Mean square consistency and improved rate of convergence of generalized subsampling estimator for nonstationary time series

Patrice Bertail, Anna E. Dudek, Lukasz Lenart

Mean square consistency and improved rate of convergence of generalized subsampling estimator for nonstationary time series

Patrice Bertail*
MODAL'X UMR CNRS 9023, Université Paris Nanterre,
200, Avenue de la République, 92001 Nanterre, France
and
Anna E. Dudek†
Department of Applied Mathematics, AGH University of Krakow,
al. Mickiewicza 30, 30-059 Krakow, Poland
and
Łukasz Lenart‡
Department of Mathematics, Kracow University of Economics,
ul. Rakowicka 27, 31-510 Krakow, Poland

Abstract

In this paper, we show the mean square consistency for a generalized subsampling estimator based on the aggregation of the mean, median, and trimmed mean of some subsampling estimators for general nonstationary time series. Consistency requires standard assumptions, including the existence of moments and α-mixing conditions. We apply our results to the Fourier coefficients of the autocovariance function of periodically correlated time series. Furthermore, as in the i.i.d. case, we show that the generalized subsampling estimator satisfies Bernstein inequality and concentrates at an improved rate (under the condition of no or small bias) compared to the original estimator. Finally, we illustrate our results with some simulation data examples.

*patrice.bertail@parisnanterre.fr,*This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).*

†aedudek@agh.edu.pl, *Anna Dudek was partially supported by the Faculty of Applied Mathematics AGH UST statutory tasks within subsidy of Ministry of Science and Higher Education.*

‡lukasz.lenart@uek.krakow.pl, *Łukasz Lenart was financed from the subsidy granted to the Krakow University of Economics.*

Keywords: Subsampling, bagging, hyper-efficiency, nonstationarity, Periodic ARMA
MSC Classification: 62G05,62M10,62M15

1 Introduction

The subsampling procedure introduced in Politis and Romano [1992] and Politis and Romano [1994] (see also Politis et al. [1999]) works under mild assumptions. Its idea is to consider subsamples $B_t = (X_t, X_{t+1}, \ldots, X_{t+b-1})$ of length b from the original sample $(X_1, X_2, \ldots, X_n)$, where $b = b(n)$ is an integer sequence such that $b/n \longrightarrow 0$. Suppose θ is the unknown parameter of interest. We then denote its estimator based on $(X_1, X_2, \ldots, X_n)$ by $\widehat{\theta}_n$, and its subsampling version based on $(X_t, X_{t+1}, \ldots, X_{t+b-1})$ by $\widehat{\theta}_{n,b,t} = \widehat{\theta}_{n,b,t}(B_t)$. Lenart [2018] generalized the subsampling approach and, to construct the estimator $\widehat{\theta}_{n,b,\mathbf{T}_k}$, proposed using a number k of subsamples simultaneously of different lengths and with a vector of starting points $\mathbf{T}_k = (t_1, t_2, \ldots, t_k)$.

The idea behind Politis and Romano's subsampling is to approximate the non-degenerate limiting distribution of $\tau_n(\widehat{\theta}_n - \theta)$, where τ_n is a normalizing sequence, using the empirical distribution

$$L_{n,b}(A) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} \mathbb{I}\{\tau_b(\widehat{\theta}_{n,b,t} - \widehat{\theta}_n) \in A\}, \tag{1}$$

where A is a Borel set, and $\mathbb{I}\{C\}$ is an indicator function of the event C. Finally, the normalizing sequence τ_b is often obtained by substituting in τ_n, b by n. In the general case of the generalized subsampling procedure, the normalizing sequence τ_b can be difficult to obtain and may need to be estimated, as is done, for example, in Bertail et al. [1999] or Bertail et al. [2024]. For simplicity, we assume here that the rate of convergence is known.

In this paper, we use the generalized subsampling procedure to construct mean square consistent subsampling estimators, where the estimator $\widehat{\theta}_{n,b,\mathbf{T}_k}$ is computed as the mean, median, or trimmed mean of the set of estimator values calculated on individual subsamples, i.e., on the set $\{\widehat{\theta}_{n,b,t_j}, j = 1, 2, \ldots, k\}$. We also generalize these results to

the case when the parameter θ depends on another unknown parameter.

Aggregating estimators to obtain improved estimators is not a new idea and has received considerable attention in the statistical learning literature since the work of Breiman [1996] on bagging (bootstrap aggregating), which is a resampling technique whose main goal is to reduce the variability of a given estimator. The idea is that averaging many estimators makes the resulting output more stable. The subsampling version of bagging, called subbagging, was analyzed in an i.i.d. framework by Bühlmann and Yu [2002]. Recall that aggregation methods are now standard procedures for reducing computational cost and variance in statistical learning (see Hastie et al. [2009], Bühlmann [2003] and references therein) for non-regular estimators. They can be applied in many different contexts and can yield substantial gain in terms of performance, see for instance Yang [2004]. We refer to the recent survey González et al. [2020] for additional references, algorithms, and a practical evaluation of aggregation methods.

In the i.i.d. case it can be proven that averaging over several subsamples may improve the original estimator based on n observations (see Bertail et al. [2018], Bertail et al. [2024], Politis [2024]).

This phenomenon has already been noticed and studied in some specific cases in the literature on subbagging. For regular M-estimators, we refer to Zou et al. [2021] and Gao et al. [2022], who proved the asymptotic normality of mean aggregated subsampling estimators in the i.i.d. case. In this regular framework, even if there is no improvement in terms of convergence rate, the authors proved that with the adequate choice of subsampling size, the efficiency can be similar and the computational cost clearly reduced for large datasets.

Recently, Bertail et al. [2024] proved some general concentration inequality in the framework of statistical learning procedures, while Politis [2024] independently studied the asymptotic normality of the subbagged estimator under general conditions. These papers show that the mean of several subsampling estimators improves the convergence rate over the original estimator when its original rate of convergence is slower than $\sqrt{n}$.

A simpler version of this result, which is based on splitting the data into non-overlapping blocks and averaging over the estimators computed on each subsample was studied in Banerjee et al. [2019] in the framework of isotonic regression. In this case, it is known that the rate of convergence of the standard monotonicity-constrained estimator is of the order $n^{1/3}$. In contrast, the pooled estimator based on subsamples has a better rate, close to $n^{1/2}$ under some additional conditions. The main drawback of this approach is that this pooled estimator is hyper-efficient, and therefore lacks uniformity (over some given class of function) in a minimax sense.

The second part of the paper is devoted to deriving some concentration inequalities in the nonstationary case for our aggregated estimators. One of our goals is to show that we have the same type of hyper-efficiency phenomenon for dependent data using the generalized subsampling estimators proposed in Fukuchi [1999] and Lenart [2018] as in the i.i.d case.

The paper is organized as follows. In Section 2 we provide a brief introduction to generalized subsampling estimators, with a particular emphasis on the hypotheses required for the main results of the paper. In Section 3, we present results on the mean square consistency for the generalized subsampling estimators. In Section 4, we obtain some Bernstein inequalities for the generalized subsampling estimators, which allow us to get an improved convergence rate. Section 5 is devoted to certain extensions of our consistency results when the estimators depend on certain nuisance parameters that must be estimated. In Section 6, we present some applications of these results to a class of cyclostationary time series. Moreover, we show empirically that aggregating subsampling estimators improves the rate of convergence for simple estimators based on the recentered periodogram. The technical proofs are given in the supplemenary materials.

2 Generalized subsampling estimator

We will first recall the idea of the generalized subsampling procedure introduced by Lenart [2018]. In the standard subsampling approach, we consider all possible

subsamples of fixed length b. In the generalized subsampling approach, we select some number of these subsamples, while further including a step to allow the subsample length b to vary in a range. Specifically, let b_1 and b_2 denote two sequences corresponding to the minimum and maximum block length, respectively ($b_1 \leq b_2$). Let k_1 and k_2 be the possible minimum and maximum number (respectively) of selected blocks such that $k_1 \leq k_2$. For a given length b with $b_1 \leq b \leq b_2$, we get that the number of all possible subsamples is $q_{n,b} = n - b + 1$. By $S_{n,b} = \{1, 2, \ldots, q_{n,b}\}$ we denote a set of indices of the first observations (abbreviated as, initial indices) in these subsamples. Let $BK = BK(n) = \{b_1, b_1 + 1, b_1 + 2, \ldots, b_2\} \times \{k_1, k_1 + 1, k_1 + 2, \ldots, k_2\}$ be a set of all possible pairs (b, k). For a fixed (b, k), $\mathbf{T}_k = (t_1, t_2, \ldots, t_k) \in S_{n,b}^k$ is an associated vector of initial indices of length k. Let (b^*, k^*) be a random vector with discrete probability distribution on BK. For each $(b, k) \in BK$, we denote $p_{(b,k)} = P^*\Big((b^*, k^*) = (b, k)\Big)$. Then, conditionally to $(b^*, k^*) = (b, k) \in BK$ let $\mathbf{T}_{k^*}^* = (t_1^*, t_2^*, \ldots, t_k^*)$ be a random vector with the support $S_{n,b}^k$ with conditional probabilities denoted by $P^*\Big(\mathbf{T}_{k^*}^* = \mathbf{T}_k \in S_{n,b}^k \Big| (b^*, k^*) = (b, k)\Big) = p_{\mathbf{T}_k|(b,k)}$. Note that $\mathbf{T}_{k^*}^*$ has a random length equal to k^*. Finally, we denote the joint probability of $(b^*, k^*, \mathbf{T}_{k^*}^*)$ as $P^*\Big((b^*, k^*, \mathbf{T}_{k^*}^*) = (b, k, \mathbf{T}_k)\Big) = p_{(b,k,\mathbf{T}_k)}$.

Let $(X_1, X_2, \ldots, X_n)$ be an univariate sample from a time series $\{X_t, t \in \mathbb{Z}\}$ and $\widehat{\theta}_n$ be an estimator of the (real or complex-valued) parameter θ. We assume that $E(\widehat{\theta}_n) \longrightarrow \theta$, as $n \longrightarrow \infty$. By $\widehat{\theta}_{n,b,t}$ we denote the classical subsampling estimator of θ based on a subsample $(X_t, X_{t+1}, \ldots, X_{t+b-1})$ of length b, where $1 \leq t \leq n - b + 1$. By $\widehat{\theta}_{n,b,\mathbf{T}_k}$ we denote the generalized subsampling estimator (GSE) based on a set of k subsamples of length b (see examples in Lenart [2018]). The vector $\mathbf{T}_k$ provides the initial indices for these subsamples.

In this paper we generalize the subsampling estimator proposed by Fukuchi [1999]. Fukuchi's estimator is based on the classical subsampling procedure and of the form

$$\Psi_{n,b}^{\text{Fukuchi}} = \frac{1}{n - b + 1} \sum_{t=1}^{n-b+1} \widehat{\theta}_{n,b,t}. \tag{2}$$

Note that $\Psi_{n,b}^{\text{Fukuchi}} = E^*(\widehat{\theta}_{n,b,t^*})$, where t^* is a random variable with uniform discrete distribution on the set $S_{n,b}$.

In our case, instead of $E^*(\widehat{\theta}_{n,b,t^*})$, we consider the expectation of the GSE, i.e., $E^*(\widehat{\theta}_{n,b^*,\mathbf{T}^*_{k^*}})$. The estimator is then of the form

$$\widehat{\theta}_{n,BK} = E^*\left[\widehat{\theta}_{n,b^*,\mathbf{T}^*_{k^*}}\right] = \sum_{(b,k)\in BK} p_{(b,k)} \sum_{\mathbf{T}_k \in S^k_{n,b}} p_{T_k|(b,k)} \widehat{\theta}_{n,b,\mathbf{T}_k}. \tag{3}$$

Remark 2.1. Note that $\widehat{\theta}_{n,BK}$ is defined as the expected value of a discrete random variable, which can take up to $\sum_{(b,k)\in BK}(n-b+1)^k$ different values. Therefore, due to the high computational cost, it may be necessary to use a Monte Carlo procedure to approximate $\widehat{\theta}_{n,BK}$. Furthermore, note that the (conditionally on the available sample) the distribution of $\widehat{\theta}_{n,b^*,\mathbf{T}^*_{k^*}}$ is a discrete distribution, which is a mixture (with weights equal to $p_{(b,k)}$) of $\#BK$ classical subsampling distributions, each with fixed k and b. By $\#BK$ we denote the cardinality of the set BK.

In this paper, we consider the following assumptions.

Assumption 2.1. *The random variables (b^*,k^*) and $\boldsymbol{T}^*_{k^*}\big|(b^*,k^*)=(b,k)$ have uniform discrete distributions on BK and $S^k_{n,b}$, respectively.*

Assumption 2.2. *The random variable $\boldsymbol{T}^*_{k^*}\big|(b^*,k^*)=(b,k)$ has a discrete distribution on $S^k_{n,b}$ such that there exists constants C_p and C_q (which do not depend on (b,k)) such that $p_{\boldsymbol{T}_k|(b,k)} \leq \frac{C_p}{q^k_{n,b}}$ and $p(b,k) \leq \frac{C_p}{(b_2-b_1+1)(k_2-k_1+1)}$, where $q_{n,b} = n-b+1$.*

Note that Assumption 2.2 is more general than 2.1. However, in both we assume that $p_{\mathbf{T}_k|(b,k)}$ is proportional to $1/q^k_{n,b}$.

Additionally, we will assume that $\{X_t : t \in \mathbb{Z}\}$ is α-mixing, i.e. $\alpha_X(k) \longrightarrow 0$ as $k \longrightarrow \infty$, 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 $\mathcal{F}_X(-\infty, t) = \sigma(\{X_s : s \leq t\})$, $\mathcal{F}_X(t+k, \infty) = \sigma(\{X_s : s \geq t+k\})$.

In what follows, we will consider three different estimators $\widehat{\theta}_{n,b,\mathbf{T}_k}$ (see formula (3)). These are:

i) sample mean

$$\widehat{\theta}^{Mean}_{n,b,\mathbf{T}_k} = \frac{1}{k}\sum_{j=1}^{k} \widehat{\theta}_{n,b,t_j},$$

ii) sample median

$$\widehat{\theta}^{Median}_{n,b,\mathbf{T}_k} = \mathrm{med}\{\widehat{\theta}_{n,b,t_j} : j = 1, 2, \ldots, k\},$$

iii) sample trimmed mean

$$\widehat{\theta}^{TM(f)}_{n,b,\mathbf{T}_k} = \mathrm{TM}^{(f)}\{\widehat{\theta}_{n,b,t_j} : j = 1, 2, \ldots, k\},$$

where the mean of the elements in list $\{\widehat{\theta}_{n,b,t_j} : j = 1, 2, \ldots, k\}$ is computed by excluding the fraction f of the smallest and the fraction f of largest elements from that list. Furthermore, we assume that there exists $n_f \in \mathbb{N}$ such that for $n > n_f$ we have that $fk(n) = fk > 2$.

Remark 2.2. For complex-valued estimators $\widehat{\theta}_{n,b,t}$, the median and trimmed mean are defined as a complex number whose real and imaginary parts are the median and trimmed mean of real and imaginary parts, respectively. Other definitions of the median are possible for complex values, such as the geometric median (computed by the Weiszfeld's algorithm in R).

Remark 2.3. Under Assumption 2.1, the estimator (3) is of the form

$$\widehat{\theta}_{n,BK} = \frac{1}{w_1 w_2} \sum_{(b,k)\in \mathrm{BK}} \frac{1}{q^k_{n,b}} \sum_{\mathbf{T}_k \in S^k_{n,b}} \widehat{\theta}_{n,b,\mathbf{T}_k}.$$

If in addition, $\widehat{\theta}_{n,b,\mathbf{T}_k} = \widehat{\theta}^{Mean}_{n,b,\mathbf{T}_k}$, then $\widehat{\theta}_{n,BK}$ simplifies to

$$\widehat{\theta}_{n,BK} = \widehat{\theta}^{Mean}_{n,BK} = \frac{1}{w_1} \sum_{b=b_1}^{b_2} \frac{1}{q_{n,b}} \sum_{j=1}^{n-b+1} \widehat{\theta}_{n,b,t},$$

where $w_1 = b_2 - b_1 + 1$, $w_2 = k_2 - k_1 + 1$, $q_{n,b} = n - b + 1$.

3 Mean square consistency

In this section, we first provide a general mean square consistency results for $\widehat{\theta}_{n,BK}$. We then give more specific conditions under which the estimators $\widehat{\theta}^{Mean}_{n,BK}$, $\widehat{\theta}^{Median}_{n,BK}$ and $\widehat{\theta}^{TM(f)}_{n,b,\mathbf{T}_k}$ are mean square consistent.

The following theorem is a generalization of the Fukuchi [1999]'s result.

Theorem 3.1. *Assume that $\{X_t, t \in \mathbb{Z}\}$ is an α-mixing time series such that Assumption 2.2 holds and*

i) there exists $\theta \in \mathbb{R}$ such that

$$E(\widehat{\theta}_{n,BK}) \longrightarrow \theta, \qquad as \quad n \longrightarrow \infty;$$

ii) there exist $\xi > 0$, $\Delta > 0$ and $K < \infty$ such that

$$\sup_{n\in\mathbb{Z},\ (b,k)\in BK,\ \boldsymbol{T}_k \in S^k_{n,b}} \|\widehat{\theta}_{n,b,\boldsymbol{T}_k}\|_{2+\xi} < \Delta \quad and \quad \sum_{k=1}^{\infty} \alpha^{\frac{\xi}{2+\xi}}(k) < K;$$

iii) $b_2^2 k_2 / n \longrightarrow 0$, as $n \longrightarrow \infty$.

Then

$$\widehat{\theta}_{n,BK} \xrightarrow{L_2} \theta, \qquad as \quad n \longrightarrow \infty.$$

Theorem 3.2. *Assume that Assumption 2.2 holds and*

i) there exists $\theta \in \mathbb{R}$ *such that for any sequence of integers* $\tilde{b}$ *such that* $b_1 \leq \tilde{b} \leq b_2$ *and any sequence of integers* t_n *such that* $1 \leq t_n \leq n - \tilde{b} + 1$ *we have convergence*

$$E(\widehat{\theta}_{n,\tilde{b},t_n}) \longrightarrow \theta, \qquad as \quad n \longrightarrow \infty;$$

ii) there exist $\xi > 0$, $\Delta > 0$ *and* $K < \infty$ *such that*

$$\sup_{n\in\mathbb{Z},\ b\in\{b_1,\ldots,b_2\},\ t\in\{1,2,\ldots,n-b+1\}} \|\widehat{\theta}_{n,b,t}\|_{2+\xi} < \Delta \quad and \quad \sum_{k=1}^{\infty} \alpha^{\frac{\xi}{2+\xi}}(k) < K;$$

iii) $b_2^2 k_2/n \longrightarrow 0$, *as* $n \longrightarrow \infty$.

Then

$$\widehat{\theta}^{Mean}_{n,BK} = \sum_{(b,k)\in BK} \sum_{\boldsymbol{T}_k \in S^k_{n,b}} p_{b,k,\boldsymbol{T}_k} \widehat{\theta}^{Mean}_{n,b,\boldsymbol{T}_k} \xrightarrow{L_2} \theta, \qquad as \quad n \longrightarrow \infty.$$

In the following result, we generalize Theorem 3.2 to the case of a complex-valued parameter.

Corollary 3.1. *Assume that assumptions ii)-iii) of Theorem 3.2 are satisfied for the real and imaginary parts of complex-valued estimator* $\widehat{\theta}_{n,b,t}$. *Additionally, assume that there exists a complex-valued parameter* θ, *such that for any sequence of integers* $\tilde{b}$ *such that* $b_1 \leq \tilde{b} \leq b_2$ *and any sequence of integers* t_n *such that* $1 \leq t_n \leq n - \tilde{b} + 1$ *we have the convergence*

$$E|\widehat{\theta}_{n,\tilde{b},t_n} - \theta| \longrightarrow 0, \qquad as \quad n \longrightarrow \infty.$$

Then the GSEs based on the following estimators $\widehat{\theta}_{n,b,\boldsymbol{T}_k}$ *of the mean*

$$\widehat{\theta}^{Mean*}_{n,b,\boldsymbol{T}_k} = \left| \frac{1}{k} \sum_{j=1}^{k} \widehat{\theta}_{n,b,t_j} \right|, \tag{4}$$

$$\widehat{\theta}^{Mean**}_{n,b,\boldsymbol{T}_k} = \frac{1}{k} \sum_{j=1}^{k} |\widehat{\theta}_{n,b,t_j}| \tag{5}$$

are mean square consistent for $|\theta|$.

If $|\widehat{\theta}_{n,b,t}|$ is an estimator of $|\theta|$, where $\widehat{\theta}_{n,b,t}$ is a complex-valued estimator of parameter θ, than we can define the GSE by taking the absolute value after or before calculating median or trimmed mean, i.e. by defining $\widehat{\theta}_{n,b,\mathbf{T}_k}$ in (3) as

$$\widehat{\theta}^{Median*}_{n,b,\mathbf{T}_k} = |\text{med}\{\widehat{\theta}_{n,b,t_j} : j = 1, 2, \ldots, k\}|, \tag{6}$$

$$\widehat{\theta}^{Median**}_{n,b,\mathbf{T}_k} = \text{med}\{|\widehat{\theta}_{n,b,t_j}| : j = 1, 2, \ldots, k\}, \tag{7}$$

$$\widehat{\theta}^{TM(f)*}_{n,b,\mathbf{T}_k} = |\text{TM}^{(f)}\{\widehat{\theta}_{n,b,t_j} : j = 1, 2, \ldots, k\}|, \tag{8}$$

$$\widehat{\theta}^{TM(f)**}_{n,b,\mathbf{T}_k} = \text{TM}^{(f)}\{|\widehat{\theta}_{n,b,t_j}| : j = 1, 2, \ldots, k\}, \tag{9}$$

where the median and trimmed mean above (under the absolute value) is a complex number whose real and imaginary parts being the median and trimmed mean of real and imaginary parts, respectively.
Finally, we also consider the following GSEs based on the geometric median GM for the bivariate parameter $(\Re(\theta), \Im(\theta))$ and for $|\theta|$, of the form

$$\widehat{\theta}^{GM}_{n,b,\mathbf{T}_k} = \text{GM}\Big\{\Big(\Re(\widehat{\theta}_{n,b,t_j}), \Im(\widehat{\theta}_{n,b,t_j})\Big) : j = 1, 2, \ldots, k\Big\}, \tag{10}$$

$$\widehat{\theta}^{GM*}_{n,b,\mathbf{T}_k} = \text{Norm}\Big[\text{GM}\Big\{\Big(\Re(\widehat{\theta}_{n,b,t_j}), \Im(\widehat{\theta}_{n,b,t_j})\Big) : j = 1, 2, \ldots, k\Big\}\Big], \tag{11}$$

respectively. By $\Re(z)$ and $\Im(z)$ we denote the real and imaginary parts of the complex number z, and Norm$[\boldsymbol{v}]$ denotes the Euclidean norm of the vector $\boldsymbol{v} \in \mathbb{R}^2$.

The following result states the consistency of the GSE in the case of the median and trimmed mean for a real-valued or complex parameter θ and its modulus $|\theta|$. In addition, the GSE based on the geometric median is considered.

Theorem 3.3. *Assume that Assumption 2.1 holds and*

i) there exists θ (real or complex-valued) such that uniformly to $b \in \{b_1, b_1+1, \ldots, b_2\}$

and $t \in \{1, 2, \ldots, n-b+1\}$ *we have a sequence* C_n *such that*

$$\sqrt{k_2} E|\widehat{\theta}_{n,b,t} - \theta| \leq C_n \longrightarrow 0,$$

as $n \longrightarrow \infty$*, where* C_n *is independent on* t*;*

Moreover, assume that the conditions ii) and iii) of Theorem 3.2 hold. Then, as $n \longrightarrow \infty$*, for the real-valued* θ*, we have that*

$$\widehat{\theta}_{n,BK}^{Median} = \sum_{(b,k)\in BK} \sum_{\boldsymbol{T}_k \in S_{n,b}^k} p_{b,k,\boldsymbol{T}_k} \widehat{\theta}_{n,b,\boldsymbol{T}_k}^{Median} \xrightarrow{L_2} \theta,$$

$$\widehat{\theta}_{n,BK}^{TM(f)} = \sum_{(b,k)\in BK} \sum_{\boldsymbol{T}_k \in S_{n,b}^k} p_{b,k,\boldsymbol{T}_k} \widehat{\theta}_{n,b,\boldsymbol{T}_k}^{TM(f)} \xrightarrow{L_2} \theta,$$

and for the complex-valued θ*, we have that*

$$\widehat{\theta}_{n,BK}^{TM(f)*} = \sum_{(b,k)\in BK} \sum_{\boldsymbol{T}_k \in S_{n,b}^k} p_{b,k,\boldsymbol{T}_k} \widehat{\theta}_{n,b,\boldsymbol{T}_k}^{TM(f)*} \xrightarrow{L_2} |\theta|,$$

$$\widehat{\theta}_{n,BK}^{Median*} = \sum_{(b,k)\in BK} \sum_{\boldsymbol{T}_k \in S_{n,b}^k} p_{b,k,\boldsymbol{T}_k} \widehat{\theta}_{n,b,\boldsymbol{T}_k}^{Median*} \xrightarrow{L_2} |\theta|,$$

$$\widehat{\theta}_{n,BK}^{GM*} = \sum_{(b,k)\in BK} \sum_{\boldsymbol{T}_k \in S_{n,b}^k} p_{b,k,\boldsymbol{T}_k} \widehat{\theta}_{n,b,\boldsymbol{T}_k}^{GM*} \xrightarrow{L_2} |\theta|.$$

Remark 3.1. Note that the conclusion of the above theorem also holds for the estimators (7) and (9) with real-valued parameter $|\theta|$ and the corresponding subsampling estimator $|\hat{\theta}_{n,b,t}|$.

Remark 3.2. Note that in order to construct our estimators, we need to calibrate the parameters b_1, b_2, k_1, and k_2. It would be really challenging to obtain the best theoretical values (according to some estimated MSE criteria) due to the general dependent framework chosen here. We will therefore give a rule of thumb for calibrating these parameters. In practice, for some theoretical reasons explained in the next section,

we will always choose k_1 and k_2 of the same order smaller than $\sqrt{n}$. The aggregation process will improve the performance of the estimator by reducing the variance, but we need some estimators based on subsampling size b that are the best possible. Thus, this also suggests selecting b_1 and b_2 of the same order by using some "optimal" subsampling size selection procedure. Several proposals have been made in the literature. On the one hand, one can use the procedure proposed by Bickel and Sakov [2008] for m out of n bootstrap, which can be easily adapted to our framework. The idea is to try to find a range of subsampling sizes for which the mean squared error estimator of the subsampling distribution is stable. This range of values could in turn give some reasonable values for b_1 and b_2. On the other hand, another solution would be to use the procedure proposed in Politis and White [2004], which assumes a subsampling size of the order $\hat{C}_M n^{1/3}$ (see the formula (9) on p. 58 of their paper), where $\hat{C}_M$ is an estimated constant based on a truncation level M. In our experience, the estimator of the constant is sensitive to the choice of the truncation level. Therefore, we suggest using several values of M to obtain b_1 and b_2 corresponding to the smallest and largest value obtained this way. The aggregation of several subsampling estimators based on these values will be more robust to the choice of this hyper-parameter. Note that with this selection rule $b_2 - b_1$ cannot be too large (which actually causes some problems in practice, as can be observed in simulation studies).

4 Rate of convergence and concentration results

It is known that averaging over several (overlapping or non-overlapping) subsamples may improve the original estimator based on n observations (see Banerjee et al. [2019], Bertail et al. [2024]). In this section, we will show that this phenomenon also applies to dependent data using the generalized mean subsampling estimators proposed in Fukuchi [1999] and Lenart [2018]. For this purpose, as in Bertail et al. [2024] who studied the i.i.d. case, we will obtain some concentration inequalities. This is possible due to the structure of the U-statistic (based on blocks of observations) of the subsampling

estimator. For simplicity, we only consider here the real-valued case. Recall that the estimators we are considering are of the form

$$\widehat{\theta}_{n,BK} = \sum_{(b,k)\in BK} p_{(b,k)} \sum_{\mathbf{T}_k \in S_{n,b}^k} p_{T_k|(b,k)} \widehat{\theta}_{n,b,\mathbf{T}_k}.$$

and in particular, under Assumption 2.1,

$$\widehat{\theta}_{n,BK} = \frac{1}{w_1 w_2} \sum_{(b,k)\in BK} \frac{1}{q_{n,b}^k} \sum_{\mathbf{T}_k \in S_{n,b}^k} \widehat{\theta}_{n,b,\mathbf{T}_k},$$

with $w_1 = b_2 - b_1 + 1$, $w_2 = k_2 - k_1 + 1$ $q_{n,b} = n - b + 1$. Moreover, in the case of the mean, median, and trimmed mean we have $\widehat{\theta}_{n,b,\mathbf{T}_k} = \widehat{\theta}_{n,b,\mathbf{T}_k}^{Mean} = \frac{1}{k} \sum_{t_j \in \mathbf{T}_k} \widehat{\theta}_{n,b,t_j}$, $\widehat{\theta}_{n,b,\mathbf{T}_k} = \widehat{\theta}_{n,b,\mathbf{T}_k}^{Median} = \underset{t_j \in \mathbf{T}_k}{\text{med}} \{\widehat{\theta}_{n,b,t_j}\}$, and $\widehat{\theta}_{n,b,\mathbf{T}_k} = \widehat{\theta}_{n,b,\mathbf{T}_k}^{TM(f)} = \text{TM}^{(f)}\{\widehat{\theta}_{n,b,t_j}, j = 1, \ldots, k\}$, respectively.

We aim first to obtain Hoeffding or Bernstein inequalities for this kind of mean aggregated estimators. The main tool for that is the Chernoff's method which consists of writing, for any $\lambda > 0$, for any $x \in \mathbb{R}$,

$$P(\widehat{\theta}_{n,BK} - \theta > x) \leq P(\exp(\lambda(\widehat{\theta}_{n,BK} - \theta)) > \exp(\lambda x)) \leq E(\exp(\lambda(\widehat{\theta}_{n,BK} - \theta - x)),$$

finding a good approximation for the Laplace transform, and then optimizing in λ.

We begin with a well-known lemma of Hoeffding [1963], see his section 5, which reduces the problem to a study of concentration inequalities for the part of the expression relating to V- and U-statistics, that is, $\frac{1}{q_{n,b}^k} \sum_{\mathbf{T}_k \in S_{n,b}^k} \widehat{\theta}_{n,b,\mathbf{T}_k}$. The result follows easily from Jensen's inequality.

Lemma 4.1 (Hoeffding,1963)**.** *We have for any $\lambda > 0$, for any $x \in \mathbb{R}$,*

$$P(\ \widehat{\theta}_{n,BK} - E\widehat{\theta}_{n,BK} > x) \leq E(\exp(\lambda(\widehat{\theta}_{n,BK} - E\widehat{\theta}_{n,BK} - x))$$
$$\leq \sum_{(b,k)\in BK} p_{(b,k)} E\left(\exp\left(\lambda\left(\sum_{\boldsymbol{T}_k \in S_{n,b}^k} p_{T_k|(b,k)}(\widehat{\theta}_{n,b,\boldsymbol{T}_k} - E\widehat{\theta}_{n,BK} - x)\right)\right)\right).$$

It follows that under Assumption 2.1, we have

$$P(\ \widehat{\theta}_{n,BK} - E\widehat{\theta}_{n,BK} > x) \leq$$
$$\frac{1}{w_1 w_2} \sum_{(b,k)\in BK} E\left(\exp\left(\lambda\left(\frac{1}{q_{n,b}^k} \sum_{\mathbf{T}_k \in S_{n,b}^k} (\widehat{\theta}_{n,b,\mathbf{T}_k} - E\widehat{\theta}_{n,BK}) - x\right)\right)\right).$$

Note that in this way we reduce the problem to controlling the Laplace transform of a V- or U-statistic of order k. The first result allowing to control the Laplace transform of the means of strong mixing sequences in a quite general framework is due to Merlevede et al. [2009], see their Theorem 2. This result was in turn used by Han [2018] to obtain Bernstein-type inequalities for V- and U-statistics, see Theorem 2.1, or more precisely his Theorem 3.1 in the appendix, in our nonstationary case. Notice that the bound is true indifferently for U- and V-statistics. For us, this means that we can sample the blocks indifferently with or without replacement (in the latter case, the term $q_{n,b}^k = (n-b+1)^k$ should simply be replaced by $q_{n,b}^k = \binom{n-b+1}{k}$). The following result is a straightforward extension of Han [2018] that allows the kernel h to depend on n as in the case of subsampling distribution. The reason for this is that in Han's results/proofs, the constants C_1 and C_2 defined below depend on the kernel size. However, it can be verified that these coefficients increase at most linearly in the degree of the U-statistic k. This dependence will have no effect if $k/\sqrt{n} \longrightarrow 0$. Recall that in such a case we have, by Stirling formula, $\frac{\binom{n}{k}}{n^k} \to 1$. The condition on k is a typical condition that ensures that the U-statistic is in an asymptotically normal regime as the degree of the kernel increases (see Heilig and Nolan [2001] in the i.i.d. case). For i.i.d. rv's we recover the standard form of Hoeffding inequality for U-statistics.

Lemma 4.2 (Han, 2018)**.** *Assume that* Z_i, $i = 1, \ldots, \infty$ *is a sequence of strongly mixing (not necessarily stationary) r.v.'s with mixing coefficients decreasing exponentially, that is such that, for* j *large enough,* $\alpha(j) \leq C\exp(-cj)$ *for some constants,* $C > 0$ *and* $c > 0$*. Define the V-statistic*

$$V_n = \frac{1}{n^k} \sum_{(i_1,\ldots,i_k)=(1,\ldots,1)}^{(n,\ldots,n)} h(Z_{i_1}, \ldots, Z_{i_k}),$$

where h *is a bounded symmetric kernel of size* k *with* $|h(z_1, \ldots, z_k)| \leq M$*, where* $M > 0$ *is an absolute constant. Assume finally that the size of the kernel is such that* $k/\sqrt{n} \longrightarrow 0$ *as* $n \longrightarrow \infty$*. Then there exist positive constants* C_1 *and* C_2 *depending only on* c*, such that, for any* λ *such that* $0 < \lambda < \frac{n}{C_1 kM \log(n) \log\log(4n)}$ *we have*

$$E(\exp(\lambda(V_n - EV_n)) \leq \exp\left(\frac{C_2\lambda^2 M^2 k n^{-1}}{1 - C_1\lambda M k n^{-1}\log(n)\log\log(4n)}\right).$$

By the Chernoff's method, this gives, in turn, for some constant $C_3 > 0$*, for any* $x > 0$ *and* n *sufficiently large*

$$P\left(|V_n - EV_n| > x\right) \leq \exp\left(-\frac{C_3 x^2 n/k}{M^2 + Mx\log(n)\log\log(4n)}\right).$$

Notice that in the i.i.d. case, we know that U-statistics (with a non-degenerate first order-gradient or influence function) concentrate at a rate close to n/k, which is almost the rate we get here. This bound is clearly not optimal. In fact, we would prefer to have a true Bernstein bound, replacing the term M^2 in the denominator with the true variance of the kernel of the U-statistic. Indeed, Bernstein bounds can yield a better convergence rate when the variance is small compared to the second term in the denominator (see Bertail et al. [2024]). In our case, because of the form of the kernel (the statistic of interest based on a block of size b), we expect the variance of the subsampling estimator to be of the order at least $1/\tau_b^2$. Thus, we expect a loss when using this bound, but we have been unable to obtain a sharp bound in a general nonstationary framework (which seems quite unrealistic to obtain). However, even with this crude bound, we will

show that, regardless of the convergence rate of the statistic of interest, averaging using the divide-and-conquer methodology for the generalized subsampling estimator yields a better rate compared to the original one, provided that the bias of the original statistic is not too large (see discussion below).

In the case of averaging the subsampling estimators $\widehat{\theta}_{n,BK}^{Mean}$, we will apply Lemma 4.2 with $h(B_{t_1}, B_{t_2}, \ldots, B_{t_k}) = \frac{1}{k} \sum_{t_j \in \mathbf{T}_k} \widehat{\theta}_{n,b,t_j}(B_{t_j})$. Similarly, in the case of $\widehat{\theta}_{n,BK}^{Median}$ we can apply the result of Lemma 4.2 with $h(B_{t_1}, B_{t_2}, \ldots, B_{t_k}) = \underset{t_j \in \mathbf{T}_k}{\text{med}}\{\widehat{\theta}_{n,b,t_j}\}$ and $h(B_{t_1}, B_{t_2}, \ldots, B_{t_k}) = \text{TM}^{(f)}\{\widehat{\theta}_{n,b,t_j} : j = 1, 2, \ldots, k\}$ for the trimmed mean case. In our case, the $Z_i's$ are the statistics evaluated on each block of random variables that are drawn with (or without) replacement. We need to ensure whether $\frac{1}{k} \sum_{t_j \in \mathbf{T}_k} \widehat{\theta}_{n,b,t_j}$ or $\text{med}\{\widehat{\theta}_{n,b,t_j}\}$ or $\text{TM}^{(f)}\{\widehat{\theta}_{n,b,t_j}\}$ are bounded to apply the result of Han [2018]. In particular, if we have that $\sup_{b_1 \leq b \leq b_2} \sup_{1 \leq t \leq n-b+1} |\widehat{\theta}_{n,b,t}|$ is bounded, then the kernel of the corresponding U-statistic will automatically be bounded. Moreover, we also need to check that the blocks are α-mixing with exponential mixing coefficient for a value large enough. This can be checked by assuming a sufficiently small kernel subsampling size which also satisfies $b/\sqrt{n} \longrightarrow 0$. This yields the following result for the mean, median and trimmed mean of the subsampling estimators.

Theorem 4.1. *Assume that Assumption 2.2 holds and that* $\sup_{b_1 \leq b \leq b_2} \sup_{1 \leq t \leq n-b+1} |\widehat{\theta}_{n,b,t}|$ *is bounded a.s. by some constant* $M > 0$. *Moreover, assume that the sequence* $\{X_i\}$ *is strongly mixing with exponentially decreasing mixing coefficients* $\alpha(j) \leq C \exp(-cj)$ *for some positive constants* C *and* c. *Assume, in addition, that* k_1, k_2 *are of the same order and are chosen such that* $k_2/\sqrt{n} \longrightarrow 0$, *as well as* b_1, b_2 *are of the same order and are chosen such that* $b_2/\sqrt{n} \longrightarrow 0$ *as* $b_2 \longrightarrow \infty$. *Then there exists a constant* $C_3 > 0$ *depending on* c *and* C, *such that, for* n *large enough,*

$$P\left(\widehat{\theta}_{n,BK} - E\widehat{\theta}_{n,BK} > x\right) \leq \exp\left(-\frac{C_3 x^2 n/k_2}{M^2 + Mx \log(n) \log\log(4n)}\right).$$

It follows that the rate of convergence of the generalized subsampling estimator is of

order

$$\widehat{\theta}_{n,BK} - E\widehat{\theta}_{n,BK} = O_P\left(\frac{k_2^{1/2}\log(n)^{1/2}\log\log(n)^{1/2}}{n^{1/2}}\right).$$

The assumption of estimator boundedness will hold for robust estimators, but can be removed at the expense of additional assumptions on the a.s. rate of convergence of the statistics.

Notice that we recenter our generalized subsampling estimator with its expectation rather than the true value of the parameter (in the unbiased case, for instance, for the mean, this will not be a problem). However, in some cases we will need to control the bias and effectively choose the "optimal subsampling size". In the unbiased case, the convergence rate can be made close to $\sqrt{n}$ by choosing a very small subsampling size. However, in the biased case, a second-order condition is needed to control the rate precisely. For instance, if the bias is of the order $1/\sqrt{k}$, as assumed in Theorem 3.3, then a straightforward calculus shows that k should be chosen slightly smaller than $n^{1/2}$. However, this also imposes that b should be larger than k. In this case, the rate of the bagged estimator is close (up to a certain $\log(n)$ term) to $n^{1/4}$ independently of the original rate of the original statistic. Therefore, it is important to debias the original estimator if its convergence rate is fast. This can be done using jackknife types of methods. The same kind of ideas were used in Bertail [1997] to obtain a better rate of convergence of extrapolated subsampling distributions.

Unlike the i.i.d. case (see Bertail et al. [2024]), we do not assume here that there exists a concentration inequality for the original estimator $\widehat{\theta}_{n,b,t_j}(B_{t_j})$ for any t_j of the form: for any t_j, for some $\eta > 0$, there exist some universal constants $C_i > 0,\ i = 1, 2$ such that

$$\Pr\left(\tau_b|\widehat{\theta}_{n,b,t_j}(B_{t_j}) - \theta| > \eta\right) \leq C_1\exp(-\frac{n}{b}C_2\eta^2). \tag{12}$$

Since our Bernstein bound is suboptimal, we do not need this assumption at the expense of a suboptimal rate for our subsampling estimator. This condition may be difficult to verify in practice, especially in the nonstationary case. It may be easier

to verify it in the stationary framework. In that case, we expect to obtain a better convergence rate by first proving the true Bernstein bound (with a variance instead of M^2). This will be the subject of our future research.

5 Extensions to estimators with nuisance parameters

In this section, we will consider a generalized version of the estimator $\widehat{\theta}_{n,b,\mathbf{T}_k}$. Sometimes an unknown parameter θ depends on another parameter Γ (say a nuisance parameter) that must be estimated first. For example, to estimate the covariance, the overall mean must be estimated (unless it is known). Therefore, below we assume that the unknown parameter θ depends on Γ, and denote it by $\theta(\Gamma)$. When Γ is known, then the GSE is of the form $\widehat{\theta}_{n,b,\mathbf{T}_k}(\Gamma)$. Otherwise, Γ needs to be estimated. By $\widehat{\Gamma}_n$ we denote its estimator based on the original sample $(X_1,\ldots,X_n)$. In this case, we define the GSE by

$$\widehat{\theta}_{n,BK}(\widehat{\Gamma}_n) = E^*\left(\widehat{\theta}_{n,b^*,\mathbf{T}^*_{k^*}}(\widehat{\Gamma}_n)\right) = \sum_{(b,k)\in BK} p_{(b,k)} \sum_{\mathbf{T}_k\in S^k_{n,b}} p_{T_k|(b,k)}\widehat{\theta}_{n,b,\mathbf{T}_k}(\widehat{\Gamma}_n). \qquad (13)$$

Below we provide the results corresponding to Theorems 3.1-3.3 stating consistency of our estimators.

Theorem 5.1. *Assume that $\{X_t, t \in \mathbb{Z}\}$ is an α-mixing time series such that Assumption 2.2 holds and*

i) there exists $\theta \in \mathbb{R}$ such that

$$E(\widehat{\theta}_{n,BK}(\widehat{\Gamma}_n)) \longrightarrow \theta$$

i') OR there exists $\theta \in \mathbb{R}$ such that for any sequences of integers $\tilde{b} = \tilde{b}(n)$ and $\tilde{k} = \tilde{k}(n)$ such that $0 < b_1 \le \tilde{b} \le b_2 < n$, $0 < k_1 \le \tilde{k} \le k_2 < n$, $b_1, b_2 \longrightarrow \infty$ as $n \longrightarrow \infty$, but $b_2/n \longrightarrow 0$ as $n \longrightarrow \infty$ and any sequence of vectors $\boldsymbol{T}_{n,\tilde{k}} = (t_{n,1}, t_{n,2}, \ldots, t_{n,\tilde{k}})$ with integers $t_{n,i}$ such that $1 \le t_{n,i} \le n - \tilde{b} + 1$, $i = 1, 2, \ldots, \tilde{k}$

we have that

$$E\left[\widehat{\theta}_{n,\tilde{b},\boldsymbol{T}_{n,\tilde{k}}}(\widehat{\Gamma}_n))\right] \longrightarrow \theta;$$

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

$$\sup_{\substack{b_1\le b\le b_2\\ k_1\le k\le k_2\\ \boldsymbol{T}_k\in S^k_{n,b},\ n\in\mathbb{Z}}} \left\|\widehat{\theta}_{n,b,\boldsymbol{T}_k}(E(\widehat{\Gamma}_n)))\right\|_{2+\xi} < \Delta \quad \textit{and} \quad \sum_{j=1}^{\infty} \alpha^{\frac{\xi}{2+\xi}}(j) < \infty,$$

iii) $b_2^2 k_2/n \longrightarrow 0$, *as* $n \longrightarrow \infty$;

iv) for any sequences of integers $\tilde{b}, \tilde{b}'$, $\tilde{k}, \tilde{k}'$ *such that* $0 < b_1 \le \tilde{b} \le b_2 < n$, $0 < b_1 \le \tilde{b}' \le b_2 < n$, $0 < k_1 \le \tilde{k} \le k_2 < n$, $0 < k_1 \le \tilde{k}' \le k_2 < n$, $b_1, b_2 \longrightarrow \infty$ *as* $n \longrightarrow \infty$, *but* $b_2/n \longrightarrow 0$ *as* $n \longrightarrow \infty$ *and any sequences of vectors* $\boldsymbol{T}_{n,\tilde{k}} = (t_{n,1}, t_{n,2}, \ldots, t_{n,\tilde{k}})$ *and* $\boldsymbol{T}'_{n,k'} = (t'_{n,1}, t'_{n,2}, \ldots, t'_{n,\tilde{k}'})$ *with integers* $t_{n,i}$ *and* $t'_{n,i}$ *such that* $1 \le t_{n,i} \le n - \tilde{b} + 1$ *for* $i = 1, 2, \ldots, k$ *and* $1 \le t'_{n,i} \le n - \tilde{b}' + 1$ *for* $i = 1, 2, \ldots, k'$ *we have that*

$$cov\left(\widehat{\theta}_{n,b,\boldsymbol{T}_k}(\widehat{\Gamma}_n) - \widehat{\theta}_{n,b,\boldsymbol{T}_k}(E(\widehat{\Gamma}_n)), \widehat{\theta}_{n,b',\boldsymbol{T}'_{k'}}(E(\widehat{\Gamma}_n))\right) \longrightarrow 0,$$
$$cov\left(\widehat{\theta}_{n,b,\boldsymbol{T}_k}(\widehat{\Gamma}_n) - \widehat{\theta}_{n,b,\boldsymbol{T}_k}(E(\widehat{\Gamma}_n)), \widehat{\theta}_{n,b',\boldsymbol{T}'_{k'}}(\widehat{\Gamma}_n) - \widehat{\theta}_{n,b',\boldsymbol{T}'_{k'}}(E(\widehat{\Gamma}_n))\right) \longrightarrow 0.$$

Then the estimator $\widehat{\theta}_{n,BK}(\widehat{\Gamma}_n)$ *is mean square consistent, i.e.,*

$$\widehat{\theta}_{n,BK}(\widehat{\Gamma}_n) \xrightarrow{L_2} \theta, \qquad \textit{as } n \longrightarrow \infty.$$

Theorem 5.2. *Assume that Assumption 2.2 holds and*

i) there exists $\theta \in \mathbb{R}$ *such that for any sequence of integers* $\tilde{b}$ *such that* $b_1 \le \tilde{b} \le b_2$ *and any sequence of integers* t_n *such that* $1 \le t_n \le n - \tilde{b} + 1$ *we have*

$$E(\widehat{\theta}_{n,\tilde{b},t_n}(\widehat{\Gamma}_n)) \longrightarrow \theta, \quad \textit{as } n \longrightarrow \infty;$$

ii) there exists $\xi > 0$, $\Delta > 0$ *and* $K < \infty$ *such that*

$$\sup_{\substack{n\in\mathbb{Z}\\ b\in\{b_1,\ldots,b_2\}\\ t\in\{1,2,\ldots,n-b+1\}}} \|\widehat{\theta}_{n,b,t}(E(\widehat{\Gamma}_n))\|_{2+\xi} < \Delta \quad \textit{and} \quad \sum_{k=1}^{\infty} \alpha^{\frac{\xi}{2+\xi}}(k) < K;$$

iii) $b_2^2 k_2/n \longrightarrow 0$, *as* $n \longrightarrow \infty$;

iv) for any sequences of integers $\tilde{b}, \tilde{b}'$, *such that* $0 < b_1 \leq \tilde{b} \leq b_2 < n$, $0 < b_1 \leq \tilde{b}' \leq b_2 < n$, $b_1, b_2 \longrightarrow \infty$ *as* $n \longrightarrow \infty$, *but* $b_2/n \longrightarrow 0$ *as* $n \longrightarrow \infty$ *and any sequences of integers* t_n *and* t'_n *such that* $1 \leq t_n \leq n - \tilde{b} + 1$ *and* $1 \leq t'_n \leq n - \tilde{b}' + 1$ *we have that*

$$cov\Big(\widehat{\theta}_{n,b,t_n}(\widehat{\Gamma}_n) - \widehat{\theta}_{n,b,t_n}(E(\widehat{\Gamma}_n)), \widehat{\theta}_{n,b',t'_n}(E(\widehat{\Gamma}_n))\Big) \longrightarrow 0,$$
$$cov\Big(\widehat{\theta}_{n,b,t_n}(\widehat{\Gamma}_n) - \widehat{\theta}_{n,b,t_n}(E(\widehat{\Gamma}_n)), \widehat{\theta}_{n,b,t'_n}(\widehat{\Gamma}_n) - \widehat{\theta}_{n,b,t'_n}(E(\widehat{\Gamma}_n))\Big) \longrightarrow 0.$$

Then

$$\widehat{\theta}^{Mean}_{n,BK}(\widehat{\Gamma}_n) = \sum_{(b,k)\in BK} \sum_{\boldsymbol{T}_k \in S^k_{n,b}} p_{b,k,\boldsymbol{T}_k} \widehat{\theta}^{Mean}_{n,b,\boldsymbol{T}_k}(\widehat{\Gamma}_n) \xrightarrow{L_2} \theta, \quad as \;\; n \longrightarrow \infty.$$

Theorem 5.3. *Let Assumption 2.2 and conditions ii)−iv) of Theorem 5.2 hold. Assume that*

i) there exists $\theta \in \mathbb{R}$ *such that uniformly at* $b \in \{b_1, b_1 + 1, \ldots, b_2\}$ *and* $t \in \{1, 2, \ldots, n - b + 1\}$ *we have a sequence* C_n *such that*

$$\sqrt{k_2} E|\widehat{\theta}_{n,b,t}(\widehat{\Gamma}_n) - \theta| \leq C_n \longrightarrow 0,$$

as $n \longrightarrow \infty$, *where the sequence* C_n *doesn't depend on* t;

Then, as $n \longrightarrow \infty$ *we have that*

$$\widehat{\theta}_{n,BK}^{Median} = \sum_{(b,k)\in BK} \sum_{\boldsymbol{T}_k \in S_{n,b}^k} p_{b,k,\boldsymbol{T}_k} \widehat{\theta}_{n,b,\boldsymbol{T}_k}^{Median}(\widehat{\Gamma}_n) \xrightarrow{L_2} \theta$$

and

$$\widehat{\theta}_{n,BK}^{TM(f)} = \sum_{(b,k)\in BK} \sum_{\boldsymbol{T}_k \in S_{n,b}^k} p_{b,k,\boldsymbol{T}_k} \widehat{\theta}_{n,b,\boldsymbol{T}_k}^{TM(f)}(\widehat{\Gamma}_n) \xrightarrow{L_2} \theta.$$

6 Application

Below we apply the results of Section 5 to almost periodically correlated (APC) time series, which are an important class of nonstationary time series. They are a generalization of periodically correlated (PC) processes, which are characterized by periodic mean and covariance functions. In the APC case, the mean and covariance functions are almost periodic functions of the time argument. To be precise, a function $f(t) : \mathbb{Z} \longrightarrow \mathbb{R}$ is called almost periodic if for any $\epsilon > 0$ there exists an integer $L_\epsilon > 0$ such that among any L_ϵ consecutive integers there is an integer p_ϵ such that $\sup_{t\in\mathbb{Z}} |f(t+p_\epsilon)-f(t)| < \epsilon$ (see Corduneanu [1989]). The PC/APC processes are used to model data with cyclic features in many fields, such as telecommunications, mechanics, vibroacoustic, economics, hydrology, and many others (see e.g., Antoni [2009], Gardner et al. [2006], Hurd and Miamee [2007], Napolitano [2012, 2016]).

Subsampling consistency has been proven for various characteristics of APC processes, both in the discrete and continuous case. The Fourier coefficients of the mean function of the APC time series were considered in Lenart [2013], while the spectral characteristics in Lenart [2011] and Lenart [2016]. Finally, the consistency for the Fourier coefficients of the autocovariance function of the continuous APC process was obtained by Dehay et al. [2014]. Below we will discuss the application of our results to the Fourier coefficients of the autocovariance function of PC time series with a deterministic trend function. This example can be easily extended to the APC case

with a finite number of significant frequencies of the mean and autocovariance functions.

Let $\{X_t\}$ be a time series with a mean function of the form

$$E(X_t) = \mu(t) = d \cdot f(t) + \sum_{k=0}^{T-1} m(\psi_k) e^{i\psi_k t}, \tag{14}$$

where $d \in \mathbb{R}$ is some constant, $f : \mathbb{R} \longrightarrow \mathbb{R}$ is some known function, $\psi_k = \frac{2k\pi}{T}$, $k = 0, 1, \ldots, T-1$ and $T > 0$ is a period length. For the function f, we assume the following.

Assumption 6.1. *Assume that the real-valued function $f : \mathbb{R} \longrightarrow \mathbb{R}$*

i) has at most a finite number of local extrema;

ii) is such that there exist constants $s > 0$ and $C_1 \neq 0$ such that $f(x)/x^s \longrightarrow C_1$, as $x \longrightarrow \infty$;

iii) there exists constant $C_2 > 0$ such that

$$\lim_{n \longrightarrow \infty} \left[\frac{1}{n^{2s+1}} \sum_{j=1}^{n} \left(f(j) - \bar{f}_n \right)^2 \right] = C_2,$$

where $\bar{f}_n = \frac{1}{n} \sum_{j=1}^{n} f(j)$.

A simple example of a function satisfying the above conditions is a polynomial function of degree $r \geq 1$. In this case, it is sufficient to note that $s = r$.

Let $B(t, \tau)$ be the autocovariance function of $\{X_t\}$. We assume that $B(t, \tau)$ is periodic in t, with the same period T that occurred in the equation (14), i.e.,

$$B(t, \tau) = \operatorname{cov}(X_t, X_{t+\tau}) = \sum_{k=0}^{T-1} a(\lambda_k, \tau) e^{i\lambda_k t}, \tag{15}$$

$\tau \in \mathbb{Z}$, $\lambda_k = 2k\pi/T$ for $k = 0, 1, 2, \ldots, T-1$. To estimate $B(t, \tau)$, we must first estimate the mean function $\mu(t)$. In this case, Γ is the vector of parameters associated

with $\mu(t)$, i.e., $\Gamma = (d, m(\psi_0), m(\psi_1), \ldots, m(\psi_{T-1}))$. Its estimator based on sample $(X_1, X_2, \ldots, X_n)$ is of the form $\widehat{\Gamma}_n = (\widehat{d}_n, \widehat{m}_n(\psi_0), \widehat{m}_n(\psi_1), \ldots, \widehat{m}_n(\psi_{T-1}))$. To estimate d we use the following natural estimator based on the idea of ordinary least squares method

$$\widehat{d}_n = \frac{\sum_{j=1}^{n} \left(f(j) - \bar{f}_n\right) X_j}{\sum_{j=1}^{n} \left(f(j) - \bar{f}_n\right)^2},$$

where $\bar{f}_n = \frac{1}{n} \sum_{j=1}^{n} f(j)$. Then the estimators of $m(\psi_k)$, $k = 0, \ldots, T-1$ are of the form

$$\widehat{m}_n(\psi_k) = \frac{1}{n} \sum_{j=1}^{n} (X_j - \widehat{d}_n f(j)) e^{-ij\psi_k}.$$

Finally,

$$\widehat{\mu}_n(t) = \widehat{d}_n f(t) + \sum_{k=0}^{T-1} \widehat{m}_n(\psi_k) e^{i\psi_k t}$$

is the estimator of the mean function $\mu(t)$, $t \in \{1, 2, \ldots, n\}$.

In the following, we focus on the estimation of $\mathrm{var}(X_t) = B(t, 0)$. The estimator of the Fourier coefficient $a(\lambda_k, 0)$ for any $k = 0, \ldots, T-1$ is of the form

$$\widehat{a}_n(\lambda_k, 0) = \frac{1}{n} \sum_{j=1}^{n} (X_j - \widehat{\mu}_n(j))^2 e^{-ij\lambda_k},$$

and its subsampling counterpart based on block B_t is defined as follows

$$\widehat{\theta}_{n,b,t}(\widehat{\Gamma}_n) = \widehat{a}_{n,b,t}(\lambda_k, \tau) = \frac{1}{b} \sum_{j=t}^{t+b-1} (X_j - \widehat{\mu}_n(j))^2 e^{-ij\lambda_k}. \tag{16}$$

Theorem below states consistency of the GSE based on (16).

Theorem 6.1. *Assume that Assumptions 2.2 and 6.1 hold. Let $\{X_t\}$ be an α-mixing time series with the mean function given by (14) and the autocovariance function by (15). Assume that there exist real numbers $\xi^* > 0$, $\Delta^* > 0$ and $K^* < \infty$ such that*

i) $\sup_{n \in \mathbb{Z}} \|X_n - \mu(n)\|_{4+2\xi^*} < \Delta^*$;

ii) $\sum_{k=1}^{\infty} k^{h^*-2}\alpha^{\frac{\xi^*}{h^*+\xi^*}}(k) < K^*$;

iii) $b_2^2 k_2/n \longrightarrow 0$, *as* $n \longrightarrow \infty$,

where h^* *is the smallest even integer such that* $h^* \geq 4+\xi^*$. *Then the estimators* $\widehat{\theta}^{Mean}_{n,b,\boldsymbol{T}_k}$, $\widehat{\theta}^{Median}_{n,b,\boldsymbol{T}_k}$ *and* $\widehat{\theta}^{TM(f)}_{n,b,\boldsymbol{T}_k}$ *based on (16) are mean square consistent.*

6.1 Some simulation results

Below is a brief illustration of our results using several simulated APC and PC time series as examples. We considered two models:

M1: a pure sinusoid convoluted with an ARMA model

$$X_t = \mu(t) + Z_t, \text{ with } \mu(t) = a\sin(\lambda_0 t) \text{ and } (1-\phi B)Z_t = (1-\theta\varepsilon_t),$$

where $\varepsilon_t,\ t = 1, \ldots n$ are i.i.d. Gaussian r.v.'s with variance 1. In our simulations, we set $\lambda_0 = 0.25$, $\phi = 0.8$, $a = 1$ and $\theta = 0.5$. Moreover, we set a sample size $n = 2000$. An example realization of such series is shown in Figure 1.

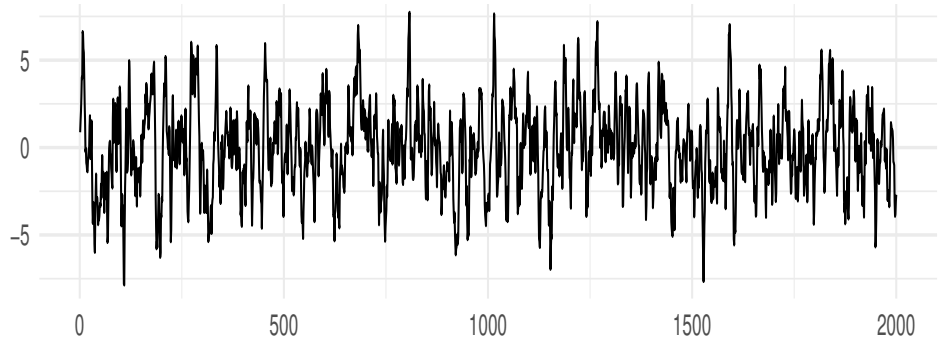

Figure 1: Model **M1**: pure sinusoidal component with ARMA noise.

M2: a PARMA(p,q) model

$$X_t = \sum_{j=1}^{p} \phi_j(t) X_{t-j} + \sum_{k=0}^{q} \theta_k(t)\varepsilon_{t-k},$$

where $\varepsilon_t,\ t = 1, \dots n$ are i.i.d. Gaussian r.v.'s with variance 1. The coefficients $\phi_j(t),\ \theta_k(t)$ are periodic functions with period d, i.e., $\phi_j(t) = \phi_j(t+d),\ \theta_k(t) = \theta_k(t+d)$. We used the package perARMA (see Dudek et al. [2016] and references therein), in which the parsimonious parameterization suggested first in Jones and Brelsford [1967] is implemented. It involves representing periodically varying parameters by Fourier series

$$\phi_j(t) = a_{j,1} + \sum_{l=1}^{\lfloor d/2 \rfloor} (a_{j,2l}\cos(2\pi lt/d) + a_{j,2l+1}\sin(2\pi lt/d)),\ j = 1, \dots, p,$$
$$\theta_k(t) = b_{k,1} + \sum_{l=1}^{\lfloor d/2 \rfloor} (b_{k,2l}\cos(2\pi lt/d) + b_{k,2l+1}\sin(2\pi lt/d)),\ k = 1, \dots, q.$$

with only a few non-zero components.

We generated two PARMA models with a period length $d = 12$.

- **M2a**: PAR(2) model with $a_{1,1} = 0.3,\ a_{2,2} = 0.5$.
- **M2b**: PARMA(2,1) with $a_{1,1} = 0.8,\ a_{1,2} = -0.9,\ a_{2,1} = 0.3,\ b_{1,1} = -0.7,\ b_{2,1} = -0.2$;

All coefficients not specified above are set to 0. We considered a sample size $n = 2004$, corresponding to 167 periods of length $d = 12$. Examples of the realization of the considered time series are shown in Figures 2 and 3.

Additionally, the corresponding sample autocorrelation functions are presented in Figures 4 and 5. Note that cyclostationarity is more difficult to detect in the case of the **M2a** model.

Our goal was to use our subsampling approach to estimate the magnitude of the Fourier coefficient of the mean function at a fixed frequency $\lambda_0 \in [0, \pi]$, i.e., $|m(\lambda_0)|$. This means that our parameter of interests θ is a complex-valued and we estimate the value of $|\theta|$. For **M1** we have $|m(\lambda_0)| = \frac{1}{2}$, while for **M2a** and **M2b** $|m(\lambda_0)| = 0$. Thus, our statistic of interest was the rescaled periodogram, i.e., $|\widehat{\theta}_n| = |1/n \sum_{j=1}^{n} X_j e^{-ij\lambda_0}|$, with a subsampling counterpart $|\widehat{\theta}_{n,b,t}| = |1/b \sum_{j=t}^{t+b-1} X_j e^{-ij\lambda_0}|$.

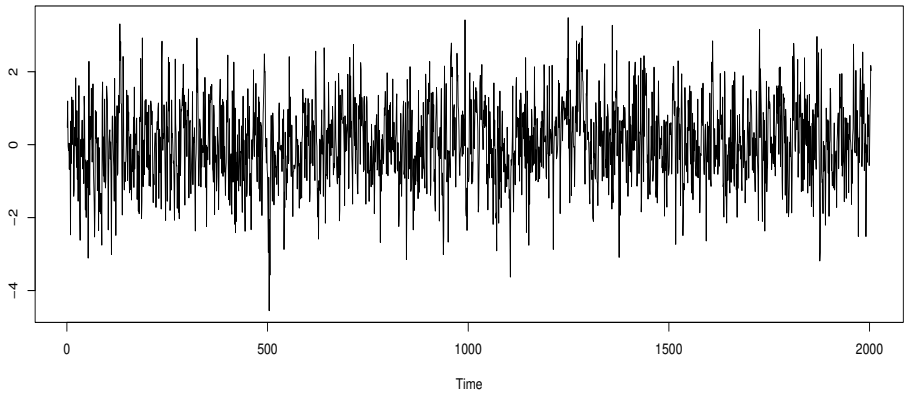

Figure 2: Model **M2a** realization with $n = 2004$.

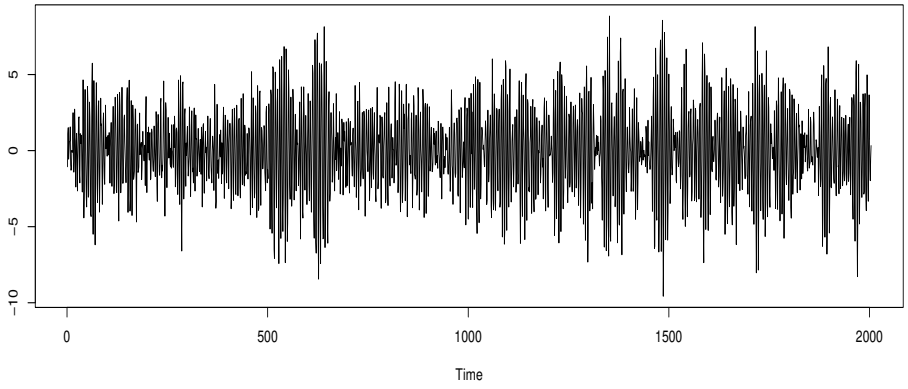

Figure 3: Model **M2b** realization with $n = 2004$.

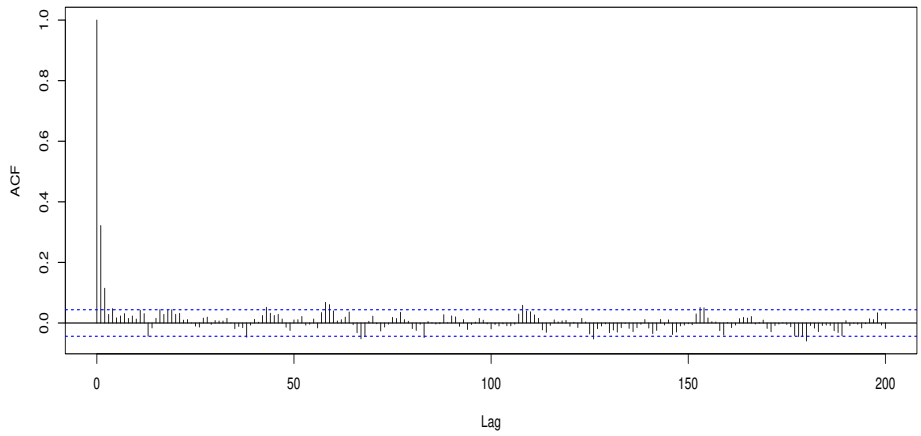

Figure 4: Sample autocorrelation function for realization of **M2a** model with $n = 2004$.

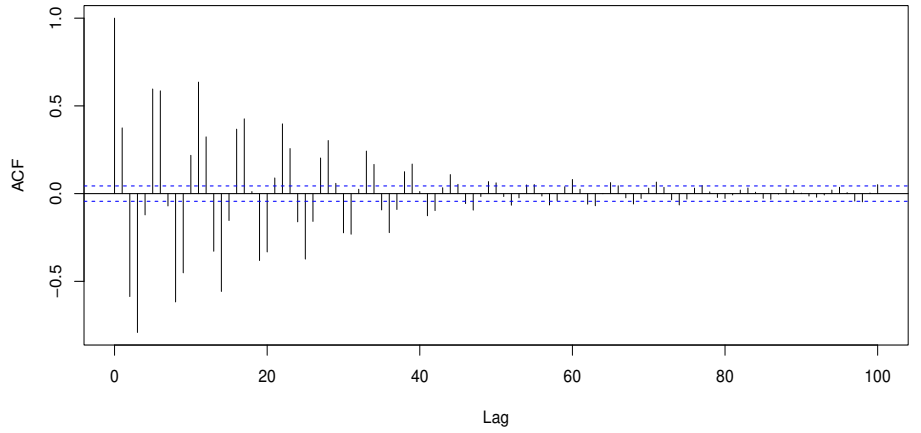

Figure 5: Sample autocorrelation function for realization of **M2b** model with $n = 2004$.

The bagged subsampling procedure was repeated 500 times, and we plotted the mean of our estimators over these repetitions. We assumed a uniform distribution for (b^*, k^*) and $\mathbf{T}^*_{k^*}|(b^*, k^*) = (b, k)$ (see Assumption 2.1). For each simulation, we considered 3 scenarios taking different values of the parameters b_1, b_2 and k_1, k_2. These are:

GESV1: (moderate b and small k): $b_1 = \lfloor\sqrt{n}\rfloor$, $b_2 = b_1 + 10$, $k_1 = 50$, $k_2 = 53$;

GESV2: (large b and moderate k): $b_1 = \lfloor n^{2/3}\rfloor$, $b_2 = b_1 + 10$, $k_1 = 100$, $k_2 = 110$;

GESV3: (large b and large k): $b_1 = \lfloor n^{2/3}\rfloor$, $b_2 = b_1 + 10$, $k_1 = 500$, $k_2 = 530$.

Figures 6-8 show results of the performed simulations for the model **M1**. Specifically, we plotted the periodogram calculated on the whole sample (black) along with the bagged estimators: mean of means (Figure 6), see formula (4), mean of medians (Figure 7) and mean of trimmed means (Figure 8), see formula (8), respectively, based on the scenarios **GESV1** (green), **GESV2** (blue) and **GESV3** (red). For the periodogram, the median was computed in three different ways, using the median of rescaled periodogram directly (see (7)), the geometric median of the complex Fast Fourier Transform (which can be viewed as the depth distance, i.e., the value that is at the smallest Euclidean distance from all points), then applying the modulus (see (11)), and the value obtained by taking the medians of the real and imaginary parts, respectively before computing (see (6)). All results were similar and very similar to the trimmed mean, except in the first case, that is for the median of the rescaled subsampled periodogram. Therefore, we present only this problematic case to highlight the possible bias in this case. Notice that for the geometric median we cannot directly obtain a Bernstein bound, but it is easy to get a Bernstein inequality for the last case, by splitting into the real and imaginary parts at the cost of some loss in the constants. Finally, the trimmed means are computed by removing 1% of the largest and 1% smallest observations of the rescaled periodogram at each frequency.

Notice that for the mean, except for the scenario **GESV1**, when k and b are small, which causes an upper bias, the rescaled periodogram based on the whole sample and its bagged estimators are quite similar. In addition, in all cases, the frequency λ_0 is well detected. Choosing the median of the rescaled periodogram seems to be the worst

choice and seems to produce a huge bias. On the contrary, using the median of the real and imaginary part respectively gives very similar results to the mean case when there are no outliers.

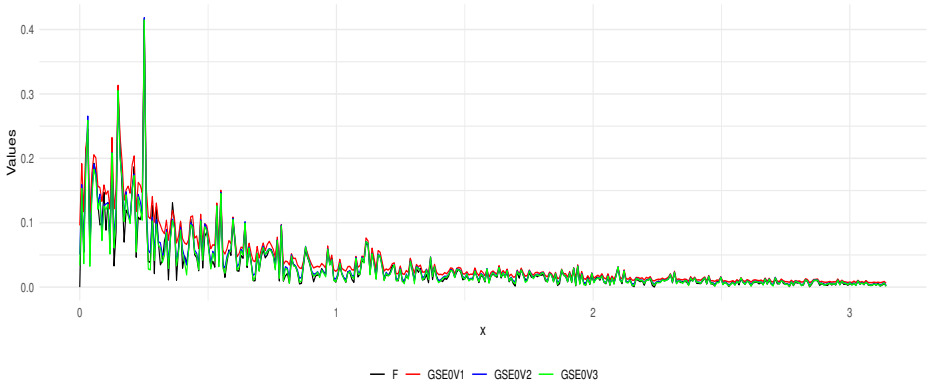

Figure 6: Model **M1**, mean case: rescaled periodogram with $n = 2000$.

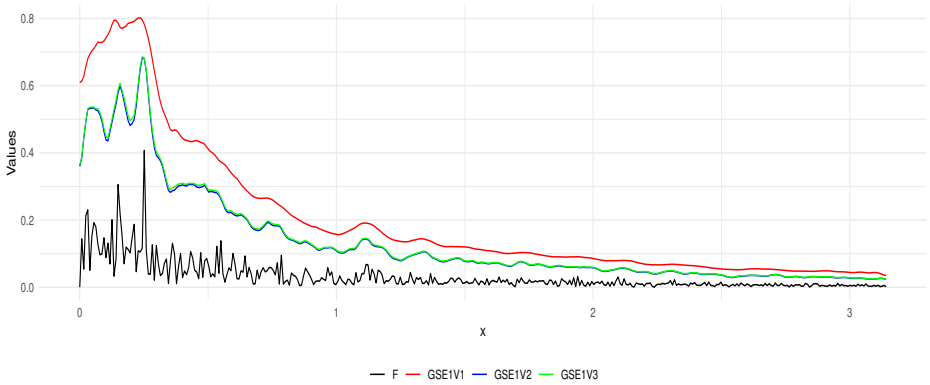

Figure 7: Model **M1**, geometric median case: rescaled periodogram with $n = 2004$.

The results for the PARMA model **M2b** are presented in Figures 9-11. We omit the results for the model **M2a** which gives similar but less clear results than **M2b**. Note that for **M2b** results for the mean of means (Figure 9) and mean of trimmed means (Figure 11) or the mean of medians are very similar and allow us to identify the cyclic nature of the data. Again, the mean of medians applied to rescaled periodgram (Figure 10) seems to be strongly biased: this is likely due to the strong nonlinear effect that

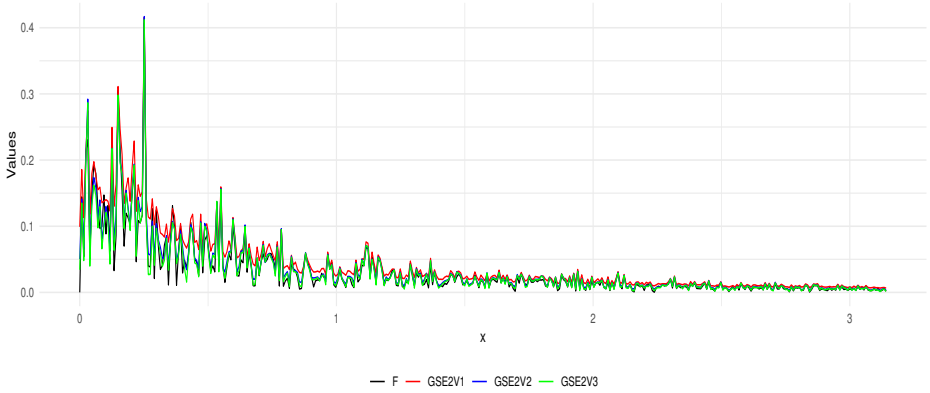

Figure 8: Model **M1**, trimmed mean case: rescaled periodogram with $n = 2004$.

occurs when taking the median of modulus of complex values, even if the method is consistent in all cases.

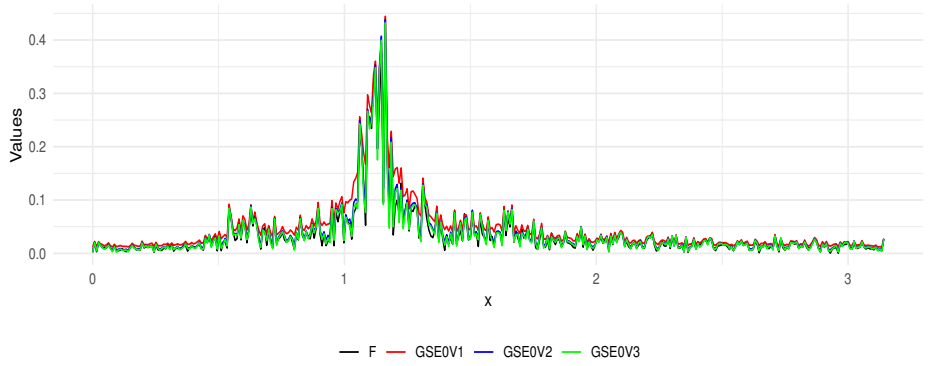

Figure 9: PARMA **M2b** model, mean case: rescaled periodogram with $n = 2004$.

To evaluate the true rate of convergence of the bagged subsampling estimators for the different scenarios, we generated realizations from models **M1**, **M2a** and **M2b** for $n \in \{2\,000, 2\,500, 3\,000, \ldots, 20\,500\}$. For each of the 40 values of n, we repeated the procedure 500 times, which resulted in 40×500 different values of the estimators taken at the frequency λ_0. Based on these repetitions, we evaluated the convergence rate of the bagged subsampling estimator of the rescaled periodogram (taken at frequency λ_0 by looking at the behavior of the estimated variance $var_{b,k,n}$ (over 500 Monte-Carlo

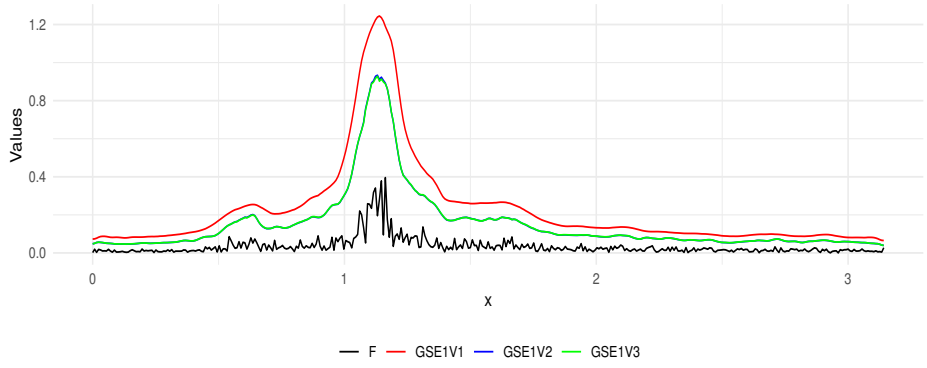

Figure 10: PARMA **M2b** model, geometric median case: rescaled periodogram with $n = 2004$.

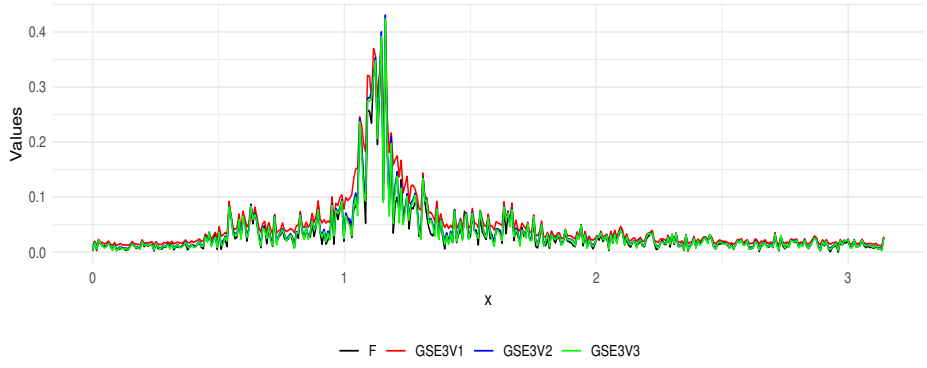

Figure 11: PARMA **M2b** model, trimmed mean case: rescaled periodogram with $n = 2004$.

repetitions of the procedure) as n varies.

Since our results show that the rate of convergence is of the order $(\frac{n}{k})^{\beta}$ with β close to $1/2$ we should have

$$\log(var_{b,k,n}) = -2\beta log(n)(1+\nu_n), \quad \text{with } \nu_n = o(1),$$

and the coefficient of the slope should be close to -1. In Table 1, we present the least squares estimates of the slope along with the standard errors based on 40 different values of n obtained for each model considered **M1**, **M2a**, **M2b**, and for each parameter: mean, median, geometric median (geoMedian) and trimmed mean (TM) under 3 scenarios: 1 ($k_1 = 50$, $b_1 = \lfloor n^{1/2} \rfloor$), 2 ($k_1 = 100$, $b_1 = \lfloor n^{2/3} \rfloor$), 3 ($k_1 = 500$, $\lfloor b_1 = n^{2/3} \rfloor$). The value of R^2 is always very high (close to 0.99) and is omitted from the table. Note that in scenario 3, when b is close to $n^{2/3}$ and k is reasonable, the bagged estimator for the mean provides an estimator with an improved rate of convergence, close to 1, (and even better for the model **M2b**) and slightly slower for the median and trimmed mean. Only for the geometric median we do not see an improvement in some cases. For the model **M1**, since in this case, the rate of convergence of the single subsampling estimator is of order $b^{1/2}$, we see that the bagged estimator has a rate of the same order. In all cases where b and k are small, there is an improvement over a single subsampling estimator, but it is smaller than in scenario 3. The rate appears to improve as k becomes larger, even if b is held fixed equal to $\lfloor n^{2/3} \rfloor$. It seems that for the trimmed mean we also get improved results, but at the cost of a slight loss in term of efficiency.

Bagged Estimator	Scenario	**M1**	**M2a**	**M2b**
Mean	1	−0.5853 (0.0068)	−0.87560 (0.0109)	−1.0512 0.0124
	2	−0.74610 0.0100	−1.0130 (0.0099)	−1.2067 (0.0133)
	3	−0.9395 (0.01088)	−1.1664 (0.0142)	−1.2765 (0.0145)
Median of r.p.	1	−0.575077 0.00939	−0.8600 (0.0104)	−0.9729 (0.0116)
	2	−0.623078 (0.0092)	−0.8454 (0.0104)	−0.9815 (0.0108)
	3	−0.7014 (0.0116)	−0.8664 (0.0124)	−0.9829 (0.0104)
Geo Median	1	−0.5367 (0.0070)	−0.84846 (0.0110)	−1.0186 (0.0123)
	2	−0.6795 (0.0110)	−0.9689 (0.0100)	−1.1708 (0.0132)
	3	−0.8357 (0.0127)	−1.0923 (0.0153)	1.2158 (0.0151)
TM	1	−0.5397 (0.0092)	0.8490 (0.011)	−1.0501 (0.0126)
	2	−0.6856 0.0110	−0.9759 (0.0101)	−1.2033 (0.0133)
	3	−0.8823 (0.0119)	−1.17654 (0.0158)	−1.2646 (0.0150)

Table 1: Convergence rate estimation results: estimated value of -2β along with standard error for each considered model **M1**, **M2a**, **M2b**, and for each parameter: mean, median, geometric median (geoMedian) and trimmed mean (TM) under 3 scenarios: 1 ($k_1 = 50,\ b_1 = \lfloor n^{1/2} \rfloor$), 2 ($k_1 = 100,\ b_1 = \lfloor n^{2/3} \rfloor$), 3 ($k_1 = 500,\ \lfloor b_1 = n^{2/3} \rfloor$).

7 Conclusions

This paper provides some general results on the mean square consistency for generalized subsampling estimators that can be viewed as bagged subsampling estimators. We also developed some Bernstein inequalities for these generalized subsampling estimators, which allow us to show that bagged subsampling estimators enjoy some improved

convergence rate properties, as in the i.i.d. case. We have extended and applied our results to some cyclostationary times series. We have investigated the properties of our estimator on some periodic ARMA models. Our simulation results illustrate the consistency of the bagged subsampling estimators and, also reflect the gain in accuracy that we have explored in theory. Using the mean of means, medians or trimmed means yields fairly similar results when applied directly to the periodogram. However, care should be taken using the median or trimmed mean when applied to the rescaled peridogram. For example, we applied the median to the rescaled periodogram before averaging and this appears to cause a significant bias. The main remaining issue is the choice of the hyperparameters, b the subsampling size, and k the number of subsampling estimators used to compute the bagged estimators. In a nonstationary setting, this is a challenging problem. We hope to solve this problem in the future by combining the tools developed in Bertail and Dudek [2024], Lenart [2011] and Lenart [2018].

References

J. Antoni. Cyclostationarity by examples. *Mech. Syst Sig Process.*, 23(4):987–1036, 2009.

M. Banerjee, C. Durot, and B. Sen. Divide and conquer in nonstandard problems and the super-efficiency phenomenon. *The Annals of Statistics*, 47(2):720 – 757, 2019. doi: 10.1214/17-AOS1633. URL `https://doi.org/10.1214/17-AOS1633`.

P. Bertail. Second-order properties of an extrapolated bootstrap without replacement under weak assumptions. *Bernoulli*, 3(2):149 – 179, 1997.

P. Bertail, D.N. Politis, and J.P. Romano. On subsampling estimators with unknown rate of convergence. *Journal of the American Statistical Association*, 94(446):569–579, 1999. doi: 10.1080/01621459.1999.10474151. URL `https://www.tandfonline.com/doi/abs/10.1080/01621459.1999.10474151`.

P. Bertail, O. Jelassi, J. Tressou, and M. Zetlaoui. Subsampling for big data : some

recent advances. In *Recent advances in Nonparametric Statistics, ISNPS 2018, Avignon, Springer Verlag, Berlin.* 2018. URL https://hal.science/hal-01707500.

P. Bertail, M. Bouchouia, O. Jelassi, J. Tressou, and M. Zetlaoui. Scaling by subsampling for big data, with applications to statistical learning. *Journal of Nonparametric Statistics*, 36(1):78–117, 2024. doi: 10.1080/10485252.2023.2219782. URL https://doi.org/10.1080/10485252.2023.2219782.

Patrice Bertail and Anna E. Dudek. Optimal choice of bootstrap block length for periodically correlated time series. *Bernoulli*, 30(3):2521 – 2545, 2024. doi: 10.3150/23-BEJ1683. URL https://doi.org/10.3150/23-BEJ1683.

Peter J. Bickel and Anat Sakov. On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica*, 18(3):967–985, 2008. ISSN 10170405, 19968507. URL http://www.jstor.org/stable/24308525.

Leo Breiman. *Machine Learning*, 24(2):123–140, Jan 1996. doi: https://doi.org/10.1023/a:1018054314350. URL https://link.springer.com/article/10.1023/A:1018054314350.

P. Bühlmann. *Bagging, Subagging and Bragging for Improving some Prediction Algorithms*, pages 19–34. 12 2003. ISBN 9780444513786. doi: 10.1016/B978-044451378-6/50002-8.

Peter Bühlmann and Bin Yu. Analyzing bagging. *The Annals of Statistics*, 30(4), Aug 2002. doi: https://doi.org/10.1214/aos/1031689014.

C. Corduneanu. *Almost Periodic Functions.* Chelsea, New York, 1989.

D. Dehay, A.E. Dudek, and J. Leśkow. Subsampling for continuous-time nonstationary stochastic processes. *J. Stat. Plan. Inf.*, 150:142–158, 2014.

A.E. Dudek, H. Hurd, and W. Wójtowicz. Periodic autoregressive moving average methods based on fourier representation of periodic

coefficients. *WIREs Computational Statistics*, 8(3):130–149, 2016. doi: https://doi.org/10.1002/wics.1380.

J. Fukuchi. Subsampling and model selection in time series analysis. *Biometrika*, 86(3): 591–604, 1999.

Y. Gao, R. Zhang, and H. Wang. On the asymptotic properties of a bagging estimator with a massive dataset. *Stat*, 11(1):e485, 2022. doi: https://doi.org/10.1002/sta4.485.

W.A. Gardner, A. Napolitano, and L. Paura. Cyclostationarity: half a century of research. *Signal Processing*, 86:639–697, 2006.

Sergio González, Salvador García, Javier Del Ser, Lior Rokach, and Francisco Herrera. A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. *Information Fusion*, 64: 205–237, Jul 2020. doi: https://doi.org/10.1016/j.inffus.2020.07.007. URL `https://www.sciencedirect.com/science/article/abs/pii/S1566253520303195`.

F. Han. An Exponential Inequality for U-Statistics Under Mixing Conditions. *Journal of Theoretical Probability*, 31(1):556–578, March 2018. ISSN 0894-9840. doi: 10.1007/s10959-016-0722-4.

T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2009. ISBN 9780387848846. URL `https://books.google.fr/books?id=eBSgoAEACAAJ`.

C. Heilig and D. Nolan. Limit theorems for the infinite-degree u-process. *Statistica Sinica*, 11(1):289–302, 2001.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Am. Statist. Assoc.*, 58(301):13–30, 1963.

H.L. Hurd and A.G. Miamee. *Periodically Correlated Random Sequences: Spectral Theory and Practice*. Wiley, Hoboken, New Jersey, 2007.

R.H. Jones and W.M. Brelsford. Time series with periodic structure. *Biometrika*, 54 (3/4):403–408, 1967.

Ł. Lenart. Asymptotic distributions and subsampling in spectral analysis for almost periodically correlated time series. *Bernoulli*, 17(1):290–319, 2011.

Ł. Lenart. Non-parametric frequency identification and estimation in mean function for almost periodically correlated time series. *J. Multivariate Anal.*, 115:252–269, 2013.

Ł. Lenart. Generalized resampling scheme with application to spectral density matrix in almost periodically correlated class of time series. *Journal of Time Series Analysis*, 37:369–404, 2016.

Ł. Lenart. Generalized subsampling procedure for non-stationary time series. *Electronic Journal of Statistics*, 12(2):3875–3907, 2018.

F. Merlevede, M. Peligrad, and E. Rio. Bernstein inequality and moderate deviations under strong mixing conditions. In *High dimensional probability V: the Luminy volume*, volume 5, pages 273–293. Institute of Mathematical Statistics, 2009.

A. Napolitano. *Generalizations of Cyclostationary Signal Processing: Spectral Analysis and Applications*. Wiley-IEEE Press, 2012.

A. Napolitano. Cyclostationarity: New trends and applications. *Signal Processing*, 120: 385–408, 2016.

Dimitris N. Politis and Halbert White. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70, Dec 2004. doi: https://doi.org/10.1081/etc-120028836.

D.N. Politis. Scalable subsampling: computation, aggregation and inference. *Biometrika*, 111(1):347–354, 03 2024. ISSN 1464-3510. doi: 10.1093/biomet/asad021. URL `https://doi.org/10.1093/biomet/asad021`.

D.N. Politis and J.P. Romano. *A General Theory for Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions*. and National Science Foundation (U.S.), 1992.

D.N. Politis and J.P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.*, 22(4):2031–2050, 1994.

D.N. Politis, J. Romano, and M. Wolf. *Subsampling*. Springer-Verlag, New York, 1999.

Yuhong Yang. Aggregating regression procedures to improve performance. *Bernoulli*, 10(1), Feb 2004. doi: https://doi.org/10.3150/bj/1077544602.

T. Zou, X. Li, X. Liang, and H. Wang. On the subbagging estimation for massive data, 2021. URL `https://arxiv.org/abs/2103.00631`.