J. (1999). Subsampling and model selection in time series analysis. *Biometrika* 86(3), 591–604.
- Gardner, W., A. Napolitano, and L. Paura (2006). Cyclostationarity: half a century of research. *Signal Processing* 86, 639–697.
- Gladyshev, E. (1961). Periodically correlated random sequences. *Soviet mathematics* 2, 385–388.
- Hurd, H. (1991). Correlation theory of almost periodically correlated processes. *J. Multivariate Anal.* 37, 24–45.

- Hurd, H. and A. Miamee (2007). *Periodically Correlated Random Sequences: Spectral Theory and Practice*. Wiley, Hoboken, New Jersey.
- Ladiray, D. (2012). Theoretical and real trading-day frequencies. In W. Bell, S. Holan, and T. McErloy (Eds.), *Economic time Series: Modeling and Seasonality*, pp. 255–279. Chapman and Hall.
- Lenart, L. (2011). Asymptotic distributions and subsampling in spectral analysis for almost periodically correlated time series. *Bernoulli* 17(1), 290–319.
- Lenart, L. (2013). Non-parametric frequency identification and estimation in mean function for almost periodically correlated time series. *J. Multivariate Anal.* 115, 252–269.
- Lenart, L. (2016). Generalized resampling scheme with application to spectral density matrix in almost periodically correlated class of time series. *Journal of Time Series Analysis* 37, 369–404.
- Lenart, L. (2018). Generalized subsampling procedure for non-stationary time series. *Electronic Journal of Statistics* 12(2), 3875–3907.
- Lii, K. and M. Rosenblatt (2006). Estimation for almost periodic processes. *Ann. Statist.* 34(3), 1115–1139.
- Loève, M. (1963). *Probability Theory*. Van Nostrand, Princeton, NJ.
- Napolitano, A. (2012). *Generalizations of Cyclostationary Signal Processing: Spectral Analysis and Applications*. Wiley-IEEE Press.
- Napolitano, A. (2016). Cyclostationarity: New trends and applications. *Signal Processing* 120, 385–408.

- Napolitano, A. and C. Spooner (2000). Median-based cyclic polyspectrum estimation. *IEEE Transactions on Signal Processing* 48(5), 1462–1466.
- Nematollahi, A. and T. Rao (2005). On the spectral density estimation of periodically correlated (cyclostationary) time series. *Sankhya: The Indian Journal of Statistics* 67(3), 568–589.
- Nowotarski, J. and R. Weron (2015). Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Computational Statistics* 30, 791–803.
- Politis, D., J. Romano, and M. Wolf (1999). *Subsampling*. Springer-Verlag, New York.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, London.
- Proakis, J. and D. Manolakis (2009). *Digital Signal Processing*. Pearson Education.
- Stoica, P. and R. Moses (2005). *Spectral Analysis of Signals*. Pearson Prentice-Hall.
- Tewes, J., D. Politis, and D. Nordman (2019). Convolved subsampling estimation with applications to block bootstrap. *Annals of Statistics* 47(1), 468–496.
- Žurbenko, J. G. (1986). *The Spectral Analysis of Time Series*. North-Holland, Amsterdam.
- Weron, R. (2006). *Modeling and Forecasting Electricity Loads and Prices. A Statistical Approach*. John Wiley & Sons, West Susse.
- Yavorskyj, I., R. Yuzefovych, I. Kravets, and I. Matsko (2012). Properties of characteristics estimators of periodically correlated random processes in preliminary determination of the period of correlation. *Radioelectronics and Communications Systems* 55(8), 335–348.