Bootstrapping the autocovariance of PC time series - a simulation study

Anna E. Dudek

Institut de Recherche Mathématique de Rennes, Université Rennes 2, Rennes, France,

AGH University of Science and Technology, Dept. of Applied Mathematics

al. Mickiewicza 30, 30-059, Krakow, Poland

aedudek@agh.edu.pl

Paweł Potorski

AGH University of Science and Technology, Dept. of Applied Mathematics

al. Mickiewicza 30, 30-059, Krakow, Poland

Abstract

In this paper a simulation comparison of the bootstrap confidence intervals for the coefficients of the autocovariance function of a periodically correlated time series is provided. Two bootstrap methods are used: the circular version of the Extension of Moving Block Bootstrap and the circular version of the Generalized Seasonal Block Bootstrap. The bootstrap pointwise and simultaneous confidence intervals for the real and the imaginary parts of the Fourier coefficients of the autocovariance function are constructed. The actual coverage probabilities, the average lengths and the average upper and lower quantiles values are calculated. A heuristic method of the block length choice is proposed.

1 Introduction

Studies of periodically correlated (PC) processes were started in 1961 by Gladyshev [14]. Time series X_t, $t \in \mathbb{Z}$ with finite second moments is called PC with the period d, if it has periodic mean and covariance function, i.e.,

$$\mathrm{E}\left(X_{t+d}\right) = \mathrm{E}\left(X_t\right),\ \mathrm{Cov}\left(X_t, X_s\right) = \mathrm{Cov}\left(X_{t+d}, X_{s+d}\right).$$

For more details concerning PC processes we refer the reader to [15].

Over the last 60 years the theory of PC processes has developed fast and found applications in many branches of vibroacoustics, mechanics, signal analysis, hydrology, climatology and econometrics (e.g., energy markets). Many motivating examples can be found in [1], [13], [15], [20] and [23]. The wide range of possible applications resulted in thousands of papers. Unfortunately, the analysis of PC processes is very difficult. One of the major problems is the estimation of the asymptotic covariance matrix for parameters of interest. In practice it is almost impossible and to construct confidence intervals and hence resampling methods need to be used.

The idea of resampling techniques is to approximate the distribution of the statistics of interest. Nowadays, bootstrap is the most popular resampling method. It was introduced in 1979 by Efron ([11]). At the beginning it was designed for i.i.d. data. In such cases the bootstrap sample is created by random sampling with the replacement of observations from the original sample. However, this approach cannot be applied for dependent data. In 1986 Carlstein ([3]) proposed to select randomly blocks of observations. This allows us to preserve the dependence structure of the data inside the blocks. If the data are weakly dependent, i.e., observations that are far from each other are almost independent, this approach provides consistent estimators for many characteristics of stationary and also nonstationary

time series. The number of block bootstrap techniques is constantly growing. New methods or modifications of existing techniques appear to provide estimators with lower bias, better rates of convergence or that mimic some specific structure of the data such as periodicity. In this paper we consider two block methods: the Extension of Moving Block Bootstrap (EMBB) and the Generalized Seasonal Block Bootstrap (GSBB). The EMBB is a modification of Carlstein's approach based on the Moving Block Bootstrap (MBB) method. It allows us to select any block of observations from the sample, while Carlstein's idea was to choose only non-overlapping blocks. On the other hand, the GSBB is designed for periodic data. To preserve periodicity during the block selection process, the set of blocks is restricted and varies in each step of the algorithm. The detailed description of the EMBB and the GSBB is presented in Section 2.

In this paper we focus on the second-order analysis of a periodically correlated (PC) time series. Time series X_t, $t \in \mathbb{Z}$ with finite second moments is called PC with the period d, if it has periodic mean and covariance function

$$\mathrm{E}\left(X_{t+d}\right) = \mathrm{E}\left(X_t\right), \, \mathrm{Cov}\left(X_t, X_s\right) = \mathrm{Cov}\left(X_{t+d}, X_{s+d}\right).$$

For more details concerning PC processes we refer the reader to [15].

Proper detection of so-called second-order frequencies is crucial in many applications like e.g., diagnostics of rotating machines ([7], [6]) and the analysis of the human walk ([10]). These problems are of great importance and mistakes can be very costly, for example a large machine in a factory might be needlessly stopped for repair. Thus, we decided to perform a study that may help practitioners to choose the optimal bootstrap approach. Recently there appeared a few theoretical results for constructing bootstrap consistent pointwise and

simultaneous confidence intervals for the coefficients of the autocovariance function of a PC time series. Unfortunately, the provided tools are not complete. There is no method for choosing the block length. Moreover, there is no study comparing properties of different approaches. For our work we decided to use computational resources available in the Polish Grid Infrastructure PL-Grid. That allowed us to consider a wide range of block length values, from small to very large sample sizes, and different period lengths. In the sequel we try to provide answers for often posed questions like:

- does using the EMBB have any advantage over the GSBB;
- do some existing methods of block length choice for stationary time series also apply for our nonstationary case;
- does preserving periodic structure by the GSBB improve, compared with the EMBB, the coverage of the confidence intervals;
- how are the coverage probabilities affected by non-optimal block length choices;
- can the same block length be used to construct all pointwise confidence intervals for the considered frequencies;
- are simultaneous confidence intervals less sensitive to non-optimal block length choice than pointwise ones.

This paper is organized as follows. In Section 2 the necessary notation and definitions are introduced. Moreover, block bootstrap algorithms and bootstrap consistency results are recalled. Section 3 is dedicated to the results of the simulation study. The performance of the two considered block bootstrap methods is compared. Different sample sizes and block lengths are used. Pointwise and simultaneous confidence intervals for the coefficients of the

autocovariance function are constructed. Additionally, the problem of the block length choice is discussed and a new heuristic approach is proposed. In Section 4 a short summary of the results is provided.

2 Problem formulation

Let $\{X_i, i \in \mathbb{Z}\}$ be a PC time series with period d. From now on we assume that $\mathrm{E}(X_t) \equiv 0$. Moreover, let $B(t,\tau) = \mathrm{Cov}(X_t, X_{t+\tau})$ be its autocovariance function. The variables t and τ represent time and shift, respectively. Note that $B(t,\tau)$ is periodic in t. Its analysis in the frequency domain is performed using the following Fourier decomposition

$$B(t,\tau) = \sum_{\lambda \in \Lambda_\tau} a(\lambda,\tau) \exp(i\lambda t),$$

where $\Lambda_\tau = \{\lambda : a(\lambda,\tau) \neq 0\}$. The set Λ_τ is finite and is a subset of the set $\overline{\Lambda} = \{2k\pi/d, k = 0,\ldots,d-1\}$. This means that there is a finite number of the second-order significant frequencies. To detect them it is enough to point out the nonzero Fourier coefficients of $B(t,\tau)$. For that purpose one needs to construct confidence intervals for $a(\lambda,\tau)$. Below we recall results from literature to show how difficult this task is.

Let $X_1,\ldots,X_n$ be a sample from the considered time series. Without loss of generality we assume that $\tau \geq 0$. An estimator of $a(\lambda,\tau)$ is of the form

$$\widehat{a}_n(\lambda,\tau) \;=\; \frac{1}{n}\sum_{t=1}^{n-\tau} X_t X_{t+\tau} \exp(-i\lambda t).$$

Moreover, the asymptotic results that we present below require the mixing assumptions. To be precise X_t is assumed to be α-mixing i.e., $\alpha_X(k) \to 0$ as $k \to \infty$, where

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

and $\mathcal{F}_X(-\infty, t) = \sigma(\{X_s : s \leq t\})$, $\mathcal{F}_X(t+k, \infty) = \sigma(\{X_s : s \geq t+k\})$. If $\alpha_X(k) = 0$ it means that the observations that are k time units apart are independent. In general mixing assumptions are introduced to ensure that observations that are far from each other are 'almost' independent. An easy example of α-mixing process is a m-dependent time series.

Let us introduce some additional notation. By $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ we denote r-dimensional vectors of frequencies and shifts of the form $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r)^{'}$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_r)^{'}$. Additionally, $a(\boldsymbol{\lambda}, \boldsymbol{\tau}) = \big(\Re(a(\lambda_1, \tau_1)), \Im(a(\lambda_1, \tau_1)), \ldots, \Re(a(\lambda_r, \tau_r)), \Im(a(\lambda_r, \tau_r))\big)^{'}$ and $\widehat{a}_n(\boldsymbol{\lambda}, \boldsymbol{\tau})$ is its estimator. By $\Re(z)$ and $\Im(z)$ we denote the real and the imaginary part of the complex number z.

Theorem 1 below states asymptotic normality of $\widehat{a}_n(\boldsymbol{\lambda}, \boldsymbol{\tau})$. For the proof we refer the reader to [26] (Theorem 2.6) and [18] (Theorem 1).

Theorem 1 *Assume that* $\{X_t, t \in \mathbb{Z}\}$ *is a PC* α*-mixing time series with* $\mathrm{E}(X_t) \equiv 0$ *and* WP(4) *such that:*

(i) $\sup_{t \in \mathbb{Z}} \mathrm{E}|X_t|^{4+2\delta} < \infty$ *for some* $\delta > 0$,

(ii) $\sum_{\tau=1}^{\infty} \tau \alpha_X^{\delta/(2+\delta)}(\tau) < \infty$.

Then

$$\sqrt{n}\left(\widehat{a}_n\left(\boldsymbol{\lambda}, \boldsymbol{\tau}\right) - a\left(\boldsymbol{\lambda}, \boldsymbol{\tau}\right)\right) \xrightarrow{d} N_{2r}\left(0, \Sigma(\boldsymbol{\lambda}, \boldsymbol{\tau})\right),$$

where $\Sigma(\boldsymbol{\lambda}, \boldsymbol{\tau}) = [\sigma_{ef}]_{e,f=1,\ldots,2r}$,

$$\sigma_{ef} = \begin{cases} \frac{1}{d}\sum_{s=1}^{d}\sum_{k=-\infty}^{\infty} C(s,k)\cos(\lambda_e s)\cos(\lambda_f(s+k)) & \textit{for} \quad e = 2g_1 - 1, f = 2g_2 - 1, \\ \frac{1}{d}\sum_{s=1}^{d}\sum_{k=-\infty}^{\infty} C(s,k)\sin(\lambda_e s)\sin(\lambda_f(s+k)) & \textit{for} \quad e = 2g_1, f = 2g_2, \\ -\frac{1}{d}\sum_{s=1}^{d}\sum_{k=-\infty}^{\infty} C(s,k)\sin(\lambda_e s)\cos(\lambda_f(s+k)) & \textit{for} \quad e = 2g_1, f = 2g_2 - 1, \\ -\frac{1}{d}\sum_{s=1}^{d}\sum_{k=-\infty}^{\infty} C(s,k)\cos(\lambda_e s)\sin(\lambda_f(s+k)) & \textit{for} \quad e = 2g_1 - 1, f = 2g_2, \end{cases}$$

where $C(s,k) = \mathrm{Cov}\left(X_s X_{s+\tau_e}, X_{s+k} X_{s+k+\tau_f}\right)$, $g_1, g_2 = 1, \ldots, r$ *and* $\det(\Sigma(\boldsymbol{\lambda}, \boldsymbol{\tau})) \neq 0$.

WP(k) denotes weakly periodic times series of order k with period d. In other words time series is WP(k) when its k-th moments are periodic.

Theorem 1 enables construction of the asymptotic pointwise confidence intervals for $a(\lambda, \tau)$. Unfortunately, the asymptotic covariance matrix $\Sigma(\boldsymbol{\lambda}, \boldsymbol{\tau})$ is very difficult to estimate. Thus, in practice resampling methods are often used to approximate the quantiles of the asymptotic distribution. Additionally, so-called percentile bootstrap confidence intervals do not require estimation of the covariance matrix (see e.g., [12]). Since PC time series are an example of the nonstationary processes, block bootstrap methods need to be used for them to keep the dependence structure contained in the data. So far consistency of two bootstrap techniques was shown for $a(\lambda, \tau)$. These are the EMBB and the GSBB. The EMBB was proposed in [7] and [8]. It is the modification of the MBB. The MBB was introduced independently in [16] and [19] and was designed for stationary time series. However, it turned out that it can be successfully applied in some nonstationary cases ([25], [4] and [5]). The main disadvantage of the MBB and its modifications like the EMBB in the context of PC time series, is the fact that these methods do not preserve the periodic structure of the original data. On the other hand, the GSBB was designed for periodic processes. The method was proposed by Dudek et al. in [9]. Applied for a PC time series it perfectly retains the periodic structure contained

in the data.

In this paper we will use a modification of the EMBB and the GSBB, whose idea is to consider data as wrapped on the circle. It was introduced in [21] to reduce edge effects caused by the MBB and it was called the Circular Block Bootstrap (CBB). The CBB guarantees that each observation is present in the same number of blocks which is not the case of the MBB, the EMBB and the GSBB. Below we recall the algorithms of the circular versions of the EMBB (cEMBB) and the GSBB (cGSBB).

Let B_i, $i = 1, \ldots, n$ be the block of observations from our sample $X_1, \ldots, X_n$, that starts with observation X_i and has the length $b \in \mathbb{N}$, i.e., $B_i = (X_i, \ldots, X_{i+b-1})$. If $i + b - 1 > n$ then the missing part of the block is taken from the beginning of the sample and we get $B_i = (X_i, \ldots, X_n, X_1, \ldots, X_{b-n+i-1})$ for $i = n - b + 2, \ldots, n$. To apply the cEMBB and the cGSBB to a PC data, we need to assume that the sample size n is an integer multiple of the period length d ($n = wd$). If $wd < n \leq (w+1)d$ then the observations $X_i, i = wd + 1, \ldots, n$ need to be removed from the considered dataset.

cEMBB Algorithm

Since the cEMBB approach is based on the CBB algorithm, as first we recall the CBB method.

1. Choose a block size $b < n$. Then our sample can be divided into l blocks of length b and the remaining part is of length r, i.e., $n = lb + r$, $r = 0, \ldots, b - 1$.

2. For $t = 1, b + 1, 2b + 1, \ldots, lb + 1$, let $B_t^{*CBB} = \left(X_t^{*CBB}, \ldots, X_{t+b-1}^{*CBB}\right) = B_{k_t}$, where k_t are i.i.d. random variables drawn from a discrete uniform distribution $P(k_t = s) = 1/n$ for $s = 1, \ldots, n$.

3. Join the selected $l+1$ blocks $\left(B_1^{*CBB}, \ldots, B_{l+1}^{*CBB}\right)$ and take the first n observations to get the bootstrap sample $\left(X_1^{*CBB}, \ldots, X_n^{*CBB}\right)$ of the same length as the original one.

The cEMBB algorithm is a simple modification of the CBB one.

1. Define a bivariate series $Y_i = (X_i, i)$ and then do the CBB on the sample $(Y_1, \ldots, Y_n)$ to obtain $(Y_1^*, \ldots, Y_n^*)$.

Note that in the second coordinate of the series $Y_1^*, \ldots, Y_n^*$ we preserve the information on the original time indices of chosen observations. This simple modification of the CBB approach allows to construct the consistent estimators for the Fourier coefficients of the autocovariance function of a PC time series.

cGSBB Algorithm

The cGSBB algorithm differs from the CBB only in the second step, i.e.,

2'. For $t = 1, b+1, 2b+1, \ldots, lb+1$, let

$$B_t^{*cGSBB} = \left(X_t^{*cGSBB}, \ldots, X_{t+b-1}^{*cGSBB}\right) = B_{k_t} \tag{1}$$

where k_t is i.i.d. random variables drawn from a discrete uniform distribution $P\left(k_t = t + vd\right) = 1/w$ for $v = 0, 1, \ldots, w-1$.

Constructing the CBB or the cEMBB sample we are choosing among n blocks. The cGSBB case is much more subtle. The idea is to perfectly mimic the periodic structure of the data. Thus, during the block selection process the additional criteria need to be fulfilled.

Remark It is well known that a PC time series with period d can be equivalently expressed as d-variate stationary time series (see e.g., [15]). Moreover, bootstrap methods and their

properties are well investigated in the stationary case (see [17]) and hence it may seem useless to consider bootstrap for univariate periodic time series (especially when one is interested in estimation of the mean). However, as pointed out in [8] (Section 2.4) bootstrap approach via one dimensional time series is preferable. There are several reasons for that. The most important is that the EMBB and the GSBB allow to use blocks that do not contain an integer number of periods. Moreover, these blocks can start with any observation. None of those two properties hold when one wants to use the multivariate stationary representation of PC time series. As a result there are no bootstrap approaches designed for d-variate stationary time series that are equivalent to the EMBB and the GSBB for univariate PC time series.

Bootstrap confidence intervals

The consistency of the cGSBB and the cEMBB for $a(\boldsymbol{\lambda}, \boldsymbol{\tau})$ was shown in [10] and [7], respectively. Since both results were obtained under the same assumptions, in order to recall them, to simplify notation, we use $\widehat{a}_n^*(\boldsymbol{\lambda}, \boldsymbol{\tau})$ instead of $\widehat{a}_n^{*cEMBB}(\boldsymbol{\lambda}, \boldsymbol{\tau})$ and $\widehat{a}_n^{*cGSBB}(\boldsymbol{\lambda}, \boldsymbol{\tau})$. This means that whenever $\widehat{a}_n^*(\boldsymbol{\lambda}, \boldsymbol{\tau})$ is used all statements are valid for both bootstrap approaches.

Theorem 2 *Assume that* $\{X_t, t \in \mathbb{Z}\}$ *is an* α*-mixing PC time series with* $\mathrm{E}(X_t) \equiv 0$ *and* WP(4) *such that:*

(i) $\sup_{t \in \mathcal{Z}} \mathrm{E}|X_t|^{8+2\delta} < \infty$ *for some* $\delta > 0$,

(ii) $\sum_{\tau=1}^{\infty} \tau \alpha_X^{\delta/(4+\delta)}(\tau) < \infty$.

If $b \to \infty$ *as* $n \to \infty$ *such that* $b = o(n)$, *then*

$$\rho\Big(\mathcal{L}\left(\sqrt{n}\left(\widehat{a}_n(\boldsymbol{\lambda}, \boldsymbol{\tau}) - a(\boldsymbol{\lambda}, \boldsymbol{\tau})\right)\right), \mathcal{L}^*\left(\sqrt{n}\left(\widehat{a}_n^*(\boldsymbol{\lambda}, \boldsymbol{\tau}) - \mathrm{E}^*\widehat{a}_n^*(\boldsymbol{\lambda}, \boldsymbol{\tau})\right)\right)\Big) \xrightarrow{p} 0,$$

where ρ *is any distance metricizing weak convergence of probability measures on* $\mathbb{R}^{2r}$.

$\mathcal{L}(\cdot)$ denotes probability low and $\mathcal{L}^*(\cdot)$ is its bootstrap counterpart. Moreover, E^* is the conditional expectation given the sample $(X_1, \ldots, X_n)$.

Using Theorem 2 one may construct bootstrap pointwise confidence intervals for the real and the imaginary part of $a(\lambda, \tau)$. For the sake of simplicity below we describe the basic ideas only for $\Re a(\lambda, \tau)$. As we mentioned before the asymptotic variance is very difficult to estimate. Since the construction of the bootstrap percentile confidence intervals does not require variance estimation, this kind of interval is the most often used for different characteristics of PC time series. To get it for a fixed frequency λ and shift τ we define the following statistic

$$K_{CBB}(x) = P^*\left(\sqrt{n}\Re\left(\widehat{a}_n^{*cEMBB}(\lambda,\tau) - \mathrm{E}^*\left(\widehat{a}_n^{*cEMBB}(\lambda,\tau)\right)\right) \le x\right),$$

where P^* is the conditional probability given the sample $(X_1, \ldots, X_n)$. Then, the equal-tailed 95% bootstrap confidence interval for $\Re a(\lambda, \tau)$ obtained using the cEMBB is of the form

$$\left(\Re\widehat{a}_n(\lambda,\tau) - \frac{u_{cEMBB}\,(0.975)}{\sqrt{n}}, \Re\widehat{a}_n(\lambda,\tau) - \frac{u_{cEMBB}\,(0.025)}{\sqrt{n}}\right),$$

where and $u_{cEMBB}\,(0.975)$ and $u_{cEMBB}\,(0.025)$ are 97.5%. and 2.5% quantiles of $K_{cEMBB}(x)$, respectively. The confidence interval for the cGSBB is defined correspondingly.

In real data applications usually many different values of λ and τ are considered. In such situation the simultaneous confidence intervals are used. To obtain them the consistency of bootstrap for the smooth functions of $a(\boldsymbol{\lambda}, \boldsymbol{\tau})$ is required. Such results for the cEMBB and the cGSBB can be found in [10] and [7]. The quantiles of the 95% equal-tailed bootstrap

simultaneous intervals can be calculated using the maximum and the minimum statistics. Let us assume that one is interested in frequencies $\lambda_1, \ldots, \lambda_r$ and shift τ. Again we describe the construction only for the cEMBB. We define:

$$K_{cEMBB,max}(x) = P^*\left(\sqrt{n}\max_i \Re\left(\widehat{a}_n^{*cEMBB}(\lambda_i,\tau) - \mathrm{E}^*\left(\widehat{a}_n^{*cEMBB}(\lambda_i,\tau)\right)\right) \le x\right),$$
$$K_{cEMBB,min}(x) = P^*\left(\sqrt{n}\min_i \Re\left(\widehat{a}_n^{*cEMBB}(\lambda_i,\tau) - \mathrm{E}^*\left(\widehat{a}_n^{*cEMBB}(\lambda_i,\tau)\right)\right) \le x\right)$$

and we get the confidence region of the form

$$\left(\widehat{a}_n(\lambda_i,\tau) - \frac{u_{cEMBB,max}\,(0.975)}{\sqrt{n}}, \widehat{a}_n(\lambda_i,\tau) - \frac{u_{cEMBB,min}\,(0.025)}{\sqrt{n}}\right), \tag{2}$$

where $i = 1, \ldots, r$ and $u_{cEMBB,max}\,(0.975)$ and $u_{cEMBB,min}\,(0.025)$ are 97.5%. and 2.5% quantiles of $K_{cEMBB,max}(x)$ and $K_{cEMBB,min}(x)$, respectively.

In the next section we compare bootstrap pointwise and simultaneous confidence intervals obtained with the cEMBB and the cGSBB. We try to provide answers for questions posed in Section 1.

3 Simulation study

The aim of our study is to compare the performance of the cEMBB and the cGSBB applied to construct the pointwise and the simultaneous confidence intervals for the real and the imaginary parts of the coefficients of the autocovariance function. For that purpose the actual coverage probabilities (ACPs), the average lengths and the average upper and lower quantiles values were calculated for all constructed intervals. Finally, confidence intervals were used to identify the significant frequencies. For our consideration we chose a few examples of PC time series that are listed below.

M1 $X_t = \cos(2\pi t/4)\varepsilon_t^1 + \cos(2\pi t/5)\varepsilon_t^2 + Z_t,$

M2 $X_t = \cos(2\pi t/4)\varepsilon_t^1 + \cos(2\pi t/6)\varepsilon_t^2 + Z_t,$

M3 $X_t = \cos(2\pi t/4)\varepsilon_t^1 + \cos(2\pi t/8)\varepsilon_t^2 + Z_t,$

M4 $X_t = 4\cos(2\pi t/4)\varepsilon_t^1 + \cos(2\pi t/5)\varepsilon_t^2 + Z_t,$

M5 $X_t = 8\cos(2\pi t/4)\varepsilon_t^1 + \cos(2\pi t/5)\varepsilon_t^2 + Z_t,$

where Z_t is zero-mean moving-average time series of the form

$$Z_t = 0.5\varepsilon_{t-3} + 0.3\varepsilon_{t-2} + 0.2\varepsilon_{t-1} + \varepsilon_t.$$

$\{\varepsilon_t\}_{t\geq 1}$, $\{\varepsilon_t^1\}_{t\geq 1}$, $\{\varepsilon_t^2\}_{t\geq 1}$ are i.i.d. from the standard normal distribution. Moreover, the initial observations in each model were generated as standard normal random variables.

Each considered model **M1**-**M5** has exactly 4 true significant frequencies. Models **M1**-**M3** differ in period length and distance between the consecutive significant frequencies. In fact we were changing period lengths of the cosine function to see what will happen if two consecutive frequencies will be close to (**M3**) or far from each other (**M2**). Better separation may improve the detection. Moreover, if two frequencies λ_1 and λ_2 are close to each other and $a(\lambda_1, \tau)$ is much bigger than $a(\lambda_2, \tau)$, the detection of λ_2 can be more difficult than in the situation when $a(\lambda_1, \tau)$ and $a(\lambda_2, \tau)$ are comparable. The frequency $\lambda = 0$ is always the strongest (value of the corresponding coefficient is always the largest among all significant frequencies). Thus, to investigate its influence on our results the distance between the three remaining frequencies and $\lambda = 0$ was set to be short (**M3**), medium (**M2**) and large (**M1**). Finally, we added to our considerations models **M4**-**M5** to check how strengthening of some frequencies will affect the detections. Note that **M4** and **M5** were obtained from **M1** by multiplying

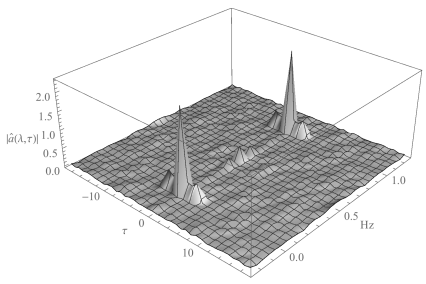
the first cosine function by an appropriate constant. That increased the value of $a(\lambda, 0)$ from 2.38 (**M1**) to 9.88 (**M4**) and 33.88 (**M5**) for $\lambda = 0\,\text{Hz}$ and from 0.5 (**M1**) to 8 (**M4**) and 32 (**M5**) for $\lambda = 0.5\,\text{Hz}$. Values of $a(\lambda, 0)$ for $\lambda = 0.4\,\text{Hz}$ and $\lambda = 0.6\,\text{Hz}$ remained unchanged and were equal to 0.25. In Figure 1 we present the estimated values of $|a(\lambda, \tau)|$ for **M1**-**M5** for sample size $n = 1920$.

For our study we took three sample sizes, namely $n \in \{120, 480, 1920\}$. We set $\tau = 0$ and we considered all frequencies $\lambda \in \{2k\pi/d : k = 0, \ldots, d-1\}$. Using the cEMBB and the GSBB we constructed the 95% equal-tailed pointwise and simultaneous confidence intervals (see Section 2) for the real and the imaginary part of $a(\lambda, \tau)$. The number of bootstrap resamples was $B = 500$ and the number of algorithm iteration was 1000. The block lengths b were chosen from the set $\{1, 2, \ldots, 100\}$. To compare the cEMBB and the cGSBB we calculated the ACPs, the average lengths and the average quantiles for all constructed confidence intervals. Since we considered 100 values of block length, the computation needed to be supported by the supercomputer available via the PL-Grid Infrastructure. For the sake of clarity we split the summary of our results into three parts.

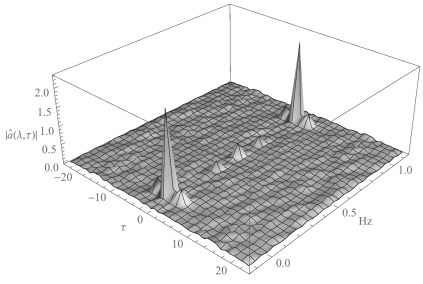
Actual coverage probabilities

For the sake of simplicity, from now on we denote by cEMBB-ACPs and cGSBB-ACPs the ACPs obtained with the cEMBB and the cGSBB, respectively. The main conclusions are the same for all models. Generally, the cEMBB provides higher ACPs than the cGSBB. This means that confidence intervals constructed using the cEMBB are wider than corresponding ones obtained with the cGSBB. Moreover, in most of the cases taking $b < 20$ we get confidence intervals too wide (using the cEMBB) or too narrow (using the cGSBB).

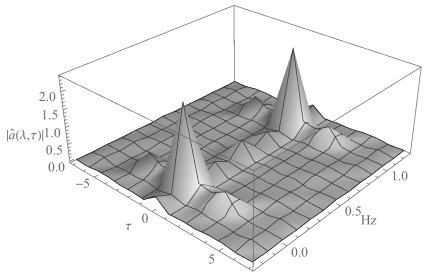
Some of the cEMBB-APC curves contain periodic structure (see Figure 2(a)). In fact for



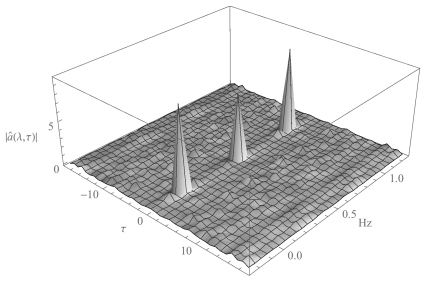
(a) **M1**



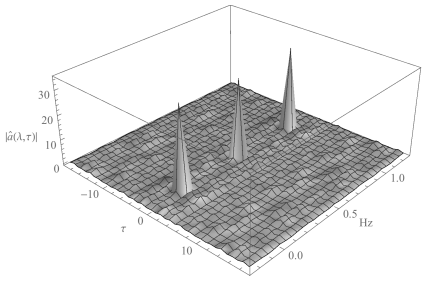
(b) **M2**



(c) **M3**

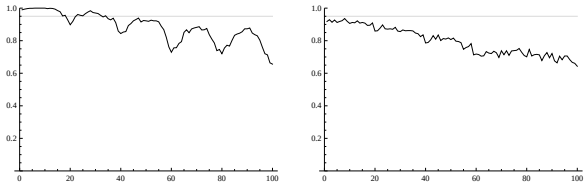


(d) **M4**



(e) **M5**

Figure 1: Estimated values of $|a(\lambda, \tau)|$ for **M1**-**M5** with sample size $n = 1920$.



(a) (b)

Figure 2: Model **M1** with $n = 120$: ACPs (black lines) of pointwise confidence intervals for $\Re a(0.05\,\mathrm{Hz}, 0)$ together with nominal coverage level (grey lines) for cEMBB (first column) and cGSBB (second column). On the horizontal axis block length $b \in \{1, 2, \ldots, 100\}$.

M1-**M5** this phenomena can be observed for frequencies that are close to $\lambda = 0\,\mathrm{Hz}$, i.e., $\lambda = 1/d$ Hz and $\lambda = 2/d$ Hz. Let us recall that $a(0,0)$ is always the largest among $a(\lambda, 0)$ values for $\lambda \in \Lambda_0$. For **M1** $a(0,0)$ is almost 5 times higher than the corresponding value for $\lambda = 0.5\,\mathrm{Hz}$ and 10 times than for $\lambda = 0.4\,\mathrm{Hz}$ and $\lambda = 0.6\,\mathrm{Hz}$. Interestingly, for models **M4** and **M5** periodicity appears also for $\lambda = 0.4\,\mathrm{Hz}$ and $\lambda = 0.45\,\mathrm{Hz}$. These two frequencies are in neighbourhood of another very strong frequency $\lambda = 0.5\,\mathrm{Hz}$. Sometimes periodicity is difficult to detect in APCs graphs. However, it can be much easier captured in figures presenting the average lengths of confidence intervals. Moreover, it is worth noticing that periodic structure appears not only in the context of the pointwise confidence intervals. We deal with it also in cEMBB-ACP graphs for the simultaneous confidence intervals. On the other hand, the cGSBB provides ACP curves without any periodicity (see Figure 2(b)). We think that the periodicity phenomena may be caused by the bias of the cEMBB estimator. Both bootstrap estimators are only asymptotically unbiased but bias of the cEMBB estimator may be strongly dependent on the chosen block length. Let us recall that the cGSBB preserves the periodic structure contained in the original data, while the cEMBB destroys it completely.

Block length choice - validity of some existing approaches

To popularize bootstrap techniques in the analysis of the PC time series, it is important to provide a method of block length choice. For now there is no such result. It concerns not only coefficients of the autocovariance function. Even in the simplest case of the overall mean estimation there are no indications for practitioners how to choose b. For stationary time series this issue is quite well investigated. Thus, we decided to check if some of the existing heuristic approaches can be applied for our non-stationary case.

In the literature it is often advised for periodic data to use the MBB with the block length equal to an integer multiple of the period length (see e.g., [24]). Our study confirms the validity of this approach. For $b = kd$, $k \in \mathbb{N}$ the cEMBB-ACP curves behave very similar to the cGSBB-ACP ones. In general median differences for the pointwise confidence intervals range from 0% to 3%, but for $n > 120$ from 0% to 0.5%. For the simultaneous confidence intervals the situation is very similar. For $n = 120$ the largest absolute difference is around 4% and for $n > 120$ all absolute differences belong to the interval $(0, 0.019)$. If periodic structure is present, a small change of the b value may result in big change of the actual coverage probability. Additionally, for large samples one may always find the block length being an integer multiple of the period length that provides cEMBB-ACPs close to nominal level.

The Minimum Volatility Method (see [22] p. 197) assumes that the block length b should be chosen from the region in which confidence interval is a stable function of b. For that purpose one may use a plot of the average confidence interval lengths or the average quantile values. We look for a region in which the curve is quite flat and there is not much variation. This method is not suitable for the cEMBB because of the periodic structure that appears

sometimes. However, we tried to use it for the cGSBB. Unfortunately, we did not manage to find rules how to select b values for the different frequencies to obtain cGSBB-ACPs close to 95%. Curves of the average lengths of pointwise and simultaneous confidence intervals obtained with the cGSBB are quite smooth and having low volatility. Some parts of the functions are quite flat, but often they do not contain block lengths corresponding to ACPs close to 95%.

The approach based on the logarithm of quantile for the subsampling method was recently discussed in [2] (see also references therein). The author proposed to find the largest b before the function becomes more erratic. Unfortunately, this methods does not work in the considered problem. When periodic structure is not present, the log of quantiles are smooth. We cannot localize regions with different behaviour.

Block length choice - new heuristic method for the cEMBB

Inspired by the aforementioned methods, to choose the value b we investigated plots of average lengths and logarithm of upper quantiles. In the latter case we managed to detect the property that allows us to choose the block length for the cEMBB. However, we would like to indicate that for now we do not have any theoretical confirmation for the proposed heuristic approach. Thus, further research needs to be done to confirm its validity. Our study shows that it seems to work quite well for the simultaneous confidence intervals.

Block length choice algorithm

1. If the log of quantile contains strong periodic structure, we look at local minima of the considered function. Their values create a string $m_1, m_2, \ldots$, which is decreasing

starting from m_o, $o \in \mathbb{N}$. Finally, we choose the block length b for which the log of the quantile is equal to m_o.

2. If the log of quantile does not contain any periodic structure, we do not consider very small or very large block lengths. From the remaining "reasonable" block lengths we choose the largest b before the first break point, i.e., point after which the function starts to decrease sharply. In fact for very small b a sharp decrease can also be observed. But we are interested in the largest b from the first region where the function is quite flat. We would like to indicate that this choice is not always easy. Sometimes the region in which the function is not decreasing is very small.

In Figure 3 we present the log of quantiles for the cEMBB simultaneous confidence intervals for $\Re a(\lambda, 0)$ and $\Im a(\lambda, 0)$ obtained for **M1** model. Using the proposed algorithm for **M1**-**M5** we chose the block lengths for the cEMBB simultaneous confidence intervals. Table 1 contains obtained values together with the ACPs generated by them for each sample size. In general, independently of the considered model and sample size, the chosen b differs for the real and imaginary part of $a(\lambda, 0)$. Those for $\Im a(\lambda, 0)$ are higher than the corresponding ones for $\Re a(\lambda, 0)$. The ACPs obtained for $\Re a(\lambda, 0)$ are usually below the nominal coverage level. The lowest value equal to 89% was got for **M3** with $n = 120$. The highest equal to 95.3% was observed for **M4** with $n = 1920$. For other cases the ACPs belong to interval $(91\%, 95\%)$. On the other hand the simultaneous confidence intervals for $\Im a(\lambda, 0)$ are too wide independently on the sample size. The ACPs change from 95.7% to 97.9%.

For $n = 1920$ in the case of the real and the imaginary part of $a(\lambda, \tau)$ sometimes we were forced to choose some $b \leq 100$ while the minima of log of quantiles in the corresponding figures never started to decrease. In such a situation we were taking the largest possible integer

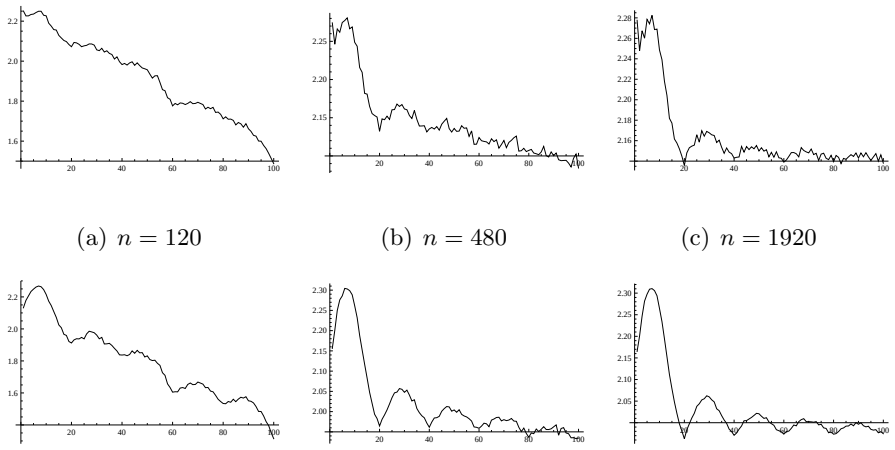
multiple of the period length. But in fact we should consider $b > 100$. For **M1**, **M4** and **M5** we performed an additional study taking $b \in \{120, 140, \ldots, 600\}$. The new choices of b values are in Table 2. One may observe that new cEMBB-ACPs are closer to 95%. Especially in the case of $\Im a(\lambda, \tau)$ the improvement is significant.

Finally, we compared results for **M1** with those for **M4** and **M5**. Let us recall that these models differ in values of $\Re a(0\,\mathrm{Hz}, 0)$ and $\Re a(0.5\,\mathrm{Hz}, 0)$. **M4** and **M5** contain two strongly significant frequencies ($\lambda = 0\,\mathrm{Hz}$ and $\lambda = 0.5\,\mathrm{Hz}$) in contrast to **M1** for which frequency $\lambda = 0\,\mathrm{Hz}$ is strongly dominating all other frequencies. For **M4**-**M5** we obtained with our algorithm b values that resulted in cEMBB-ACPs that are closer to 95% than the corresponding ones for **M1**. Strengthening of some frequencies improved their detection.

The proposed algorithm works quite well for the simultaneous confidence intervals obtained with the cEMBB. Unfortunately, for the cGSBB we did not manage to find any properties of the log of quantiles curves that would allow us to choose the block length. For the cEMBB we take advantage of the fact that for $b = 1$ it provides too wide confidence intervals and hence the first stabilization area usually contains b close to optimal one. Using the cGSBB with small b results sometimes in too low and sometimes too high ACP and hence, for example, local maximum can be very misleading.

Simplification of algorithm for the pointwise confidence intervals case

Choice of the block length for the pointwise confidence intervals is more problematic. In practice one would prefer to use one value of b for all considered frequencies. Moreover, when period length is long, the amount of calculations to get all plots is too big. Thus, we decided to check what would happen if we choose b only on the basis of curves that contain



(a) $n = 120$ (b) $n = 480$ (c) $n = 1920$

(d) $n = 120$ (e) $n = 480$ (f) $n = 1920$

Figure 3: Model **M1**: logarithms of the upper quantiles of simultaneous confidence intervals for $\Re a(\lambda, 0)$ (first row) and for $\Im a(\lambda, 0)$ (second row), $\lambda \in \overline{\Lambda}$ for cEMBB. On the horizontal axis block length $b \in \{1, 2, \ldots, 100\}$.

periodic structure. And to make procedure even simpler for each model we restricted our consideration only to two consecutive frequencies from $\overline{\Lambda}$ after $\lambda = 0\,\mathrm{Hz}$, i.e., $\lambda = 1/d$ Hz and $\lambda = 2/d$ Hz. For the $n = 480$ and $n = 1920$ the absolute deviation from 95% of majority of APC values is maximally 2.3% and 2.1% in the $\Re a(\lambda, \tau)$ and $\Im a(\lambda, \tau)$ case, respectively. It seems that our method of block length works quite well also for the pointwise confidence intervals.

4 Conclusions

To obtain bootstrap confidence intervals for the coefficients of the autocovariance function of PC time series we can use two bootstrap approaches. These are the EMBB and the GSBB or their circular versions. We performed an extensive study to compare their behaviour and we are unable to state, which of them is better. The ACPs obtained with the cEMBB are usu-

Table 1: For **M1**-**M5** and each sample size n and block length choice $b \leq 100$ for the cEMBB used to construct simultaneous confidence intervals for $\Re a(\lambda, 0)$ (column 3) and $\Im a(\lambda, 0)$ (column 5), where $\lambda \in \overline{\Lambda}$. In columns 4 and 6 are the obtained cEMBB-ACPs.

		Real part		Imaginary part	
model	n	b	cEMBB-ACP	b	cEMBB-ACP
M1	120	10	0.935	20	0.968
	480	40	0.912	60	0.957
	1920	100	0.934	100	0.976
M2	120	5	0.918	24	0.962
	480	5	0.947	48	0.970
	1920	48	0.935	84	0.975
M3	120	8	0.890	16	0.966
	480	16	0.931	24	0.974
	1920	60	0.949	96	0.964
M4	120	8	0.944	20	0.965
	480	30	0.938	60	0.975
	1920	60	0.953	100	0.979
M5	120	5	0.947	20	0.957
	480	30	0.938	60	0.966
	1920	100	0.949	100	0.973

Table 2: For **M1**, **M4**, **M5** and sample size $n = 1920$ and block length choice $b > 100$ for the cEMBB used to construct simultaneous confidence intervals for $\Re a(\lambda, 0)$ (column 3) and $\Im a(\lambda, 0)$ (column 5), where $\lambda \in \overline{\Lambda}$. In columns 4 and 6 are the obtained cEMBB-ACPs.

		Real part		Imaginary part	
model	n	b	cEMBB-ACP	b	cEMBB-ACP
M1	1920	200	0.943	320	0.972
M4	1920	60	0.953	340	0.952
M5	1920	180	0.946	280	0.955

ally higher than the corresponding ones for the cGSBB. It means that the cEMBB confidence intervals are often wider than the cGSBB ones. For very large samples, the performance of both bootstrap methods is very comparable.

The cEMBB-ACPs curves sometimes contain periodic structure. It seems this happens for frequencies that are close to the true strong frequencies. This phenomena may be considered as a disadvantage of the cEMBB, because a slight change in the chosen block length has a significant affect on the confidence interval. However, it can also be used in the future to construct a test for the significant frequency detection. Moreover, we used this feature to propose a heuristic method of block length choice that seems to work well in the case of pointwise and the simultaneous confidence intervals. In fact for the pointwise confidence intervals we simplified it to make the choice only on the basis of 1-2 figures, which substantially reduces the amount of computations and provides satisfactory results. We also checked a few heuristic approaches for the block length choice that were designed for the stationary data, but none of them is working for PC processes. We did not succeed in finding any indication

how the block length should be chosen when the cGSBB is used. Finally, we managed to confirm the suggestion appearing in the literature that for the cEMBB block lengths that are integer multiples of the period length should be considered. If we do not have any other method of block length choice, taking $b = kd$ allows us to avoid the periodicity effect and can provide the ACP close to the nominal one.

Acknowledgement This is a pre-print of a contribution published in *Cyclostationarity: Theory and Methods – IV; Contributions to the 10th Workshop on Cyclostationary Systems and Their Applications, February 2017, Grodek, Poland; Editors: Fakher Chaari, Jacek Leskow, Radoslaw Zimroz, Agnieszka Wylomanska, Anna Dudek; Part of the Applied Condition Monitoring book series (ACM, volume 16)* published by Springer, Cham. The final authenticated version is available online at https://link.springer.com/book/10.1007/978-3-030-22529-2.
This research was partially supported by the PL-Grid Infrastructure. Anna Dudek has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 655394. The authors express their sincere gratitude to Prof. Patrice Bertail for his comments and ideas.

References

[1] Antoni, J.: Cyclostationarity by examples. Mechanical Systems and Signal Processing **23**(4), pp. 987–1036, (2009)

[2] Bertail, P.: Comments on: Subsampling weakly dependent time series and application to extremes. TEST **20**(3), pp. 487–490, (2011)

[3] Carlstein, E.: The use of subseries methods for estimating the variance of a general statistic from a stationary time series. Annals of Statist. **14**, pp. 1171–1179, (1986)

[4] Dehay, D., Dudek, A.E.: Bootstrap method for Poisson sampling almost periodic process. J Time Ser Anal **36**(3), pp. 327–351, (2015)

[5] Dehay, D., Dudek, A.E.: Bootstrap for the second-order analysis of Poisson-sampled almost periodic processes. Electron. J. Statist. **11**(1), pp. 99—147, (2017)

[6] Dehay, D., Dudek, A.E. and Elbadaoui, M.: Bootstrap for almost cyclostationary processes with jitter effect, Digital Signal Processing, **73**, pp. 93–105, (2018)

[7] Dudek, A.E.: Circular block bootstrap for coefficients of autocovariance function of almost periodically correlated time series. *Metrika* **78**(3), pp. 313–335, (2015)

[8] Dudek, A.E.: Block bootstrap for periodic characteristics of periodically correlated time series. *Journal of Nonparametric Statistics* **30**(1), 87–124, (2018).

[9] Dudek, A.E, Leśkow, J., Paparoditis, E., Politis, D.N.: A generalized block bootstrap for seasonal time series. J Time Ser Anal. **35**, pp. 89–114, (2014a)

[10] Dudek, A.E., Maiz, S., Elbadaoui, M.: Generalized Seasonal Block Bootstrap in frequency analysis of cyclostationary signals. Signal Process. **104C**, pp. 358–368, (2014b)

[11] Efron, B.: Bootstrap Methods: Another look at the Jacknife. Ann Statist. **7**, pp. 1–26, (1979)

[12] Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Chapman & Hall/CRC, (1993)

[13] Gardner, W.A., Napolitano, A., Paura, L.: Cyclostationarity: Half a century of research. Signal Process. **86**(4), pp. 639–697, (2006)

[14] Gladyshev, E.G.: Periodically correlated random sequences. Soviet Math. **2**, pp. 385–388, (1961)

[15] Hurd, H.L., Miamee, A.G.: Periodically Correlated Random Sequences: Spectral. Theory and Practice. Wiley, (2007)

[16] Künsch, H.: The jackknife and the bootstrap for general stationary observations. Ann Statist. **17**, pp. 1217–1241, (1989)

[17] Lahiri, S.N.: Resampling Methods for Dependent Data, Springer, New York (2003)

[18] Lenart, Ł., Leśkow, J., Synowiecki, R.: Subsampling in testing autocovariance for periodically correlated time series. J. Time Ser. Anal. **29**, pp. 995–1018, (2008)

[19] Liu, R., Singh, K.: Moving block jackknife and bootstrap capture weak dependence. In: Le Page R, Billard L. *Exploring the Limits of Bootstrap*. New York: Wiley; pp. 225–248, (1992)

[20] Napolitano, A.: Generalizations of Cyclostationary Signal Processing: Spectral Analysis and Applications. Wiley-IEEE Press, (2012)

[21] Politis, D.N., Romano, J.P.: A circular block-resampling procedure for stationary data. Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., pp. 263–270, Wiley, New York (1992)

[22] Politis, D.N., Romano, J.P., Wolf, M.: Subsampling. New York: Springer, (1999)

[23] Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L.: Applied Modeling of Hydrologic Time Series. Water Resources Publications, (1980)

[24] Srinivasa, V.V., Srinivasan, K.: Hybrid moving block bootstrap for stochastic simulation of multi-site multi-season streamflows. J. Hydrol., **302**, pp. 307–330, (2005)

[25] Synowiecki, R.: Consistency and application of moving block bootstrap for nonstationary time series with periodic and almost periodic structure. Bernoulli **13**(4), pp. 1151-1178, (2007).

[26] Synowiecki, R.: Metody resamplingowe w dziedzinie czasu dla niestacjonarnych szeregów czasowych o strukturze okresowej i prawie okresowej. PhD Thesis at the Departement of Applied Mathematics, AGH University of Science and Technology, Krakow, Poland. http://winntbg.bg.agh.edu.pl/rozprawy2/10012/full10012.pdf (2008). Accessed 7 Aug 2017.