



**HAL**  
open science

# Optimal choice of bootstrap block length for periodically correlated time series

Patrice Bertail, Anna E. Dudek

► **To cite this version:**

Patrice Bertail, Anna E. Dudek. Optimal choice of bootstrap block length for periodically correlated time series. 2022. hal-03472199v2

**HAL Id: hal-03472199**

**<https://hal.science/hal-03472199v2>**

Preprint submitted on 12 Nov 2022

**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.

# Optimal choice of bootstrap block length for periodically correlated time series

Patrice Bertail<sup>1</sup>, Anna E. Dudek<sup>2</sup>

<sup>1</sup>MODAL'X, UPL Université Paris Nanterre, e-mail: [patrice.bertail@gmail.com](mailto:patrice.bertail@gmail.com)

<sup>2</sup>Department of Applied Mathematics, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Krakow, Poland, e-mail: [aedudek@agh.edu.pl](mailto:aedudek@agh.edu.pl)

**Abstract:** This paper discusses the problem of choosing the optimal block length for two block bootstrap methods designed for periodically correlated processes. These are the Generalized Seasonal Block Bootstrap and the Extension of Moving Block Bootstrap. Two estimation problems are considered: the overall mean and the seasonal means. In both cases, the optimal block length is obtained by minimizing the mean squared error of the corresponding bootstrap variance estimator and in all cases it is proportional to the cube root of the sample size and should be a multiple of the period length plus one observation to avoid some bias. Finally, the results of the performed simulation are presented, in which optimal blocks lengths are estimated for several periodically correlated time series.

**Keywords and phrases:** asymptotic expansion, bifrequency square, periodic time series, spectral estimation.

## 1. Introduction

The class of periodically correlated time series (PC) is an important example of nonstationary processes. To be precise  $\{X_t, t \in \mathbb{Z}\}$  is a PC (or equivalently cyclostationary) real-valued time series with period  $d$ , when  $X_t$  has periodic mean and covariance functions, i.e. for any  $t$  and  $s$ ,

$$E(X_t) = E(X_{t+d}) \quad \text{and} \quad \text{Cov}(X_t, X_s) = \text{Cov}(X_{t+d}, X_{s+d}).$$

The study of PC sequences were initiated by Gladyshev (1961) and has been extensively developed ever since. Processes that are generated by the mixing of randomness and periodicity usually have the structure of periodic correlation. Consequently, many real data can be described by PC sequences. Many motivating examples concerning economical, climatological, telecommunication and mechanical time series can be found e.g. in Hurd and Miamee (2007); Antoni (2009); Napolitano (2012, 2016, 2019); Gardner (1994); Gardner et al. (2006).

Bootstrap for periodic data has been developed over the past 20 years. The first approach called the Seasonal Block Bootstrap (SBB) was proposed by Politis (2001). The method assumes that the block length is always an integer multiple of the period length. On the other hand, the Periodic Block Bootstrap of Chan et al. (2004) is based on blocks, whose lengths are always smaller than the period length. These two bootstrap techniques offer much flexibility in selecting the block length. For the SBB we need data that contain many periods, but the period can be short, while the PBB requires long periods. In 2014 Dudek et al. introduced the Generalized Seasonal Block Bootstrap (see Dudek et al. (2014)). It was the first approach for periodic data that did not assume any relation between the block and the period lengths. Moreover, for particular block length choices it reduces to the PBB and the SBB. Finally, in Dudek et al. (2016) Dudek et al. showed that the Tapered Block Bootstrap (TBB) can be adapted for PC sequences. All mentioned methods assume that the period length is known. But sometimes it may happen that the period length is unknown or the signal under consideration is a composition of two components with incommensurable periods. In this case, a good candidate for bootstrapping data could be the Moving Block Bootstrap (MBB), which is one of the oldest and most universal block bootstrap techniques designed for stationary time processes. It was proposed independently by Künsch (1989) and Liu and Singh (1992). When applied to PC sequence, it completely destroys the periodic structure in the data. Therefore, Dudek (2015) and Dudek (2018) introduced the Extension of Moving Block Bootstrap (EMBB), which retains all the advantages of the MBB while being valid for many characteristics of PC time series and can even be applied to generalizations of PC processes.

The main problem when one wants to apply a block bootstrap method is the choice of block length. Unfortunately, till now in the literature there is no result showing how this can be done for PC data. One of typical approaches in the stationary case is to minimize the mean squared error (MSE) of the bootstrap variance estimator (see e.g. Lahiri (2003)). The problem, however, is that the usual proving techniques for PC time series did not give the expansion of the MSE. In addition, extensive simulation studies to verify whether certain heuristic approaches for stationary data can be adapted to nonstationary PC case have shown that neither is suitable (see Dudek and Potorski (2020)). Finally, at least in some cases one could use the fact that one-dimensional PC time series with period  $d$  can be equivalently represented as  $\mathbb{R}^d$ -valued stationary time series (see e.g. Proposition 1.1 in Hurd and Miamee (2007)). Then one could bootstrap the  $d$ -variate stationary data and apply one of the

known block length selection methods. However, it turns out that it is preferable to bootstrap PC sequence. The application of the SBB to one-dimensional PC time series is equivalent to using the MBB for  $\mathbb{R}^d$ -valued stationary time series. There is no equivalent approach to the GSBB and the EMBB in the multivariate stationary case, while these two methods are more universal and have higher overlap than the SBB (for more details we refer the reader to Section 2.4 in [Dudek \(2018\)](#)). Summarizing, we have a few bootstrap approaches for PC sequences and no method of choosing the block length.

While we cannot really exploit the relationship between the PC time series and a multivariate stationary sequence, we still expect that the methods for obtaining optimal block lengths for stationary data can be adapted to the periodic case. In this paper we generalize the approach of [Nordman \(2009\)](#). Under the same kind of assumptions on cumulants, summability of autocovariances and behavior of the block length when the sample size is growing to infinity, we show that in the case of the overall mean and the seasonal means the optimal block length for the GSBB and the EMBB is of order  $n^{1/3}$ , where  $n$  is the sample size.

Paper is organized as follows. Section 2 provides a brief summary of Nordman's results. In Section 3 the GSBB and the EMBB algorithms are recalled. Bootstrap estimators of seasonal means and overall mean are also presented. Section 4 is dedicated to PC time series and their properties. Finally, Section 5 contains our main results. In particular, the MSE expansions for the GSBB and the EMBB variance estimators are obtained in the case of the overall mean and the seasonal means. A brief summary of the obtained results with some remarks is given in Section 6. Results of the performed simulation study and all proofs can be found in the supplementary materials.

## 2. Stationary time series case

In this section, we first briefly introduce the MBB and CBB algorithms. Then, we recall the MSE expansion of the bootstrap variance estimator.

Let  $\{X_t, t \in \mathbb{Z}\}$  be a stationary time series and  $(X_1, \dots, X_n)$  be an observed sample. Moreover, let  $B_i = (X_i, \dots, X_{i+b-1}), i = 1, \dots, n - b + 1$  be a block of observations, which starts with observation  $X_i$  and has length  $b, b \in \mathcal{N}$ .

### MBB algorithm

1. Choose a block size  $b < n$ . Then our sample can be divided into  $l$  (non-overlapping) blocks of length  $b$  and the remaining part is of length  $r$ , i.e.  $n = lb + r, r = 0, \dots, b - 1$ .
2. From the set  $\{B_1, \dots, B_{n-b+1}\}$  choose randomly with replacement  $l + 1$  blocks  $B_1^*, \dots, B_{l+1}^*$ . Probability of choosing any block is  $1/(n - b + 1)$ .
3. Join the selected  $l + 1$  blocks  $(B_1^*, \dots, B_{l+1}^*)$  and take the first  $n$  observations to get the bootstrap sample  $(X_1^*, \dots, X_n^*)$  of the same length as the original one.

Note that the MBB approach is based on  $n - b + 1$  blocks of length  $b$ . Most of observations from the original sample  $(X_1, \dots, X_n)$  belong to exactly  $b$  different blocks. But the observations from the beginning and the end of the sample appear more rarely. For example the first observation  $X_1$  is present only in the block  $B_1$ . This fact often causes bias in MBB estimators. Therefore, [Politis and Romano \(1992\)](#) introduced the Circular Block Bootstrap (CBB), which ensures that each observation belongs to  $b$  blocks. The idea behind this approach is to treat the data as wrap on a circle, and thus  $n$  different blocks can be defined. For  $i = n - b + 2, \dots, n$  the block  $B_i$  is created by combining the observations from the end and the beginning of the sample to obtain a block of the required length. To be precise

$$B_i = (X_i, \dots, X_n, X_1, \dots, X_{b-n+i-1}) \quad \text{for } i = n - b + 2, \dots, n.$$

### CBB algorithm

The CBB, comparing with the MBB, differs only in the second step. Now we will select randomly blocks from the set of  $n$  blocks. Step 2 of the CBB method is as follows:

- 2'. From the set  $\mathcal{B} = \{B_1, B_2, \dots, B_n\}$  choose randomly with replacement  $l + 1$  blocks  $B_1^*, \dots, B_{l+1}^*$ . Probability of choosing any block is  $1/n$ .

The idea behind the CBB is very universal. It can be easily adapted to other types of block bootstrap methods, also for nonstationary cases.

From now on, we denote a probability measure (conditional on the data  $X_1, \dots, X_n$ ) by  $P^*$  and expectation and variance with respect to  $P^*$  by  $E^*$  and by  $\text{Var}^*$  respectively.

Let  $\mu = EX_1$ , its estimator  $\hat{\mu} = 1/n \sum_{i=1}^n X_i$  and the bootstrap estimator (obtained using the MBB or the CBB)  $\hat{\mu}^* = 1/n \sum_{i=1}^n X_i^*$ . Moreover, let  $\sigma^{2*} = \text{Var}^*(\sqrt{n}\hat{\mu}^*)$ .

To choose the optimal block length for the MBB and the CBB in the overall mean estimation problem, we may minimize the MSE of the bootstrap variance estimator  $\sigma^{2*}$ . For this purpose we need to have an expansion of the bias and variance of  $\sigma^{2*}$ . These kinds of results were obtained e.g., in [Carlstein \(1986\)](#), [Bühlman and Künsch \(1999\)](#), [Hall et al. \(1995\)](#), [Lahiri \(1999\)](#) and recently in [Nordman \(2009\)](#). Proving technique proposed by [Nordman \(2009\)](#) is very general. He provided a very nice and elegant way for obtaining the variance expansion of  $\sigma^{2*}$  for different block bootstrap methods like the MBB, the CBB, the Nonoverlapping Block Bootstrap (NBB), the Tapered Block Bootstrap (TBB) and the Stationary Bootstrap (SB). Below we focus only on the MBB/CBB case as we will later generalize it to PC time series.

Let  $\gamma(k) = \text{Cov}(X_1, X_{1+k})$  be the autocovariance function of the considered process  $\{X_t\}$  and

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \hat{\mu})(X_{t+k} - \hat{\mu})$$

be its estimator.

Nordman's approach is based on the fact that  $\sigma^{2*}$  can be expressed (exactly or approximately) as a weighted sum of autocovariance estimators, i.e. depending on the bootstrap approach we have

$$\sigma^{2*} = \sum_{k=0}^{n-1} a_{k,n} \hat{\gamma}_n(k) \quad \text{or} \quad \sigma^{2*} \approx \sum_{k=0}^{n-1} a_{k,n} \hat{\gamma}_n(k), \quad (1)$$

where the symbol ' $\approx$ ' denotes that  $\text{Var}(\sigma^{2*}) = \text{Var}(\sum_{k=0}^{n-1} a_{k,n} \hat{\gamma}_n(k)) + o(b/n)$ .

The coefficients  $a_{k,n}$  for the MBB are of the form

$$a_{k,n} = 2(1 - k/b) \text{ for } 1 \leq k < b \quad \text{and} \quad a_{k,n} = 0 \text{ for } k \geq b,$$

and for the CBB

$$\begin{aligned} a_{k,n} &= 2(1 - k/b) \text{ for } 1 \leq k < b, \quad a_{k,n} = 0 \text{ for } b \leq k \leq n - b \\ \text{and} \quad a_{k,n} &= a_{n-k,k} \text{ for } n - b < k < n. \end{aligned}$$

In the case of the MBB  $\sigma^{2*}$  is only approximately equal to a weighted sum of  $\hat{\gamma}_n(k)$ . The same fact applies to the TBB. Values of  $a_{k,n}$  for the other bootstrap approaches mentioned above can be found in [Nordman \(2009\)](#). Nordman showed the following very general variance result (see Theorem 2(i) in [Nordman \(2009\)](#)). Denote

$$T_n = \sum_{k=0}^{n-1} a_{k,n} \hat{\gamma}(k), \quad (2)$$

where  $a_{k,n}$  are such that

$$A_n = \sum_{k=1}^{n-1} (a_{k,n})^2 \left(1 - \frac{k}{n}\right)^2 > 0. \quad (3)$$

Moreover, let  $f(\omega) = 1/(2\pi) \sum_{k=-\infty}^{\infty} \gamma(k) \exp(-ik\omega)$  be the spectral density function and  $\text{cu}(\cdot)$  denotes the cumulant. Let us recall that the  $p$ th order joint cumulant of  $p$  variate random variable  $(Z_1, \dots, Z_p)$  such that  $E|Z_i|^p < \infty$  for  $i = 1, \dots, p$  is given by

$$\text{cu}(Z_1, \dots, Z_p) = \sum_{\pi \in \Pi_p} (|\pi| - 1)! (-1)^{|\pi|-1} \prod_{B \in B_\pi} E \left( \prod_{i \in B} Z_i \right),$$

where  $\Pi_p$  is the set of all partitions of  $\{1, \dots, p\}$  and  $|\pi|$  is the number of parts in the partition  $\pi$ .  $B_\pi$  is the set of blocks induced by the partition (see e.g. [Brillinger \(2001\)](#)).

The following theorem holds.

**Proposition 2.1.** *Let  $\{X_t, t \in \mathcal{Z}\}$  be a strictly stationary time series such that  $\gamma(0) + \sum_{k=1}^{\infty} k|\gamma(k)| < \infty$ . Moreover, assume that  $\sum_{t_1, t_2, t_3 \in \mathcal{Z}} |\text{cu}(X_0, X_{t_1}, X_{t_2}, X_{t_3})| < \infty$  and  $\max_{0 \leq k \leq n-1} |a_{k,n}| < \infty$ . Then*

$$\text{Var}(T_n) = \frac{A_n(2\pi)^2}{n} \int_{-\pi}^{\pi} K_n(\omega) f^2(\omega) d\omega + O \left( \left( \frac{A_n B_n}{n} \right)^{1/2} + B_n \right), \quad (4)$$

where

$$\begin{aligned} B_n &= \frac{1}{n} + \frac{\log(n)}{n^2} \left( \sum_{k=1}^{n-1} |a_{k,n}| \right)^2 + \frac{1}{n} \sum_{k=1}^{n-1} |a_{k,n}| \frac{k}{n} \left( 1 - \frac{k}{n} \right), \\ K_n(\omega) &= \frac{H_n(\omega)H_n(-\omega)}{2\pi A_n} = \frac{|H_n(\omega)|^2}{2\pi A_n}, \\ H_n(\omega) &= \sum_{k=1}^{n-1} a_{k,n} \left( 1 - \frac{k}{n} \right) \exp(-ik\omega). \end{aligned}$$

Finally, using the form of the weights  $a_{k,n}$  the following result can be obtained (see Theorem 1 in Nordman (2009)).

**Proposition 2.2.** *Let  $\{X_t, t \in \mathcal{Z}\}$  be a strictly stationary time series such that  $\gamma(0) + \sum_{k=1}^{\infty} k|\gamma(k)| < \infty$ .*

1. *Let  $\sum_{t_1, t_2, t_3 \in \mathbb{Z}} |\text{cu}(X_0, X_{t_1}, X_{t_2}, X_{t_3})| < \infty$ . Moreover, let  $b \rightarrow \infty$  and  $b \log n/n \rightarrow 0$  as  $n \rightarrow \infty$ . Then the variances of the MBB and the CBB estimators of  $n\text{Var}(\bar{X}_n)$  are*

$$\text{Var}(\sigma_{MBB}^{2*}) = \text{Var}(\sigma_{CBB}^{2*}) = \frac{b}{n}D + o\left(\frac{b}{n}\right),$$

where  $D = 4/3(2\pi f(0))^2$ .

2. *If  $b^{-1} + b^2/n = o(1)$  as  $n \rightarrow \infty$ , then the bias of the MBB and the CBB estimators are*

$$\text{Bias}(\sigma_{MBB}^{2*}) = \text{Bias}(\sigma_{CBB}^{2*}) = -\frac{1}{b}G + o\left(\frac{1}{b}\right),$$

where  $G = \sum_{k=-\infty}^{\infty} |k|\gamma(k)$ .

Proposition 2.2 allows to get the MSE expansion for  $\sigma^{2*}$ . Then, the optimal block length  $b_{opt}$  is obtained by minimization of the MSE i.e.,

$$b_{opt} = \text{argmin}_b \{MSE(\sigma^{2*})\}.$$

It is easy to calculate that for large samples

$$b_{opt} = \sqrt[3]{\frac{2G^2}{D}} \sqrt[3]{n}$$

and hence for the MBB and the CBB  $b_{opt}$  is proportional to  $n^{1/3}$ . In fact, for the NBB and the SB, one gets the same result, only with a different constant. In the case of the TBB,  $b_{opt}$  is proportional to  $n^{1/5}$  (see e.g., Nordman (2009), Lahiri (2003)).

**Remark 2.1.** *In Propositions 2.1 and 2.2 the condition on a strict stationarity of  $\{X_t, t \in \mathcal{Z}\}$  can be weakened. It may be replaced by the assumption that  $\{X_t, t \in \mathcal{Z}\}$  is weakly stationary up to order 4. Therefore, in our considerations, we do not use the concept of strict periodicity and we formulate all our results for weakly periodic time series.*

### 3. Bootstrap methods for PC time series

In this section we recall shortly the GSBB and the EMBB algorithms. For the sake of simplicity of the presentation we discuss only their circular versions, i.e. we treat data as wrapped on the circle. This simplifies in particular the GSBB algorithm.

Let  $\{X_t, t \in \mathbb{Z}\}$  be a periodic (strictly or weakly) time series with a period  $d$  and let  $(X_1, \dots, X_n)$  be a considered sample. Without loss of generality we take  $n = v_1 d$ ,  $v_1 \in \mathbb{N}$ .

Let  $B_i, i = 1, \dots, n$  be a block of observations that starts with observation  $X_i$  and has length  $b$ . We have

$$B_i = (X_i, \dots, X_{i+b-1}), \quad i = 1, \dots, n - b + 1.$$

Additionally, for  $i = n - b + 2, \dots, n$  block  $B_i$  is created by joining the observations from the end and the beginning of the sample to get a block of the required length. To be precise

$$B_i = (X_i, \dots, X_n, X_1, \dots, X_{b-n+i-1}), \quad i = n - b + 2, \dots, n.$$

Let  $n = bl + r$ , where  $l \in \mathbb{N}$  and  $r \in \{0, \dots, b - 1\}$ , i.e., our sample can be divided into  $l$  disjoint blocks of the length  $b$  and a remaining shorter block of length  $r$ .

The Generalized Seasonal Block Bootstrap (GSBB) was introduced in [Dudek et al. \(2014\)](#). It is dedicated to periodic data with known and fixed period  $d$ . It is worth to indicate that it is a very general bootstrap algorithm because it does not require any relation between the block length  $b$  and the period length  $d$ . Moreover, it completely preserves the periodic structure of the original data.

From now on by  $\langle t \rangle$ , we denote the season associated with a time index  $t$ , where  $t = 1, \dots, n$ . To be precise

$$\langle t \rangle = (t \bmod d) \quad \text{for } (t \bmod d) \neq 0$$

and  $\langle t \rangle = d$ , otherwise.

### Circular Generalized Seasonal Block Bootstrap (CGSBB) algorithm

1. Choose a (positive) integer block size  $b < n$ .
2. For  $t = 1, b + 1, 2b + 1, \dots, lb + 1$ , let

$$(X_t^*, X_{t+1}^*, \dots, X_{t+b-1}^*) = (X_{k_t}, X_{k_t+1}, \dots, X_{k_t+b-1})$$

where  $k_t$  is a discrete uniform random variable taking values in the set

$$S_{t,n} = \{\langle t \rangle, \langle t \rangle + d, \dots, \langle t \rangle + (v_1 - 1)d\}.$$

As usual, the random variables  $k_1, k_2, \dots$  are taken to be independent.

3. Join the  $l + 1$  blocks  $(X_{k_t}, X_{k_t+1}, \dots, X_{k_t+b-1})$  thus obtained together to form a new series of bootstrap pseudo-observations  $X_1^*, X_2^*, \dots, X_n^*, \dots, X_{(l+1)b}^*$  from which only the first  $n$  points  $X_1^*, X_2^*, \dots, X_n^*$  are retained so that the bootstrap series has same length as the original.

The GSBB/CGSBB can be applied only when the period length is known  $d$ . In fact, if one needs to estimate  $d$ , the problem is that the estimate must be very accurate. In the event of any deviation of the estimate from the true value of  $d$ , the GSBB/CGSBB fails. Therefore, it is convenient to have an alternative approach that works when the period length is unknown and is consequently suitable for more general classes of time series, such as almost periodically correlated (APC) processes. The MBB of [Künsch \(1989\)](#) and [Liu and Singh \(1992\)](#), which is one of the most versatile block bootstrap technique, might be a good candidate, but it has been found to be generally inconsistent for PC time series and their generalizations. In fact it is valid only for the overall mean of the PC/APC sequences (see [Synowiecki \(2007\)](#)). However, it turned out that it can be modified (see [Dudek \(2015\)](#), [Dudek \(2018\)](#)) to ensure consistent results for different characteristics of PC and APC processes. The idea of the Extension of Moving Block Bootstrap (EMBB) is very simple. Instead of bootstrapping the original data, one bootstraps a new bivariate series. The key feature of this approach is the fact that it preserves information on seasons of the observations. Below we present the circular version of the EMBB algorithm.

### Circular Extension of Moving Block Bootstrap (CEMBB) algorithm

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

Note that in the second coordinate of the series  $Y_1^*, \dots, Y_n^*$  we preserve the information on the original time indices of chosen observations. Below we show how this information can be included in the definitions of the bootstrap estimators of the different parameters of  $\{X_t\}$  to provide the consistency of the EMBB. We start with periodic characteristics of PC time series.

Let  $\{i_1, \dots, i_{l+1}\}$  be the selected block numbers in step 2 of the CBB algorithm performed on  $(Y_1, \dots, Y_n)$ . Moreover, let us recall that  $n = lb + r$ , where  $r \in \{0, 1, \dots, b - 1\}$ . Using the first and the second coordinates of  $Y_1^*, \dots, Y_n^*$  we form a vector containing the selected observations

$$(X_1^*, \dots, X_n^*) = (X_{i_1}, \dots, X_{i_1+b-1}, \dots, X_{i_l}, \dots, X_{i_l+b-1}, X_{i_{l+1}}, \dots, X_{i_{l+1}+r-1})$$

and the vector of their original time indices

$$TI = (i_1, \dots, i_1 + b - 1, \dots, i_l, \dots, i_l + b - 1, i_{l+1}, \dots, i_{l+1} + r - 1).$$

By  $v_s^*$  we denote the number of elements in the EMBB sample that are from season  $s$  i.e.

$$\begin{aligned} v_s^* &= \#TI_s = \#\{t : t \in TI, t \bmod d = s\} \quad \text{for } s = 1, \dots, d - 1, \\ v_d^* &= \#TI_d = \#\{t : t \in TI, t \bmod d = 0\}. \end{aligned}$$

Note that  $n = v_1^* + \dots + v_d^*$  and that  $TI_s$  contains those time indices from  $TI$  that represent season  $s$ .

Since in this paper we focus on means, below we recall the bootstrap estimators of the overall mean and the seasonal means. We denote by  $\mu$  the overall mean of the considered time series and by  $\mu_1, \dots, \mu_d$  the seasonal means. Under the assumption  $n = v_1 d$  the estimators of  $\mu_s$  and  $\mu$  are of the following form

$$\hat{\mu}_s = \frac{1}{v_1} \sum_{m=0}^{v_1-1} X_{s+md} \quad \text{and} \quad \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{d} \sum_{s=1}^d \hat{\mu}_s. \quad (5)$$

The GSBB/CGSBB counterparts of (5) are very natural. They have exactly the same form and are simply computed using the bootstrap sample. This property is a direct consequence of the fact that the GSBB/CGSBB perfectly mimics the periodic structure contained in the data, so in the bootstrap sample each season is represented by exactly the same number of observations as in the original sample. Thus, for the GSBB/CGSBB we have

$$\hat{\mu}_s^* = \frac{1}{v_1} \sum_{m=0}^{v_1-1} X_{s+md}^* \quad \text{and} \quad \hat{\mu}^* = \frac{1}{n} \sum_{i=1}^n X_i^* = \frac{1}{d} \sum_{s=1}^d \hat{\mu}_s^*. \quad (6)$$

The EMBB/CEMBB estimators are generally more complicated. For seasonal means we have (see [Dudek \(2018\)](#))

$$\hat{\mu}_s^* = \frac{1}{v_s^*} \sum_{i \in TI_s} X_i^* \quad \text{for} \quad s = 1, \dots, d. \quad (7)$$

If  $v_s^* = 0$  for some  $s$  then we take  $\hat{\mu}_s^* = 0$ .

The number  $v_s^*$  in (7) is equal to the number of observations from season  $s$  in the bootstrap sample. Unlike the GSBB/CGSBB, the EMBB/CEMBB destroys the periodic structure of the data, and  $v_s^*$  is generally not equal to  $v_1$ . Moreover, when the block length is smaller than the period length ( $b < d$ ), it may happen that in the bootstrap sample there will be no observations from certain seasons. Overall, some seasons may be underrepresented and others may be overrepresented.

The case of the overall mean is much simpler. The EMBB/CEMBB estimator of  $\mu$  is simply  $\hat{\mu}^* = 1/n \sum_{i=1}^n X_i^*$ . Thus, in this case we do not use information from which seasons the observations in the bootstrap sample come from, so the EMBB/CEMBB method is equivalent to the usual MBB/CBB.

#### 4. PC time series - short overview and some properties

In this section, we introduce some additional notation and we briefly recall some basic definitions and properties of PC time series. We focus only on those that are essential for our results. For more details on PC sequences we refer the reader to the book of [Hurd and Miamee \(2007\)](#).

Let  $B(t, k)$  be the autocovariance function of a PC time series  $\{X_t, t \in \mathbb{Z}\}$  (with period  $d$ ), i.e.  $B(t, k) = \text{Cov}(X_t, X_{t+k})$ . Arguments  $t$  and  $k$  represent time and shift, respectively. Note that  $B(t, k)$  is a periodic function of the time argument. Moreover, the seasonal variances  $\text{Var}(X_s)$  ( $s = 1, \dots, d$ ) are values of the autocovariance function at lag 0, i.e.,  $\text{Var}(X_s) = B(s + md, 0)$  for  $m \in \mathbb{N}$ .

The seasonal sample autocovariances are of the form (for  $s = 1, \dots, d$  and  $k \in \mathbb{Z}$ )

$$R_n(s, k) = \frac{1}{v_{s,n}} \sum_{m=0}^{v_{s,n}-k-1} (X_{s+md} - \hat{\mu}_s)(X_{s+md+k} - \hat{\mu}_{\langle s+k \rangle}), \quad (8)$$

where  $\langle t \rangle = (t \bmod d)$  for  $t \neq kd, k \in \mathbb{Z}$  and  $\langle t \rangle = d$  otherwise, denotes the season associated with  $t$  (see also Section 3). Moreover,  $v_{s,n}$  and  $v_{s,n-k}$  are the numbers of observations from the season  $s$  in the samples  $(X_1, \dots, X_n)$  and  $(X_1, \dots, X_{n-k})$ , respectively.

The circular version of (8) is defined as follows

$$R_{n,cir}(s, k) = \frac{1}{v_{s,n}} \sum_{m=0}^{v_{s,n}-1} (X_{s+md} - \hat{\mu}_s)(X_{s+md+k} - \hat{\mu}_{\langle s+k \rangle}).$$

Finally, we denote

$$R_n(k) = \frac{1}{d} \sum_{s=1}^d R_n(s, k) \quad \text{and} \quad R_{n,cir}(k) = \frac{1}{d} \sum_{s=1}^d R_{n,cir}(s, k).$$

One of the key steps in our considerations will be to obtain an analogous relation to (1), that is, we will show in all cases that the bootstrap variance can be expressed (exactly or at least approximately) as some linear combination of  $R_n(s, k)$ . In fact, in the case of circular versions of the bootstrap approaches,  $R_{n,cir}(s, k)$  will naturally appear in the expansions and we therefore provide below the relation between  $R_n(s, k)$  and  $R_{n,cir}(s, k)$ .



**Lemma 4.1.** *Let  $\{X_t, t \in \mathbb{Z}\}$  be a PC time series with period  $d$ . Assume  $n = v_1 d$  for  $v_1 \in \mathbb{N}$ . Then for any  $s = 1, \dots, d$  and  $k = 1, 2, \dots$*

$$\begin{aligned} R_{n,cir}(s, k) &= R_n(s, k) && \text{for } 0 \leq s + k < d, \\ R_{n,cir}(s, k) &= R_n(s, k) + R_n(\langle s + k \rangle, n - k) && \text{for } s + k \geq d. \end{aligned}$$

The autocovariance function  $B(t, k)$  (for any fixed  $k \in \mathbb{Z}$ ) can be represented as Fourier series

$$B(t, k) = \sum_{\lambda \in \Lambda} a(\lambda, k) \exp(i\lambda t), \quad (9)$$

where the set  $\Lambda = \{\lambda \in [0, 2\pi) : a(\lambda, k) \neq 0\} \subset \{2m\pi/d, m = 0, 1, \dots, d-1\}$ . PC time series are harmonizable in the sense of Loève (see [Loève \(1963\)](#)). When the period length is equal to  $d$ , then the spectral measure is concentrated in the bifrequency square  $[0, 2\pi)^2$  on  $2d-1$  equally distanced diagonal lines (see [Hurd and Miamee \(2007\)](#)). Consequently, one may associate with  $\{X_t\}$  a spectral matrix  $\mathbf{f} = [f_{jl}]$ ,  $j, l = 0, \dots, d-1$  of the form

$$f_{jl}(\omega) = \frac{1}{d} f_{l-j}((\omega - 2\pi j)/d), \quad j, l = 0, \dots, d-1,$$

where  $\omega \in [0, 2\pi)$  and

$$\begin{aligned} f_l(\omega) &= \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} a\left(\frac{2l\pi}{d}, \tau\right) \exp(-i\tau\omega), \\ a\left(\frac{2l\pi}{d}, \tau\right) &= \frac{1}{d} \sum_{t=1}^d \text{Cov}(X_t, X_{t+\tau}) \exp\left(-\frac{2\pi i l t}{d}\right). \end{aligned}$$

To estimate elements of the matrix  $\mathbf{f}$  the dual-frequency periodogram is used.

Let  $\{\tilde{X}_t, t \in \mathbb{Z}\}$  be zero-mean version of  $\{X_t\}$ , i.e.  $\tilde{X}_t = X_t - \mu_{\langle t \rangle}$ . The dual-frequency periodogram based on the sample  $(\tilde{X}_1, \dots, \tilde{X}_n)$  is defined as follows

$$I_n(\omega_1, \omega_2) = \frac{1}{2\pi n} \tilde{d}_n(\omega_1) \tilde{d}_n(-\omega_2), \quad (10)$$

where

$$\tilde{d}_n(\omega) = \sum_{t=1}^n \tilde{X}_t \exp(-it\omega), \quad \omega \in [0, 2\pi). \quad (11)$$

#### 4.1. Relationship to multivariate stationary time series

One-dimensional PC time series  $\{X_t\}$  with period  $d$  can be equivalently expressed as  $d$ -variate stationary time series  $\{\tilde{X}_m\}$  (see Proposition 1.1 in [Hurd and Miamee \(2007\)](#)), where

$$\tilde{X}_m = (X_{1+(m-1)d}, \dots, X_{d+(m-1)d})'$$

is the vector consisting of observations from the  $m$ -th period. Consequently, working with one-dimensional PC time series may seem unnecessary in view of the fact that the theory for multivariate processes is very well established. In the following, we explain why this approach is not often used in the literature and direct analysis of PC series is preferred.

Applying any block bootstrap approach to  $\{\tilde{X}_m\}$  results in bootstrap blocks that are always an integer multiple of the period length and that always start with observation from the first season. In this way, the MBB for  $\{\tilde{X}_m\}$  corresponds to the SBB method used for a one-dimensional PC series. The GSBB and the EMBB are much more general. The block length is not related to the period length and it may start with observation from any season. Moreover, the GSBB and the EMBB have higher overlap than the SBB among blocks to be bootstrapped. Finally, if the considered data have few periods, then the choice of block length in the multivariate case is very limited. Consequently, bootstrapping  $\{X_t\}$  is preferred to bootstrapping  $\{\tilde{X}_m\}$ .

Approach via multivariate stationary time series requires knowledge of the period length, which is not always known for real data. To estimate it, one usually performs a time and frequency domain analysis on the original periodic sequence. Moreover, it may happen that the considered data is a composite of two signals with incommensurable periods and then the output signal is no longer PC, but almost periodically correlated (APC). Such data cannot be transformed to a multivariate stationary sequence. In addition, methods developed for discrete PC time series can often be generalized to discrete APC and continuous PC/APC processes. This



fact is also true for bootstrap. For example, in [Dudek \(2015\)](#) is shown the EMBB consistency for the Fourier coefficients of the mean and autocovariance function of an APC time series. Direct analysis of PC data also allows easy switching between between time and frequency domain. In mechanics, for example, data are usually first demeaned by subtracting estimated seasonal means, and then analysis is performed in the frequency domain to detect possible machine faults. For more details and some additional arguments on the advantages of the original periodic sequence analysis, we refer the reader to [Hurd and Miamee \(2007\)](#).

## 5. Main results

In this section we provide the bias and variance expansions for the GSBB/CGSBB and the EMBB/CEMBB variance estimators in the case of overall mean and seasonal means. As a consequence, we obtain the optimal block lengths. To derive our results we adapt and generalize ideas of [Nordman \(2009\)](#). Additionally, they have a general form that can be applied with minor modifications to other bootstrap approaches in the future.

A time series  $\{X_t, t \in \mathbb{Z}\}$  is said to be a weakly periodic process of order  $k$  (say WP( $k$ )) with period  $d$ , iff  $\{X_t\}$  has periodic moments of order  $k$ , i.e.  $E|X_t|^k < \infty$  and for any  $t, \tau_1, \dots, \tau_{k-1} \in \mathbb{Z}$

$$E(X_t X_{t+\tau_1} \dots X_{t+\tau_{k-1}}) = E(X_{t+d} X_{t+\tau_1+d} \dots X_{t+\tau_{k-1}+d}).$$

Note that  $\{X_t\}$  is a PC time series iff it is WP(1) and WP(2).

In the sequel we assume the following conditions:

- A0**  $\{X_t, t \in \mathbb{Z}\}$  is a PC real-valued time series with period  $d$  that is WP(3) and WP(4);
- A1**  $n = v_1 d$ ,  $v_1 \in \mathbb{N}$ ;
- A2**  $\sum_{s=1}^d \text{Var}(X_s) + \sum_{s=1}^d \sum_{\tau=-\infty}^{\infty} |\tau| |\text{Cov}(X_s, X_{s+\tau})| < \infty$ ;
- A3**  $\sup_{s=1, \dots, d} \sum_{t_1 \in \mathbb{Z}} \sum_{t_2 \in \mathbb{Z}} \sum_{t_3 \in \mathbb{Z}} |\text{cu}(X_s, X_{t_1}, X_{t_2}, X_{t_3})| < \infty$ ;
- A4**  $b \rightarrow \infty$  and  $b \log(n)/n \rightarrow \infty$  as  $n \rightarrow \infty$ .

The condition **A4** is exactly the same as the corresponding condition of [Nordman \(2009\)](#). The assumptions **A2** and **A3** are modified versions of the Nordman's conditions for the considered PC case. **A0** denotes that  $\{X_t\}$  is weakly periodic up to order 4. Periodicity of the 3rd and the 4th order moments in **A0** allows us to provide results without assuming periodicity in distribution. Moreover, it is not a restrictive condition. For example, let the time series  $\{A_t, t \in \mathbb{Z}\}$  be weakly stationary up to order 4 and  $f(t)$  be a periodic real valued function. Then time series  $B_t = f(t)A_t$  satisfies the condition **A0**. In general, if the second order moments are periodic, this is a consequence of the existence of a periodic mechanism in the generation of the process. Therefore, it is very likely that such a periodicity is present also in the distribution and in the higher-order statistics. For instance, in a case of PC communications signals, higher-order moments and cumulants can be computed analytically and they are usually periodic. Moreover, Lemma A.4 in [Lenart \(2011\)](#) provides moment and mixing conditions under which **A2** holds ( $\sup_t E|X_t|^{6+3\delta} < \infty$  and  $\sum_{k=0}^{\infty} (k+1)^2 \alpha(k)^{\delta/(2+\delta)} < \infty$  for some  $\delta > 0$ , where  $\alpha(\cdot)$  is  $\alpha$ -mixing sequence corresponding to  $\{X_t\}$ ). Finally, **A1** is necessary for the circular versions of the block bootstrap methods in the periodic case. Thus, to make our results fully comparable for the usual and circular methods, we assume **A1** in all cases. This also allows us to simplify the computation and notation for the non-circular cases, which are already very complicated. In practice, if  $n = v_1 d + r$ , where  $r \in 1, \dots, d-1$  then it is sufficient to remove the last  $r$  observations from the considered dataset.

In the following  $(X_1^*, \dots, X_n^*)$  is a bootstrap sample. Depending on the case under consideration, it will be obtained by means of the GSBB, CGSBB, EMBB or CEMBB methods. By  $b \in \mathbb{N}$  we denote the bootstrap block length. For the sake of simplicity without loss of generality we assume that  $n = lb$ , where  $l \in \mathbb{N}$  is the number of selected blocks to form the bootstrap sample.

### 5.1. Overall mean case

At first we introduce some additional notations. By  $\sigma_{as}^2$  we denote the asymptotic variance of  $\sqrt{n}\hat{\mu}$ , i.e. as  $n \rightarrow \infty$

$$\text{Var}(\sqrt{n}\hat{\mu}) \rightarrow \sigma_{as}^2 = \frac{1}{d} \sum_{s=1}^d \sum_{k=-\infty}^{\infty} \text{Cov}(X_s, X_k). \quad (12)$$

The following results are divided into two parts. We first discuss the GSBB/CGSBB case and then we focus on the EMBB/CEMBB case. For all bootstrap approaches we first obtain an analogous relation to (1), i.e.

we show that the bootstrap variance  $\text{Var}^*(\sqrt{n}\widehat{\mu}^*)$  can be expressed (exactly or approximately) as some linear combination of  $R_n(s, k)$ . Then we provide the bias and variance expansions for  $\text{Var}^*(\sqrt{n}\widehat{\mu}^*)$ , which ultimately allow us to calculate the MSE and find the optimal block length.

### 5.1.1. Results for the GSBB/CGSBB

In the case of the GSBB/CGSBB the bootstrap variance  $\text{Var}^*(\sqrt{n}\widehat{\mu}^*)$  can be expressed in terms of  $R_n(s, k)$  as follows.

**Theorem 5.1.** *Let  $\{X_t, t \in \mathbb{Z}\}$  be a PC real-valued time series with period  $d$  such that  $\sum_{s=1}^d \text{Var}(X_s) + \sum_{s=1}^d \sum_{k=-\infty}^{\infty} |\text{Cov}(X_s, X_{s+k})| < \infty$ . Moreover, assume that **A1** holds. Then for the CGSBB we get that*

$$\begin{aligned} \text{Var}^*(\sqrt{n}\widehat{\mu}^*) &= \frac{1}{d} \sum_{s=1}^d R_{n, \text{cir}}(s, 0) \\ &+ 2 \frac{1}{n} \sum_{m=0}^{l-1} \sum_{s=1}^d \sum_{k=1}^{b-1} \left( v_{b-1} + \mathbb{I}_{\{r_{b-1} \geq s\}} - \left\lfloor \frac{k-1}{d} \right\rfloor \right) R_{n, \text{cir}}(\langle mb + s \rangle, k) \end{aligned} \quad (13)$$

and for the GSBB we have that

$$\begin{aligned} \text{Var}^*(\sqrt{n}\widehat{\mu}^*) &\approx \frac{1}{d} \sum_{s=1}^d R_n(s, 0) \\ &+ 2 \frac{1}{n} \sum_{m=0}^{l-1} \sum_{s=1}^d \sum_{k=1}^{b-1} \left( v_{b-1} + \mathbb{I}_{\{r_{b-1} \geq s\}} - \left\lfloor \frac{k-1}{d} \right\rfloor \right) R_n(\langle mb + s \rangle, k), \end{aligned} \quad (14)$$

where

$$\begin{aligned} \mathbb{E}(\text{Var}^*(\sqrt{n}\widehat{\mu}^*)) &= \mathbb{E} \left( \frac{1}{d} \sum_{s=1}^d R_n(s, 0) \right) \\ &+ \mathbb{E} \left( 2 \frac{1}{n} \sum_{m=0}^{l-1} \sum_{s=1}^d \sum_{k=1}^{b-1} \left( v_{b-1} + \mathbb{I}_{\{r_{b-1} \geq s\}} - \left\lfloor \frac{k-1}{d} \right\rfloor \right) R_n(\langle mb + s \rangle, k) \right) \\ &+ o \left( \frac{1}{b} \right), \end{aligned} \quad (15)$$

$$\begin{aligned} \text{Var}(\text{Var}^*(\sqrt{n}\widehat{\mu}^*)) &= \text{Var} \left( \frac{1}{d} \sum_{s=1}^d R_n(s, 0) \right) \\ &+ 2 \frac{1}{n} \sum_{m=0}^{l-1} \sum_{s=1}^d \sum_{k=1}^{b-1} \left( v_{b-1} + \mathbb{I}_{\{r_{b-1} \geq s\}} - \left\lfloor \frac{k-1}{d} \right\rfloor \right) R_n(\langle mb + s \rangle, k) + o \left( \frac{b}{n} \right), \end{aligned} \quad (16)$$

and  $b-1 = v_{b-1}d + r_{b-1}$ ,  $r_{b-1} \in \{0, \dots, d-1\}$ .

For definition of the symbol ' $\approx$ ' see (1). In fact, the approximate variance expansion for the GSBB is a consequence of the bias resulting from the fact that the observations at the beginning and end of the sample appear in fewer blocks than the other observations. The edge effect disappears for the CGSBB. The same observations can be made for the MBB and the CBB (see Section 5.1.2).

In the formulas (13)-(14)  $k$  is the distance between observations in a block of length  $b$ . Thus, the maximal value of  $k$  equals  $b-1$ . The expression  $\mathbb{I}_{\{r_{b-1} \geq s\}}$  is a consequence of the fact that  $b-1$  may be not an integer multiple of the period length ( $r_{b-1} \neq 0$ ).

In the GSBB case, the formula (14) allows us to define a weighted average of the seasonal autocovariances that we will consider and which is analogous to (2) in the stationary case. We denote

$$T_n = \frac{1}{d} \sum_{s=1}^d R_n(s, 0) + 2 \frac{1}{n} \sum_{m=0}^{l-1} \sum_{s=1}^d \sum_{k=1}^{b-1} \left( v_{b-1} + \mathbb{I}_{\{r_{b-1} \geq s\}} - \left\lfloor \frac{k-1}{d} \right\rfloor \right) R_n(\langle mb + s \rangle, k).$$

$T_n$  can be equivalently rewritten as follows

$$T_n = \frac{1}{d} \sum_{s=1}^d a_{s,0} R_n(s, 0) + \frac{1}{d} \sum_{s=1}^d \sum_{k=1}^{b-1} a_{s,k} \frac{1}{l} \sum_{m=0}^{l-1} R_n(\langle mb + s \rangle, k), \quad (17)$$

where  $a_{s,0} = 1, s = 1, \dots, d, a_{s,k} = 2d/b(v_{b-1} + \mathbb{I}_{\{r_{b-1} \geq s\}} - \lfloor \frac{k-1}{d} \rfloor), s = 1, \dots, d, k = 1, \dots, b-1$  and  $a_{s,k} = 0$  otherwise. Note that for  $d = 1$  these coefficients are matching those obtained for the MBB in stationary case (see Section 2). The coefficients  $a_{s,k}$  depend of  $n$ , but to simplify the notation we do not introduce the subscript  $n$ .

**Remark 5.1.** In the case of the CGSBB, applying Lemma 4.1, we obtain that the coefficients  $a_{s,k}$  in (17) are of the form:  $a_{s,0} = 1, s = 1, \dots, d, a_{s,k} = 2d/b(v_{b-1} + \mathbb{I}_{\{r_{b-1} \geq s\}} - \lfloor \frac{k-1}{d} \rfloor), s = 1, \dots, d, k = 1, \dots, b-1, a_{s,k} = 0$  for  $s = 1, \dots, d, b \leq k \leq n-b$  and  $a_{s,k} = a_{s,n-k}$  for  $s = 1, \dots, d, n-b+1 \leq k \leq \min\{n-d+s, n-1\}$ . However, without loss of generality, to simplify the computation of bias and variance, the approximation of  $T_n$  is obtained by taking  $\tilde{a}_{s,k} = a_{s,k}$  for  $s = 1, \dots, d, 1 \leq k \leq n-b$  and  $\tilde{a}_{s,k} = a_{s,n-k}$  for  $s = 1, \dots, d, n-b+1 \leq k \leq n-1$  can be used.

Theorem below generalizes results of Nordman (see Proposition 2.1 in Section 2).

**Theorem 5.2.** Assume that **A0** - **A3** hold. Moreover, let the coefficients  $a_{s,\tau} \in \mathbb{R}, 1 \leq s \leq d, 0 \leq \tau \leq n-1$  be such that

- (i)  $a_{s,\tau} = a_{s,m_\tau d}$ , where  $\tau = s_\tau + m_\tau d$  with  $s_\tau \in \{0, \dots, d-1\}, m_\tau \in \mathbb{N}$ ;
- (ii)  $\sup_n \max_{1 \leq s \leq d} \max_{0 \leq \tau \leq n-1} |a_{s,\tau}| < \infty$ .

Then we have that

$$\text{Var}(T_n) = \sum_{k=0}^{d-1} \frac{(2\pi)^2 A_{k,n}}{n} \int_0^{2\pi} K_n(k, \omega) f_k^2(\omega) d\omega + O\left(\left(\frac{A_n B_n}{n}\right)^{1/2} + B_n\right), \quad (18)$$

where

$$\begin{aligned} A_n &= \sum_{k=0}^{d-1} A_{k,n} = \frac{d}{2} \sum_{\tau=1}^{n-1} \left(1 - \frac{\tau}{n}\right)^2 \frac{1}{l} \sum_{j=-l+1}^{l-1} \left(1 - \frac{|j|}{l}\right) \sum_{s_1=1}^d a_{s,\tau} (a_{\langle s-jb \rangle, \tau} + a_{\langle s+jb \rangle, \tau}), \\ B_n &= \frac{1}{n} + \frac{\log(n)}{n^2} \left( \sum_{s=1}^d \sum_{k=1}^{n-1} |a_{s,k}| \right)^2 + \frac{1}{n} \left( \sum_{s=1}^d \sum_{k=1}^{n-1} \frac{k}{n} \left(1 - \frac{k}{n}\right) |a_{s,k}| \right), \\ A_{k,n} &= \sum_{\tau_1=1}^{n-1} \left(1 - \frac{\tau_1}{n}\right)^2 \sum_{s_1=1}^d \sum_{s_2=1}^d a_{s_1, \tau_1} a_{s_2, \tau_1} \cos\left(\frac{2\pi k}{d}(s_1 - s_2)\right) \\ &\quad \cdot \frac{1}{l^2} \sum_{j_1=0}^{l-1} \sum_{j_2=0}^{l-1} \cos\left(\frac{2\pi k}{d}(\langle j_1 - j_2 \rangle b)\right), \\ K_n(k, \omega) &= \frac{W_n(k, \omega) W_n(-k, -\omega)}{2\pi A_n} = \frac{|W_n(k, \omega)|^2}{2\pi A_{k,n}}, \\ W_n(k, \omega) &= \sum_{s=1}^d \sum_{\tau=1}^{n-1} a_{s,\tau} \left(1 - \frac{\tau}{n}\right) \frac{1}{l} \sum_{j=0}^{l-1} \exp\left(i \frac{2\pi k}{d}(\langle jb + s \rangle + \tau)\right) \exp(-i\omega\tau). \end{aligned}$$

The assumption (i) implies that the coefficients  $a_{s,\tau}$  do not depend on  $s_\tau$ , i.e. they do not depend on season from which  $\tau$  originates. Note that since  $\tau$  can be equal to 0, we took  $s_\tau \in \{0, \dots, d-1\}$ . In this form the condition (i) is satisfied for all bootstrap approaches for PC time series considered in this paper. For instance, in the case of the GSBB/CGSBB, the only term in the expression for  $a_{s,\tau}$  that depends on  $\tau$  is  $\lfloor \frac{\tau-1}{d} \rfloor$ . Indeed, we get that  $\lfloor \frac{s_\tau + m_\tau d - 1}{d} \rfloor = m_\tau$  does not depend on  $s_\tau$ . The condition (i) is used only in the final part of the proof of Theorem 5.2 and allows for a considerable simplification of calculations. However, as this is not a necessary condition, it could be replaced by another assumption for other bootstrap approaches.

Theorem 5.2 is very general. Below we state the form of variance in the case of the GSBB/CGSBB. In fact, we managed to simplify the expression for  $A_{k,n}$  and thus the obtained formula more resembles Nordman's formula (see (4)) obtained for the stationary case.

**Theorem 5.3.** Let conditions **A0** - **A3** hold. Then for the GSBB and the CGSBB, (18) can be expressed as follows

$$\text{Var}(T_n) = \frac{(2\pi)^2 \tilde{A}_n}{n} \sum_{k=0}^{d-1} \int_0^{2\pi} K_n(k, \omega) f_k^2(\omega) d\omega + O\left(\left(\frac{A_n B_n}{n}\right)^{1/2} + B_n\right), \quad (19)$$

where

$$\tilde{A}_n = d^2 \sum_{\tau_1=1}^{n-1} \left(1 - \frac{\tau_1}{n}\right)^2 (\tilde{a}_{\tau_1, n})^2.$$

Note that  $\tilde{A}_n$  matches the expression for  $A_n$  of Nordman (see 3). The only difference is  $d^2$ , which is due to the periodic structure of the PC time series considered in this paper. For  $d = 1$  the coefficients  $\tilde{a}_{\tau_1, n}$  for the GSBB and the CGSBB are equal to  $a_{k, n}$  for the MBB and the CBB, respectively.

The next two theorems specify the form of the bias and the variance of  $\text{Var}^*(\sqrt{n}\hat{\mu}^*)$ .

**Theorem 5.4.** *Under conditions A0 - A4 and as  $n \rightarrow \infty$ , the variance of the GSBB and the CGSBB variance estimators is of the form*

$$\text{Var}(\text{Var}^*(\sqrt{n}\hat{\mu}^*)) = \frac{b}{n} \frac{4(2\pi d)^2}{3} \sum_{k=0}^{d-1} |f_k(0)|^2 + o\left(\frac{b}{n}\right). \quad (20)$$

For  $d = 1$ , which corresponds to the stationary case, the formula (20) is exactly the same as that given in Proposition 2.2.

**Theorem 5.5.** *Assume that A0 - A2 hold. Moreover, let  $1/b + b^2/n = o(1)$  as  $n \rightarrow \infty$ . Then for the GSBB and the CGSBB we have*

$$\begin{aligned} \mathbb{E}(\text{Var}^*(\sqrt{n}\hat{\mu}^*)) &= \sigma_{as}^2 - \frac{1}{db} \sum_{s=1}^d \sum_{k=-\infty, k \neq 0}^{\infty} \left( r_{b-1} + d \left\lfloor \frac{k}{d} \right\rfloor \right) \text{Cov}(X_s, X_{s+k}) \\ &+ \frac{1}{db} \sum_{s=1}^d (\mathbb{I}_{\{r_{b-1} \geq s\}}) \sum_{k=-\infty, k \neq 0}^{\infty} \sum_{\bar{s}=0}^{d-1} \text{Cov}(X_{<\bar{s}b+s>}, X_{<\bar{s}b+s>+k}) + o\left(\frac{1}{b}\right), \end{aligned} \quad (21)$$

where  $\sigma_{as}^2$  is given by (12) and  $b-1 = v_{b-1}d + r_{b-1}$  with  $r_{b-1} \in \{0, \dots, d-1\}$ ,  $v_{b-1} \in \mathbb{N}$ .

**Remark 5.2.** *Note that the bias expansion, given by formula (21) with  $d = 1$ , matches the one obtained for the stationary case (see Proposition 2.2). However, in general (21) has two additional summands that are result of the periodic structure contained in the data and which depend on  $r_{b-1}$ , i.e. the remainder obtained from dividing  $b-1$  by the period length  $d$ . Moreover, the covariances  $\text{Cov}(X_{<\bar{s}b+s>}, X_{<\bar{s}b+s>+k})$  also depend on  $b$  or more precisely on  $r_b$ . The only way to obtain a MSE expansion that is not an implicit equation of  $b$  is to set  $b$  such that  $r_{b-1} = 0$ . This means that  $b = v_b d + 1$ ,  $v_b \in \mathbb{N}$ . Consequently, in the GSBB/CGSBB algorithm we will use all possible blocks of observations. To explain that let us assume that the first observation in the sample  $(X_1, \dots, X_n)$  is from season  $s = 1$ . Then in the first step of the algorithm one considers all blocks of the length  $b$ , whose first observation comes from season 1. Since  $b = v_b d + 1$ , in the second step one needs all blocks of the length  $b$  that start with an observation from season 2 etc. Finally, already in  $d$  consecutive steps of the GSBB algorithm one uses all possible blocks of observations. On the contrary, when  $b = v_b d$ , only blocks whose first observation is from season 1 are used in each step.*

Denote by  $\sigma_{b,GSBB}^{2*}$  and  $\sigma_{b,CGSBB}^{2*}$ , the bootstrap variance  $\text{Var}^*(\sqrt{n}\hat{\mu}^*)$  obtained with the block length  $b$  for the GSBB and the CGSBB, respectively. Applying Theorems 5.4 and 5.5 and setting the block length  $b = v_b d + 1$ ,  $v_b \in \mathbb{N}$  (see Remark 5.2), one gets the following expansion for the mean squared error

$$MSE(\sigma_{b,GSBB}^{2*}) = MSE(\sigma_{b,CGSBB}^{2*}) = \frac{b}{n} \frac{4(2\pi d)^2}{3} \sum_{k=0}^{d-1} |f_k(0)|^2 + \frac{G(d)^2}{b^2} + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b^2}\right),$$

where  $G(d) = \sum_{s=1}^d \sum_{k=-\infty}^{\infty} \left\lfloor \frac{k}{d} \right\rfloor \text{Cov}(X_s, X_{s+k})$ .

Taking

$$b_{opt} = \text{argmin}_b \{MSE(\sigma_b^{2*})\}.$$

one gets that

$$b_{opt,GSBB} = b_{opt,CGSBB} = \sqrt[3]{\frac{2G^2(d)}{D(d)}} \sqrt[3]{n}, \quad (22)$$

where  $D(d) = \frac{4(2\pi d)^2}{3} \sum_{k=0}^{d-1} |f_k(0)|^2$ . Thus, for both the GSBB and the CGSBB in the overall mean estimation problem,  $b_{opt}$  is proportional to  $n^{1/3}$ . Thus, as in stationary case (see Section 2) the optimal block length is proportional to  $n^{1/3}$ . Note also that for  $d = 1$  the constants  $D$  and  $G$  are equal to the corresponding constants obtained in the stationary case for the MBB/CBB.

**Remark 5.3.** Our proposed value of  $b_{opt}$  is in fact suboptimal. In the bias term (21), the dependence on  $b$  reduces to a dependence on  $r_{b-1}$  and  $r_b$ . Thus, in the general case where  $r_{b-1}$  need not be equal to 0, (21) can be expressed as

$$\mathbb{E}(\text{Var}^*(\sqrt{n}\hat{\mu}^*)) = \sigma_{as}^2 - \frac{1}{db}(G + G_1(r_{b-1}) - G_2(r_{b-1})) + o\left(\frac{1}{b}\right),$$

where  $G(d)$  is defined as before,  $G_1(r_{b-1}) = \sum_{s=1}^d \sum_{k=-\infty, k \neq 0}^{\infty} r_{b-1} \text{Cov}(X_s, X_{s+k})$  and  $G_2(r_{b-1}) = \sum_{s=1}^d (\mathbb{I}_{\{r_{b-1} \geq s\}}) \sum_{k=-\infty, k \neq 0}^{\infty} \sum_{\tilde{s}=0}^{d-1} \text{Cov}(X_{\langle \tilde{s}b+s \rangle}, X_{\langle \tilde{s}b+s \rangle+k})$ . Note that since  $r_{b-1}$  fully determines the value of  $r_b$ , we write  $G_2(r_{b-1})$  instead of  $G_2(r_{b-1}, r_b)$ . Furthermore,  $r_{b-1}$  depends on  $d$ , but to simplify notation we write  $G_i(r_{b-1}), i = 1, 2$ .

The constants  $G_1(r_{b-1})$  and  $G_2(r_{b-1})$  can take only a finite number of values  $G_1(0), \dots, G_1(d-1)$  and  $G_2(0), \dots, G_2(d-1)$ . We can therefore calculate the MSE for each value of  $r_{b-1}$  and define  $r^{opt} \in \{0, \dots, d-1\}$  such that

$$MSE(r^{opt}) = \min\{MSE(0), \dots, MSE(d-1)\}.$$

Then the optimal block size is

$$b_{opt,GSBB} = b_{opt,CGSBB} = \sqrt[3]{\frac{2\tilde{G}^2(r^{opt})}{D(d)}} \sqrt[3]{n},$$

where  $\tilde{G}(r^{opt}) = G + G_1(r^{opt}) - G_2(r^{opt})$ . Note that estimation of  $G_2(r_{b-1})$  is particularly difficult because the sum does not include all covariances. Since for now we do not have an efficient method to estimate  $\tilde{G}(0), \dots, \tilde{G}(d-1)$ , in the paper we always assume  $r_{b-1} = 0$ .

### 5.1.2. Results for the MBB/CBB

Let us recall that in the overall mean case the EMBB/CEMBB reduces to the MBB/CBB (see Section 3). Below we provide all corresponding results to those presented in the previous section for the GSBB/CGSBB. In particular, we give the form of  $\text{Var}^*(\sqrt{n}\hat{\mu}^*)$  as the weighted average of  $R_{n,cir}(s, k)$ . The expression is exact for the CBB and approximate for the MBB. Then we present the bias and the variance expansions for the bootstrap variance estimator. Finally, we obtain the optimal block length.

**Theorem 5.6.** Let  $\{X_t, t \in \mathbb{Z}\}$  be a PC real-valued time series with period  $d$  such that  $\sum_{s=1}^d \text{Var}(X_s) + \sum_{s=1}^d \sum_{k=-\infty}^{\infty} |\text{Cov}(X_s, X_{s+k})| < \infty$ . Moreover, assume that **A1** holds. Then for the CBB we get that

$$\begin{aligned} \text{Var}^*(\sqrt{n}\hat{\mu}^*) &= \frac{1}{d} \sum_{s=1}^d R_{n,cir}(s, 0) \\ &+ \frac{2}{b} \frac{1}{d} \sum_{s=1}^d \sum_{s_1=1}^d \sum_{k=1}^{b-1} \left( v_{b-1} + \mathbb{I}_{\{r_{b-1} \geq s\}} - \left\lfloor \frac{k-1}{d} \right\rfloor \right) R_{n,cir}(\langle s + s_1 - 1 \rangle, k) \end{aligned} \quad (23)$$

and for the MBB we have that

$$\begin{aligned} \text{Var}^*(\sqrt{n}\hat{\mu}^*) &\approx \frac{1}{d} \sum_{s=1}^d R_n(s, 0) \\ &+ \frac{2}{b} \frac{1}{d} \sum_{s=1}^d \sum_{s_1=1}^d \sum_{k=1}^{b-1} \left( v_{b-1} + \mathbb{I}_{\{r_{b-1} \geq s\}} - \left\lfloor \frac{k-1}{d} \right\rfloor \right) R_n(\langle s + s_1 - 1 \rangle, k), \end{aligned} \quad (24)$$

where

$$\begin{aligned} \mathbb{E}(\text{Var}^*(\sqrt{n}\hat{\mu}^*)) &= \mathbb{E}\left(\frac{1}{d} \sum_{s=1}^d R_n(s, 0)\right) \\ &+ \mathbb{E}\left(\frac{2}{b} \frac{1}{d} \sum_{s=1}^d \sum_{s_1=1}^d \sum_{k=1}^{b-1} \left( v_{b-1} + \mathbb{I}_{\{r_{b-1} \geq s\}} - \left\lfloor \frac{k-1}{d} \right\rfloor \right) R_n(\langle s + s_1 - 1 \rangle, k)\right) \\ &+ o\left(\frac{1}{b}\right), \end{aligned} \quad (25)$$

$$\begin{aligned}
\text{Var}(\text{Var}^*(\sqrt{n}\widehat{\mu}^*)) &= \text{Var}\left(\frac{1}{d}\sum_{s=1}^d R_n(s,0)\right) \\
&+ \frac{2}{b}\frac{1}{d}\sum_{s=1}^d\sum_{s_1=1}^d\sum_{k=1}^{b-1}\left(v_{b-1} + \mathbb{I}_{\{r_{b-1}\geq s\}} - \left\lfloor\frac{k-1}{d}\right\rfloor\right) R_n(\langle s+s_1-1\rangle, k) \\
&+ o\left(\frac{b}{n}\right),
\end{aligned} \tag{26}$$

and  $b-1 = v_{b-1}d + r_{b-1}$  with  $r_{b-1} \in \{0, \dots, d-1\}$ ,  $v_{b-1} \in \mathbb{N}$ .

**Theorem 5.7.** *Under conditions **A0** - **A4** and as  $n \rightarrow \infty$ , the variance of the CBB and the MBB variance estimators is of the form*

$$\text{Var}(\text{Var}^*(\sqrt{n}\widehat{\mu}^*)) = \frac{b}{n}\frac{4(2\pi d)^2}{3}\sum_{k=0}^{d-1}|f_k(0)|^2 + o\left(\frac{b}{n}\right). \tag{27}$$

Note that the variance expansion is exactly the same as the one for the GSBB/CGSBB (see Theorem 5.4). Thus, for  $d=1$ , it matches the formula obtained in the stationary case (see Proposition 2.2).

**Theorem 5.8.** *Assume that **A0** - **A2** hold. Moreover, let  $1/b + b^2/n = o(1)$  as  $n \rightarrow \infty$ . Then for the MBB and the CBB we have*

$$\begin{aligned}
\mathbb{E}(\text{Var}^*(\sqrt{n}\widehat{\mu}^*)) &= \sigma_{as}^2 - \frac{1}{db}\sum_{s=1}^d\sum_{k=-\infty, k\neq 0}^{\infty}\left(r_{b-1} + d\left\lfloor\frac{k}{d}\right\rfloor\right)\text{Cov}(X_s, X_{s+k}) \\
&+ \frac{1}{b}\sum_{s=1}^d(\mathbb{I}_{\{r_{b-1}\geq s\}})\left(\sigma_{as}^2 - \frac{1}{d}\sum_{s_1=1}^d\text{Var}(X_{s_1})\right) + o\left(\frac{1}{b}\right),
\end{aligned} \tag{28}$$

where  $\sigma_{as}^2$  is given by (12) and  $b-1 = v_{b-1}d + r_{b-1}$  with  $r_{b-1} \in \{0, \dots, d-1\}$ ,  $v_{b-1} \in \mathbb{N}$ .

**Remark 5.4.** *As in the case of the GSBB/CGSBB the bias expansion, given by formula (28) with  $d=1$ , matches the one obtained for stationary case (see Proposition 2.2). However, again in general case we have two additional terms that are a consequence of the periodic structure contained in the data and that depend on  $r_{b-1}$ , i.e. the remainder obtained from dividing  $b-1$  by the period length  $d$ . Thus, one should consider  $b = v_b d + 1$ ,  $v_b \in \mathbb{N}$ . Then, minimizing the MSE results in  $b_{opt}$  exactly the same as for the GSBB/CGSBB, i.e.  $b_{opt}$  is proportional to  $n^{1/3}$  (see (22)).*

**Remark 5.5.** *Note that the third summand on the right-hand side of (28) is equal to  $\frac{1}{b}\sum_{s=1}^d(\mathbb{I}_{\{r_{b-1}\geq s\}})\left(\sum_{s_1=1}^d\sum_{k=-\infty, k\neq 0}^{\infty}\text{Cov}(X_{s_1}, X_{s_1+k})\right)$ . The bias expansions (28) matches the corresponding expansion for the GSBB/CGSBB (21) only if  $r_{b-1} = 0$ . If this is not the case, which bootstrap approach produces a larger bias of the variance estimator depends on the values of the individual covariances.*

## 5.2. Seasonal means case

Let the season  $s$  be fixed. We consider estimation of the seasonal mean  $\mu_s$ . Under the condition **A1** our sample  $(X_1, \dots, X_n)$  contains  $v_1$  observations from the season  $s$ . The asymptotic variance of  $\sqrt{v_1}\widehat{\mu}_s$  is of the form

$$\sigma_{s,as}^2 = \sum_{k=-\infty}^{\infty}\text{Cov}(X_s, X_{s+kd}). \tag{29}$$

Let  $\widehat{\mu}_i^*$  be the bootstrap counterpart of  $\widehat{\mu}_i$ . Since the distance between two consecutive observations from the season  $s$  is always equal to  $d$ , we express  $\text{Var}^*(\sqrt{v_1}\widehat{\mu}_s^*)$ , depending on the bootstrap approach, in terms of  $R_{n,cir}(s, kd)$  or  $R_n(s, kd)$ . Note that series  $\{X_{s+kd}, k \in \mathbb{Z}\}$  is stationary. Let us recall that we bootstrap the original sample  $(X_1, \dots, X_n)$ , not just the stationary subsample.

We denote the spectral density function of  $\{X_{s+kd}, k \in \mathbb{Z}\}$  by

$$f_0^s(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\text{Cov}(X_s, X_{s+kd})\exp(-ikd\omega), \quad \omega \in [0, 2\pi).$$

### 5.2.1. Results for the CGSBB/GSBB

Similarly to the overall mean case, the bootstrap variance  $\text{Var}^*(\sqrt{v_1}\widehat{\mu}_s^*)$  can be expressed as the weighted average of  $R_n(s, k)$ . Depending on the bootstrap approach, the relationship is either exact or approximate.

**Theorem 5.9.** *Let  $\{X_t, t \in \mathbb{Z}\}$  be a PC real-valued time series with period  $d$  such that  $\sum_{s=1}^d \text{Var}(X_s) + \sum_{s=1}^d \sum_{k=-\infty}^{\infty} |\text{Cov}(X_s, X_{s+k})| < \infty$ . Moreover, assume that **A1** holds. Then for the CGSBB we get that*

$$\begin{aligned} \text{Var}^*(\sqrt{v_1}\widehat{\mu}_s^*) &= R_{n, \text{cir}}(s, 0) + 2 \frac{d}{n} \sum_{m=0}^{l-1} \sum_{k=1}^{v_{b-1}} (v_{b-1} - k + \mathbb{I}_{\{r_{b-1} \geq \langle s - mb \rangle\}}) R_{n, \text{cir}}(s, kd) \end{aligned} \quad (30)$$

and for the GSBB we have that

$$\text{Var}^*(\sqrt{v_1}\widehat{\mu}_s^*) \approx R_n(s, 0) + 2 \frac{d}{n} \sum_{m=0}^{l-1} \sum_{k=1}^{v_{b-1}} (v_{b-1} - k + \mathbb{I}_{\{r_{b-1} \geq \langle s - mb \rangle\}}) R_n(s, kd), \quad (31)$$

where

$$\begin{aligned} \text{E}(\text{Var}^*(\sqrt{v_1}\widehat{\mu}_s^*)) &= \text{E} \left( R_n(s, 0) + 2 \frac{d}{n} \sum_{m=0}^{l-1} \sum_{k=1}^{v_{b-1}} (v_{b-1} - k + \mathbb{I}_{\{r_{b-1} \geq \langle s - mb \rangle\}}) R_n(s, kd) \right) + o\left(\frac{1}{b}\right), \end{aligned} \quad (32)$$

$$\begin{aligned} \text{Var}(\text{Var}^*(\sqrt{v_1}\widehat{\mu}_s^*)) &= \text{Var} \left( R_n(s, 0) + 2 \frac{d}{n} \sum_{m=0}^{l-1} \sum_{k=1}^{v_{b-1}} (v_{b-1} - k + \mathbb{I}_{\{r_{b-1} \geq \langle s - mb \rangle\}}) R_n(s, kd) \right) + o\left(\frac{b}{n}\right). \end{aligned} \quad (33)$$

Note that  $k$  is the distance between observations in a block of length  $b$ . In the formulas for the overall mean case (Theorem 5.1)  $k = 1, \dots, b-1$ , while in (30)-(31)  $k = 1, \dots, v_{b-1}$ . In the overall mean case we use all observations in the block and the maximal distance is  $b-1$ . Here, we focus only on observations from the season  $s$  and hence the maximal value of  $k$  is  $v_{b-1}$ . The expression  $\mathbb{I}_{\{r_{b-1} \geq \langle s - mb \rangle\}}$  in the formulas (30)-(31) is a consequence of the fact that  $b-1$  may not be an integer multiple of the period length ( $r_{b-1} \neq 0$ ). Additionally, among all blocks of a fixed length  $b$ , some may have more observations from the season  $s$  than others. For instance, if  $b$  is an integer multiple of a period length ( $b = v_b d$ ), then all blocks have exactly the same number of observations from the season  $s$ . But if  $b = v_b d + r_b$  with  $r_b \neq 0$ , then some blocks will contain  $v_b$  observations from the season  $s$ , while others  $v_b + 1$ .

Below we provide the variance and bias expansions for  $\text{Var}^*(\sqrt{v_1}\widehat{\mu}_s^*)$ . We omit the details of the proofs because the reasoning proceeds exactly as presented by Nordman (2009) for the overall mean of a stationary time series with some adaptations to the case of the seasonal mean case considered here.

**Theorem 5.10.** *Under conditions **A0** - **A4** and as  $n \rightarrow \infty$ , the variance of the GSBB and the CGSBB seasonal mean variance estimators is of the form*

$$\text{Var}(\text{Var}^*(\sqrt{v_1}\widehat{\mu}_s^*)) = \frac{b}{n} \frac{4(2\pi d)^2}{3} (f_0^s(0))^2 + o\left(\frac{b}{n}\right). \quad (34)$$

When  $d = 1$ , i.e. data are stationary, the estimation of  $\mu_s$  reduces to the estimation of  $\mu$  and the formula (34) exactly matches the formula in Proposition 2.2. In general case, the extra  $d^2$  in the first summand of (34) is a consequence of the periodicity in the data.

**Theorem 5.11.** *Assume that **A0** - **A2** hold. Moreover, let  $1/b + b^2/n = o(1)$  as  $n \rightarrow \infty$ . Then for the GSBB/CGSBB we have*

$$\begin{aligned} \text{E}(\text{Var}^*(\sqrt{v_1}\widehat{\mu}_s^*)) &= \sigma_{s, as}^2 - \frac{1}{b} \sum_{k=-\infty, k \neq 0}^{\infty} (r_{b-1} + dk) \text{Cov}(X_s, X_{s+dk}) \\ &+ \frac{1}{b} \sum_{\tilde{s}=0}^{d-1} (\mathbb{I}_{\{r_{b-1} \geq \langle s + d - \tilde{s}b \rangle\}}) \sum_{k=-\infty, k \neq 0}^{\infty} \text{Cov}(X_s, X_{s+dk}) + o\left(\frac{1}{b}\right), \end{aligned} \quad (35)$$

where  $\sigma_{as}^2$  is given by (29) and  $b-1 = v_{b-1}d + r_{b-1}$  with  $r_{b-1} \in \{0, \dots, d-1\}$ ,  $v_{b-1} \in \mathbb{N}$ .



As with the overall mean, (35) contains summands that depend on  $r_{b-1}$ . In fact the expansion (35) depends also on  $r_b$ , because  $\langle s + d - \tilde{s}b \rangle = \langle s + d - \tilde{s}r_b \rangle$ , where  $b = v_b d + r_b$ ,  $v_b \in \mathbb{N}$  and  $r_b \in \{0, \dots, d-1\}$ . To be able to minimize the MSE, as for the overall mean, we take  $b = v_b d + 1$  and hence we get  $r_{b-1} = 0$ .

We denote by  $\sigma_{b,s,GSBB}^{2*}$  and  $\sigma_{b,s,CGSBB}^{2*}$ , the bootstrap variance  $\text{Var}^*(\sqrt{v_1} \hat{\mu}_s^*)$  obtained with the block length  $b$  for the GSBB and the CGSBB, respectively. Applying Theorems 5.10 and 5.11 and setting the block length  $b = v_b d + 1$ ,  $v_b \in \mathbb{N}$ , we get the following expansion for the mean squared error

$$MSE(\sigma_{b,s,GSBB}^{2*}) = MSE(\sigma_{b,s,CGSBB}^{2*}) = \frac{b}{n} D_s(d) + \frac{G_s^2(d)}{b^2} + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b^2}\right),$$

where  $D_s(d) = \frac{4}{3} (2\pi df_0^s(0))^2$  and  $G_s(d) = \sum_{k=-\infty}^{\infty} d|k| \text{Cov}(X_s, X_{s+kd})$ . Taking

$$b_{s,opt} = \text{argmin}_b \{MSE(\sigma_{b,s}^{2*})\}.$$

we have that

$$b_{s,opt,GSBB} = b_{s,opt,CGSBB} = \sqrt[3]{\frac{2G_s^2(d)}{D_s(d)}} \sqrt[3]{n}.$$

Thus, for both the GSBB and the CGSBB in the seasonal mean estimation problem,  $b_{s,opt}$  is proportional to  $n^{1/3}$ .

In practice one does not estimate the single seasonal mean, but all seasonal means  $\mu_s$ ,  $s = 1, \dots, d$  at the same time. The data are bootstrapped once using some fixed  $b$  and all bootstrap estimates are computed. Thus, to find  $b_{opt}$  in this case, we propose instead minimizing  $MSE(\sigma_{b,s}^{2*})$  to minimize  $\sum_{s=1}^d MSE(\sigma_{b,s}^{2*})$ . Then

$$b_{all,opt} = \text{argmin}_b \left\{ \sum_{s=1}^d MSE(\sigma_{b,s}^{2*}) \right\},$$

which gives

$$b_{all,opt,GSBB} = b_{all,opt,CGSBB} = \sqrt[3]{\frac{2 \sum_{s=1}^d G_s^2(d)}{\sum_{s=1}^d D_s(d)}} \sqrt[3]{n}.$$

Finally,  $b_{all,opt}$  can be used to construct the simultaneous confidence intervals for the seasonal means.

**Remark 5.6.** In Section 4.1 we showed that the SBB for  $\{X_t\}$  is equivalent to the MBB for the  $d$ -variate stationary time series. Since we consider  $r_{b-1} = 0$  and consequently  $b = v_b d + 1$ , one might think that we can simply perform the MBB for the stationary time series  $\{X_{s+kd}, k \in \mathbb{Z}\}$ , obtain the optimal block length  $v_b^{opt}$ , and then use it to compute  $b_{opt}$  for  $\{X_t\}$ . However, we would like to point out that this approach is not correct. Note that by using the GSBB with  $b = v_b d + 1$ , we obtain a bootstrap sample consisting of blocks that contain  $v_b$  or  $v_b + 1$  observations from a fixed season  $s$ . More specifically, for example, if  $s = 1$ , then the first selected block contains  $v_b + 1$  observations from the first season. Subsequent  $d - 1$  blocks each contain  $v_b$  observations from the first season, and so on, i.e., blocks numbered  $1, d+1, 2d+1, \dots$  have  $v_b + 1$  observations from the first season, and all others have only  $v_b$ . On the other hand, when we apply the MBB to the stationary time series  $\{X_{s+kd}, k \in \mathbb{Z}\}$ , all bootstrap blocks have  $v_b$  observations from the first season.

### 5.2.2. Results for the EMBB/CEMBB

Below we provide all the results for the EMBB/CEMBB. We only comment on those that differ from the corresponding ones in Section 5.2.1.

**Theorem 5.12.** Let  $\{X_t, t \in \mathbb{Z}\}$  be a PC real-valued time series with period  $d$  such that  $\sum_{s=1}^d \text{Var}(X_s) + \sum_{s=1}^d \sum_{k=-\infty}^{\infty} |\text{Cov}(X_s, X_{s+k})| < \infty$ . Moreover, assume that **A1** holds. Then for the CEMBB we get that

$$\begin{aligned} & \text{Var}^*(\sqrt{v_1} \hat{\mu}_s^*) \\ & \approx R_{n,cir}(s, 0) + 2 \frac{1}{b} \sum_{s_1=1}^d \sum_{k=1}^{v_b-1} (v_b-1 - k + \mathbb{I}_{\{r_{b-1} \geq \langle s-s_1 \rangle\}}) R_{n,cir}(s, kd), \end{aligned} \quad (36)$$

where for the CEMBB

$$\begin{aligned} & \text{E}(\text{Var}^*(\sqrt{v_1} \hat{\mu}_s^*)) \\ & = \text{E} \left( R_{n,cir}(s, 0) + 2 \frac{1}{b} \sum_{s_1=1}^d \sum_{k=1}^{v_b-1} (v_b-1 - k + \mathbb{I}_{\{r_{b-1} \geq \langle s-s_1 \rangle\}}) R_{n,cir}(s, kd) \right) + o\left(\frac{1}{b}\right), \end{aligned} \quad (37)$$

$$\begin{aligned} & \text{Var}(\text{Var}^*(\sqrt{v_1}\hat{\mu}_s^*)) \\ &= \text{Var}\left(R_{n,cir}(s,0) + 2\frac{1}{b}\sum_{s_1=1}^d\sum_{k=1}^{v_{b-1}}(v_{b-1}-k + \mathbb{I}_{\{r_{b-1}\geq <s-s_1>\}})R_{n,cir}(s,kd)\right) + o\left(\frac{b}{n}\right). \end{aligned} \quad (38)$$

Moreover, for the EMBB (36), (37), (38) hold with  $R_{n,cir}(\cdot, \cdot)$  replaced by  $R_n(\cdot, \cdot)$ .

As in the CGSBB/GSBB case, the expression  $\mathbb{I}_{\{r_{b-1}\geq <s-s_1>\}}$  is a consequence of the fact that a block of length  $b-1$  may not be an integer multiple of the period length ( $r_{b-1} \neq 0$ ) and furthermore that the considered block starts with an observation from the season  $s_1$  ( $s_1$  can take any value in the set  $\{1, \dots, d\}$ ). Since  $s = s_1 + (s - s_1)$  and we are interested only in the observations from the season  $s$ , we consider only the observations  $X_{s_1+(s-s_1)+wd}$ ,  $w \in \mathbb{N}$ , belonging to this block. Finally, the term  $\mathbb{I}_{\{r_{b-1}\geq <s-s_1>\}}$  is nonzero, if the observation from the season  $s$  is among the last  $r_{b-1}$  observations in the block starting with the observation from the season  $s_1$  and having length  $b-1$ . Recall that we are considering a block of length  $b-1$ , since we are interested in the distance between the observations. In the case of the overall mean, all observations in a block of length  $b$  were used and hence the maximal distance between them was  $b-1$ . Here, we have only observations from one season and the maximal distance is  $v_{b-1}$ .

Note that the variance expansions of both the EMBB and the CEMBB for the seasonal means are approximate. Recall that for the overall mean, the variance expansions for the CGSBB and the CBB are exact. The same phenomenon is observed for the CGSBB in the case of seasonal means (see Theorem 5.10. Since the CGSBB preserves the periodicity of the data in the bootstrap sample we have exactly the same number of observations from each season as in the original sample. It no longer true for the CEMBB. The expansion would only be exact if the block length is an integer multiple of the period length.

Similarly to the GSBB/CGSBB case one may obtain the following results.

**Theorem 5.13.** *Under conditions A0 - A4 and as  $n \rightarrow \infty$ , the variance of the EMBB and the CEMBB seasonal mean variance estimators is of the form*

$$\text{Var}(\text{Var}^*(\sqrt{v_1}\hat{\mu}_s^*)) = \frac{b}{n}\frac{4(2\pi d)^2}{3}(f_0^s(0))^2 + o\left(\frac{b}{n}\right). \quad (39)$$

**Theorem 5.14.** *Assume that A0 - A2 hold. Moreover, let  $1/b + b^2/n = o(1)$  as  $n \rightarrow \infty$ . Then for the EMBB/CEMBB we have*

$$\begin{aligned} \mathbb{E}(\text{Var}^*(\sqrt{v_1}\hat{\mu}_s^*)) &= \sigma_{s,as}^2 - \frac{1}{b}\sum_{k=-\infty, k \neq 0}^{\infty}(r_{b-1} + dk)\text{Cov}(X_s, X_{s+dk}) \\ &+ \frac{1}{b}\sum_{s_1=1}^d(\mathbb{I}_{\{r_{b-1}\geq <s-s_1>\}})\sum_{k=-\infty, k \neq 0}^{\infty}\text{Cov}(X_s, X_{s+dk}) + o\left(\frac{1}{b}\right). \end{aligned} \quad (40)$$

As in the overall mean case (see Theorem 5.8) the bias expansion depends on  $r_{b-1}$ . To minimize the MSE, we take  $b = v_b d + 1$  to get  $r_{b-1} = 0$  (as we did for the GSBB/CGSBB). This leads to the same results as obtained for the GSBB/CGSBB (see Section 5.2.1), i.e.

$$b_{s,opt,EMBB} = b_{s,opt,CEMBB} = \sqrt[3]{\frac{2G_s^2(d)}{D_s(d)}}\sqrt[3]{n}.$$

and

$$b_{all,opt,EMBB} = b_{all,opt,CEMBB} = \sqrt[3]{\frac{2\sum_{s=1}^d G_s^2(d)}{\sum_{s=1}^d D_s(d)}}\sqrt[3]{n},$$

where all constants are defined in Section 5.2.1.

**Remark 5.7.** *Note that the bias expansions (40) coincides with the corresponding expansion for the GSBB/CGSBB (35) only when  $r_{b-1} = 0$ . In all other cases, the indicator functions appearing in both equations are not equal. Therefore, which bootstrap method produces larger bias of the variance estimator depends on the values of the individual covariances.*

## 6. Summary and conclusions

In this paper, we considered four block bootstrap approaches designed for periodic time series. These are the GSBB, CGSBB, EMBB and CEMBB methods. We focused on two parameters: the overall mean and the

seasonal means. In all cases, we tried to solve the problem of the block length choice. First of all, our results indicated choice of the block length that is equal to an integer multiple of the period length plus one. Such choice guarantees for the GSBB and the CGSBB that in the bootstrap algorithm all possible blocks will be used. Furthermore, in all cases, we found the optimal bootstrap block length  $b_{opt}$  minimizing the MSE of the corresponding bootstrap variance estimator and we obtained  $b_{opt}$  proportional to  $n^{1/3}$ . Thus, the order is the same as in the stationary case for the overall mean when the MBB or the CBB is used. One may expect that this property will hold for other bootstrap methods for PC time series.

In Dudek et al. (2016) the authors showed that the Tapered Block Bootstrap can be adapted to PC processes. In the stationary case,  $b_{opt}$  for the TBB is proportional to  $n^{1/5}$ . However, since already in the stationary case this result was technically more challenging than the corresponding one for the MBB, we expect it for PC time series to be much more difficult than proofs presented in this paper. Also it seems possible to modify the Stationary Bootstrap for periodic sequences. In this case we believe that the MSE expansion will match (up to some constants) to that obtained by Nordman (2009).

## Funding

Research of Patrice Bertail has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01). Anna Dudek acknowledges support from the King Abdullah University of Science and Technology (KAUST) Research Grant OSR-2019-CRG8-4057.2.

## References

- Antoni, J. (2009). Cyclostationarity by examples. *Mech. Syst Sig Process.*, **23**(4), 987–1036.
- Bühlman, P. and Künsch, H. R. (1999). Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis*, **31**, 295–310.
- Brillinger, D.R. (2001). *Time Series: Data Analysis and Theory, Society for Industrial and Applied Mathematics*, Philadelphia.
- 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.
- Chan, V., Lahiri, S.N., and Meeker, W.Q (2004). Block bootstrap estimation of the distribution of cumulative outdoor degradation. *Technometrics*, **46**, 215–224.
- Dudek, A.E. (2015). Circular block bootstrap for coefficients of autocovariance function of almost periodically correlated time series. *Metrika*, **78**(3), 313–335.
- Dudek, A.E. (2018). Block bootstrap for periodic characteristics of periodically correlated time series. *Journal of Nonparametric Statistics*, **30**(1), 87–124.
- Dudek, A.E., Leśkow, J., Paparoditis, E. and Politis, D. (2014). A generalized block bootstrap for seasonal time series. *J. Time Ser. Anal.*, **35**, 89–114.
- Dudek, A.E., Paparoditis, E. and Politis, D. (2016). Generalized Seasonal Tapered Block Bootstrap, *Statistics and Probability Letters*, **115**, 27–35.
- Dudek, A.E. and Potorski P. (2020). Bootstrapping the Autocovariance of PC Time Series - A Simulation Study - *Cyclostationarity: Theory and Methods - IV; Contributions to the 10th Workshop on Cyclostationary Systems and Their Applications, February 2017, Grodek, Poland*; Editors: F. Chaari, J. Leskow, R. Zimroz, A. Wylomanska, A. Dudek; Part of the Applied Condition Monitoring book series (ACM, volume 16) published by Springer, Cham, pp 41–55.
- Gardner, W.A. (1994). *Cyclostationarity in communications and and signal processing*. IEEE Press.
- Gardner, W.A., Napolitano, A. and Paura, L. (2006). Cyclostationarity: half a century of research. *Signal Processing*, **86**, 639–697.
- Gladyshev, E.G. (1961). Periodically Correlated Random Sequences. *Soviet mathematics*, **2**, 385–388.
- Hall, P., Horowitz, J.L. and Jing, B.Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, **82**, 561–574.
- Lenart, L. (2011). Asymptotic distributions and subsampling in spectral analysis for almost periodically correlated time series. *Bernoulli*, **17**(1), 290–319.
- Hurd, H.L. and Miamee, A.G. (2007). *Periodically Correlated Random Sequences: Spectral. Theory and Practice*. Wiley.
- Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, **17**, 1217–1241.
- Lahiri, S.N. (1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist.*, **27**(1), 386–404.
- Lahiri, S.N. (2003) *Resampling Methods for Dependent Data*, Springer, New York.
- Liu, R. and Singh, K. (1992). Moving block jackknife and bootstrap capture weak dependence. Exploring the Limits of Bootstrap, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist. Wiley, New York, pp 225–248.

- Loève, M. (1963). *Probability Theory*. Van Nostrand, New York
- Napolitano, A. (2012). *Generalizations of Cyclostationary Signal Processing: Spectral Analysis and Applications*. Wiley-IEEE Press.
- Napolitano, A. (2016). Cyclostationarity: New trends and applications. *Signal Process.*, **120**, 385–408.
- Napolitano, A. (2019). *Cyclostationary Processes and Time Series. Theory, Applications, and Generalizations*. Academic Press.
- Nematollahi, A.R. and Subba Rao, T. (2005). On the Spectral Density Estimation of Periodically Correlated (Cyclostationary) Time Series. *Sankhya Ser A*, **67(3)**, 568–589.
- Nordman, D. (2009). A note on the stationary bootstrap. *Ann. Stat.*, **37**, 359–370.
- Politis, D.N. (2001). Resampling time series with seasonal components, in *Frontiers in Data Mining and Bioinformatics: Proceedings of the 33rd Symposium on the Interface of Computing Science and Statistics*, Orange County, California, June 13-17, pp. 619–621.
- Politis, D.N. and Romano, J.P. (1992). A circular block-resampling procedure for stationary data. *Exploring the Limits of Bootstrap*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist. Wiley, New York, pp 263–270.
- Synowiecki, R. (2007). Consistency and application of moving block bootstrap for nonstationary time series with periodic and almost periodic structure. *Bernoulli*, **13(4)**, 1151–1178.