Title: Estimation of the parameters for non-stationary time series with long memory and heavy tails using weak dependence condition

Author: Elżbieta Gajecka-Mirek

University of Silesia in Katowice

Institute of Mathematics

Doctoral Thesis

Estimation of the parameters for non-stationary time series with long memory and heavy tails using weak dependence condition

Elżbieta Gajecka - Mirek

Supervisor
dr hab. Jacek Leśkow

Katowice 2015

Acknowledgments

I wish to express my gratitude to my supervisor Professor Jacek Leśkow for his help, support and patience. I have been extremely lucky to have a supervisor who cared so much about my work. His inspirational guidance, enthusiasm and encouragement helped to make my thesis possible.

Dziękuję moim Rodzicom i Mężowi - za wszystko.

Contents

1 Abstract

Statistical inference for unknown distributions of statistics or estimators may be based on asymptotic distributions. Unfortunately, in the case of dependent data the structure of such statistical procedures is often ineffective. There are different reasons that, e.g. too small a number of observations, the unknown form of the asymptotic distribution or too slow convergence to the asymptotic distribution. In the last three decades we can observe an intensive development the of so-called resampling methods.

Using these methods, it is possible to directly approximate the unknown distributions of statistics and estimators. The idea of resampling is simple, i.e. we calculate the estimator replications and the replications determine the empirical distribution called resampling distribution. A problem that needs to be solved during the study of the resampling procedures is their consistency, i.e. whether the resampling distribution is close enough to the true one? There are many resampling methods.

Their consistency for independent observations has been extensively studied. The case of the stationary data with strong mixing dependence structure has also been well investigated. Resampling for time series with a specific non-stationarity, i.e. the periodic and almost periodic also been the subject of research. Recent research on resampling methods focus mainly on the time series with the weak dependency structure, defined by Paul Doukhan.
The thesis presents a time series model with specific features i.e.: long memory, heavy tails (stable or GED) and periodic structure. Such a model can be naturally used in many areas like energy, vibromechanics, telecommunications, climatology and economics.

The objective of this thesis is to present several consistency theorems for the resampling method for the estimator of the mean function in the above-mentioned time series. Only one of the resampling techniques can be used for the long-range dependent data. This method is subsampling. It involves selecting from the observation all possible subsequences of a some length and calculate the estimator on these subsequences.

In the thesis,we introduce and prove theorems that are necessary to establish consistency of resampling. Moreover, a brief overview of the previous results in inference for non-stationary time series is presented.

Key words: resampling methods, subsampling, periodically correlated time series, strong mixing, weak dependence, consistency of subsampling, heavy tails, long range dependence

2 Streszczenie

Wnioskowanie statystyczne dla nieznanych rozkładów statystyk lub estymatorów można oprzeć na rozkładach asymptotycznych. Niestety, w przypadku danych zależnych, takie procedury statystyczne są niejednokrotnie nieefektywne. Różne są tego przyczyny, np. zbyt mała liczba danych, nieznana postać rozkładu asymptotycznego, zbyt wolna zbieżność do rozkładu asymptotycznego. Od początku lat osiemdziesiątych ubiegłego wieku intensywnie prowadzone są badania nad rozwojem tzw. metod resamplingowych ([30], [80]). Za pomocą tychże metod można bezpośrednio przybliżać nieznane rozkłady statystyk i estymatorów.

Idea resamplingu jest prosta. Obliczamy replikacje estymatora i z tych replikacji wyznaczamy rozkład empiryczny tzw. rozkład resamplingowy.
Problem, z którym trzeba się zmierzyć badając procedury resamplingowe to ich zgodność, tzn. czy rozkad resamplingowy jest bliski prawdziwemu rozkładowi? Metod resamplingowych jest wiele. Ich zgodność w przypadku obserwacji niezależnych została dogłębnie zbadana ([30]). Przypadek danych stacjonarnych ze swoistą strukturą zależności tzn. silnie mieszających także został zbadany ([61], [62], [71]). Przedmiotem intensywnych prac badaczy był również resampling dla niestacjonarnych szeregów czasowych ze specyficzną formą niestacjonarności tzn. okresowych i prawie okresowych ([19], [27], [66], [67], [68], [77]). Ostatnie badania nad metodami resamplingowymi koncentrują się głównie na szeregach czasowych ze zdefiniowaną przez Paula Doukhana słabą zależnością ([25]).

W niniejszej pracy został przedstawiony model dla szeregów czasowych, które mają bardzo specyficzne własności tzn.: posiadają długą pamięć ([7]), ciężkie ogony (stabilne lub GED) ([55], [89], [90], [92]) oraz strukturę okresową ([3], [36], [37], [47]). Taki model może mieć naturalne zastosowanie w wielu dziedzinach np.: energetyce, wibromechanice, telekomunikacji, klimatologii jak również w ekonomii.

Celem pracy jest pokazanie twierdzeń dotyczących zgodności estymatora jednej z metod resamplingowych dla funkcji średniej we wspomnianych powyżej szeregach czasowych. Okazuje się, że jedyną metodą resamplingową, którą można zastosować do danych z długą pamięcią jest subsampling. Polega ona na wyborze z obserwacji wszystkich możliwych podciągów o pewnej długości i wyznaczaniu estymatora na tych podciągach. W pracy sformułowano i udowodniono centralne twierdzenia graniczne, niezbędne do udowodnienia zgodności subsamplingu. Ponadto przedstawiony został przegląd dotychczasowych rezultatów dotyczących metod resamplingowych w szeregach czasowych.

Słowa kluczowe: metody resamplingowe, subsampling, szeregi czasowe okre-

sowo skorelowane, własność słabego mieszania, zgodność subsamplingu, ciężkie ogony, długa pamięć

3 Introduction

Possibility to construct sampling distributions of estimators for time series is very important in statistical studies. Traditional statistical inference based on asymptotic distributions does not always lead to effective statistical procedures. There are several reasons for this, e.g.:

- the convergence of the estimator to the asymptotic distribution is slow and often requires a large collection of observations. In practice, there is not always the possibility to receive enough data because of the costs or technical restrictions.
- The asymptotic distribution is often very complicated and depends on the unknown parameters, which in the case of dependent data is difficult to estimate.

In such situations, the resampling methods are helpful. Moreover, in many cases these methods are the only effective technique. Resampling methods include: jackknife, bootstrap methods, subsampling and model based resampling, e.g. sieve bootstrap. These methods allow us to approximate the unknown distributions (or characteristics) of the statistics and estimators without a reference to the form of the distribution. These approximations are used to construct the confidence intervals for the parameters and testing statistical hypothesis.
The development of resampling methods started in the eighties of the last century from the Efron's [30] work, dedicated to independent data.
The main idea of resampling is based on sampling from some distribution $\hat{P}$ that corresponds to data. In the case of i.i.d. observations the most popular resampling technique is the nonparametric bootstrap for which $\hat{P}$ is simply an empirical distribution function. For dependent data, however, the construction of $\hat{P}$ is more complicated involving blocks of data.

Consistency of subsampling means that the method generates valid quantiles for confidence intervals in non-stationary models. One can compute the confidence intervals and critical values from the subsampling distributions instead of the asymptotic distributions.
In the nineties of the last century the research was focused on stationary time series. At the present time, the efforts of researchers are concentrated on the non-stationary series, with discrete and continuous time [27], [34], [49], [66], [87].
One of the specific form of non-stationarity is periodicity. Gladyshev [36] initiated the development of research on periodicity in time series and stochastic process. In 2006 Gardner et al. [35] have provided a general overview of research on pe-

riodicity and time series, considering over 1500 published papers on the topic. It was shown that the models with periodic structure are widely applicable e.g. in communication, signal processing, vibromechanics, econometrics, climatology and biology.
The resampling methods for periodic time series is an open research area, where many fundamental properties have yet to be proven.
In the thesis, we will deal only with one form of resampling - subsampling, since we will work with long memory time series.
The attention will be focused on the class of time series which simultaneously deals with three features: periodic structure, heavy tails and long memory. The motivation for this approach comes from the fact that in many applications of time series one is confronted with large probabilities of extremal events i.e. heavy tailed behavior.
The heavy-tailed random variables are variables with distributions whose extreme values are "more probable than normal". Examples of such distributions are the Generalized Error Distribution (GED) distributions or stable distributions. Both classes will be discussed in this thesis. Additionally, in real data sets one has to deal with long range dependence as well.

The presence of long range dependence in time series means that there exists dependence between observations which are distant in time from each other.

In 2007 Politis and McElroy [75] have proposed the model, based on sub-Gaussian vectors [89], that was the combination of the two features: heavy tails and long memory. This particular model was the starting point for studies included in this thesis.
It is obvious that among the observations of the time series there is a relationship - the dependence. Over the years, the most popular way for studying this dependence have been the mixing conditions like:

- strong mixing ($\alpha-$mixing) - the most popular condition of all mixing conditions, introduced by Rosenblatt [83],
- absolute regularity ($\beta-$mixing) introduced by Volkonski and Rozanov [91],
- uniform mixing ($\phi-$mixing) introduced by Ibragimov [52] .

The fact is that under natural restrictions on the process parameters, many processes of interest fulfill mixing conditions [24]. On the other hand, there is a large class of processes for which mixing properties do not hold. It turned out that mixing conditions are, in many cases, too strong to define dependence in time series.

Simple example of such a time series is a stationary $AR(1)$ process:

$$X_t = aX_{t-1} + \epsilon_t,$$

where the innovations are independent with $P(\epsilon_t = 1) = P(\epsilon_t = -1) = 1/2$ and $0 < |a| \leq 1/2$.
This process has a stationary distribution on $[-2, 2]$ and X_t has always the same sign as ϵ_t. It is possible to recover $X_{t-1}, X_{t-2}, ...$ from X_t, it means that the process $\{X_t\}_{t\in\mathbb{Z}}$ is purely deterministic going backwards in time, so it cannot be strong mixing (proof is given in [1]).
In 1999 Doukhan and Louhichi [25] and simultaneously Bickel and Bühlmann [13] proposed an alternative condition for the dependence in time series called weak dependence and $\nu-$dependence, respectively. This kind of dependence property is obtained from the convergence to zero of covariances of the process. They called a process weakly dependent if the covariances of smooth functions of blocks of random variables separated by a time gap tend to zero as the time gap increases. It has been shown that many classes of processes for which mixing does not hold satisfy weaker conditions - the weak dependence condition [18]. The definition of weak dependence in comparison to, for example, mixing is very general. It includes very general data sets and models like causal, non causal linear, bilinear, strong mixing processes, dynamical systems or Markov processes driven by discrete innovations.
The main objective of this thesis is to introduce the theoretical results describing the consistency of subsampling method and to show how to use them in statistical inference for time series with periodic behavior.
Three specific features of time series will be studied: heavy tails (stable and GED), long memory and periodic behavior. The construction of described in the thesis process entails the weak dependence property.

The central limit theorem for the mean estimator will be given. The subsampling method to estimate the mean vector will be presented and the applications of the central limit theorem to prove the consistency of subsampling method will be shown.

The structure of the thesis is as follows. In the fourth chapter of the thesis the definitions and the main ideas will be introduced. The purpose of the fifth chapter is reviewing the existing resampling methods for periodic time series. Sixth chapter contains the construction of our long memory, heavy tailed and periodically stationary model and its properties. Moreover, the new central limit theorems are presented. Moreover, the consistency of one of the resampling method - subsampling is shown. In the seventh chapter the applications are presented.

Main original results of the dissertation are presented in Sections 6 and 7.

4 Basic concepts and definitions

In this chapter the basic concepts and definitions will be presented. Some of them will be illustrated by examples.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(\mathbb{R}, \Sigma)$ a measurable space.

Let also $\{X_t : t \in \mathbb{I}\}$ be a real-valued stochastic process and $\mathbb{I}$ is the set of time indexes. A real-valued stochastic process $\{X_t : t \in \mathbb{I}\}$ with $\mathbb{I} = \mathbb{Z}$ is called time series.

Below the definitions of strictly and weakly stationary processes are introduced. The first is referred to invariance of the moments of adequate orders and second is referred to invariance of distributions. In both definitions $T-$the length of the period and $r-$order parameter are natural numbers.

Definition 4.1. *([47], p. 3) The time series $\{X_t\}_{t\in\mathbb{Z}}$ is called strictly stationary if for each $t_1, t_2, t_3, ..., t_n \in \mathbb{Z}$ we have*

$$(X_{t_1}, X_{t_2}, ..., X_{t_n}) \stackrel{d}{=} (X_{t_1+1}, X_{t_2+1}, ..., X_{t_n+1}).$$

Definition 4.2. *The time series $\{X_t\}_{t\in\mathbb{Z}}$ is called weakly stationary of order r, $(WS(r))$, if $E|X_t|^r < \infty$ and for each $t, \tau_1, \tau_2, ..., \tau_{r-1} \in \mathbb{Z}$ and $h \in \mathbb{Z}$,*

$$E(X_t X_{t+\tau_1} ... X_{t+\tau_{r-1}}) = E(X_{t+h} X_{t+\tau_1+h} ... X_{t+\tau_{r-1}+h}).$$

Comment 4.1. *For $r = 2$ we obtain classical weak stationarity.*
It means that the mean of the time series is constant and autocovariance function depends only on h.

Definition similar to Def. 4.2 can be found in [88], p. 9 and in [81], p. 105.

4.1 Data with the periodic structure

The case of stationary stochastic models was quite well investigated in the past century. For non-stationary models there is a need to classify the type of non-stationarity at hand.
Many real life phenomena are characterized by a seasonal behavior which, obviously, is not non-stationary. Seasonal data appear in such fields as: economics, biology, climatology, telecommunications and many others. If seasonality is not easily removable it means that we are dealing with a particular type of non-stationarity, for example the periodic structure. In such cases it is not just the mean that has a periodic rhythm. A periodic rhythm also describes the behavior

of covariance.
Popular models used for describing such phenomena are periodically nonstationary processes. Synonyms for periodically non-stationary are periodically stationary, cyclostationary, processes with periodic structure and many others. The pioneer of research on periodically non-stationarity was Gladyshev [36], [37]. For a review the research of Dehay and Hurd [20], Hurd et al. [48], [47], Hurd and Leśkow [49] can be referred. Development of these research brought many theoretical results (Gardner et. al. [35], Leśkow et al. [66]).

Below are introduced the formal definitions of periodicity of the time series.

Definition 4.3. *([47], p. 3) A time series $\{X_t\}_{t\in\mathbb{Z}}$ is called (strictly) periodically stationary (PS) with period T if, for every n, any collection of times $t_1, ..., t_n \in \mathbb{Z}$, and Borel sets $A_1, ..., A_n \subset \mathbb{R}$,*

$$P_{t_1+T,...,t_n+T}(A_1, ..., A_n) = P_{t_1,...,t_n}(A_1, ..., A_n),$$

and there are no smaller values of $T > 0$ for which above equation holds. $\tau \in \mathbb{Z}$.

For the time series $\{X_t\}_{t\in\mathbb{Z}}$ we define the autocovariance of the pair (X_t, X_{t+h}) to be

$$\gamma_X(t, h) = Cov(X_t, X_{t+h}).$$

Definition 4.4. *([47], p. 5) Time series $\{X_t\}_{t\in\mathbb{Z}}$ is periodically correlated (PC) in Gladyshev sense, if the mean $\mu_X(t)$ is periodic $(\mu_X(t) = \mu_X(t + T))$ and the autocovariance function $\gamma_X(t, h)$ is periodic in t for all $h \in \mathbb{Z}$ $(\gamma_X(t, h) = \gamma_X(t + T, h))$.*

If there is no ambiguity, we will write $\gamma(t, h)$ (or $\gamma(h)$ if we deal with classical weak stationarity) instead of $\gamma_X(t, h)$ for the autocovariance function of time series $\{X_t\}_{t\in\mathbb{Z}}$.

4.2 Long range dependence

In many periodic phenomena the existence of the long range dependence is observed [46], [76], [7]. The presence of long range dependence in time series means that there exists a relationship between observations which are far away from each other in time. Classical fields where long range dependence occurs are dendrochronology and hydrology. Long memory occurs in the sense that a hyperbolic behavior of the autocorrelations holds for almost all lags and frequencies respectively.

The investigation of long range dependence in time series data was started in the seventies by Lawrance and Kottegoda [65], McLeod and Hipel [74], and then in eighties by Hosking ([45]).

The long range dependence can be defined as long memory. Note that the definition of long memory introduced below is one of many possible definitions.

Definition 4.5. *([16], p.520) A stationary, in the sense of the Definition 4.2, time series $\{X_t\}_{t\in\mathbb{Z}}$ has long memory if its autocovariance function γ satisfies the following formulas:*

$$\sum_{0<|h|<n} \gamma(h) \sim Cn^{\beta}$$

where $\beta \in [0,1)$, and $C \neq 0$.

Definition 4.6. *A PC or PS time series $\{X_t\}_{t\in\mathbb{Z}}$ has a long memory if the autocovariance function $\gamma(s)(h) = Cov(X_{s+qT}, X_{s+(q+h)T})$ for each $q \in \mathbb{Z}$ satisfies the following formula*

$$\sum_{0<|h|<n} \gamma(s)(h) \sim C(s)n^{\beta}, \quad s \in \{1,\ldots,T\}$$

where $\beta \in [0,1)$. For each $s \in \{0,\ldots,T-1\}$ $C(s)$ is the finite constant such that

$$C(s) = 2 \cdot \lim_{n\to\infty} \frac{\sum_{h=1}^{n-1} \gamma(s)(h)}{n^{\beta}} > 0.$$

Let us assume that the notation for the long memory with parameter $\beta \in [0,1)$ will be $LM(\beta)$.

Granger and Joyeux [38] and Hosking [44] proposed the use of fractional differencing in modeling this kind of data. Fractional differencing is related to the so called Hurst phenomenon in hydrology (Hosking [44]). $\beta-$ the long memory parameter is related to $H-$the Hurst parameter: $H = \frac{\beta+1}{2}$.

4.2.1 Gegenbauer process

Many data sets presenting the long range dependence also exhibit some form of periodicity. In the case when seasonality is not difficult to remove (for example by removing seasonal means), such phenomena can be modeled via stationary processes - so-called seasonal fractional models presented e.g. by Gray et al. [39], [40]. Following the suggestion of Hosking [44], that so-called Gegenbauer processes, can deal with long memory and seasonal behavior, by suitable selection of the coefficient. Gray studied the so-called GARMA (Gegenbauer ARMA) processes.

Hui and Li have in [46] considered the use of fractional differencing in modeling persistence phenomenon in a periodic process. They mix together periodicity and long-memory i.e. they propose a process consisting of two independent fractional long-memory components. However, the processes are covariance stationary [44].

Notice that the Gegenbauer processes are suitable tool to describe the long memory behavior [44], [32]. This kind of process will be used in modeling the long range dependence in my dissertation.

Definition 4.7. *([39]) Let us assume that ε_t is i.i.d. innovation process. The process $\{G_t\}_{t\in\mathbb{Z}}$ defined by the equation:*

$$\Pi_{1\leq i\leq k}(I - 2\nu_i B + B^2)^{d_i} G_t = \varepsilon_t, \tag{1}$$

is the k-factor Gegenbauer process.
$0 < d_i < 1/2$ if $|\nu_i| < 1$ or $0 < d_i < 1/4$ if $|\nu_i| = 1$ for $i = 1, ..., k$ and I is identity operator, B is backshift operator.

Theorem 4.1. *([41]) Process defined by the Definition 4.7 is long memory, stationary, causal and invertible and has a moving average representation:*

$$G_t = \sum_{j\geq 0} \psi_j(d,\nu)\epsilon_{t-j},$$

with $\sum_{j=0}^{\infty} \psi_j^2(d,\nu) < \infty$, where $\psi_j(d,\nu), j \geq 0$, is defined by:

$$\psi_j(d,\nu) = \sum_{\substack{0\leq l_1,...,l_n\leq j \\ l_1+...+l_n=j}} C_{l_1}(d_1,\nu_1)\cdot ... \cdot C_{l_k}(d_k,\nu_k),$$

where $C_{l_i}(d_i,\nu_i)$ are the Gegenbauer polynomials defined as follows:

$$(1 - 2\nu z + z^2)^{-d} = \sum_{j\geq 0} C_j(d,\nu) z^j, \ |z| \leq 1, \ |\nu| \leq 1.$$

Moreover, if $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ in the Definition 4.7 is the Gaussian white noise, then $\{G_t\}_{t\in\mathbb{Z}}$ is Gaussian time series.

The Gegenbauer processes are stationary, seasonal fractional models [39], [40]. It is enough to take, for example $\nu = cos\varpi_t$, with $\varpi_t = 2t\pi/T$, where T is a season.

4.3 Mixing and Weak dependence

The time series is a sequence of dependent observations. In 1999 Doukhan and Louhichi [25] and simultaneously Bickel and Bühlmann [13] introduced a

new way of describing data dependency - the weak dependence. Until then, the most widely used methods to describe the dependence in the time series were mixing techniques.

In the literature there are several concepts of mixing: α, β, φ, $\psi-$ mixing [24], [15]. The most general and the most widely used is $\alpha-$mixing.
Let $\{X_t : \ t \in \mathbb{Z}\}$ be time series and $\mathcal{F}(t_1, t_2)$ be $\sigma-$algebra generated by the observations $\{X_t : t_1 \leq t \leq t_2\}$.

Definition 4.8. *([24]) We define $\alpha-$mixing sequence as follows*

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

where $\tau \in \mathbb{N}$.
The time series $\{X_t\}_{t\in\mathbb{Z}}$ is $\alpha-$mixing if $\alpha_X(\tau) \to 0$ for $\tau \to \infty$.

Comment 4.2. *P, in Definition 4.8, is the measure corresponding to the whole process $\{X_t\}_{t\in\mathbb{Z}}$.*

Definition 4.9. *([51], p. 111) Let us define*

$$r_X(\tau) = \sup_{\xi_1, \xi_2} |corr(\xi_1, \xi_2)|,$$

and ξ_1 and ξ_2 are respectively measurable to the $\sigma-$algebras $\mathcal{F}_X(-\infty, t)$ and $\mathcal{F}_X(t+\tau, \infty)$.
The time series $\{X_t\}_{t\in\mathbb{Z}}$ satisfy the completely regular condition if $lim_{\tau\to\infty} r_X(\tau) = 0$.

Definition 4.10. *([51], p. 111) Let us define*

$$\rho_X(\tau) = \sup_{\substack{\xi_1 \in L^t_{-\infty} \\ \xi_2 \in L^\infty_{t+\tau}}} |corr(\xi_1, \xi_2)|,$$

L^m_n is the closure in L^2 of the vector space spanned by $X_n, ..., X_m$. The time series $\{X_t\}$ satisfy the completely linear regular condition if $lim_{\tau\to\infty}\rho_X(\tau) = 0$.

It is known [60], that for Gaussian processes the relationship between above coefficient is as follows:

$$\rho_X(\tau) = r_X(\tau),$$

and

$$\alpha_X(\tau) \leq r_X(\tau) \leq 2\pi\alpha_X(\tau).$$

As we see in the formal definition of mixing, the distant observations are almost independent random variables. As a consequence, we can obtain limit results, like limit theorems. But the mixing conditions are dependence conditions in terms of the $\sigma-$algebras generated by the initial time series. This means that we need to consider conditions which are often unverifiable or very difficult to verify in practice.
To quote Bardet [6] the mixing notions are adapted in areas where history, that is the $\sigma-$algebra generated by the past is very important.
Moreover, there are also time series which do not fulfill any of the mixing conditions.
Using weak dependence instead of mixing conditions provides us the whole spectrum of new statistical possibilities. The definition of weak dependence includes very general data sets and models like causal, non causal linear (e.g. Bernoulli shifts), bilinear, strong mixing processes or dynamical systems. Moreover, properties of dependence are independent of the marginal distribution of the time series, that can be the discrete one e.g. Markov processes driven by discrete innovations. The weak dependence provides the tools in the analysis of various statistical procedures with very general data sets.

Below, the two definitions of weak dependence are introduced. In the subsequent considerations it is completely sufficient to use the second definition, which is a special, more simple case of the first one.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $\mathcal{X}$ be a Polish space. Let

$$\mathcal{F} = \bigcup_{u\in\mathbb{N}^*} \mathcal{F}_u \quad and \quad \mathcal{G} = \bigcup_{u\in\mathbb{N}^*} \mathcal{G}_u,$$

where $\mathcal{F}_u$ and $\mathcal{G}_u$ are two classes of functions from $\mathcal{X}^u$ to $\mathbb{R}$.

Definition 4.11. *([18], p. 11) Let X and Y be two random variables with values in $\mathcal{X}^u$ and $\mathcal{X}^v$ respectively. If Ψ is some function from $\mathcal{F}\times\mathcal{G}$ to $\mathbb{R}_+$, define the $(\mathcal{F},\mathcal{G},\Psi)-$dependence coefficient $\epsilon(X,Y)$ by*

$$\epsilon(X,Y) = \sup_{\substack{f\in\mathcal{F}_u \\ g\in\mathcal{G}_v}} \frac{|Cov(f(X),g(Y))|}{\Psi(f,g)}.$$

Let $\{X_n\}_{n\in\mathbb{Z}}$ be a sequence of $\mathcal{X}-$valued random variables. Let $\Gamma(u,v,k)$ be the set of (i,j) in $\mathbb{Z}^u\times\mathbb{Z}^v$ such that $i_1 < ... < i_u \leq i_u + k \leq j_1 < ... < j_v$. The

dependence coefficient $\epsilon(k)$ *is defined by*

$$\epsilon(k) = \sup_{u,v} \sup_{u,v \in \Gamma(u,v,k)} \epsilon((X_{i_1}, ..., X_{i_u}), (X_{j_1}, ..., X_{j_v})).$$

The sequence $\{X_n\}_{n\in\mathbb{Z}}$ *is* $(\mathcal{F}, \mathcal{G}, \Psi)-$*dependent if the sequence* $(\epsilon(k))_{k\in\mathbb{N}}$ *tends to zero. If* $\mathcal{F} = \mathcal{G}$ *we simply denote this as* $(\mathcal{F}, \Psi)-$*dependence.*

Fact 4.1. *([18], p. 11) The Definition 4.11 can be easily extended to general metric sets of indexes* T *equipped with a distance* δ *(e.g.* $T = \mathbb{Z}^d$ *yields the case of random fields). The set* $\Gamma(u, v, k)$ *is then the set of* (i, j) *in* $T^u \times T^v$ *such that*

$$k = min\{\delta(i_l, j_m)/1 \leq l \leq u,\ 1 \leq m \leq v\}.$$

In the thesis it will be sufficient to assume the simpler version of the Definition (4.11), as below.

Let $(E, \| \cdot \|)$ be a normed space and $u \in \mathbb{N}^*$. We assume that a function $h : E^u \longrightarrow \mathbb{R}$ belongs to the class $\mathcal{L} = \{h : E^u \to \mathbb{R},\ \| h \|_\infty \leq 1, Lip(h) < \infty\}$, where $Lip(h) = \sup_{x\neq y} \frac{|h(x)-h(y)|}{\|x-y\|_1}$ and $\| x \|_1 = \sum_{i=1}^u \| x_i \|$.

Definition 4.12. *([6]) A sequence* $\{X_t\}_{t\in\mathbb{Z}}$ *of random variables taking values in* $E = \mathbb{R}^d$ *(*$d \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$*) is* $(\epsilon, \mathcal{L}, \Psi)-$*weakly dependent if there exists* $\Psi : \mathcal{L} \times \mathcal{L} \times \mathbb{N}^* \times \mathbb{N}^* \to \mathbb{R}$ *and a sequence* $\{\epsilon_r\}_{r\in\mathbb{N}}$ $(\epsilon_r \to 0)$ *such that for any* $(f, g) \in \mathcal{L} \times \mathcal{L}$*, and* $(u, v, r) \in \mathbb{N}^{*2} \times \mathbb{N}$

$$|Cov(f(X_{i_1}, ..., X_{i_u}), g(X_{j_1}, ..., X_{j_v}))| \leq \Psi(f, g, u, v)\epsilon_r$$

whenever $i_1 < i_2 < ... < i_u \leq r + i_u \leq j_1 < j_2 < ... < j_v$.

The weak dependence notions are related to the initial time series and are measured in terms of covariance of the functions.
The asymptotic behavior of the covariance shows us the independence between "past" and "future". Intuitively, the weak dependence is "forgetting" in time series.

Just as there are different notions of mixing, so there are several concepts of weakly dependent processes. Generally, one can identify following types of weak dependence:

- $\lambda-$weak dependence, if $\mathcal{F}_u$, $\mathcal{G}_u$ bounded and X_t is $\mathbb{L}^1$ integrable
- $\kappa-$weak dependence, if $\mathcal{F}_u$ bounded and X_t is $\mathbb{L}^1$ integrable

- $\theta-$weak dependence, if $\mathcal{F}_u$, $\mathcal{G}_u$ bounded and X_t is $\mathbb{L}^1$ integrable
- $\zeta-$weak dependence, if $\mathcal{F}_u$ bounded and X_t is $\mathbb{L}^1$ integrable
- $\eta-$weak dependence, if $\mathcal{F}_u$, $\mathcal{G}_u$ bounded and X_t is $\mathbb{L}^1$ integrable

Note there are other cases of weakly dependent, that are not quoted here.

The form of the Ψ function corresponds to particular cases of weak dependence.
The coefficient λ corresponds to:

$$\Psi(f,g,u,v) = uvLip(f)Lip(g) + uLip(f) + vLip(g),$$

the coefficient η corresponds to:

$$\Psi(f,g,u,v) = uLip(f) + vLip(g),$$

the coefficient θ corresponds to:

$$\Psi(f,g,u,v) = vLip(g),$$

the coefficient ζ corresponds to:

$$\Psi(f,g,u,v) = min(u,v)Lip(f)Lip(g),$$

and the coefficient κ corresponds to:

$$\Psi(f,g,u,v) = uvLip(f)Lip(g).$$

There exist following dependences between particular cases of weak dependence:

$$\left\{ \begin{array}{l} \theta \Rightarrow \eta \\ \zeta \Rightarrow \kappa \end{array} \right\} \Longrightarrow \lambda$$

In the definition of weak dependence we denote respectively λ_r, η_r, ζ_r, κ_r or θ_r instead of ϵ_r.

Examples of weakly dependent sequences

First example is $\alpha-$mixing time series, which of course is also weakly dependent. We give the definition of which will also be used in the sequel of the paper thesis.

Definition 4.13. *([6], p. 4) The stationary sequence $\{X_t\}_{t\in\mathbb{Z}}$ is said to be $m-$dependent if X_s and X_t are independent whenever $|s-t| > m$.*

Example 4.1. *The $m-$dependent time series defined as in Definition 4.13 is both $\alpha-$mixing and weakly dependent. It follows strictly from the Definitions respectively 4.8 and 4.12. It is enough to take respectively τ and r more than m.*
The most popular example of $m-$dependent sequence is $MA(m)$.
If $\{Z\}_{n\in\mathbb{N}}$ is an independent sequence then for any finite non zero sequence $(a_1, ..., a_m)$ the moving average ($MA(m)$) process $X_n = a_1Z_1 + ... + a_mZ_{n-m+1}$ is $m-$dependent.

Below non mixing but weakly dependent sequences are given:

Example 4.2. *Bernoulli shift: $X_n = H(\xi_n, \xi_{n-1}, ...)$, where $n \in \mathbb{N}$ (with $H(x) = = \sum_{k=0}^{\infty} 2^{-(k+1)}x_k$) is not mixing but is weakly dependent.*

Indeed: $X_n = \sum_{k=0}^{\infty} 2^{-(k+1)}\xi_{n-k}$, where ξ_{n-k} is the $k-th$ digit in the binary representation of the uniformly chosen number $X_n = 0.\xi_n\xi_{n-1}... \in [0, 1]$.
Such X_n is deterministic function of X_1, so the event $A = (X_1 \leq \frac{1}{2})$ belongs to the $\sigma-$ algebras: $\sigma(X_t, t \leq 0)$ and $\sigma(X_t, t \geq n)$. From the definition:

$$\alpha(n) \geq \mid P(A \cap A) - P(A)P(A) \mid = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

From the lemma below follows that Bernoulli shift $\{X_n\}_{n\in\mathbb{N}}$ is weakly dependent.

Lemma 4.1. *([18]) Bernoulli shifts are $\theta-$weakly dependent with $\theta(r) \leq 2\delta_{[r/2]}$, where $\{\delta_r\}_{r\in\mathbb{N}}$ is defined by: $E \mid H(\xi_{t-j}, j \in \mathbb{Z}) - H(\xi_{t-j}1_{|j|\leq r}, j \in \mathbb{Z}) \mid$.*

□

Example 4.3. *Example of the model which satisfies the weak dependence definition (but not fulfills mixing conditions) is AR(1) model defined as: $X_t = aX_{t-1} + \varepsilon_t$, where $|a| < 1$ and innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are i.i.d. Bernoulli variables with parameter $s = P(\varepsilon_t = 1) = 1 - P(\varepsilon_t = 0)$.*

Example 4.4. *One of the useful theorems to construct the stationary time series with weak dependence properties is given below:*

Theorem 4.2. *([23]) Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be centered i.i.d innovations and let linear process $\{X_t\}_{t\in\mathbb{Z}}$ be defined as*

$$X_t = \sum_{k=0}^{\infty} b_k\varepsilon_{t-k},$$

where $k \in \mathbb{Z}$ and the series b_k is square summable. Then X_t is $\eta-$weakly dependent, where $\eta_{2r}^2 = \sum_{k>r} b_k^2$.

The above models are weakly dependent but they do not exhibit mixing properties [26].
More examples of weakly dependence sequences can be found in the research of Doukhan et al. [18], [41].

4.4 Heavy tailed random variables

The heavy-tailed random variables are variables with distributions whose extreme values are "more probable than normal". The heavy tail phenomena occur frequently in real life. In contradiction to Gaussian phenomena which do not allow for large fluctuations, "heavy tails" can be used to describe high variability. The data with "heavy tails" appear in such different fields as economics, telecommunications, meteorology, physics and signal processing.

If we define kurtosis as μ_4/σ^4 and μ_4 is the fourth central moment (if it exists), while σ is the standard deviation then we can say that heavy-tailed variables are those with kurtosis greater than three, and whose tails go to zero slower than in the normal distribution.

4.4.1 Stable random variables

The stable variables are very specific group of the heavy tails variables. Although they have infinite moments, they are very convenient to use in many applications. Stability plays very important role in the theory of stochastic processes and time series. It is connected with insensitivity character of the process to change of the scale.

One of the main objectives of statistics, is to find the equation that best describes the set of observed points. In the 18th and 19th centuries pioneers of the statistics also considered "the best fit" problem. They found the least squares method very suitable. They considered generating functions and the distribution of the errors and found the importance of the normal distribution. Laplace and Poisson applied the theory of Fourier series and integral as a new tool for analysis of the probability problems. For example the Gaussian density:

$$f_2(x) = \frac{1}{\pi}\int_0^\infty exp(-ct^2)cos(tx)dt$$

is the Laplace's Fourier transform. In the 1853 Augustin Cauchy has discovered that the function f_α, where α is not necessarily equal to 2, satisfying the equation:

$$\int_{-\infty}^{\infty} exp(itx)f_\alpha(x)dx = exp(-\sigma^\alpha|t|^\alpha), \quad \alpha > 0 \tag{2}$$

has the convolution property

$$(Af_\alpha(A\cdot)) * (Bf_\alpha(B\cdot)) = Cf_\alpha(C\cdot)$$

for some $C = C(A, B)$ and all $A,\ B > 0$. To show the non-negativity of the function f_α is sufficient to show that the function $exp(-|t|^\alpha)$ is a characteristic function. Cauchy succeeded in proving that $f_\alpha(x) \geq 0$ for all x only in the cases: $\alpha = 1$ and $\alpha = 2$.

In 1923 Pólya presented the condition for a function to be the characteristic function:
The $\{\psi(t),\ t \in \mathbb{R}\}$ *is characteristic function if* $\psi(t)$ *is real, non-negative,* $\psi(0+) = \psi(0) = 1$, $\psi(t) = \psi(-t)$ *and* ψ *is convex on* $(0, \infty)$.
Above condition implies for $0 < \alpha < 1$ that the function f_α is non-negative. One year after Pólya, in 1924, Lévy proved that for all $\alpha \in (0, 2]$ functions f_α in (2) are non-negative.

The definitions, adopted from the book of Taquu [89] clarify the concept of stable distributions.

Definition 4.14. *([89], p.2) A random variable* X *is said to have a stable distribution if for any positive numbers* A *and* B, *there is a positive* C *and real number* D *such that*

$$AX_1 + BX_2 \stackrel{d}{=} CX + D,$$

where X_1 *and* X_2 *are independent copies of* X, *and "*$\stackrel{d}{=}$*" denotes equality in distributions.*

We also have the following equivalent definition.

Definition 4.15. *([89], p.3) A random variable* X *is said to have a stable distribution if for any* $n \geq 2$, *there is a positive number* C_n *and a real number* D_n *such that*

$$X_1 + X_2 + ... + X_n \stackrel{d}{=} C_nX + D_n,$$

where $X_1, X_2, ..., X_n$ *are independent copies of* X.

Definition 4.16. *([89], p.5) A random variable* X *is said to have a stable distribution if it has a domain of attraction, i. e., if there is a sequence of i.i.d. random variables* $Y_1, Y_2, ...$ *and sequence of positive numbers* $\{d_n\}$ *and real numbers* $\{a_n\}$, *such that*

$$\frac{Y_1 + Y_2 + ... + Y_n}{d_n} + a_n \stackrel{d}{\Rightarrow} X,$$

where "$\stackrel{d}{\Rightarrow}$*" denotes convergence in distribution.*

The Definition (4.16) implies that limits of the normalized sums of i.i.d. random variables can only be the stable distributions.
It is also possible to define the stable random variable by the characteristic function. This kind of definition is equivalent to the definition "in distribution", but it is analytically more tractable.

Definition 4.17. *([89], p.5) Random variable X has a stable distribution if there exist the parameters such that random variable X has the form $\tau > 0$ and $\mu \in \mathbb{R}$: characteristic function of the variable distribution X has a form:*

$$\varphi(t) = \begin{cases} \exp\{-\tau^\alpha |t|^\alpha (1 - i\beta(t) tg\frac{\pi\alpha}{2}) + i\mu t\}, & \alpha \neq 1 \\ \exp\{-\tau|t|(1 + i\beta\frac{2}{\pi}(t)\ln|t|) + i\mu t\}, & \alpha = 1 \end{cases}.$$

While considering stable distributions one usually takes the "enough regular" case of $\alpha-$stability, that is the case when α belongs to the interval $(0, 2]$. Observe that the stable distribution is well defined only when $\alpha \in (0, 2]$. For $\alpha > 2$ the real-valued $\alpha-$stable random variables do not exist ([29]).

Recall that $\alpha \in (0, 2]$ is the stability index, $\beta \in [-1, 1]$ is the skewness parameter, $\tau > 0$ is the scale parameter and $\mu \in \mathbb{R}$ is the location parameter.

If $\alpha = 2$, then the random variable X is Gaussian. In the case when $\alpha \in (0, 2)$, we obtain distributions with tails much heavier than Gaussian. Moreover, when $\alpha \in (0, 2)$, there is no second moment in the distribution and when $\alpha \in (0, 1]$, there is no even first one.

As we have noticed $\beta \in [-1, 1]$ is the skewness parameter. If $\beta > 0$, then distribution of the random variable is skewed to the right (the right tail is heavier), in the case when $\beta < 0$, we have skewness to the left. In the case $\beta = 0$ we deal with the *symmetric* distribution if $\mu = 0$ or about μ, if $\mu \neq 0$.

In the case when $\beta = \pm 1$ we call the distribution of the random variable X *totally skewed* or *completely asymmetric*. The last term was proposed by Weron [92]. Finally, if $\beta = 1$ and $\alpha \in (0, 1)$, then the random variable is positive, real valued. In case if $\beta = -1$ and $\alpha \in (0, 1)$, then the random variable is negative, real valued.

The scale parameter τ plays similar role as the variance in Gaussian case, e.g. if some stable random variables X, X_1 have the scale parameters $\tau_X, 1$, respectively, then variables X and $\tau \cdot X_1$ have the same distributions.

The fourth parameter μ is responsible for the shift.

Instead of writing that a random variable X has the $\alpha-$stable distribution with the parameters: α, β, τ and μ we write $X \sim S_\alpha(\tau, \beta, \mu)$.

It is well-known that the many stable distributions have the self-similarity property.

Definition 4.18. *([89], p. 311)*
The time series $\{X_t\}_{t\in\mathbb{Z}}$ *is H-self similar with index* $H > 0$ *if the finite-dimensional distributions of:* X_{ct} *and* $c^H X_t$ *are the same for all* $c > 0$.

For more informations and examples the reader is referred to Weron et al. [55], p. 135.

The self similarity index H is related to the Hurst parameter H. While $1/2 < H < 1$, the asymptotic behavior of the autocovariance function defines the long range dependence [89], p. 310.

Examples of stable and non-stable distributions It is clear to see that the normal distribution is α-stable with $\alpha = 2$.

In the following example we consider the lack of the stability of the uniform distribution.

Example 4.5. *Non-stability of the uniform distribution.*
Let $X_1,\ X_2-$ *be independent random variables with the uniform distribution on* $[0,1]$. *The density function of uniform distribution on* $[0,1]$ *is:*

$$f(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases} .$$

The density function of $X_1 + X_2$ *is equal:*

$$f_{X_1+X_2}(x) = \begin{cases} 0, & x \notin [0,2] \\ x, & x \in [0,1] \\ 2-x, & x \in [1,2] \end{cases} .$$

It is clear that distribution function of the sum $X_1 + X_2$ *and the uniform distribution of* X_1 *or* X_2 *are completely different. It means that uniform random variable is not stable.*

Notice that all stable distributions that are non degenerate are continuous.

Below analogous definitions for the stability but in the multivariate case will be presented. It leads to the definition of the stability of the time series.

Definition 4.19. *([89], p.57) A random vector $X = (X_1, X_2, ..., X_n)$ is said to be a stable random vector in $\mathbb{R}^n$ if for any positive numbers A and B there is a positive number C and vector $D \in \mathbb{R}^n$ such that*

$$AX^{(1)} + BX^{(2)} \stackrel{d}{=} CX + D, \tag{3}$$

where $X^{(1)}$ and $X^{(2)}$ are independent copies of X, and $\stackrel{d}{=}$ denotes equality in distribution. The vector X is called strictly stable if the equation (3) holds with $D = 0$, $\forall A, B > 0$.

Fact 4.2. *([89], p.58) A random vector X is stable if and only if for any $k \geq 2$, there is a constant $\alpha \in (0, 2]$ and a vector D_k such that*

$$X^{(1)} + X^{(2)} + ... + X^{(n)} \stackrel{d}{=} k^{1/\alpha} X + D_k,$$

where $X^{(1)}, X^{(2)}, ..., X^{(n)}$ are independent copies of X. The α is called stability index.

The finite-dimensional distributions of time series $\{X_t\}_{t\in\mathbb{Z}}$ are the distributions of the vectors

$$(X_{t_1}, X_{t_2}, ..., X_{t_n}), \qquad t_2, t_2, ..., t_n \in \mathbb{Z}, \qquad D \geq 1.$$

Definition 4.20. *([89], p.112) A time series $\{X_t\}_{t\in\mathbb{Z}}$ is stable if all its finite-dimensional distributions are stable.*

Comment 4.3. *([89], p.112) If the finite-dimensional distributions of the stable time series $\{X_t\}_{t\in\mathbb{Z}}$ are stable then they must all have the same index of stability α. We use the term $\alpha-$stable time series when we want to specify the index of stability.*

It is also possible to define the stable random vector by the characteristic function.

Let $(X_1, X_2, ..., X_n)$ be a $\alpha-$stable vector in $\mathbb{R}^n$ and let

$$\Phi_\alpha(\theta) = \Phi_\alpha(\theta_1, \theta_2, ..., \theta_n) = E \exp\big\{i \sum_{k=1}^{n} \theta_k X_k\big\}.$$

$\Phi_\alpha(\theta)$ is characteristic function of $\alpha-$stable random vector.

The interesting question is: are the coordinates of the $\alpha-$stable random vector $(X_1, ..., X_n)$ also $\alpha-$stable? Are linear combinations of $\alpha-$stable variables are $\alpha-$stable as well? The following Lemma provides the answers.

Lemma 4.2. *(Theorem 2.1.2, ([89], p. 58)) Let $X = (X_1, ..., X_n)$ be a stable vector in $\mathbb{R}^n$. Then, in the Definition 3 the constants A, B and C can be chosen in such a way that $C = (A^\alpha + B^\alpha)^{1/\alpha}$. Moreover, any linear combination $Y = \sum_{k=1}^n b_k X_k$ of the components of X is an $\alpha-$stable random variable. We take $b_k \in \mathbb{R}$.*

Another natural question: does the conversion of the Lemma 4.2 hold? If all linear combinations of the coordinates of the random vector are $\alpha-$stable, is the vector $\alpha-$stable? In the Gaussian case, $\alpha = 2$, the answer is yes. What with $\alpha < 2$? Below lemma gives the answer.

Lemma 4.3. *(Theorem 2.1.5, ([89], p. 59)) Let X be a random vector in $\mathbb{R}^n$.*

(a) If all linear combinations $Y = \sum_{k=1}^n b_k X_k$ have strictly stable distributions, then X is a strictly stable random vector in $\mathbb{R}^n$.

(b) If all linear combinations $Y = \sum_{k=1}^n b_k X_k$ are $\alpha-$stable, where $\alpha \geq 1$, then X is a stable vector in $\mathbb{R}^n$.

The following useful Lemma considers the case of sums of independent α-stable variables.

Lemma 4.4. *Let $X_1, X_2, ..., X_n$ be independent random variables with $X_i \sim S_\alpha(\tau_i, \beta_i, \mu_i)$, $i = 1, 2, ..., n$, where $n < \infty$. Then $X_1 + ... + X_n \sim S_\alpha(\tau, \beta, \mu)$, with*

$$\tau = (\tau_1^\alpha + ... + \tau_n^\alpha)^{1/\alpha}, \qquad \beta = \frac{\beta_1 \tau_1^\alpha + ... + \beta_n \tau_n^\alpha}{\tau_1^\alpha + ... + \tau_n^\alpha}, \qquad \mu = \mu_1 + ... + \mu_n.$$

Proof of the Lemma 4.4

- $\alpha \neq 1$

$$lnE \exp i\theta(X_1 + ... + X_n) = ln(E \exp i\theta X_1) + ... + ln(E \exp i\theta X_n) =$$
$$= -(\tau_1^\alpha + ... + \tau_n^\alpha)|\theta|^\alpha + i|\theta|^\alpha sign(\theta)\big(\tan\frac{\pi\alpha}{2}\big)(\beta_1\tau_1^\alpha + ... + \beta_n\tau_n^\alpha) + i\theta(\mu_1 + ... + \mu_n) =$$
$$= -(\tau_1^\alpha + ... + \tau_n^\alpha)|\theta|^\alpha\big[1 - i\frac{\beta_1\tau_1^\alpha + ... + \beta_n\tau_n^\alpha}{\tau_1^\alpha + ... + \tau_n^\alpha} sign(\theta)\tan\frac{\pi\alpha}{2}\big] + i\theta(\mu_1 + ... + \mu_n).$$

- $\alpha = 1$

$$lnE \exp i\theta(X_1 + ... + X_n) = ln(E \exp i\theta X_1) + ... + ln(E \exp i\theta X_n) =$$
$$-(\tau_1 + \tau_2)|\theta|(1 + i(\beta_1 + \beta_2)\frac{2}{\pi} sign(\theta) ln|\theta|) + i(\mu_1 + \mu_2\theta)$$

□

Lemma 4.5. *(Property 1.2.3, ([89]), p. 11) Let $X \sim S_\alpha(\tau, \beta, \mu)$ and let a be a non-zero real constant. Then*

$$aX \sim S_\alpha(|a|\tau, sign(a)\beta, a\mu), \qquad if \quad \alpha \neq 1$$

$$aX \sim S_1(|a|\tau, sign(a)\beta, a\mu - \frac{2}{\pi}a(ln|a|)\tau\beta), \qquad if \quad \alpha = 1.$$

4.4.2 The Generalized Error Distribution

The Generalized Error Distribution (GED) is a parametric model of a heavy tailed distribution. Unlike α-stable distributions, all moments of the GED are finite and the GED has a relatively simple form of a probability density function.

The Generalized Error Distribution is a symmetric unimodal member of the exponential family. The domain of the probability distribution function is $(-\infty, \infty)$. The original concept of the GED was introduced by Subbotin in 1923 [86], so it is known as *"Subbotin's family of distributions"*. However, Subbotin proposed a two-parameters GED model:

$$f(x; h, m) = \frac{mh}{2\Gamma(1/m)} exp\{-h^m|x|^m\}, \tag{4}$$

where $x \in \mathbb{R}$ and $h > 0$ and $m \geq 1$ are scale and shape parameters, respectively. In 1963 Lunetta ([72]) has defined a three-parameters GED class, as follows:

$$f(x; \mu, \sigma, \alpha) = \frac{1}{2\sigma\alpha^{1/\alpha}\Gamma(1 + 1/\alpha)} exp\{-\frac{1}{2}|\frac{x - \mu}{\sigma}|^\alpha\}, \tag{5}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\tau > 0$ is the scale and $\alpha > 0$ is the shape (power).
Of course, the m in the equation (4) is equal to the α in the equation (5) while $h = (\alpha^{1/\alpha}\tau)^{-1}$. Taking into account the fact that the Euler gamma function Γ satisfies the formula $r\Gamma(r) = \Gamma(r + 1)$, the equations (4) and (5) are equivalent, whenever the location parameter μ in (5) is equal to zero.
The GED is also called the generalized normal class. The reason is that for the random variable X with density function as in the formula (5) we have the following equation:

$$\tau = \{E|X - \mu|^\alpha\}^{1/\alpha},$$

which for $\alpha = 2$, gives the standard deviation in the normal case. Note that if $\alpha \neq 2$, τ must not be confused with the standard deviation of X.
Below we give the definition in which τ stands for the standard deviation:

Definition 4.21. *([90]) The random variable X has GED distribution ($X \sim G(\mu, \tau, \alpha)$) if the density function, $f(x)$, of X is given by the equation:*

$$f(x; \mu, \tau, \alpha) = (2\Gamma(1 + 1/\alpha)A(\tau, \alpha))^{-1} exp\{-|\frac{x - \mu}{A(\tau, \alpha)}|^{\alpha}\} \tag{6}$$

with $A(\tau, \alpha) = \tau\sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)}$.

In the sequel of this dissertation it is enough to consider the case $\tau = 1$ in the equation (6). Therefore, we will be considering the density function

$$f(x; \mu, \alpha) = \frac{\alpha}{2A(\alpha)\Gamma(1/\alpha)} exp\{-|\frac{x - \mu}{A(\alpha)}|^{\alpha}\}, \tag{7}$$

where $A(\alpha) = \sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)}$, $\alpha > 0$, $\mu \in (-\infty, \infty)$, and $x \in \mathbb{R}$.

Our definition of the GED is as follows:

Definition 4.22. *The random variable X has a GED distribution $X \sim G(\mu, 1, \alpha)$ if the density function, $f(x)$, of X follows the equation (7).*

The rth central moment of a random variable $X \sim G(\mu, 1, \alpha)$ can be calculated as

$$E(X - EX)^r = \frac{1}{\sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)}\Gamma(1 + 1/\alpha)} \int_{-\infty}^{\infty} (x - EX)^r e^{-\frac{1}{2}|x - EX|^{\alpha}} dx,$$

where $r \in \mathbb{N}$. When r is odd then the rth moments are equal to zero, by symmetry. For r even the rth moments are as follow:

$$EX^r = (\sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)})^r \frac{\Gamma(1/\alpha(r + 1))}{\Gamma(1/\alpha)}.$$

Notice that in the consequence of the Definition 4.22 the first four moments of GED distribution are: mean $= \mu$, variance $= 1$ skewness $= 0$, kurtosis $= \frac{\Gamma(5/\alpha)\Gamma(1/\alpha)}{\Gamma^2(3/\alpha)}$.
Recall that the GED distribution has heavy tails, if $\alpha < 2$. When $\alpha > 2$ we get tails lighter than normal.

It is clear to see that the normal distribution is GED with $\alpha = 2$. Below are other examples of the GED distribution.

Example 4.6. *If we choose $\alpha = 1$ in the Definition 4.21 then the GED distribution is so-called Double Exponential, or Laplace, distribution, i.e. $G(\mu, \tau^2, 1) = L(\mu, 4\tau^2)$.*

Comment 4.4. *Note that the Subbotin's model (4) does not allow for the tails heavier than those in the Laplace distribution. Unlike the formula (5), where the tails heavier than those in the Laplace distribution are allowed.*

Example 4.7. *If we consider the Definition (4.21) and $\alpha \to 0$, then the GED distribution tends to the uniform distribution $U(\mu - \tau, \mu + \tau)$.*

5 Brief overview of the inference for non-stationary time series

There are many results for the stationary and weakly dependent time series, for example Doukhan et al. [22], [26], Bardet et al. [6]. The results will be briefly introduced in the Subsection 3.3.
The main objectives of interest of this Section are resampling methods in time domain for non-stationary time series with periodic structure.
We focus on periodic (seasonal) time series, because periodicity is a common feature of the real life data sets. We can find seasonality in many fields like telecommunication, economy, climatology, vibromechanics. The most popular approach to model the periodic data is the notion of periodic correlation. This approach was introduced by Gladyshev [36] and developed among others by Hurd and Miamee [47], Hurd and Leśkow [49], [50], Hurd, Makagon and Miamee [48], Leśkow and Weron [68], Leśkow and Dehay [21], Leśkow and Synowiecki [66], [67].

5.1 Linear filtration methods

In this Section the results of Javorskyj et al. [56], [57] will be presented. The results are dedicated to the inference of the periodically correlated (PC) process with continuous time.

In the case of discrete time the realization of the process is the time series. Hence the process is, in that sense, the generalization of the time series.

The tool used in the article to estimate the mean and the covariance functions is the linear filtration theory.

Recall that for the PC processes $\{X_t\}_{t\in\mathbb{R}}$ the mean function and the covariance function of the process fulfill the conditions:

$$EX_t = m(t) = m(t+T),$$

$$Cov(X_t, X_{t+u}) = \gamma(t,u) = \gamma(t+T,u).$$

The process of interest in this Section is the PC process in the sense of Gladyshev (the Definition 4.4 with $\mathbb{Z}$ replaced by $\mathbb{R}$). Hence there exists the second moment of the process. Moreover, we assume a summability for the autocovariance function, it means that $\int_{t=-\infty}^{\infty} |\gamma(t,u)|du < \infty$, for all $t \in \mathbb{R}$.

Remark 5.1. *Notice that the last condition shows that if we take $\{X_t\}$ with $t \in \mathbb{Z}$,*

there will not be the long memory in the sense of the Definition 4.5 in considered process.

Fact 5.1. *([56])*
Let us assume that $\int_{t=0}^{T} |m(t)|dt < \infty$ *and* $\int_{t=0}^{T} |\gamma(t,u)|dt < \infty$, *then we use the following Fourier representation for the mean and the autocovariance, respectively:*

$$m(t) = \sum_{k\in\mathbb{Z}} m_k e^{ik\omega_0 t},$$

$$\gamma(t,u) = \sum_{k\in\mathbb{Z}} B_k(u) e^{ik\omega_0 t},$$

where $|m_k| \to 0,\ |B_k(u)| \to 0,$ *if* $k \to \infty$ *and* $\omega_0 = 2\pi/T$.

Functions m_k and $B_k(u)$ are called the mean and the autocovariance components.

5.1.1 Coherent and component methods

In this Section the traditional methods of linear filtration for statistical analysis of PC process will be introduced. This methods are the coherent and the component methods [58], [59]. Moreover, the generalization of this procedures will be used to estimate the mean and the covariance functions and the comparison of the methods will be done.

Without a loss of the generality we assume, in all the Section, that the length of the signal is $n = NT$, where T is the known period and N is a number of periods that are averaged.
As we have mentioned above, the coherent and the component procedures are traditional methods for statistical inference of the PC time process. The coherent method is based on synchronous averaging. Therefore, the estimators of the first and second order components are defined as follows:

$$\hat{m}_N(t) = \frac{1}{N}\sum_{p=0}^{N-1} X_{t+pT},$$

$$\hat{b}_N(t,u) = \frac{1}{N}\sum_{p=0}^{N-1} [X_{t+pT} - \hat{m}_N(t+pT)][X_{t+u+pT} - \hat{m}_N(t+u+pT)].$$

On the other hand, the component method is based on trigonometric polynomials:

$$\hat{m}_n(t) = \sum_{k=-N_1}^{N_1} \hat{m}_{k,n} e^{ik\omega_0 t},$$

$$\hat{\gamma}_n(t,u) = \sum_{k=-N_2}^{N_2} \hat{B}_{k,n}(u) e^{ik\omega_0 t},$$

where

$$\hat{m}_{k,n} = \frac{1}{n} \int_{s=1}^{n} X_s e^{-ik\omega_0 s} ds,$$

$$\hat{B}_{k,n}(u) = \frac{1}{n} \int_{s=1}^{n} [X_s - \hat{m}_n(s)][X_{s+u} - \hat{m}_n(s+u)] e^{-ik\omega_0 s} ds,$$

N_1, N_2 are number of harmonics and $\omega_0 = 2\pi/T$.

In the case when the number of harmonics is large ($N_1 \to \infty,\ N_2 \to \infty$), the properties of both methods are similar. Otherwise the component method gives better results.

Note that the coherent estimation is based on synchronous averaging and only one value on period T is averaged overall the realization length, meanwhile in the component method the integral realization transformations are used. The component method gives more precise results, in the case of mixture of harmonic function and a white noise.

Let us denote the weight function by h. For the estimator $\hat{m}(t)$ we get

$$\hat{m}(t) = \int_0^{NT} X_{t-s} h(s) ds, \tag{8}$$

The coherent or the component estimators are obtained by using the different form of the weight functions.

Theorem 5.1. *([56]) If the weight function h is in the form:*

$$h(s) = \frac{1}{N} \sum_{p=0}^{N-1} \delta(s - pT), \tag{9}$$

where δ is the Dirac function then we obtain a coherent estimator:

$$\hat{m}_N(t) = \frac{1}{N} \sum_{p=0}^{N-1} X_{t-pT}. \tag{10}$$

If the weight function h is in the form:

$$h(s) = \frac{1}{n}\sum_{k=-N_1}^{N_1} e^{ik\omega_0 s} = \frac{sin[(N_1+\frac{1}{2})\omega_0 s]}{nsin\frac{\omega_0 s}{2}}, \tag{11}$$

then we obtain a component estimator:

$$\hat{m}_n(t) = \int_0^n X_{t-s}\big[\frac{1}{n}\sum_{k=-N_1}^{N_1} e^{ik\omega_0 s}\big]ds.$$

Of course functions (9) and (11) are periodic and they satisfy the unbiasedness condition:

$$\int_0^n h(s)e^{-ik\omega_0 s}ds = 1, \tag{12}$$

where $k = -N_1, ..., N_1$, since

$$E\hat{m}_n(t) = \int_0^n m(t-s)h(s)ds = \sum_{k=-N_1}^{N_1} m_k e^{ik\omega_0 t}\big[\int_0^n h(s)e^{-ik\omega_0 s}ds\big]$$

and

$$E\hat{m}_N(t) = \int_0^n m(t-s)h(s)ds.$$

Remark 5.2. The weight function is determined by the properties of the process i.e. it depends on probability characteristics parameters of the process. Only a priori information about these parameters can provide its usefulness.

It is possible to compare the coherent and component estimation procedures using the form (8) of the mean estimator.

We can write (8), as follows:

$$\hat{m}(t) = \sum_{p=0}^{N-1}\int_0^T X_{t-s-pT}h(s)ds, \tag{13}$$

where the interval $[0, n]$ was divided into subintervals $[pT, (p+1)T],\ p = 0, ..., N-1$, and $h(s+pT) = h(s)$.
To get the coherent statistic (10) it is enough to put the following limit of component weight function

$$\lim_{N_1\to\infty} h(s) = \lim_{N_1\to\infty}\frac{\sin\big[\big(N_1+\frac{1}{2}\big)\omega_0 s\big]}{n\sin[\frac{\omega_0 s}{2}]} = \frac{1}{N}\delta(s),$$

where

$$\delta(s) = \frac{1}{T}\sum_{k\in\mathbb{Z}} e^{ik\omega_0 s},\ s\in[0,T]$$

into the equation (13).

Remark 5.3. The component method uses an information about the number of mean function components - N_1, while the coherent method provides its infinite number. The component method becomes more complex as the number of harmonic components grows.

The equation (8) can be also considered from the different perspective. Assume that (8) was obtained as a result of transmitting the process X_s through the filter with a transfer function $H(\omega)$, where

$$H(\omega) = \int_{-\infty}^{\infty} h(s)e^{-i\omega s}ds,$$

and $\omega \in (-\infty, \infty)$.

Fact 5.2. *([56]) In the case of a coherent estimate the transfer function of the filter has the form*

$$H(\omega) = \frac{1}{N}\sum_{p=0}^{N-1} e^{i\omega T/2}(N-1)\frac{\sin[\frac{NaT}{2}]}{N\sin[\frac{\omega T}{2}]},$$

and in the case of a component estimation

$$H(\omega) = \sum_{k=-N_1}^{N_1} e^{-i(\omega-k\omega_0)(n/2)}\frac{\sin[(\omega-k\omega_0)\frac{n}{2}]}{(\omega-k\omega_0)\frac{n}{2}}.$$

According to the harmonic spectral representation of the PC process ([56]):

$$X_t = \int_{-\infty}^{\infty} e^{i\omega t}dZ(\omega),$$

the estimators (8) can be represented as follows:

$$\hat{m}_n(t) = \int_{-\infty}^{\infty} H(\omega)e^{i\omega t}dZ(\omega), \hat{m}_N(t) = \int_{-\infty}^{\infty} H(\omega)e^{i\omega t}dZ(\omega). \tag{14}$$

The characteristics of $dZ(\omega)$ are defined by the equations

$$EdZ(\omega) = \sum_{k=-N_1}^{N_1} m_k\delta(\omega-k\omega_0)d\omega, \tag{15}$$

$$EdZ(\omega_1)dZ^{\star}(\omega_2) = f(\omega_1,\omega_2)d\omega_1 d\omega_2, \tag{16}$$

the star denotes a complex conjugate,

$$f(\omega_1,\omega_2) = \sum_{k=-N_1}^{N_1} \tilde{f}_k(\omega_1)\delta(\omega_2-\omega_1+k\omega_0)d\omega_1 d\omega_2,$$

where

$$\tilde{f}_k(\omega_1) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{B}_k(u) e^{-i\omega u} du.$$

There is following relationship connected with the Fourier decomposition between $\tilde{B}_k(u)$ and correlation function of the process $\{X_t\}$

$$\tilde{\gamma}(s,u) = EX_s X_{s+u} = \sum_{k=-N_2}^{N_2} \tilde{B}_k(u) e^{ik\omega_0 s}.$$

The correlation components $\tilde{B}_k(u)$ are different from covariance components $B_k(u)$ in tails.

Based on (14), we can get the properties of the estimator of th mean function. The estimator of the mean function is as follows

$$E\hat{m}(t) = \sum_{k=-N_1}^{N_1} m_k H(k\omega_0) e^{ik\omega_0 t}.$$

This property was obtained by the calculation using the equation (15). It is necessary to mention that $E\hat{m}(t)$ is equal to the mean function only if $H(k\omega_0) = 1,\ k = -N_1, ..., N_1$. It is a condition for unbiasedness.

Let us consider the variance of the estimator $\hat{m}(t)$:

$$D[\hat{m}(t)] = \int_{-\infty}^{\infty} \sum_{-\infty}^{\infty} H(\omega_1 H^{\star}(\omega_2) e^{i(\omega_1 - \omega_2)t} E dZ(\omega_1) dZ^{\star}(\omega_2) - m^2(t) =$$

$$= \sum_{k=-N_2}^{N_2} e^{ik\omega_0 t} \int_{-\infty}^{\infty} H(\omega) H^*(\omega - k\omega_0) f_k(\omega) d\omega.$$

Since

$$H(\omega - k\omega_0) = H(\omega), \tag{17}$$

the variance of the coherent estimator is as follows:

$$D[\hat{m}(t)] = \sum_{k=-N_2}^{N_2} e^{ik\omega_0 t} \times \int_{-\infty}^{\infty} |H_k(\omega)|^2 f_k(\omega) d\omega.$$

For the component filter its transfer function does not satisfy (17) but for a sufficiently long signal and $\omega - k\omega_0 \in [-N_2\omega_0, N_2\omega_0]$ it can be used with some approximation.
If the number N_1 is larger then the difference between two estimation methods is smaller. Better selectivity of the component filter for small N_1 brings advantages

of the component method over the coherent method. Moreover, correct selection of the component filter for small N_1 leads to the advantage of component method over coherent one.

Let us consider the following example: assume that the multiplicative model

$$X_t = s_t Y_t$$

is the PC process, where Y_t is a stationary narrow band white noise with a zero mean and the covariance function equal to $2D(sin\Omega u/u)$ and the $s$is periodic function $s_t = \sum_{k=-N-1}^{N-1} s_k e^{ik\omega_0 t}$. The covariance function $b(t,u)$ is as follows

$$\gamma(t,u) = 2D\frac{sin\Omega u}{u} \sum_{r=-2N_1}^{2N_1} e^{ir\omega_0 t} \sum_{l=-N_1}^{N_1} s_l s^{\star}_{l-r} e^{il\omega_0 u}. \tag{18}$$

Assuming that $\omega_1 = \Omega + N_1\omega_0$ and $N_2 = 2N_1$ we obtain

- for the component estimator

$$D[\hat{m}(t)] = \sum_{k=-N_2}^{N_2} f_k e^{ik\omega_0 t} \int_{-\omega_1}^{\omega_1} H(\omega)H^{\star}(\omega - k\omega_0)d\omega, \tag{19}$$

- for the coherent estimator

$$D[\hat{m}(t)] = \sum_{k=-N_2}^{N_2} f_k e^{ik\omega_0 t} \int_{-\omega_1}^{\omega_1} |H(\omega)|^2 d\omega, \tag{20}$$

where

$$f_k(\omega) = D \sum_{l=-N_1}^{N_1} s_l s^{\star}_{l-r} e^{il\omega_0 u} \int_{-\Omega-l\omega_0}^{\Omega+l\omega_0} \delta(\omega_1 - \omega)d\omega_1.$$

Moreover, if $\Omega \gg N_1\omega_0$ then $\omega_1 \approx \Omega$ and it leads to the correlation function in the form as follows:

$$\gamma(t,u) \approx 2s_t^2(sin\Omega u/u).$$

Remark 5.4. The process with such correlation function are called locally stationary random processes, see [84].

In general case, when the equation (18) is not fulfilled, the number N_1 defines the measure of difference between the coherent and the component estimators. The estimators will be better if the number of observations increases.

The linear filtration theory can be also successively used to the estimation of the covariance function of the PC process. For simplicity assume that the mean

function is known.
Estimator of the covariance function $\hat{\gamma}(t,u)$ is as follows

$$\hat{\gamma}(t,u) = \int_0^n (X_{t-s} - m(t-s))(X_{t-s+u} - m(t-s+u))h(s)ds.$$

Let us consider the mean of the covariance estimator

$$E\hat{\gamma}(t,u) = \sum_{k=-N_2}^{N_2} e^{ik\omega_0 t} B_k(u) \int_0^n e^{-ik\omega_0 s} h(s)ds. \tag{21}$$

It is equal to $\gamma(t,u)$ only if the unbiasedness condition (12) is fulfilled. For the Gaussian processes it is also possible to calculate the variance of the estimator (21), see [56].

Based on the linear transformation method it is possible to introduce the optimal estimation technique, which gives the minimal value of the estimator variance [56]. This technique is determined by the form of the weight function and is based on finding its optimal form. The weight function depends on probability characteristics parameters of the process. It implies that we need to have a priori information about the parameters of the process.
The time averaged variance of the mean function estimator (8) is as follows:

$$D_T[\hat{m}(t)] = \int_0^n \int_0^n B_0(t-s)h(t)h(s)dtds. \tag{22}$$

Using Euler equation

$$\int_0^n B_0(t-s)h(s)ds = \sum_{k=0}^{2N_1} \mu_k \varphi_k(t)$$

we can write (22) as follows:

$$D_T[\hat{m}(t)] - \sum_{k=0}^{2N_1} \mu_k \int_0^n h(t)\varphi_k(t)dt =$$

$$= \int_0^n h(t)\Big[\int_0^n B_0(t-s)h(s)ds - \sum_{k=0}^{2N_1} \mu_k \varphi_k(t)\Big]dt,$$

where $\varphi_0(t) = 1$, $\varphi_k(t) = cos(k\omega_0 t)$, $\varphi_{k+N_1}(t) = sin(k\omega_0 t)$, $k = 1, ..., N_1$.

The goal is to find the function h which minimizes the equation (22) under the condition

$$\int_0^n h(s)\varphi_k(s)ds = \begin{cases} 1, & k = 0, ..., N_1, \\ 0, & k = N_1 + 1, ..., 2N_1 \end{cases}, \tag{23}$$

which guarantees the unbiasedness of the estimator of the mean function e.g. $E\hat{m}(t) = m(t)$.

The form of the weight function is

$$h(t) = \sum_{k=0}^{2N_1} \mu_k h_k(t),$$

where the functions $h_k(t)$ for each $k = 0, ..., 2N_1$ satisfy the equation:

$$\int_0^n B_0(t-s)h_k(s)ds = \varphi_k(t).$$

This equation can be solved by applying the appropriate techniques (see [56]) and the constants μ_k can be obtained from the formula

$$\sum_{k=0}^{2N_1} \mu_k \int_0^n h_k(t)\varphi_l(t)dt = \begin{cases} 1, & l = 0, ..., N_1, \\ 0, & l = N_1 + 1, ..., 2N_1 \end{cases}.$$

The problem of finding optimal weight function for covariance estimation (21) can be solved in very similar way.

If we repeat the reasoning such as for the the variance of the mean estimator, but we replace N_1 by N_2 in the equation (23), we obtain a method of finding an estimator with the minimal value of variance for the covariance estimator.

Conclusion of this Section is that estimation theory for the PC process can be developed using of the linear filtration methods.

The main advantage of the component technique over the coherent one consists in the finite number of transfer function. This number is equal to the number of harmonic components. Notice that in the case of coherent filter it is infinite. The advantage becomes really significant, if we estimate the characteristics of the PC process with small number of harmonic components. If the length of realization grows the effectiveness of both methods is similar.
The component and the coherent approach can be used in e.g. telecommunication and everywhere else, where the structure of the process is known.

5.1.2 Hilbert's transformation-based method and a frequency shift method

In previous Section traditional statistical methods for estimates characteristics of the PC processes were analyzed.

Definition 5.1. *The PC process* $\{X_t\}_{t\in\mathbb{R}}$ *has a finite power if the following condition holds*

$$\lim_{n\to\infty}\frac{1}{2n}\int_{-n}^{n}E(X_t-m(t))^2dt<\infty.$$

Let us introduce the fact:

Fact 5.3. *If PC process* $\{X_t\}_{t\in\mathbb{R}}$ *has a finite power, then it has a harmonic series representation:*

$$X(t)=\sum_{k\in\mathbb{Z}}X_k(t)e^{ik\omega_0 t}, \tag{24}$$

where $X_k(t)$ *are jointly stationary random signals.*

The linear filtration methods lead to new estimation techniques, which are based on stationary components extraction. The methods use the harmonic series representation of the PC process (24), which is the generalization of the Fourier series representation for periodic functions, where the Fourier coefficients are replaced by jointly stationary processes.

In this Section we will consider an approach based on stationary modulated processes. Two methods will be discussed: method based on a frequency shift and method based on the Hilbert transformation.

In order to extract stationary components the considered region is divided into the bands $[(k-1/2)\omega_0,(k+1/2)\omega_0],\ k\in\mathbb{Z}$.
First method consists in each shifted each band $k\omega_0$ value and low-band filtration in the interval $[-\omega_0/2,\omega_0/2]$.
The second method is filtering each part, and uses the Hilbert transformation for the components estimation. Probability characteristics of these components are determined by probability characteristics of the PC process. They are jointly stationary random processes, and their spectral density functions are located in the interval $[-\omega_0/2,\omega_0/2]$.

If the process is PC then its mean and covariance functions have the following Fourier series representation:

$$m(t)=EX_t=\sum_{k\in\mathbb{Z}}m_k e^{ik\omega_0 t},$$

$$\gamma(t,u)=\sum_{k\in\mathbb{Z}}B_k(u)e^{ik\omega_0 t},$$

where $\omega_0 = 2\pi/T$.
If the process is represented by (24) we obtain the mean

$$m(t) = \sum_{k\in\mathbb{Z}} EX_k(t)e^{ik\omega_0 t} = \sum_{k\in\mathbb{Z}} m_k e^{ik\omega_0 t},$$

and the autocovariance function

$$\gamma(t,u) = \sum_{k,l\in\mathbb{Z}} E\hat{X}_k^*(t)\hat{X}_l(t+u)e^{i(k-l)\omega_0 t}e^{ik\omega_0 t} = \sum_{k\in\mathbb{Z}} e^{ik\omega_0 t}\sum_{l\in\mathbb{Z}} R_{l-k,l}(u)e^{il\omega_0 u},$$

where

$$R_{k,l}(u) = E\hat{X}_k^*(t)\hat{X}_l(t+u), \quad \hat{X}_k(t) = X_k(t) - m_k.$$

Thus

$$B_k(u) = \sum_{l\in\mathbb{Z}} R_{l-k,l}(u)e^{il\omega_0 u}. \tag{25}$$

If $R_{k,l}(u) \neq 0$, for some k, l with $k \neq$, then X_t is a PC signal. If $R_{k,l}(u) = 0$, for all k, l with $k \neq l$, then X_t is a stationary one.
The characteristics $m(t)$ and $B_k(u)$ can be used for statistical analysis of the considered data.

If the process is PC then the components of autocovariance function are determined by auto-and crosscovariance functions of a stationary processes $X_k(t)$, defined in (24).

Notice that from the equation (25) it follows that the covariance components $B_k(u)$ are determined by $R_{l-n,l}(u)-$ the sums of covariance functions of stationary components. If we could calculate $R_{l-n,l}(u)$, we automatically obtain $B_k(u)$.
On the other hand, if we have calculated the components $B_k(u)$, we are able to estimate their probability characteristics and in consequence estimate the PC processes characteristics. But there appears the problem of calculation the direct components $X_k(t)$ from the real-life data.
Taking into consideration the harmonic spectral representation of the PC process ([57])

$$X_t = \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega),$$

we obtain

$$X_t = \sum_{k\in\mathbb{Z}} \eta_k(t)e^{ik\omega_0 t}.$$

Here low-band processes $\eta_k(t)$ are defined as

$$\eta_k(t) = \int_{-\omega_0/2}^{\omega_0/2} e^{i\omega t} dZ_k(\omega), \tag{26}$$

where $dZ_k(\omega) = dZ(\omega + k\omega_0)$.

Fact 5.4. *([57]) For all $k \in \mathbb{Z}$ the mean of the processes $X_k(t)$ from the representation (24) and the mean of the processes $Z_k(t)$ from (26) are the same.*

Fact 5.5. *(Fact 2.2, [57]) The cross-correlation functions of the processes $\eta_k(t)$ depend only on time lag, therefore, the processes $\eta_k(t)$, $k \in \mathbb{Z}$, are jointly stationary.*

Fact 5.6. *(Fact 2.3, [57]) Every PC process can be represented by the equation:*

$$X(t) = \sum_{k\in\mathbb{Z}} \eta_k(t) e^{ik\omega_0 t}, \tag{27}$$

where $\eta_k(t)$ defined by (26) are jointly stationary random processes.

Fact 5.7. *(Fact 2.4, [57]) Representation (24) and (27) are equivalent in terms of probability characteristics of the PC processes.*

The processes $\eta_k(t)$ can be extracted using the transformation:

$$\tilde{\eta}_k(t) = \int_{-\infty}^{\infty} h(t-s) X_s e^{-ik\omega_0 s} ds, \tag{28}$$

where $h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) e^{i\omega t} d\omega$ is a weight function with a transfer function

$$H(\omega) = \begin{cases} 1, & \omega \in [-\frac{\omega_0}{2}, \frac{\omega_0}{2}] \\ 0, & \omega \notin [-\frac{\omega_0}{2}, \frac{\omega_0}{2}] \end{cases}.$$

The form of the transfer function implies that

$$h(s) = \frac{sin(\frac{\omega_0 s}{2})}{\pi s}.$$

The (28) is interpreted like a shifting of a low-band filtration.

Now let us focus on the second linear filtration methods based on stationary components extraction, it is the extraction of quadrature components using the Hilbert transformation for stationary components estimation.
The division considered interval into into the bands $[(k - \frac{1}{2})\omega_0, (k + \frac{1}{2})\omega_0]$ leads to representation (27).

We can write the process $\eta(t)$ as

$$\eta(t) = \eta_c(t)cos(\omega_0 t) + \eta_s(t)sin(\omega_0 t), \tag{29}$$

where $\eta_c(t) = \eta_k(t) + \eta_k^*(t),\ eta_s(t) = i[\eta_k^*(t)\eta_k(t)]$.

Extracted process $\eta(t)$ is the PC process with characteristics that are determined by characteristics of stationary components $\eta_k^c(t)$ and $\eta_k^s(t)$.

Applying the Hilbert transformation to the (29) we get:

$$\zeta(t) = H\eta(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\eta(s)}{t-s}ds.$$

Quadrature components η_k^c and η_k^s obtained by Hilbert transformation

$$\zeta(t) = \eta_k^c(t)sin(\omega_0 t) + \eta_k^s(t)cos(\omega_0 t)$$

are as follows:

$$\eta_k^c(t) = \zeta(t)sin(\omega_0 t) + \eta(t)cos(\omega_0 t)$$
$$\eta_k^s(t) = \eta(t)sin(\omega_0 t) + \zeta(t)cos(\omega_0 t).$$

Processes $\eta_k^c(t)$ and $\eta_k^s(t)$ are stationary and jointly stationary. Because of non-stationarity of the $\eta(t)$ process their auto-correlation functions are not equal. Notice that the means of extracted processes $m_k = E\zeta_k(t)$ are equal and correlation components are determined by their auto-and cross-correlation functions.

Now we briefly investigate the properties of PC process mean and correlation functions estimated using the harmonic series representation.
The m_k are estimated by

$$\hat{m}_k = \frac{1}{n}\int_0^n \eta_k(t)dt.$$

Notice that $\hat{E}_k = m_k$. And for the variance

$$Var(\hat{m}_k) = \frac{1}{n}\int_{-n}^{n}(1 - \frac{|u|}{n})R_{k,k}^{(\eta)}(u)du,$$

where $R_{k,k}^{(\eta)}(u)$ is auto-correlation function of the process η.
The processes $\eta_k(t)$ are stationary. It implies that that $\int_{-\infty}^{\infty}|R_{k,k}^{(\eta)}(u)|du = M < \infty$.

In that case $Var(\hat{m}_k) \to 0$, with $n \to \infty$. It means that the estimator of the mean is consistent.

The correlation component for PC can be built as follows:

$$B_n(u) = \begin{cases} \sum_{l=n-L}^{L} R^{(\eta)}_{l-n,l}(u)e^{il\omega_0 u}, & n \geq 0 \\ \sum_{l=-L}^{L+n} R^{(\eta)}_{l-n,l}(u)e^{il\omega_0 u}, & n < 0 \end{cases}$$

where L is the number of spectral bands and $R^{(\eta)}_{k,l}(u)$ are the cross-correlation functions of the η_k and η_l.
If $n \geq 0$ then we have estimator of the statistic $B_n(u)$ in the form:

$$\hat{B}_n(u) = \sum_{l=n-L}^{L} \hat{R}^{(\eta)}_{l-n,l}(u)e^{il\omega_0 u}$$

where

$$\hat{R}^{(\eta)}_{l-n,l}(u) = \frac{1}{\theta}\int_0^{\theta} [\eta^*_{l-n} - \hat{m}^*_{l-n}][\eta_l(t) - \hat{m}_l]dt.$$

After averaging we obtain:

$$E\hat{R}^{(\eta)}_{l-n,l}(u) = R^{(\eta)}_{l-n,l}(u) - \frac{1}{\theta}\int_0^{\theta}(1 - \frac{|u|}{\theta})R^{(\eta)}_{l-n,l}(u)du.$$

The cross-correlation functions $R^{(\eta)}_{l-n,l}(u)$ are absolutely integrable, it means that the estimator is asymptotically unbiased.

The variance of the $B_n(u)$ is as follows:

$$Var(\hat{B}_n(u)) = E|\hat{B}_n(u) - E\hat{B}_n(u)|^2 \approx$$

$$\approx \sum_{l,k=-N}^{N} E\hat{R}^{(\eta)}_{l-n,l}(u)\hat{R}^{(\eta)}_{k-n,k}(u)e^{i(l-k)\omega_0 u} - |B_k(u)|^2.$$

If for each k the processes $\eta_k(t)$ are Gaussian then the variance tends to zero. It means that for the Gaussian PC process, estimators of the correlation function components obtained by Hilbert transformation are consistent.

The method based on harmonic series representation by stationary components concerns the structure of the process, in contrast to the coherent and component methods which deal only with characteristics of the process.

This technique is very useful for the engineers, specially in vibration diagnostic.

5.2 Resampling methods

In this Section the description of resampling methods and their consistency for time series with periodic structure will be presented.
The resampling methods are based on multiple times resampling from the sample to obtain approximation of the distribution of the investigated estimator. The development of the resampling methods is strictly connected with evolution of computers and technology. The reason of this fact is that the implementation of this method require advanced calculations tools. The first resampling method was the bootstrap method. It was described by Efron at the end of the seventies [30] of the twentieth century.

Let $x_n = (X_1, ..., X_n)$ be the sample of the time series $\{X_t\}_{t\in\mathbb{Z}}$. Let P be the joint distribution of the time series $\{X_t\}_{t\in\mathbb{Z}}$. The statistic $T_n = t_n(x_n, P)$ is calculated and T_n can be for example

$$T_n = a_n(\hat{\theta}_n(X_1, ..., X_n) - \theta(P)),$$

where a_n is normalizing sequence.
Our aim is to approximate the distribution $G_n = \mathcal{L}(T_n)$ as well as to estimate the important characteristics of such a distribution.

The bootstrap version of T_n is in a form $T^*_{m,n} = a_n(\hat{\theta}_m(X^*_1, ..., X*_m) - \hat{\theta}_n(X_1, ..., X_n))$. F_n is a empirical distribution function based on the bootstrap sample $x^*_m = (X^*_1, ..., X^*_m)$.
The estimator of G_n is in a form $\hat{G}_n = G^*_{m,n} = \mathcal{L}^*(T^*_{m,n})$, where x_n is known, hence the distribution F_n of the variables X^*_i is also known. Theoretically it is possible to obtain the distribution $\hat{G}_n$, but unfortunately getting $\hat{G}_n$ is very difficult. The reason of this complication is that the number of bootstrap samples x^*_m grows very fast.
In practice we approximate the distribution of $T^*_{m,n}$ by resampling from the empirical distribution F_n.

Development of resampling method is strictly connected with the development of technology. That is why recently we observed explosion of interest in this kind of methodology. At the beginning the statisticians investigated the resampling methods for independent random variables. More recently, the focus has shifted to dependent data.

There exist a big difference between resampling methods for dependent and independent observations. The following example [85] is the illustration of the differences.

Example 5.1. *Let $x_n = (X_1, ..., X_n)$ be the sample from a m-dependent zero mean time series $\{X_t\}_{t\in\mathbb{Z}}$, that is second ordered stationary. Assume that the subsample $(X_1^*, ..., X_n^*)$ was generated from x_n by drawing single observations with return. It is the case of Efron's bootstrap.*
We can calculate the bootstrap variance of the bootstrap subsample

$$Var^*(\sqrt{n}\bar{X}_n^*) = \frac{1}{n} Var^*\big(\sum_{j=1}^{n} X_j^*\big) =$$

$$Var^* X_1^* = \frac{1}{n}\sum_{t=1}^{n} X_t^2 - \big(\frac{1}{n}\sum_{t=1}^{n} X_t\big)^2 \xrightarrow{P} Var(X_1).$$

It is known that $Var(\sqrt{n}\bar{X}_n) \to \sum_{\tau=-m}^{m} Cov(X_1, X_{1+\tau})$, hence the bootstrap estimator do not estimate the real asymptotic variance.

The above example shows that resampling for dependent data needs to be defined in a different way. If we want to obtain consistency of resampling methods we need to resample not the single values but the blocks of values. The length of the block needs to increase with the increasing the length of the sample.

5.2.1 The Moving Block Bootstrap

In this subsection the moving block bootstrap (MBB) will be described. This procedure was independently introduced by Liu and Singh [71] and Künsch [61]. The results were developed for the strictly stationarity case.

Let $(X_1, ..., X_n)$ be the observed sample from the time series and $B(j, b) = (X_j, ..., X_{j+b-1})$ be $b-$block of the data. The length of the $b-$block is $b = b_n$. Assume, without loss of generality, that $k = n/b \in \mathbb{N}$.

Algorithm

- Let the i.i.d. random variables $i_1, i_2, ..., i_k$ come from the distribution

$$P(i_j = t) = \frac{1}{n-b+1} \; for \; t = 1, ..., n-b+1.$$

- To obtain the MBB resample

$$(X_1^*, X_2^*, ..., X_n^*)$$

the blocks $(B(i_1, b), B(i_2, b), ..., B(i_k, b))$ are concatenated.

Consistency of the moving block bootstrap

It has been proved by Lahiri [63] that if we admit the overlapping the blocks then we get the better effectiveness of block bootstrap. The MBB is the method with the maximal overlap the blocks.
The research of Liu and Singh [71], Künsch [61] and Radulović [82] was focused on the stationary case. Liu and Singh have showed the consistency of the MBB for $m-$dependent time series with finite moments of order more than 4 and the length of the block $b = o(n^{1/2})$. In the research of Künsch we find consistency of the MBB for $\alpha-$mixing time series with additional assumptions: for $\delta > 0$

$$E|X_1|^{6+\delta} < \infty \text{ and } \sum_{\tau=1}^{\infty} \tau^2 \alpha_X^{\frac{\delta}{6+\delta}}(\tau) < \infty.$$

The research of Radulović has summarized all the previous results. The author got the results of consistency of MBB for $\alpha-$mixing and strictly stationary case with weaker assumptions, it means with the standard conditions for the mixing moments:

$$E|X_1|^{2+\delta} < \infty \qquad \text{and} \qquad \sum_{\tau=1}^{\infty} \alpha_X^{\frac{\delta}{2+\delta}}(\tau) < \infty,$$

than those in Künsch's research.

First result, for the non-stationary case but with constant expectation value, was derived by Fitzenberger [33]. Under assumptions:

$$\sup_t E|X_t|^{4+\delta+\epsilon} < \infty \qquad \text{and} \qquad \alpha_X(\tau) = O(\tau^{\frac{\delta}{4+\delta}})$$

author has obtained consistency of the MBB for the length of the block b as follows: $b = o(n^{1/2})$.
Politis et al. [80] also have obtained the consistency of the MBB, but with the stronger assumptions, as follows:

$$\sup_t E|X_t|^{4+2\delta} < \infty \qquad \text{and} \qquad \sum_{\tau=1}^{\infty} \tau^2 \alpha_X^{\frac{\delta}{4+\delta}}(\tau) < \infty.$$

The known and the most general results have been obtained by Synowiecki [87]. The parameter of interest was $M\{EX_t\}-$ the mean over the time variable t, where

$$M(f(t)) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=s}^{s+n-1} f(j).$$

The estimator of the parameter M_t is $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$. The author has shown the consistency of moving block bootstrap for $\alpha-$mixing, nonstationary time series

with strictly periodic structure with the length of the block $b = o(n)$ under the assumption of summability of the covariance function

$$\sum_{\tau=0}^{\infty} |Cov(X_t, X_{t+\tau})| < \infty \qquad \text{for all } t = 1, ..., T.$$

Notice that Synowiecki in his research has not assumed noting about the order of convergence of the $\alpha-$mixing sequence. The theorem with the assumptions

$$E|X_1|^{2+\delta} < \infty \qquad \text{and} \qquad \sum_{\tau=1}^{\infty} \alpha_X^{\frac{\delta}{2+\delta}}(\tau) < \infty,$$

for $\alpha-$mixing and strictly periodic time series is the simple conclusion from the Synowiecki result [87]. And this is a generalization of the strictly stationary case considered by Radulović [82].
The mentioned results of Synowiecki are dedicated to wider class than PC time series. They are dedicated to the APC - almost periodically correlated time series. (For more information on the definition and examples of APC, we refer the reader to the Besicovitch [12], Antoni [3] and Gardner et al. [35].) The almost periodic function cannot be easily estimated and subtracted from the time series. That is the reason why the Fourier analysis needs to be applied to this problem. It helps to identify the frequencies of the mean or the covariance function. The estimators of the mean and covariance function and their asymptotic properties for APC time series have been investigated in Hurd and Leśkow [50] and Dehay and Leśkow [21]. Synowiecki in his research (see [88]) has extended the results of the consistency of MBB for the overall mean of the time series into the consistency of MBB for the coefficient of Fourier representation of mean and autocovariance function.

5.2.2 The Circular Block Bootstrap

Described in this section the circular block bootstrap (CBB) first was introduced by Politis and Romano [79]. The CBB is a modification of the MBB procedure. The idea of this method is to "wrap" the data around in a circle, which helps to avoid the edge effect.

Let $(X_1, \ldots, X_n)$ be a sample from the periodic time series with period T. We consider periodicity in the sense of distribution or in the sense of moments. The block of the length b is $B(j, b) = (X_j, ..., X_{j+b-1})$. Moreover we assume that $X_{n+j} = X_j$

Algorithm

- Let the i.i.d. random variables $i_1, i_2, ..., i_n$ come from the distribution

$$P(i_j = t) = \frac{1}{n} \; for \; t = 1, ..., n.$$

- To obtain the CBB resample

$$(X_1^*, X_2^*, ..., X_n^*)$$

the blocks $(B(i_1, b), B(i_2, b), ..., B(i_n, b))$ are concatenated.

Consistency of the circular block bootstrap

In the stationary case consistency of the CBB technique for the mean was derived by Politis and Romano [79].

In the paper of Dudek [28] the CBB method has been applied to non-stationary data with the almost periodic structure (APC).

The statistical inference for APC time series is based on the Fourier representation of the mean and the autocovariance functions. The form of the estimators of the Fourier coefficients was defined by Hurd and Leśkow [49], [50].

Dudek in [28] has introduced the circular block bootstrap version of the estimators of the Fourier coefficients for the autocovariance functions. She has proved the consistency of this estimator in the multidimensional case under the same assumptions as Synowiecki in the MBB case. Dudek's approach has helped to construct bootstrap simultaneous confidence intervals for the coefficients of the autocovariance function.

The main idea of resampling methods is to preserve the character of the sample. We want to find the most effective method that will reflect the data structure. Hence, if we deal with the periodic data we want to get periodicity in the replicated sample. Unfortunately the MBB procedure destroys the periodic structure of the considered time series.
For the periodic data, using the MBB procedure, we obtain the replication with $E^*X_t^* \neq E^*X_{t+T}^*$. Even if we modify the MBB into the CBB we get only conditional stationarity $E^*X_t^* = \bar{X}_n$.
Below we provide methods that preserve periodicity in replications. These methods are:

- the periodic block bootstrap (PBB)
  where $E^*X_t^* = E^*X_{t+T}^*$,

- the seasonal block bootstrap (SBB)
  where $E^*X_t^* = E^*X_{t+bT}^*$,
- the generalized block bootstrap (GSBB)
  where $E^*X_t^* = E^*X_{t+bT}^*$.

This methods will be described in the next Subsections.

5.2.3 The Periodic Block Bootstrap

In this Section the periodic block bootstrap method will be described. Procedure was introduced by Chan et al. [17] in 2004 and it was dedicated to periodic time series.
The authors proposed dividing the observations into the blocks of length b and resampling these blocks in a way that new periods of observations are generated. Joining together the periods generated in such way we obtain a new series of observations.

Let $(X_1, \ldots, X_n)$ be a sample from the periodic time series with period T. Here, we consider periodicity in the sense of distribution or in the sense of moments. The block of the length b is $B(j, b) = (X_j, ..., X_{j+b-1})$.

Algorithm

- Period T is divided into $L \in \mathbb{N}$ parts of the length b,
  if $T/L \notin \mathbb{N}$ then we leave the last, shorter interval; for simplicity assume that $T = Lb$
- assume that $n = rT,\ r \in \mathbb{N}$
- let us define random variable $i_{u,1}, i_{u,2}, ..., i_{u,r}$ as independent for all u and i.i.d. from the distribution

$$P(i_{u,j} = bu + tT + 1) = \frac{1}{r} \qquad \text{for} \qquad t = 0, ..., r-1,$$

  and $u = 0, ..., L-1$.
- joining the blocks

$$(B(i_{0,1}, b), B(i_{1,1}, b)..., B(i_{L-1,1}, b), B(i_{0,2}, b), B(i-1, 2, b), ...$$

$$..., B(i_{L-1,2}, b), ..., B(i_{0,r}b), B(i_{1,r}, b), ..., B(i_{L-1,r}, b))$$

  we get the resample of the PBB: $(X_1^*, X_2^*, ..., X_T^*, ..., X_n^*)$.

Disadvantage of the PBB is that the procedure is designed for periodic time series that have long periodicities, since it is assumed that the block length b is much smaller compared to the period T.
Notice that the PBB procedures requires knowledge about the length of the period.

Consistency of the periodic block bootstrap

Chan et al. in their research [17] focused on the time series $\{X_t\}_{t\in\mathbb{Z}}$ which is $SP(1)$ and PC with the period T.

In the paper [17] it was assumed that X_t fulfills the following conditions:
The considered time series is $m-$dependent with m much less than T, i.e. $m \ll T$ and the autocovariance function $B(t,\tau)$ is nonnegative.

The bootstrap techniques for dependent data will lead to the problems with consistency, if we do not assume that the block length goes to infinity [71], [61].

Leśkow and Synowiecki [66] , [88] have extended the applicability of the Chan procedure to triangular arrays of periodic random variables, using more general dependence structures - $\alpha-$mixing structure. Moreover, Leśkow and Synowiecki [66] have showed the consistency of the PBB procedure for the mean estimator assuming that the period length T tends to infinity as the sample size n increases. The consequence of this assumption is that for the fixed period the PBB procedure is not consistent.

The assumption that the period length T_n tends to infinity when $n \to \infty$ follows from the justification of the Leśkow and Synowiecki [66] results, which requires fulfilling the following condition:

$$Cov\Big(\frac{1}{\sqrt{b_n}}\sum_{j=ub_n+tT_n+1}^{(u+1)b_n+tT_n} X_j, \frac{1}{\sqrt{b_n}}\sum_{j=(u+1)b_n+tT_n+1}^{(u+2)b_n+tT_n} X_j\Big) \to 0.$$

The condition above holds only if $b_n \to \infty$, which implies $T_n \to \infty$.
Taking into consideration that in many applications we have fixed periodicity the results assuming that $T_n \to \infty$ are too restrictive and do not indicate interesting applications.

5.2.4 The Seasonal Block Bootstrap

In this Subsection the seasonal block bootstrap (SBB) will be briefly described. The SBB method was introduced in 2001 by Politis [77] and that was a version of Künschs Block Bootstrap [61] with blocks whose size and starting points are restricted to be integer multiples of the period T. It is clear that the SBB poses

a restriction on the relative size of the period and block size, i.e. the block size b must be at least of the order of the period T and only integer multiples of T.

Let $(X_1, ..., X_n)$ be the observed sample from the time series and assume that the sample is periodic with the period T, for example in distribution. The length of the block is $b = b_n$. Without a loss of generality we can assume that $n = bTk$, where $k \in \mathbb{N}$.

Algorithm

- Let random variables $i_1, i_2, ..., i_k$ be i.i.d. from a distribution

$$P(i_j = 1 + tT) = \frac{1}{(k-1)b + 1}$$

  for $t = 0, ..., (k-1)b$.

- To obtain the SBB replication $(X_1^*, X_2^*, ..., X_n^*)$ we join the blocks

$$(B(i_1, bT), B(i_2, bT), ..., B(i_k, bT)).$$

The SBB method is the modification of the MBB procedure. The modification consists of taking only the blocks of the length and initial point equal to the multiple of the period. This is related to the disadvantages of the SBB procedure, which are:

- minimal block length is equal to period length,
- block length is always an integer multiple of the period length.

Note that the SBB procedure like the PBB method requires knowledge of the period.

Consistency of the Seasonal Block Bootstrap

Politis [77] has obtained the consistency of the SBB procedure for the special case of PC time series i.e for the model $X_t = f(t) + Z_t$, where f is a periodic function and the time series $\{Z_t\}_{t\in\mathbb{Z}}$ is strictly stationary. Synowiecki [88] has introduced the generalization of this result. He has obtained the consistency the SBB procedure for the mean for the PC and $\alpha-$mixing time series under the following assumptions:

1. there exists a summable sequence $\{c_\tau\}_{\tau=0}^{\infty}$ such that $|Cov(X_t, X_{t+\tau})| < c_\tau$,

2. there exist $\zeta > 0$ such that

$$\sup_{t=1,...,n-b+1} E\Big|\frac{1}{\sqrt{b}}\sum_{j=t}^{t+b-1}(X_j - EX_j)\Big|^{2+\zeta} \leq K,$$

   where K is constant,

3. the central limit theorem holds

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

   where $= M_t(EX_t)$.

Note that the SBB procedure is consistent, if the block length is Tb, where $b \to \infty$, but $b = o(n)$.

5.2.5 The Generalized Seasonal Block Bootstrap

In the previous sections the MBB, SBB and PBB methods were introduced. This procedures have undergone modifications, improvements and generalizations. An important requirement is that the procedure preserves the periodic structure of the data. Such method is GSBB introduced by Dudek et al. [27] in 2014.
The authors were looking for a new block bootstrap method that is suitable for periodic time series with fixed arbitrary period T. They proposed to divide the series of observations into blocks of desired length, independent from periodicity, and resample these blocks in a way that keeps the periodicity.
The block size in the GSBB procedure is chosen independent from the length of the periodicity. It implies that the method avoids the inconsistency problems of the PBB procedure, and the lack of fine-tuning in block size choice problems of the SBB method.

Let $(X_1, ..., X_n)$ be a sample from the periodic time series with period T. Let $b = b_n$ be the block length and $n = \omega T$, where $\omega \in N$.

Algorithm

- We choose an positive integer block size $b < n$ such that $n = lb$, $l \in \mathbb{N}$.
- For $t = 1, b+1, ..., (l-1)b+1$ we define B_t^* as follows

$$B_t^* = (X_t^*, ..., X_{t+b-1}^*) = (X_{\tau_t}, ..., X_{\tau_{t+b-1}}),$$

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

$$\{t-TR_{1,n}, t-T(R_{1,n}-1), ..., t-2T, t-T, t, t+T, t+2T, ..., t+T(R_{2,n}-1), t+TR_{2,n}\}.$$

Here $R_{1,n} = [(t-1)/T]$ and $R_{2,n} = [(n-b-t)/T]$. Random variables $\tau_1, \tau_2, ..., \tau_l$ are i.i.d. from the distribution

$$P(\tau_\omega = 1 + (\omega - 1)b + tT) = \frac{1}{\omega}, \qquad t = 0, ..., l-1.$$

Here τ_t is the beginning of the block B_t^*, and it is restricted to be randomly chosen from a set containing only periodic shifts of t.

- Joining $l+1$ blocks $(X_{\tau_t}, X_{\tau_t+1}, ..., X_{\tau_t+b-1})$ we get the bootstrap sample $(X_1^*, X_2^*, ..., X_{(l+1)b}^*)$.

  The first n points $X_1^*, X_2^*, ..., X_n^*$ are retained and this implies that the bootstrap series has the same length as the original one. If n is an integer multiple of b, then the whole last block is superfluous.

One may notice that if $d = \tau b$, $\tau > 1$, $\tau \in \mathbb{N}$ the GSBB procedure is identical to the PBB, if $b = \tau d$, $\tau \in \mathbb{N}$ the GSBB is identical to the SBB and if $d = 1$ the GSBB is identical to the MBB.

Consistency of the GSBB procedure

The characteristics of interest are seasonal means and the overall mean. Let us define the estimators of seasonal means $\mu_i,\ i = 1, ..., T$ and overall mean $\bar{\mu} = \frac{1}{T}\sum_{i=1}^{T}\mu_i$ as follows

$$\hat{\mu}_i = \frac{1}{\omega}\sum_{j=0}^{\omega-1} X_{i+jT},$$

$$\hat{\mu} = \frac{1}{T}\sum_{i=1}^{T}\hat{\mu}_i,$$

where ω is the number of observed periods.

Dudek et al. [27] obtained consistenc theorem for the characteristics generated by the GSBB procedure. The results are quoted below.

Theorem 5.2. *([27]) Let us take $\delta > 0$, $\sup_t E|X_t|^{4+\delta} < \infty$ and $\sum_{\tau=1}^{\infty}\tau\alpha_X^{\delta/(4+\delta)}(\tau) < \infty$, where $\alpha_Y(\tau)$ is the strong mixing coefficients for stationary series $\{X_t\}$. If $b \to \infty$ as $n \to \infty$, but with $b = o(n)$, then*

$$\sup_{x\in\mathbb{R}} |P(\sqrt{n}(\hat{\bar{\mu}} - \bar{\mu}) \le x) - P^*(\sqrt{n}(\hat{\bar{\mu}}^* - E^*\hat{\bar{\mu}}^*) \le x)| \xrightarrow{P} 0,$$

Moreover

$$\sup_t \left| P(\sqrt{\omega} \max_i |\hat{\mu}_i - \mu_i| \leq t) - P^*(\sqrt{\omega} \max_i |\hat{\mu}_i^* - E^*\hat{\mu}_i^*| \leq t) \right| \xrightarrow{p} 0.$$

where $\hat{\mu}_i^* = \frac{1}{\omega} \sum_{j=0}^{\omega-1} X_{i+jd}^*$ *and* $\hat{\bar{\mu}}^* = \frac{1}{T} \sum_{i=1}^{T} \hat{\mu}_i^*$.

Theorem 5.3. *([27]) Let us take* $\delta > 0$, $\sup_t E|X_t|^{4+\delta} < \infty$ *and* $\sum_{\tau=1}^{\infty} \tau \alpha_X^{\delta/(4+\delta)}(\tau) < \infty$, *where* $\alpha_Y(\tau)$ *is the strong mixing coefficients for stationary series* $\{X_t\}$. *If* $b \to \infty$ *as* $n \to \infty$,*but with* $b = o(n)$, *then*

$$\sup_{x \in \mathbb{R}^d} \left| P(\sqrt{\omega}(\hat{\mu} - \mu) \leq x) - P^*(\sqrt{\omega}(\hat{\mu}^* - E^*\hat{\mu}^*) \leq x) \right| \xrightarrow{p} 0$$

where $\hat{\mu}_i^* = \frac{1}{\omega} \sum_{j=0}^{\omega-1} X_{i+jd}^*$ *and* $\mu = (\mu_1, ..., \mu_T)$ *is the column vector.*

Notice that assumptions in two above theorems are stronger than the standard conditions for the mixing moments.
It is also worth noting that the GSBB procedure preserves the periodic structure of the original data.
The authors have also got the results for the second ordered statistics, for more informations reader is referred to [27].

5.2.6 Subsampling

In this subsection the subsampling method will be analyzed. The procedure was introduced by Politis in 1994. The full description of the method and its consistency is contained in the monograph by Politis et al. [80].
There is a fundamental difference between the methods like GSBB, MBB, CBB, PBB on one hand and subsampling on the other hand. In subsampling there is no post-sample randomization before recalculation of the estimator. Instead, the window of the length b moves along the initial sample and so the repeated values of the estimator are obtained.

Algorithm

Let $(X_1, ..., X_n)$ be the observed sample.

- The statistic $\vartheta_n(\hat{\theta}_n - \theta)$ is recomputed over "short", overlapping blocks of length b (b depends on $n-$the length of the sample)
- $n - b + 1$ statistics are obtained: $\vartheta_b(\hat{\theta}_{n,b,t} - \hat{\theta}_n)$ where $\hat{\theta}_{n,b,t}$ is subsampling version of the estimator $\hat{\theta}_n$ calculated using $(X_t, ..., X_{t+b-1})$

- then empirical distributions:

$$L_{n,b}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} 1_{\{\vartheta_b(\hat{\theta}_{n,b,t} - \hat{\theta}_n) \leq x\}}$$

  are used to approximate the asymptotic distribution of the estimator $\vartheta_n(\hat{\theta}_n - \theta)$

The subsampling procedure is very general. The advantage of it is its insensitivity to the form of the asymptotic distribution. We do not need to know the form of the distribution, we only need to know that it exists and that is non degenerate.

Note that the subsampling procedure does not require knowledge of the length of the period.

Consistency of the subsampling procedure

Following the [80] we provide here the meaning of consistency for subsampling.
Denote the asymptotic distribution of $\vartheta_n(\hat{\theta}_n - \theta)$ as J, and the distribution function of this distribution in the point $x \in \mathbb{R}$ as $J(x)$.
The subsampling procedure is consistent if the following conditions hold:

- if x is continuity point of $J(\cdot)$, then

$$L_{n,b}(x) \xrightarrow{P} J(x)$$

- if $J(\cdot)$ is continuous, then

$$\sup_{x \in \mathbb{R}} |L_{n,b}(x) - J(x)| \xrightarrow{P} 0$$

- if $J(\cdot)$ is continuous in the point $c(1-\alpha)$, then

$$P(\vartheta_n(\hat{\theta}_n - \theta) \leq c_{n,b}(1-\alpha)) \to 1-\alpha,$$

  where $\alpha \in (0,1)$ and

$$c_{n,b}(1-\alpha) = \inf\{x : L_{n,b}(x) \geq 1-\alpha\},$$

$$c(1-\alpha) = inf\{x : J(x) \geq 1-\alpha\}.$$

The latest research on subsampling method for $\alpha-$mixing non-stationary time series with the moments of order more than two is e.g. in Lahiri [62], Synowiecki

[88]. He has estimated, by the subsampling method, the parameters of the non-stationary time series proposed by Hurd and Leśkow [49], [50] and Dehay and Leśkow [21]. To show consistency Synowiecki has used the general sufficient condition for the non-stationary time series, which was formulated by Politis in Theorem 4.2.1 [80]. Moreover he assumed that the block length goes to infinity: $b = b_n \to \infty$, but $b/n \to 0$.
Synowiecki [88] has introduced modification of Politis's estimator ([80]) and has proved that for this modification consistency is true.

The research on generalization of resampling methods are in progress. It is worth to notice that the subsampling procedure in continuous time was also investigated. For the results for random fields see Bertail, Politis and Rhomari, [10].

For non-stationary stochastic processes with periodic covariance structure the consistency problem of subsampling has been solved by Dehay, Dudek and Leśkow [19]. Moreover the authors have constructed the subsampling-based confidence intervals for the relevant characteristics of considered non-stationary processes. The mentioned article is dedicated to the much wider class than the PS or the PC. It is dedicated to the HAPC processes - harmonizable almost periodically correlated processes. Definition and properties of HAPC can be found in [69], [70], [3].

5.3 Comparison of resampling techniques

In this part of the dissertation a very brief comparison of the resampling methods will be introduced.

Moreover, the justification of the choice of the subsampling method as the estimation technique, in the later part of the thesis will be introduced. We start from the comparison of bootstrap methods.
The PBB method behaves much more stable than the MBB in terms of the length of the block. The difference between the techniques is less evident when the block length is a multiple of the period. In that case the MBB procedure interferes less with the periodicity of the data. If the block length is equal to the period, then the MBB procedure seems to be better. It is associated with overlapping of the blocks in the MBB methods. [88]
The SBB and the MBB methods can be compared only at the points of the exact multiples of the period. At those points, both methods are asymptotically equivalent. [88]

In the GSSB method, the block size is chosen independent from the length of the periodicity. It implies that this method avoids the inconsistency problems of the PBB procedure, and the lack of fine-tuning in block size choice problems of the SBB method.
For the periodic time series the research of Synowiecki [88] have showed that both methods: bootstrap and subsampling are similar. The bootstrap methods works mainly for the statistics with the normal asymptotic distribution. It is worth noting, however, that the subsampling is more effective than the bootstrap in the sense of the sensitivity to the asymptotic distribution of the statistics.
The versatility of the subsampling method can be seen if one deals with the periodic data and additionally with heavy tails and/or long range dependent structure. It will be more fully described in the next section.
Note that the above-described results have been proven under the assumption that the time series is strongly mixing. There is no analogous results for the weak dependent time series so far.

Figure 1 presents a summary of the results achieved so far for the the resampling methods used to $\alpha-$ mixing, periodically correlated time series. The APC means the Almost Periodically Correlated time series [50], which is the wider class than PC series.

5.3.1 Resampling methods for heavy tailed, long-range and weakly dependent data sets

The bootstrap methods give a correct and high quality approximations of considered statistics.
Bertail, in [18], has given the conditions for first-order correct MBB confidence intervals in general spaces for non-i.i.d. sequences. Moreover in the independent identically distributed case and in a strong-mixing setting author has shown that the bootstrap distribution obtained by a modified version of resampling without replacement (as considered by Politis and Romano [78]) leads to second-order correct confidence intervals, if the resampling size is chosen adequately. The results are also extended to a modified version of the MBB for $\alpha-$mixing random fields.

If one wants to estimate the sample mean of independent data with heavy tails, the usual form of the bootstrap does not work. The reason of this fact is that the size of the sample mean is determined by the values of a small number of extreme order statistics. This problem can be solved by using so-called the "subsample bootstrap", where resamples size is much smaller than the original sample [5].

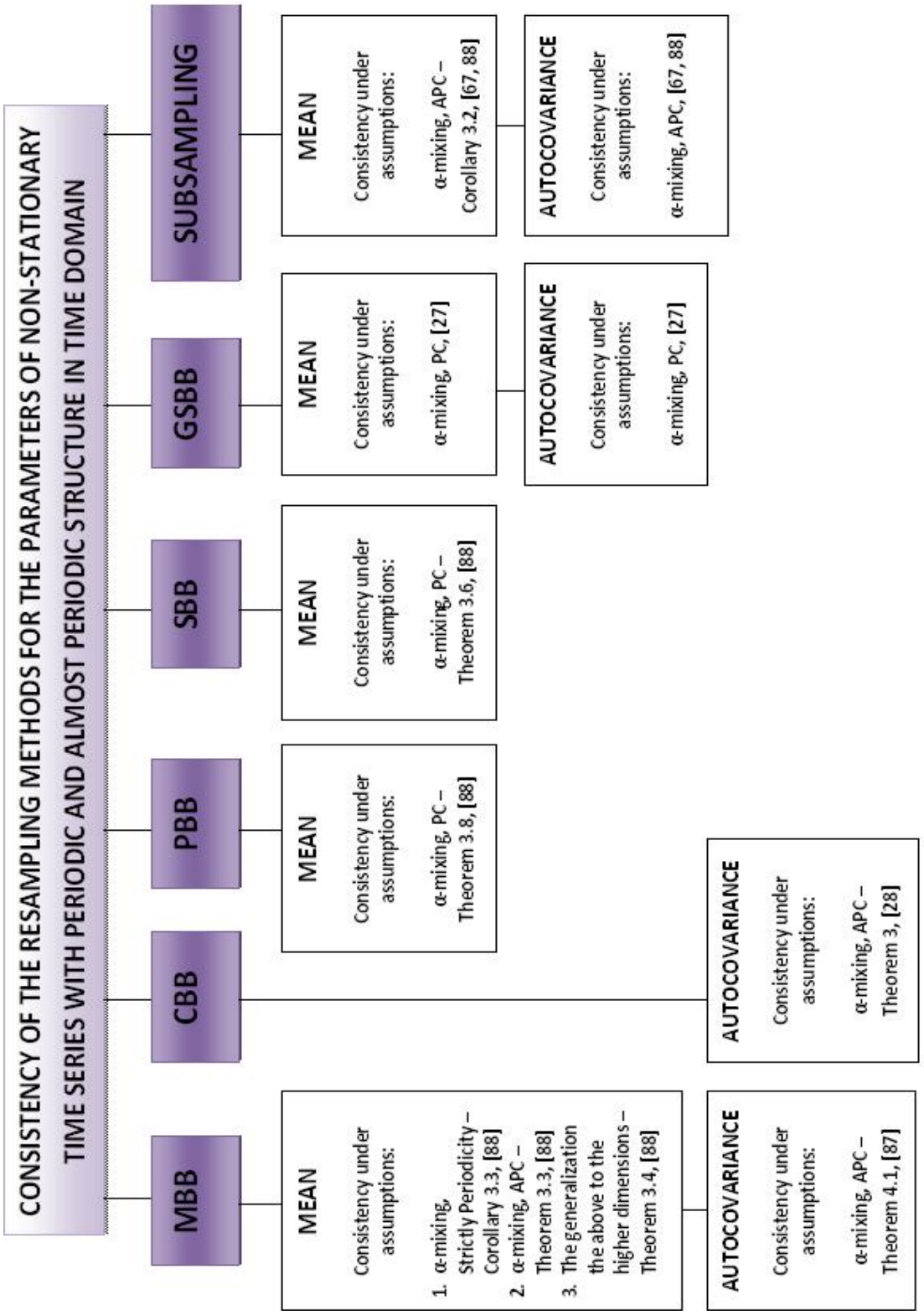

Figure 1: Consistency of the resampling methods for the parameters of non-stationary time series with periodic and almost periodic structure in time domain.

The result does not indicate whether the "subsample bootstrap" provides a more accurate estimation of the mean than more standard, asymptotic methods. Hall et al. [42] have showed that even if the subsample size is chosen optimally, the error between the bootstrap approximation and the true distribution is often larger than that of an asymptotic approximation.
But the authors have introduced a hybrid approach. It is based on a mixture of asymptotic and subsample bootstrap methods and it is shown that it improves standard results.

The resampling method which does work for heavy tailed dependent data, in the stable case, is subsampling. The theorems from the book of Politis, Section 11 [80] show the consistency of the subsampling estimators for the data with heavy tails.

From research of Hall and Lahiri [43], [64] we know that for long-range dependent data the known resampling methods, other than subsampling, do not work. It has been shown by Lahiri [64] that the block bootstrap is not consistent if we deal with long-range dependent sequences. The conclusion of Lahiri's paper [64] is: "the block bootstrap fails to capture the limit law of the normalized sample mean $\bar{X}_n$ of long-range dependent data, whenever $\bar{X}_n$ has a non-normal limit law". The ineffectiveness of block bootstrap methods implies form that the joining independent bootstrap blocks to define the bootstrap sample fails to reproduce the long-range dependence of time series.
Hall et al. [43] discussed the subsampling procedure in the context of long memory data. The authors showed that under some regularity conditions, the subsampling method produces consistent estimators of the distribution of the normalized sample mean in both normal and non-normal limit laws. They introduced a method for studentizing $\bar{X}_n$, and they showed that for this kind of statistic the subsampling method is consistent, also in both normal and non-normal limit laws.

The facts from [43] and [64] are the reasons that in the Section 4 the subsampling, not other resampling method, is used to estimate the sample mean of the observations from the process defined in Section 4.

Doukhan et al. in [22] have investigated properties of the subsampling estimators for distributions of converging and extreme statistics for the stationary time series which are weakly dependent. This results are extensions of the research of Bertail et al. [9] for the strong mixing case. In the article [22] subsampling the distribution of the normalized sample maximum for weakly dependent data is introduced. The authors have considered a sequence of statistics $S_n = s_n(X_1, ..., X_n)$

for $n = 1, 2, ...$. They have assumed that S_n is a sequence of converging statistics in the sense that $\mathbb{K}_n = P(S_n \leq x)$ has a limit $\mathbb{K}$. Let us assume that the statistics satisfy the condition as follows:

Assumption A:

$$r_n = \sup_{x \in \mathbb{R}} |\mathbb{K}_n(x) - \mathbb{K}(x)| \to_{n \to \infty} 0, \qquad \| \mathbb{K}' \|_\infty < \infty. \tag{30}$$

where $\mathbb{K}'$ denotes the density of this limit distribution.

Let us consider the subsampling scheme with overlapping samples for the observations $X_1, ..., X_n$, as follows;

$$Y_{b,i} = (X_{i+1}, ..., X_{i+b}), \quad N = n - b,$$

and let us introduce, quote [22], the so-called rough subsampling estimator for $\mathbb{K}$:

$$\hat{\mathbb{K}}_{b,n}(x) = \frac{1}{N} \sum_{i=0}^{N-1} \mathbb{I}(s_b(Y_{b,i}) \leq x), \qquad \text{rough subsampled statistics,} \tag{31}$$

where b is the length of the subsampling sample and $\lim_{n \to \infty} n/b = \infty$.

Moreover, let us defined the value $Lip(h)$ for the $h : \mathbb{R}^u \to \mathbb{R}$, $(u \in \mathbb{N}^*)$ as follows:

$$Lip(h) = \sup_{(y_1,...,y_u) \neq (x_1,...,x_u) \in \mathbb{R}^u} \frac{|h(y_1, ..., y_u) - h(x_1, ..., x_u)|}{\| y_1 - x_1 \| + ... + \| y_u - x_u \|}.$$

The most interesting theorem from my research viewpoint is theorem quoted below, in which the authors have got convergence of the estimator (31).

Theorem 5.4. *(Theorem 2, [22]) Let the* **Assumption A** *be fulfilled. If the overlapping setting is used and one among the following relations hold*

- $\eta-$*dependence:*

$$\sum_{t=0}^{\infty} \eta(t)^{\frac{1}{2}} < \infty, \qquad \lim_{n \to \infty} \frac{b}{n} (1 \vee \frac{Lip(s_b)}{\sqrt{b}}) = 0$$

- $\lambda-$*dependence:*

$$\sum_{t=0}^{\infty} \lambda(t)^{\frac{2}{3}} < \infty, \qquad \lim_{n \to \infty} \frac{b}{n} (1 \vee (\frac{(Lip(s_b))^4}{b})^{\frac{1}{3}} \vee (\frac{Lip(s_b)}{b})^{\frac{2}{3}}) = 0,$$

then

$$\lim_{n\to\infty} |E[\hat{\mathbb{K}}_{b,n}(x) - E[\hat{\mathbb{K}}_{b,n}(x)]]^2| = 0$$

and

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |\hat{\mathbb{K}}_{b,n}(x) - \mathbb{K}(x)| = 0,$$

in probability.

Let us take a look for a moment on the Lipschitz constant $Lip(s_b)$. Of course its form depends on the statistics $S_b = s_b(X_1, ..., X_b)$. To clarify the concept of $Lip(s_b)$ let us consider the following two examples.

1° The case where $s_b(X_1, ..., X_b)$ is a maximum.
If we consider the sequence of extreme statistics

$$S_b = s_b(X_1, ..., X_b) = max_{1\le i\le b} X_i,$$

then it is easy to calculate that the $Lip(s_b)$ is equal to 1.
Indeed:

$$Lip(s_b) = \sup_{(X_1,...,X_b)\neq(Y_1,...,Y_b)} \frac{\mid max_{1\le i\le b}X_i - max_{1\le i\le b}Y_i \mid}{\sum_{i=1}^{b} \mid X_i - Y_i \mid}$$

Let the $max_{1\le i\le b}X_i = X_{i_\star}$ and $max_{1\le i\le b}Y_i = Y_{j_\star}$ then

$$\frac{|X_{i_\star}-Y_{j_\star}|}{\sum_{i=1}^{b}|X_i-Y_i|} \le \frac{|X_{i_\star}-Y_{j_\star}|}{|X_{i_\star}-Y_{i_\star}|+|X_{j_\star}-Y_{j_\star}|} \le \frac{|X_{i_\star}-Y_{i_\star}|}{|X_{i_\star}-Y_{i_\star}|+|X_{j_\star}-Y_{j_\star}|} \le \frac{|X_{i_\star}-Y_{i_\star}|}{|X_{i_\star}-Y_{i_\star}|} \le 1.$$

2° The case where $s_b(X_1, ..., X_b)$ is a standardized mean.
If we consider the sequence of statistics

$$S_b = s_b(X_1, ..., X_b) = b^{-1/2}\sum_{i=1}^{b}(X_i - EX_i),$$

then $Lip(s_b)$ is equal to $b^{-1/2}$.
Indeed:

$$Lip(s_b) = \sup_{(X_1,...,X_b)\neq(Y_1,...,Y_b)} \frac{\mid b^{-1/2}\sum_{i=1}^{b}(X_i - EX_i) - b^{-1/2}\sum_{i=1}^{b}(Y_i - EY_i) \mid}{\sum_{i=1}^{b} \mid X_i - Y_i \mid}$$

$$\frac{\mid b^{-1/2}\sum_{i=1}^{b}(X_i - EX_i) - b^{-1/2}\sum_{i=1}^{b}(Y_i - EY_i) \mid}{\sum_{i=1}^{b} \mid X_i - Y_i \mid} \le$$

$$\le b^{-1/2}\frac{|\sum_{i=1}^{b} X_i - \sum_{i=1}^{b} Y_i|}{\sum_{i=1}^{b}|X_i - Y - i|} \le b^{-1/2}\frac{|\sum_{i=1}^{b}(X_i - Y_i)|}{|\sum_{i=1}^{b}(X_i - Y_i)|} \le b^{-1/2}.$$

In the articles of Bardet et al. [6] or Doukhan et al. [26] one can find applications of the classical Lindeberg central limit theorem to the class of weakly dependent processes, introduced by Doukhan and Louhichi [25].

Let $\{X_t\}_{t\in\mathbb{N}}$ be a series of zero mean random variables with values in $\mathbb{R}^d$. Let

$$\| X_t \|^2 = X_{t,1}^2 + ... + X_{t,d}^2$$

be the Euclidean norm for $X_t = (X_{t,1}, ..., X_{t,d})$. Let us denote by $\mathcal{C}_b^3$ the set of bounded functions $\mathbb{R}^d \to \mathbb{R}$ with bounded and continuous partial derivatives up to order 3.

Assumption H_δ:

It exists $0 < \delta \leq 1$ such that $\forall t \in \mathbb{N}, \quad E\|X_t\|^{2+\delta} < \infty$ and $\forall k \in \mathbb{N}^*$, define

$$A_k = \sum_{t=1}^{k} E\|X_t\|^{2+\delta}.$$

Theorem 5.5. *(Lindeberg C.L.T., [6]) Assume that the sequence $\{X_{t,k}\}_{t\in\mathbb{N}}$ satisfies Assumption H_δ, and $A_k \to_{k\to\infty} 0$, and there exists Σ a positive matrix such that $\Sigma_k = \sum_{t=1}^{k} Cov(X_{t,k}) \to_{k\to\infty} \Sigma$. Moreover, assume that for $j = 1, 2$ and $f \in \mathcal{C}_b^3$*

$$T_j(k) = \sum_{t=1}^{k} |Cov(f^{(j)}(X_1 + ... + X_{t-1}), X_t^j)| \to_{k\to\infty} 0$$

or

$$T(k) = \sum_{j=1} k|Cov(e^{i<t,X_1+...,X_{j-1}>}, e^{i<t,X_j>})| \to_{k\to\infty} 0.$$

Then,

$$S_k = \sum_{i=1}^{k} X_{i,k} \xrightarrow{d}_{k\to\infty} \mathcal{N}_d(0, \Sigma).$$

The classical Lindeberg theorem above was the base to obtain the CLT for weakly dependent time series. Doukhan et al. [26] have proved a $(2+\delta)-$order moment inequality which implies the conditions $A(k_n) \to 0$ and $T(k_n) \to 0$, when $k_n \to \infty$ and thereby they have obtained the following theorem for weakly dependent series:

Theorem 5.6. *([26]) Let $\{X_t\}_{i\in\mathbb{N}}$ be a sequence of stationary zero mean $(2+\delta)-$order random variables, with $\delta > 0$. Assume that $\{X_t\}_{i\in\mathbb{N}}$ is a $\lambda-$(or $\theta-$)weakly dependent time series satisfying $\lambda_r = O(r^{-c})$ (or $\theta_r = O(r^{-c})$) when $r \to \infty$, with*

$c > 4 + 2/\delta$.
Then it exists $0 < \sigma^2 < \infty$ *such that*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad k \to \infty.$$

It is clear that Theorem 5.6 does not work when we are dealing with long memory time series, in the sense of the Definition 4.5 ($\beta - 1 = c$, where $\beta \in [0, 1)$ is from the Definition 4.5). However, it shows the necessary prerequisites to obtain the central limit theorem in the model considered in the sequel of this dissertation.

The next Section - Section 6 is dedicated to the presentation the main results of the Author's PhD thesis.

6 Subsampling, weak dependence, heavy tails and long memory

6.1 The Model and the Central Limit Theorems

6.1.1 The model of the weakly dependent time series

In many applications of time series analysis one is confronted separately with heavy tailed and long memory behavior. The non-stationarity of the time series, and its special case - the periodicity is also a feature that researchers are dealing with.
Below we present a model which will simultaneously be dealing with three features: periodicity, long memory and heavy tails. We build it by adjusting a long memory and a heavy tailed stationary model to the T-variate process similarly as in [4], [73].

Let the time series $\{X_t\}_{t\in\mathbb{Z}}$ be defined as:

$$X_t = \sigma_t GG_t + \eta_t, \tag{32}$$

where

A1 The volatility time series σ_t and the Gaussian-Gegenbauer time series GG_t are independent

A2 The sequence of random variables σ_t is i.i.d and its marginal distribution comes either from a stable family or a GED family.

A3 GG_t is periodic Gaussian-Gegenbauer time series. We put that $GG_t = f_t \cdot G_t$, where G_t is Gaussian-Gegenbauer mean zero time series with $k = 1$, $\mid \nu \mid \leq 1$, $LM(\beta)$ with $\beta \in [0, 1)$. The function f_t is a periodic, deterministic, bounded with a known period T. The autocovariance of G_t is γ_G.

A4 The deterministic function η_t is periodic with the same period T as f_t.

Comment 6.1. $\beta = 2d$, *where d is a memory parameter from the Definition 4.7.*

The stationary case of the model (32) was considered by Politis and McElroy in [75].

6.1.2 Properties of the model

Fact 6.1. *The process* $\{X_t\}_{i\in\mathbb{Z}}$ *defined by the equation (32) is a long memory process in the sense of Definition 4.6, with* $\beta \in [0,1)$.

Indeed:

$$\sum_{0<|h|<n} \gamma(s)(h) = \sum_{0<|h|<n} (E\sigma)^2 f_h^2 \gamma_G(h) \sim C(s) n^{\beta}.$$

The last asymptotic equivalence follows from Theorem 5.4. □

Lemma 6.1. *([41]) The long memory stationary Gaussian - Gegenbauer time series is not strong mixing.*

In the proof of Lemma 6.1 authors have used the property of the Gaussian time series. They have studied the relationship as follows [60]:

$$\rho(k) = r(k), \quad \alpha(k) \leq r(k) \leq 2\pi\alpha(k),$$

between the coefficients ρ, α (not to be confused with a stable coefficient α) and r related to completely linear regular condition, $\alpha-$ mixing condition and completely regular condition, respectively. From [41] the 1-factor Gegenbauer time series is not completely linearly regular (see 4.10), hence it is not completely regular (see, Definition 4.9). And hence it follows that the 1-factor Gaussian - Gegenbauer time series can not be strong mixing.

Fact 6.2. *Assume* **A1** *through* **A4**. *Then* X_t *defined by the equation (32) is* $\lambda-$*weakly dependent.*

Proof of Fact 6.2

It follows from Lemma 6.1 that the Gaussian - Gegenbauer time series $G_{tt\in\mathbb{Z}}$, is not strong mixing.

From Theorem 4.1 we know that Gegenbauer (in the sense of the Definition 4.7) time series has a long memory. And from the [6], p.8 follows that the stationary Gaussian long memory time series has the $\lambda-$weak dependence properties.

Finally the $\lambda-$weak dependency of X_t is implied from the Proposition 1 in [53]. □

Fact 6.3. *Assume* **A1** *through* **A4**. *Then the weak dependence coefficients of the model defined by the equation (32) satisfy the following relationship:*

$$\lambda_r = O(r^{\beta-1}), \quad \beta \in [0,1).$$

Proof of above fact follows from [6].

The Fact 6.2 provide clear motivation to study weakly dependent structures. In the next Subsection we will consider the model (32) with two different heavy tails cases of volatility.

Stable volatility

Define the volatility process in the model (32) as follows:

$$\sigma_t = \sqrt{\epsilon_t}, \tag{33}$$

where ϵ_t are i.i.d. $\alpha/2$-stable, $\alpha \in (1,2)$, with the skewness parameter equal to one, the scale parameter equal to $(cos(\pi\alpha/4))^{2/\alpha}$, and the location parameter $E\epsilon_t = 0$, for each t. From the research of Taqqu ([89], Prop.1.2.16 and Prop.1.2.17), we know that $E(\sigma) = E(\sigma_t)$ is finite and non negative.
The construction of X_t is based on sub-Gaussian vector [89], p.77.

The following fact provides the information about the marginal distribution of X_t :

Fact 6.4. *Assume* **A1** *through* **A4**. *Then X_t has a symmetric about the mean $f_t E(\sigma)$ $-$stable marginal distribution with scale parameter*

$$\tau(X_t) = |f_t|\sqrt{\gamma_G(0)/2}.$$

The proof of the Fact 6.4 follows from the results of Taqqu ([89], Prop.1.2.3 and Prop. 1.3.1).
If the volatility in model (32) is stable with the stability coefficient $\alpha \in (1,2)$ then the time series is $WS(1)$. Indeed, the first moment of X_t is finite (Proposition 1.2.16, [89]) and periodic (from the construction of the model (32)).
The second moment of the time series X_t is infinite (Proposition 1.2.16, [89]), hence it is not second-order process. Thus the model (32) is not periodically correlated in the sense of Gladyshev but it still is periodically stationary (PS) as stated below.

Comment 6.2. *The X_t defined by the equation (32) does not the have the variance, but it still has finite periodic autocovariance function $\gamma(t+T,h) < \infty$ for $h \neq 0$, and is periodically stationary (PS).*

Indeed, the time series X_t has periodic covariance:

$$\gamma(t+T,h) = f_{t+T} f_{t+T+h} (E\sigma)^2 \gamma_G(t+T,h) =$$

$$= f_{t+T}f_{t+T+h}(E\sigma)^2\gamma_G(h) = f_t f_{t+h}(E\sigma)^2\gamma_G(t,h) = \gamma(t,h).$$

And it is PS, since $\sigma_t G_t$ is strictly stationary and the function f is periodic:

$$(\sigma_{t_1}G_{t_1}, \sigma_{t_2}G_{t_2}, ..., \sigma_{t_n}G_{t_n}) \stackrel{d}{=} (\sigma_{t_1+T}G_{t_1+T}, \sigma_{t_2+T}G_{t_2+T}, ..., \sigma_{t_n+T}G_{t_n+T})$$

$$(f_{t_1}\sigma_{t_1}G_{t_1}, f_{t_2}\sigma_{t_2}G_{t_2}, ..., f_{t_n}\sigma_{t_n}G_{t_n}) \stackrel{d}{=}$$

$$(f_{t_1}\sigma_{t_1+T}G_{t_1+T}, f_{t_2}\sigma_{t_2+T}G_{t_2+T}, ..., f_{t_n}\sigma_{t_n+T}G_{t_n+T})$$

$$(f_{t_1}\sigma_{t_1}G_{t_1}, f_{t_2}\sigma_{t_2}G_{t_2}, ..., f_{t_n}\sigma_{t_n}G_{t_n}) \stackrel{d}{=}$$

$$(f_{t_1+T}\sigma_{t_1+T}G_{t_1+T}, f_{t_2+T}\sigma_{t_2+T}G_{t_2+T}, ..., f_{t_n+T}\sigma_{t_n+T}G_{t_n+T}).$$

In such a way we have succeeded in constructing a heavy tailed, long memory, weakly dependent and not α-mixing time series with periodic structure.

Another interesting case of heavy tailed distributions is provided by the GED family. We focus on that case in the subsection below.

GED volatility

Assume that the volatility process σ_t in the model (32) comes from a GED distribution, as in Definition 4.22, i.e. $\sigma_t \sim G(\mu, 1, \alpha)$).

The following fact provides the information about the heaviness of X_t tails in the GED case:

Fact 6.5. *Assume* **A1** *through* **A4**. *Then X_t defined by the equation (32) has a heavy tailed marginal distribution for $\alpha > 0$.*

Proof of the Fact 6.5.
From direct calculations we can obtain the strict formula for the kurtosis of the model X_t, which is:

$$\frac{E(X_t - EX_t)^4}{(E(X_t - EX_t)^2)^2} = \frac{E\sigma_t^4 EG_t^4}{(E\sigma_t^2)^2(EG_t^2)^2} = 3\frac{\Gamma(5/\alpha)\Gamma(1/\alpha)}{\Gamma^2(3/\alpha)}.$$

If we use the Stirling's formula for Γ function we will obtain the approximation as follows:

$$kurtosis \approx 3 \cdot 1.4 \cdot 4.3^{1/\alpha}.$$

The kurtosis is more than 3 for all $\alpha > 0$. □

We do not have information about the marginal distribution of X_t, but since all the moments of both distribution GG_t and σ_t exist the model X_t is $(2+\delta)$-order, with $\delta > 0$.

In particular X_t is second-order, so it is periodically correlated in the sense of Gladyshev. The formalization of this statement is as follows:

Comment 6.3. *X_t defined by the equation (32) with σ_t coming from the GED has a periodic mean, a periodic variance and a periodic autocovariance. Moreover, the autocovariance of X_t has a form:*

$$\gamma(t,h) = (Cov(\sigma_t, \sigma_{t+h}) + \varphi^2)|f_t||f_{t+h}|\gamma_G(h).$$

Indeed:

The mean of X_t is η_t, so it is periodic. The variance is periodic:

$$\gamma(t+T,0) = (f_{t+T})^2(1+\varphi^2)\gamma_G(t+T,0) =$$

$$= (f_{t+T})^2(1+\varphi^2)\gamma_G(0) = (f_t)^2(1+\varphi^2)\gamma_G(t,0) = \gamma(t,0).$$

The autocovariance also:

$$\gamma(t+T,h) = |f_{t+T}f_{t+T+h}|\varphi^2\gamma_G(t+T,h) =$$

$$= |f_{t+T}f_{t+T+h}|\varphi^2\gamma_G(h) = |f_tf_{t+h}|\varphi^2\gamma_G(t,h) = \gamma(t,h).$$

The form of the variance and autocovariance follows from the form of the variance of variable with the GED distribution. □

6.1.3 The estimator and its properties

For the model defined by the equation (32) one of the resampling method - subsampling - is considered to approximate an asymptotic distribution of the seasonal trend components, the overall mean and the vector of the seasonal trend components.

We start with the definition.

Definition 6.1. *We define the estimator of the seasonal trend components $\eta(s)$ as follows:*

$$\hat{\eta}_N(s) = \frac{1}{N}\sum_{p=0}^{N-1} X_{s+pT}, \quad s = 1, 2, \ldots, T, \tag{34}$$

where T is the known period.

Definition 6.2. *We define the estimator of the overall mean* $\bar{\eta} = \frac{1}{T}\sum_{s=1}^{T}\eta(s)$ *as follows:*

$$\hat{\bar{\eta}} = \frac{1}{T}\sum_{s=1}^{T}\hat{\eta}_N(s) \tag{35}$$

where T *is the known period.*

For the case of stable heavy tails define $\zeta = max\{1/\alpha, (\beta+1)/2\}$, where α is the heavy tails parameter and β is a long memory parameter. For the case of the GED model define $\zeta = 1/2$.

6.1.4 Central Limit Theorems in the stable case

Let us define:

$$A_N(s) = N^{1-\zeta}(\hat{\eta}_N(s) - \eta(s)).$$

Theorem 6.1. *(Central Limit Theorem - for the seasonal means)*
Assume **A1** *through* **A4** *and the volatility process* σ_t *is defined by the equation (33). Then the following weak convergence holds:*

$$A_N(s) \stackrel{d}{\Rightarrow} \begin{cases} S(s), & if\ 1/\alpha > (\beta+1)/2 \\ V(s), & if\ 1/\alpha < (\beta+1)/2 \\ S(s)+V(s), & if\ 1/\alpha = (\beta+1)/2. \end{cases} \tag{36}$$

The variables $S(s)$ *and* $V(s)$ *are independent. Here* $S(s)$ *is a* $S\alpha S$ *variable* ($\alpha-$*stable, symmetric around zero), and the scale parameter equal to* $|f_s|\sqrt{\gamma_G(0)/2}$, $s = 1, \ldots, T$. *Moreover,* $V(s)$ *is a mean zero Gaussian variable with variance* $\tilde{C}(s)(E\sigma)^2/(\beta+1)$, *where* $\tilde{C}(s) = |f_t|(C(s) - \gamma_G(0)\mathbb{I}_{\{\beta=0\}})$ *and* $C(s)$ *is the constant from the Definition 4.6.*

Proof of Theorem 6.1
The proof of the Theorem 6.1 follows from the Theorem 1 from [75] and Lemma 4.5. Assume that we are in the space $\mathbb{L}^1$.
Let $\mathcal{E}$ and $\mathcal{G}$ be the $\sigma-$field defined as follows $\mathcal{E} = \sigma(\epsilon) = \sigma(\epsilon_t, t \in \mathbb{Z})$, $\mathcal{G} = \sigma(G) = \sigma(G_t, t \in \mathbb{Z})$ respectively. From the assumption **A1** $\mathcal{E}$ and $\mathcal{G}$ are independent with respect to the probability measure P.
Let us assume that $\beta = 0$, it follows $\zeta = 1/\alpha$.

From Theorem 4 in [54], p. 132 the characteristic function of $A_N(s)$ can be written as (similarly as in [89], p. 20)

$$Eexp\{i\nu N^{-1/\alpha}\sum_{p=0}^{N-1} Y_{s+pT}\} = E[E[exp\{i\nu N^{-1/\alpha}\sum_{p=0}^{N-1}\sigma_{s+pT}GG_{s+pT}\}|\mathcal{E}]]$$

where ν is any real number and $s = 1, 2, ..., T$. Let us investigate the inner conditional characteristic function. From the properties of Gaussian characteristic function we get that

$$E[exp\{i\nu N^{-1/\alpha}\sum_{p=0}^{N-1}\sigma_{s+pT}GG_{s+pT}\}|\mathcal{E}] =$$

$$= exp\{-\frac{(\nu N^{-1/\alpha})^2}{2} f_s^2 \sum_{p,q=0}^{N-1}\sigma_{s+pT}\sigma_{s+qT}\gamma_G(T(p-q))\},\ s = 1, ..., T.$$

The double sum is divided into the diagonal and the off-diagonal terms:

$$N^{-\frac{2}{\alpha}}\big(f_s^2\sum_{p=0}^{N-1}\sigma_{s+pT}^2\gamma_G(0) + f_s^2\sum_{p\neq q}\sigma_{s+pT}\sigma_{s+qT}\gamma_G((p-q)T)\big) \tag{37}$$

The second part of (37) from the Markov inequality tends to 0 in probability as $N \to \infty$ and hence, as it is known, in distributions. Indeed:

$$E|N^{-\frac{2}{\alpha}} f_s^2\sum_{p\neq q}\sigma_{s+pT}\sigma_{s+qT}\gamma_G((p-q)T)| \leq N^{-\frac{2}{\alpha}} f_s^2 (E(\sigma))^2\sum_{p\neq q}|\gamma_G((p-q)T)|$$

$$\leq N^{1-\frac{2}{\alpha}} f_s^2 (E(\sigma))^2\sum_{h\leq N}(1-\frac{|h|}{N})|\gamma_G(h)|.$$

For the sum, the assumptions of the dominated convergence principle ([31], p. 111) are satisfied. Hence the sum tends to $\sum_{h\in\mathbb{Z}}|\gamma_G(h)|$, which, from the condition $LM(\beta)$ (Definition 4.5), for $\beta = 0$ is finite. Consequently

$$N^{1-\frac{2}{\alpha}} f_s^2 (E(\sigma))^2\sum_{h\leq N}(1-\frac{|h|}{N})|\gamma_G(h)| \to 0,$$

as $N \to \infty$, since $\alpha < 2$ and function f is bounded.

Stability of the σ_t makes the distribution of the first part of the (37) stable. Due to the boundedness of $\exp\{-\nu^2/2\cdot\}$ the assumption of the Theorem 25.8 in [14] are satisfied and hence we have the weak convergence result:

$$Eexp\{i\nu N^{-\frac{1}{\alpha}} f_s^2\sum_{p=0}^{N-1}\sigma_{s+pT}^2\gamma_G(0)\} \to Eexp\{-\frac{\nu^2}{2} f_s^2\gamma_G(0)\epsilon\}.$$

The ϵ has the same distribution as ϵ_t. Therefore the expression
$Eexp\{-\frac{\nu^2}{2}f_s\gamma_G(0)\epsilon\}$ is the characteristic function of S variable with the scale parameter $|f_s|\sqrt{\gamma_G(0)/2}$.

In the case $1/\alpha > (\beta+1)/2$ the second term of (37) is $O(N^{1-2/\alpha}N^{\beta})$ which tends to zero as $N \to \infty$ and f is bounded. The rest of the proof is identical as in the case $\beta = 0$.
In the case $1/\alpha < (\beta+1)/2$ the formula (37) becomes

$$N^{-(\beta+1)}\big(f_s^2\sum_{p=0}^{N-1}\sigma_{s+pT}^2\gamma_G(0) + f_s^2\sum_{q\neq p}\sigma_{s+pT}\sigma_{s+qT}\gamma_G((p-q)T)\big). \tag{38}$$

The first term is $O_P(N^{2/\alpha-(\beta+1)})$ and tends to zero as $N \to \infty$ and f is bounded. From the Lemma 1, [75] and Lemma 3.1, [89] the limiting characteristic function of the second term is

$$Eexp\{-\frac{\nu^2}{2}f_s^2\frac{\tilde{C}(t)(E\sigma)^2}{\beta+1}\} = exp\{-\frac{\nu^2}{2}f_s^2\frac{\tilde{C}(s)(E\sigma)^2}{\beta+1}\}$$

which is characteristic function of a mean zero Gaussian with variance
$f_s^2\tilde{C}(s)(E\sigma)^2/(\beta+1)$.
The case $1/\alpha = (\beta+1)/2$ is the combination of the two above cases. From the Slutsky's Theorem we get the weak convergence of the sum of two independent random variables. The characteristic function is in the form: $Eexp\{-\frac{\nu^2}{2}\big(f_s^2\gamma_G(0)\epsilon + f_s^2\tilde{C}(t)(E\sigma)^2/(\beta+1)\big)\} = exp\{-|\nu|^{\alpha}(f_s^2\gamma_G(0)/2)^{\alpha/2}\}\cdot exp\{-\frac{\nu^2}{2}\frac{f_s^2\tilde{C}(s)(E\sigma)^2}{\beta+1}\}$, and indeed this is a characteristic function of the sum of a stable S and a stable/normal V variables.

The similar proof can also be found in Gajecka-Mirek [34]. □

Let us define

$$A = n^{1-\zeta}(\hat{\bar{\eta}} - \bar{\eta}).$$

The following Theorem holds for the term A.

Theorem 6.2. *(Central Limit Theorem - for the overall mean)*
Assume **A1** *through* **A4** *and the volatility process σ_t is defined by the equation (33). Then the following weak convergence holds:*

$$A \stackrel{d}{\Rightarrow} \begin{cases} S & if\ 1/\alpha > (\beta+1)/2 \\ V & if\ 1/\alpha < (\beta+1)/2 \\ S+V & if\ 1/\alpha = (\beta+1)/2. \end{cases} \tag{39}$$

The variables S and V are independent. Here S is a $S\alpha S$ variable with the scale parameter

$$(\sum_{s=1}^{T} |f_s|^{\alpha})^{1/\alpha} \sqrt{\gamma_G(0)/2}.$$

Moreover S is a sum of T $S\alpha S$ independent random variables as follows: $S = \sum_{s=1}^{T} S(s)$, V is a mean zero Gaussian variable with variance $\tilde{C}(E\sigma)^2/(\beta+1)$, where $\tilde{C} = \sum_{s=1}^{T} f_s^2(C - \gamma_G(0)\mathbb{I}_{\{\beta=0\}})$.

Proof of Theorem 6.2
From the construction of the model X_t defined by the equation (32) it is enough to repeat the method shown in the Theorem 6.1 together with Lemma 4.4 and Lemma 4.5. Note that f is assumed to be bounded. □

Let us define

$$A_N = N^{1-\zeta}(\hat{\eta} - \eta).$$

The following Theorem holds for A_N.

Theorem 6.3. *(Central Limit Theorem - for the mean's vector)*
Assume **A1** *through* **A4** *and the volatility process σ_t is defined by the equation (33). Then the following weak convergence holds:*

$$A_N \stackrel{d}{\Rightarrow} \begin{cases} S_N & if\ 1/\alpha > (\beta+1)/2 \\ V_N & if\ 1/\alpha < (\beta+1)/2 \\ S_N + V_N & if\ 1/\alpha = (\beta+1)/2. \end{cases} \tag{40}$$

The vectors S_N and V_N are independent. Here S_N is a $S\alpha S$ vector with zero location parameter, and scale vector

$$\sqrt{\gamma_G(0)/2}[|f_1|, ..., |f_T|].$$

V_N is a mean zero Gaussian variable with variance $\tilde{C}(E\sigma)^2/(\beta+1)$, where $\tilde{C}(s) = [f_1^2, ..., f_T^2](C - \gamma_G(0)\mathbb{I}_{\{\beta=0\}})$.

Proof of Theorem 6.3
The proof of the Theorem 6.3 follows from the Cramer-Wald Theorem together with the Theorem 6.2. S_N is a S vector from the Theorem 2.1.5, p.59, [89]. □

6.1.5 Central Limit Theorems in the GED case

In this subsection, the central limit theorems for the seasonal and overall mean for the variables from the GED are introduced.

Let us define $\Sigma_s = Var(X_s)$. In the GED case $\Sigma_s = |f_s|^2\gamma_G(h), \quad s = 1, ..., T$. Let us define:

$$B_N(s) = N^{-1/2}(\hat{\eta}_N(s) - \eta(s)).$$

For a sequence $\{m_N\}_{N\in\mathbb{N}}$ such that $m_N \to \infty,\ if\ N \to \infty$ let us define $\{k_N\}_{N\in\mathbb{N}}$ such that

$$k_N = [\frac{N}{m_N}] \to \infty,\ if\ N \to \infty$$

and for $\lambda-$weakly dependent model $\{X_t\}_{t\in\mathbb{Z}}$ defined by the equation (32) following condition holds

$$\lambda_{m_N} k_N^{\frac{3}{2}} \to 0, \qquad N \to \infty.$$

Let us consider a subsample $(X_{m_N}, \ldots, X_{k_N m_N})$ of $(X_s, \ldots, X_{s+(N-1)T})$.

Theorem 6.4. *(Central Limit Theorem - for the seasonal means)*
Assume **A1** *through* **A4** *and the volatility process σ_t is as in the Definition 4.22. Then, for a sequence $\{m_N\}_{N\in\mathbb{N}}$ such that $m_N \to \infty$ and $k_N = [\frac{N}{m_N}] \to \infty,\ if\ N \to \infty$ and for each $s = 1, \ldots, T$ following convergence holds:*

$$B_{k_N}(s) \to \mathcal{N}(0, \Sigma_s), \qquad N \to \infty.$$

Proof of Theorem 6.4
The proof follows from Proposition 4.1, [6]. □

Since the model defined by the equation (32) satisfies property $LM(\beta)$ (Definition 4.5), for $\beta \in [0, 1)$ the time series X_t defined by the equation (32) does not satisfy the Central Limit Theorem 5.6. The reason for this fact is that it has the long memory property. But the Central Limit Theorem will be satisfied if we choose a subsample of the observation with the appropriate asymptotic step of sampling. In the model (32) the subsampled time series $\{Y_{sm_N} = X_{sm_N} - \eta_{sm_N}\}$, with a "subsampling" step m_N such that

$$o(m_N) = N^{3/(2\beta+1)}, \tag{41}$$

satisfies the Central Limit Theorem with a convergence rate $o(N^{(1-\beta)/(4\beta+2)})$.
Following [6] there are objections to this "subsampling" method: only a part of the sample is used and the choice of the convergence rate of the "subsampling" implies the knowledge of the convergence rate of λ_r. But the convergence rate of λ_r, in long memory processes is connected with the the long memory parameter β. It could give us a step of "subsampling". The problem of the estimation of the long memory parameter β will be discussed in the next Section.

Let us define:

$$B_{k_N,N} = k_N^{-1/2}(\hat{\eta} - \eta).$$

Theorem 6.5. *(Central Limit Theorem - for the vector of the means) Assume* **A1** *through* **A4**. *Moreover, assume that*

$$\lambda_{m_N} k_N^{\frac{3}{2}} \to 0, \qquad N \to \infty.$$

Let $Z_N = (Y_{s+pT}, ..., Y_{s+(p+1)T})$, $p = 0, ..., N-1$ *be a sequence of zero mean random variables with values in* $\mathbb{R}^T$, *where* T *is the period. Then, for a sequence* $\{m_N\}_{N\in\mathbb{N}}$ *such that* $m_N \to \infty$ *and* $k_N = [\frac{N}{m_N}] \to \infty$, $if\ N \to \infty$

$$B_{k_N,N} = \frac{1}{\sqrt{k_N}} \sum_{i=1}^{k_N} Z_{im_N} \to \mathcal{N}_T(0, Cov(X_0)), \qquad N \to \infty.$$

Proof of Theorem 6.5
The proof of Theorem 6.5 implies from the Proposition 4.1, [6], but applied to the vectors. □

Let us define:

$$B = n^{-1/2}(\hat{\bar{\eta}} - \bar{\eta}).$$

Theorem 6.6. *(Central Limit Theorem - for the overall mean) Assume* **A1** *through* **A4**. *Moreover assume that*

$$\lambda_{m_N} k_N^{\frac{3}{2}} \to 0, \qquad N \to \infty.$$

Then, for a sequence $\{m_N\}_{N\in\mathbb{N}}$ *such that* $m_N \to \infty$ *and* $k_N = [\frac{N}{m_N}] \to \infty$, $if\ N \to \infty$

$$B = \frac{1}{\sqrt{k_N T}} \sum_{s=1}^{T} (\sum_{p=0}^{k_N-1} (X_{sm_N+pT} - \eta(s))) \to \mathcal{N}(0, \Sigma), \qquad N \to \infty.$$

Proof of the Theorem 6.6
The proof of the Theorem 6.6 implies from the Proposition 4.1, [6] for the vectors and the Cramer - Wald theorem. □

6.2 Consistency of the subsampling method for the mean

From research of Hall and Lahiri ([43], [64]) we know that, for long-range dependent process, the bootstrap (MBB, CBB, PBB, SBB, GSBB) do not work,

whereas subsampling still works asymptotically.

All we need to know to use subsampling is if there exists a non-degenerated asymptotic distribution of the statistic (we do not have to know the form of the asymptotic distribution) - we need to have the Central Limit Theorem.
Recall that our sample size n is equal $n = NT$.
The idea of subsampling in our model for the seasonal components is as follows:

Step 1 For each $s = 1, \ldots, T$ the estimator $\hat{\eta}_N(s)$ is recomputed from the $(X_s, \ldots, X_{s+(N-1)T})$ over "short" overlapping blocks of length b_s (b_s depends on N-the length of the sample)

Step 2 From Step 1 $N - b_s + 1$ statistics are obtained for each s. In our context those will be $a_{b_s}(\hat{\eta}_{N,b_s,i}(s) - \hat{\eta}_N(s))$ where $\hat{\eta}_{N,b_s,i}(s)$ is subsampling version of the estimator $\hat{\eta}_N(s)$ and a_{b_s} is the normalize sequence. In the stable case $a_N = N^{1-\zeta}$ and in the GED case $a_N = N^{-1/2}$.

Step 3 then the empirical distributions:

$$L_{N,b_s}(x,s) = \frac{1}{N - b_s + 1} \sum_{i=1}^{N-b_s+1} 1_{\{a_{b_s}(\hat{\eta}_{N,b_s,i}(s) - \hat{\eta}_N(s)) \leq x\}}$$

are used to approximate the asymptotic distribution $L(s)(x)$ of the estimator $a_N(\hat{\eta}_N(s) - \eta(s))$.

The idea of subsampling in our model for the vector of seasonal components is as follows:

Step 1 For each $s = 1, \ldots, T$ the estimator $\hat{\eta_N}$ is recomputed from the $(X^1, \ldots, X^N)$, where $X^i = (X_{1+(i-1)T}, ..., X_{iT})$ over "short" overlapping blocks of length b (b depends on N-the length of the sample)

Step 2 From Step 1 $N - b + 1$ statistics are obtained. In our context those will be $a_b(\hat{\eta}_{N,b} - \hat{\eta}_N)$ where $\hat{\eta}_{N,b}$ is subsamplig version of the estimator $\hat{\eta}_N$ and a_b is the normalizing sequence.

Step 3 then the empirical distributions: $L_{N,b}(x) = \frac{1}{N-b+1} \sum_{i=1}^{N-b+1} 1_{\{a_b(\hat{\eta}_{N,b} - \hat{\eta}_N) \leq x\}}$ are used to approximate the asymptotic distribution $L(x)$ of the estimator $a_N(\hat{\eta}_N - \eta_N)$.

The main problem with the subsampling procedure is its consistency. We need to prove that the finite sample quantiles generated by the subsampling procedure

converge asymptotically ($N \to \infty$) to the quantiles of the asymptotic distribution.
Now we consider the problem of consistency of the Subsampling.
To prove the consistency of the subsampling procedure we need to know if there exists a non-degenerated asymptotic distribution of the statistic which means that we need to have the central limit theorem.
In both the stable and the GED case we have obtained a weakly convergence to the limit random variables.
Denote the cumulative distribution functions of this limit random variables by $L(s)$ and L for the seasonal means and vector of the seasonal means, respectively.
To fulfill the Step 3 of the subsampling procedure the empirical distribution functions $L_N(s)(x) = P(P_N(s) \leq x)$ are computed from the subsamples
$(X_s, ..., X_{s+(N-1)T})$, for each $s = 1, \ldots, T$ where X_t is defined by the equation (32).
The conclusion from the central limit theorems is that the empirical distribution functions converge weakly to the cumulative distribution functions of the limit random variables

$$L_N(s)(x) \to L(s)(x) \qquad if \qquad N \to \infty, \qquad s = 1, \ldots, T.$$

Denote the density of the limit distribution by $L'(s)$. It is obvious that in the GED case $\| L'(s) \|_\infty < \infty$. Let us consider the stable case.
We can look at the sub-Gaussian time series $\{Y_t = X_t - \eta_t,\ t \in \mathbb{Z}\}$ as defined on the product of two probability spaces:
$(\Omega, \mathcal{G}, P)$ and $(\Omega, \mathcal{E}, P)$ on which the series $\{GG_t,\ t \in \mathbb{Z}\}$ and $\{\epsilon_t = \sigma_t^2,\ t \in \mathbb{Z}\}$ are defined, respectively. The $\mathcal{G}$ and the $\mathcal{E}$ are the $\sigma-$fields as follows:

$$\mathcal{G} = \sigma(GG_t, t \in \mathbb{Z}),$$

$$\mathcal{E} = \sigma(\epsilon_t = \sigma_t^2, t \in \mathbb{Z}).$$

From the Assumption **A1** in the definition of the model (32) the $\sigma-$fields $\mathcal{E}$ and $\mathcal{G}$ are independent with respect to the probability measure P. If we "fix" the values of ϵ_t then the series $\{Y_t = \epsilon^{1/2} GG_t,\ t \in \mathbb{Z}\}$ becomes a zero mean Gaussian time series on the probability space $(\Omega, \mathcal{G}, P)$.
Generally, the sub-Gaussian S series are conditionally centered Gaussian ([89], Section 3.11). It implies that we can view $\{Y_t = X_t - \eta_t,\ t \in \mathbb{Z}\}$ as
$N(0, \gamma_G(0) f_t^2 \epsilon_t)$, i.e., a normal with the variance $\gamma_G(0) f_t^2 \epsilon_t$. Note that by the definition in the Section 4.2.1 of this thesis ϵ_t is, for each $t \in \mathbb{Z}$, positive value random variable. It follows that in the stable case inequality $\| L'(s) \|_\infty < \infty$ also holds.

A5 $\sum_{r=0}^{\infty} \lambda_r^{\frac{2}{3}} < \infty$, where λ_r is the weak dependence sequence of the model (32), and $\lim_{n\to\infty} \frac{b_s}{N} = 0$,
b_s is the length of subsampling subseries:

$$X_{s+pT}, X_{s+(p+1)T}, \ldots, X_{s+(p+b_s-1)T}, \qquad p = 0, \ldots, N - b_s.$$

Theorem 6.7. *(Consistency theorem for seasonal means in both stable and GED cases)*
Assume **A1** *through* **A5** *and consider the subsample of the sample withe subsample step as in the equation (41), then consistency of the subsampling method holds:*

1. *If x is the point of the continuity of $L(s)$, then $L_{N,b_s}(s)(x) \xrightarrow{P} L(s)(x)$.*
2. *If L is continuous then $\sup_x |L_{N,b_s}(s)(x) - L(s)(x)| \xrightarrow{P} 0$.*
3. *If $L(s)$ is continuous in $c(1-q)$ (where $c(1-q)$ is a $q-$quantile) then if $N \to \infty$*
$$P[N^{1-\zeta}(\hat{\eta}_N(s) - \eta(s)) \leq c_{N,b}(1-q)] \to 1-q$$
in stable case or
$$P[N^{-1/2}(\hat{\eta}_N(s) - \eta(s)) \leq c_{N,b}(1-q)] \to 1-q$$
in GED case.

Where $\alpha \in (0,1)$ and

$$c_{N,b}(1-\alpha) = inf\{x : L_{N,b}(s)(x) \geq 1-\alpha\},$$

$$c(1-\alpha) = inf\{x : L(s)(x) \geq 1-\alpha\}.$$

The $\xrightarrow{P}$ denotes convergence in probability.

Proof of Theorem 6.7
Let us consider a sequence of statistics $A_N(s)$, for fixed $s = 1, 2, ..., T$ and $N = 1, 2, ...$ (or $B_N(s)$ in the GED case).
$L_N(s)(x) = P(A_N(s) \leq x)$ is cumulative distribution function of $A_N(s)$.
($L_N(s)(x) = P(B_N(s) \leq x)$ is cumulative distribution function of $B_N(s)$ in GED case.)
From the assumptions

$$sup_{x\in\mathbb{R}}|L_N(s)(x) - L(s)(x)| \longrightarrow 0, \quad N \to \infty$$

For overlapping samples the number of subsamples:

$Y_{b,q}(s) = (X_{s+qT}, X_{s+(q+1)T}..., X_{s+(q+b-1)T})$, $q = 0, 1, ..., N-b$ and the number of subsampling statistics:
$A_{N,b,q}(s) = \sqrt{b}(\hat{\eta}_{N,b,q}(s) - \hat{\eta}_N(s))$ is $N - b + 1$.
$(B_{N,b,q}(s) = \sqrt{b}(\hat{\eta}_{N,b,q}(s) - \hat{\eta}_N(s))$ is $N - b + 1$ in GED case.)
Above statistics are used to approximate the distributions $L_N(s)(x)$ by empirical distribution functions: $L_{N,b,q}(s)(x) = \frac{1}{N-b+1} \sum_{q=0}^{N-b} \mathbb{I}_{\{A_{N,b,q}(s) \leq x\}}$.
$(L_{N,b,q}(s)(x) = \frac{1}{N-b+1} \sum_{q=0}^{N-b} \mathbb{I}_{\{B_{N,b,q}(s) \leq x\}}$ in GED case.)
Let us define rough subsampled distribution:

$$U_{N,b,q}(s)(x) = \frac{1}{N-b+1} \sum_{q=0}^{N-b} \mathbb{I}_{\{\sqrt{b}(\hat{\eta}_{N,b,q}(s) - \eta_N(s)) \leq x\}}.$$

From Theorem 11.3.1 [80] for Heavy Tails (or the Theorem 2.2.1 [80] in the GED case) it is known that

$$\forall x \in \mathbb{R} \ |L_{N,b,q}(s)(x) - U_{N,b,q}(s)(x)| \xrightarrow{p} 0.$$

It follows that it is enough to investigate only the variance of $U_{N,b,q}(s)$, $s = 1, ..., T$
By Theorem (5.4), under theorem 6.7 assumptions we obtain:

$$lim_{N\to\infty} |E[U_{N,b,q}(s)(x) - E[U_{N,b,q}(s)(x)]]^2| = 0.$$

It implies that $Var(U_{N,b,q}(s)(x))$ tends to zero, it proves point 1. of the Theorem 6.7.
To prove the point 2. of the Theorem 6.7 we also use the Theorem 2 from [22].

$$lim_{N\to\infty} sup_{x\in\mathbb{R}} |U_{N,b,q}(s)(x) - L(s)(x)| = 0,$$

in probability.
The proof of point 3. If point 1. holds and under assumption of the model (32) is very similar to the proof of 3. in the Theorem 11.3.1, [80] (or the Theorem 2.2.1 [80] in the GED case). □

Theorem 6.8. *(Consistency theorem for vector of the seasonal means in both stable and GED cases)*
Assume **A1** *through* **A5**, *and consider the subsample of the sample with subsample step as in the equation (41), then consistency of the subsampling method holds:*

1. *If x is the point of the continuity of L, then $L_{N,b}(x) \xrightarrow{p} L(x)$.*
2. *If L is continuous then* $\sup_x |L_{N,b}(x) - L(x)| \xrightarrow{p} 0$.

Proof of Theorem 6.8
For any vector of constants $c \in \mathbb{R}^T$ we have the equation for the subsampling version of the characteristic functions of the distributions:

$$\phi^*_{A_{N,b,q}}(c) = \phi^*_{c^T A_{N,b,q}}(1) \text{ in stable case}$$

$$\phi^*_{B_{N,b,q}}(c) = \phi^*_{c^T B_{N,b,q}}(1) \text{ in GED case}$$

Let $Z_{s+pT} = c_s X_{s+pT}$, where $p = 0, ..., N-1$ and $s = 1, ..., T$. The series $\{Z_t\}$ fulfills the assumptions of Theorem 6.8, which means that subsampling is consistent for the mean $(\eta_N)_Z$. By Theorem A in Athreya [5] we have: in the stable case

$$\phi^*_{c^T A_{N,b,q}}(1) \xrightarrow{p} \begin{cases} \phi_{S_N}(c), & 1/\alpha > (\beta+1)/2 \\ \phi_{V_N}(c), & 1/\alpha < (\beta+1)/2 \\ \phi_{S_N+V_N}(c), & 1/\alpha = (\beta+1)/2 \end{cases}$$

where S_N and V_N are like in Theorem 6.6.
In GED case

$$\phi^*_{c^T B_{N,b,q}}(1) \xrightarrow{p} \phi_{N(\eta, c^T \Sigma c)}(1) = \phi_{N(\eta,\Sigma)}(c).$$

Moreover, in the stable case

$$P^*(A_{N,b,q} \leq x)(1) \xrightarrow{p} \begin{cases} F_{S_N}(x), & 1/\alpha > (\beta+1)/2 \\ F_{V_N}(x), & 1/\alpha < (\beta+1)/2 \\ F_{S_N+V_N}(x), & 1/\alpha = (\beta+1)/2 \end{cases}$$

for any $x \in \mathbb{R}^T$, where $F_{S_N}(x)$, $F_{V_N}(x)$ and $F_{S_N+V_N}(x)$ are the cumulative distribution function of $S\alpha S$, Gaussian and the sum of $S\alpha S$ and Gaussian random vectors, respectively.
In the GED case

$$P^*(B_{N,b,q} \leq x) \xrightarrow{p} F_{N(\eta,\Sigma)}(x),$$

for any $x \in \mathbb{R}^T$, where $F_{N(\eta,\Sigma)}(x)$ is the cumulative distribution function of $N(\eta, \Sigma)$.
The second point of the thesis of the Theorem 6.8 follows then from Polyas theorem. □

7 Applications

7.1 The rate of convergence

If the rate of convergence of the statistic is a priori unknown, then it can be estimated by subsampling and used in the sampling distribution approximation, see Bertail et al. [9] or Politis et al. [80], Chapter 8.

Recall that

$$L_{N,b}(s)(x) = \frac{1}{N-b+1} \sum_{i=1}^{N-b+1} I_{\{a_b(\hat{\eta}_{N,b}(s)-\hat{\eta}_N(s))\leq x\}},$$

$$L_N(s)(x) = \frac{1}{N} \sum_{i=1}^{N} I_{\{a_N(\hat{\eta}_N(s)-\eta(s))\leq x\}}$$

and

$$L(s)(x) = P(a_N(\hat{\eta}_N(s) - \eta(s)) \leq x)).$$

Let us define

$$\bar{L}_{N,b}(s)(x) = \frac{1}{N-b+1} \sum_{i=1}^{N-b+1} I_{\{(\hat{\eta}_{N,b}(s)-\hat{\eta}_N(s))\leq x\}}$$

$$\tilde{L}_{N,b}(s)(x) = P((\hat{\eta}_{N,b}(s) - \hat{\eta}_N(s)) \leq x).$$

The normalize sequence a is in the form: $N^{1-\zeta}$ in stable case and $N^{-1/2}$ in GED case.

Lemma 7.1. *Assume* **A1** *through* **A5** *and consider the subsample of the sample with subsample size as in the equation (41), then*

$$\bar{L}_{N,b}(s)(x) = \tilde{L}_{N,b}(s)(x) + o_P(1),$$

$N \to \infty$. The above Lemma is a simple corollary from the consistency of the subsampling method Theorem 6.7.

Theorem 7.1. *Assume* **A1** *through* **A5** *and consider the subsample of the sample with subsample size as in the equation (41). Let* $k_0 = sup\{x : L(s)(x) = 0\}$ *and* $k_1 = inf\{x : L(s)(x) = 1\}$ *and assume that* $L(s)(x)$ *is continuous and strictly increasing on* (k_0, k_1) *as a function of* x*. If the consistency of the subsampling method theorem is fulfilled then*

$$a_b \bar{L}_b^{-1}(s)(x) = L^{-1}(s)(x) + o_P(1), \tag{42}$$

for any $x \in (0, 1)$ *and* $N \to \infty$.

Proof of the Lemma 7.1
The proof of the Lemma 7.1 strictly follows from Lemma 2, [9] and Theorem 5.4. □

To estimate the rate of convergence we can use simple empirical tool proposed by Bertail [8].
Our time series is PS but for each $s = 1, ..., T$ and $p = 0, ..., N-1$ time series X_{s+pT} is stationary. Assuming that $\delta_N = N^{-\zeta}$ and taking proper logarithm in the equation (42) we get:

$$log(|\bar{L}_b^{-1}(s)(x)|) = log(L^{-1}(s)(x)) + \zeta log(b) + o_P(1).$$

If we take any $p_i \neq p_j \in (0,1)$ and draw log of some quantile range of subsampling distribution

$$|\bar{L}_b^{-1}(s)(p_i) - \bar{L}_b^{-1}(s)(p_j)| = \zeta log(b) + |L^{-1}(s)(p_i) - L^{-1}(s)(p_j)| + o_P(1)$$

If we consider different subsample size $b_{i,n}$, $i = 1, ..., I > 1$ we can use the least squares estimator of slope, see [9]:

$$\gamma_I = \frac{\sum_{i=1}^{I}(y_i - \bar{y})(log(b_{i,n}) - \bar{log})}{\sum_{i=1}^{I}(log(b_{i,n}) - \bar{log})^2},$$

where for given $t \in 0,1$ $y_i = log(|\bar{L}_{b_{i,n}}^{-1}(s)(x)|)$, $\bar{y} = I^{-1}\sum_{i=1}^{I} y_i$, and $\bar{log} = I^{-1}\sum_{i=1}^{I} log(b_{i,n})$.

Unfortunately consistency of this estimator for weakly dependent time series is unknown. There exist other, known, methods to estimate the parameters of the tails and long memory and this method can be used in the estimation of rate of convergence.

7.2 Choosing length of the block

One needs to be careful in choosing length of the block: b_N. It can't be to small of course, but also it can't be to big else the subsampling method do not work, see [8].

To choose the block length we used simple empirical tool proposed by Bertail [8].
Assuming that $\delta_N = N^{-\zeta}$ and taking proper logarithm in equation in the equation (42) we get:

$$log(|\bar{L}_b^{-1}(s)(x)|) = log(L^{-1}(s)(x)) + \zeta log(b) + o_P(1).$$

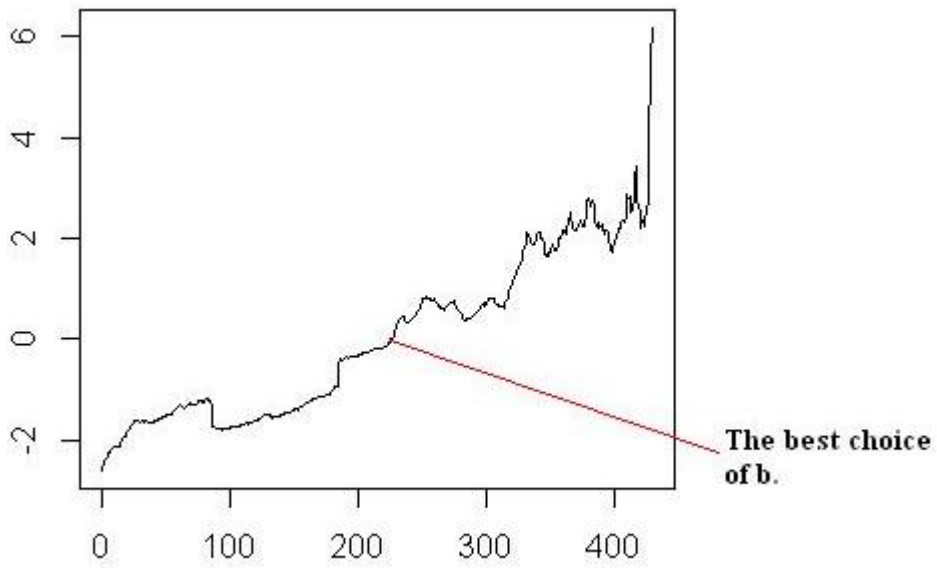

Figure 2: The example of choosing b for the process X_{s+p+T} with parameters $\alpha = 1.5$, $\beta = 0.3$ and $s = 5$, $T = 24$.

If we take any $p_i \neq p_j \in (0, 1)$ and draw log of some quantile range of subsampling distribution

$$|\bar{L}_b^{-1}(s)(p_i) - \bar{L}_b^{-1}(s)(p_j)| = \zeta log(b) + |L^{-1}(s)(p_i) - L^{-1}(s)(p_j)| + o_P(1)$$

we will see that the best choice of b is the largest one before the "unstable" behavior.

Bellow the simulation study for the mean in the stable case are introduced. We assume that the mean value $\eta(s)$ is 0, for all $s = 1, ..., T$.
For the simulation study we chose the Gaussian Gegenbauer process with k=1, innovations with mean zero and variance 1, $\nu = 1$ In this case the autocorrelation function is equal as follow, [89]:

$$\gamma_{GG}(h) = \frac{\Gamma(1-\beta)}{\Gamma(\beta/2)\Gamma(1-\beta/2)} h^{\beta-1}$$

$$\gamma_{GG}(0) = \frac{\Gamma(1-\beta)}{\Gamma^2(1-\beta/2)}.$$

The constant C in the definition of long memory for each $s = 1, \ldots, T$ is:

$$C(s) = \mu^2 f_s^2 \frac{\Gamma(1-\beta)}{\beta/2\Gamma(\beta/2)\Gamma(1-\beta/2)}.$$

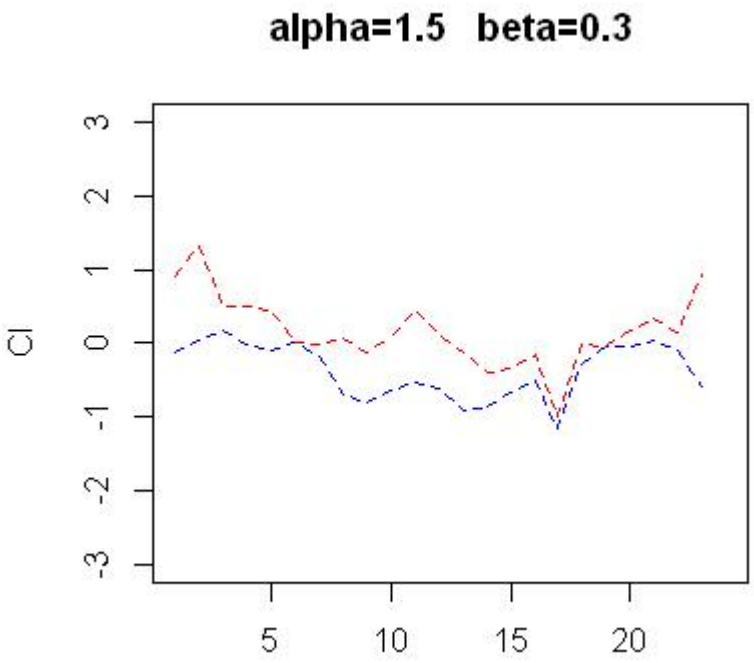

Figure 3: Equal tailed confidence interval for the mean parameter of the process X_{s+p+T} with parameters $\alpha = 1.5$, $\beta = 0.3$ and $T = 24$.

For the ϵ_t we chose $\alpha/2-$stable i.i.d. random variables with the skewed parameter 1, the location parameter 0 and the scale $(cos(\pi\alpha/4))^{2/\alpha}$.
The number of observations is $NT = 10320$, period $T = 24$. In the first case we took $\beta = 0.3$ and $\alpha = 1.5$.
This is the "tail" case.
For each s = 1,...,24, we found subsample size by the method described before and then draw the equal-tailed and symmetric 95% confidence intervals.

In the second case we took $\beta = 0.4$ and $\alpha = 1.6$.
This is the "memory" case.
And for each $s = 1, \ldots, 24$, we have done the same as in previous case.

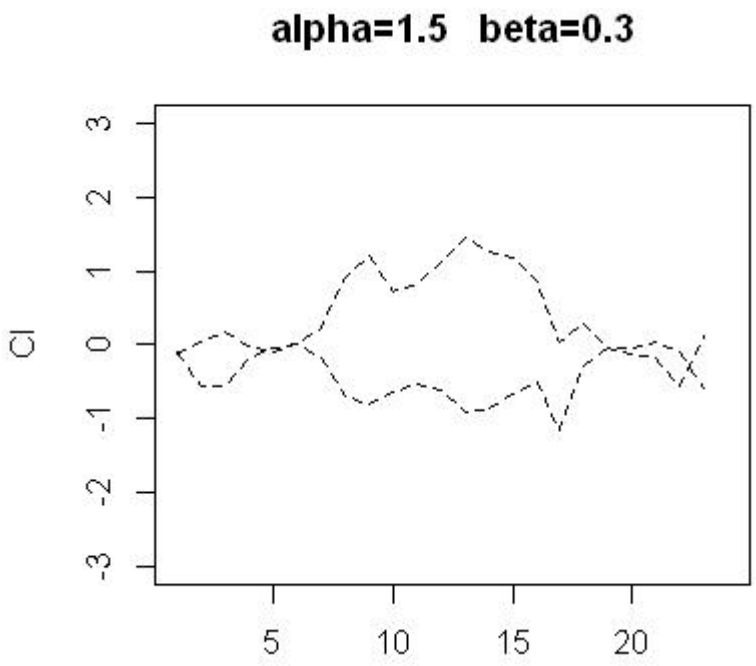

Figure 4: Symmetric confidence intervals for the mean parameter of the process X_{s+p+T} with parameters $\alpha = 1.5$, $\beta = 0.3$ and $T = 24$.

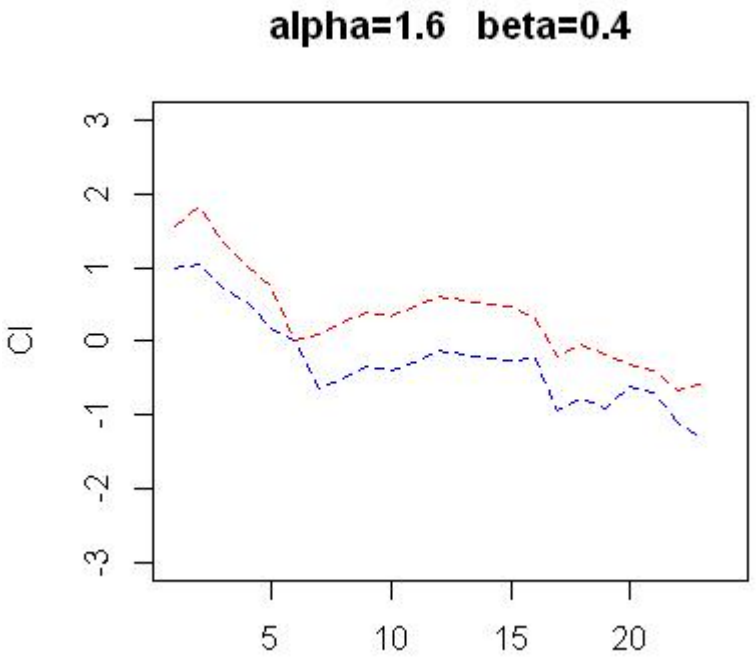

Figure 5: Equal-tail confidence interval for the mean parameter of the process X_{s+p+T} with parameters $\alpha = 1.6$, $\beta = 0.4$ and $T = 24$.

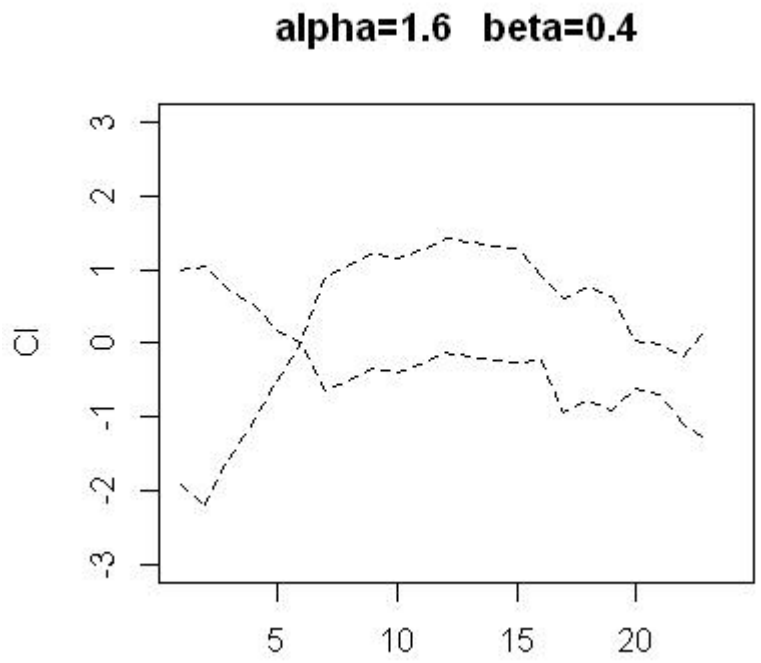

Figure 6: Symmetric confidence intervals for the mean parameter of the process X_{s+p+T} with parameters $\alpha = 1.6$, $\beta = 0.4$ and $T = 24$.

8 Conclusions and open questions

In the thesis non-stationary time series with the periodic structure and specific features: heavy tail, long memory have been considered. The consistency of the subsampling method for the mean has been obtained. The techniques used in the proofs were completely different in two different cases of heavy tails. In the stable case the theory of stable laws was used and in the GED case methods based on the Lindeberg theorem was used. Because the model (32) does not holds the mixing conditions, the new condition of dependence in time series - weak dependency has to be taken into consideration.

When working with the model defined by the equation (32) one may be faced with the following problems:

- estimations of the convergence rate which is associated with the estimation of parameters of the long memory and the tails and
- identification of the parameters of the model.

First problem was discussed in the previous Section. The second can be easily explained. In the dissertation we do the analysis of the first order and the knowl-

edge of the periodic function f it is not necessary.

Now let us consider the future research areas in this topic.
First question for the future is how to generalize the model 32 in two directions. Firstly, in the thesis only two classes of volatility series σ_t, representing the heavy tails were introduced. Which of course, does not cover all cases of heavy tails processes. Moreover, the special case of periodic models, i.e. the multiplicative model were featured. Hence, secondly, how to generalize the subsampling method to the wider class of periodic time series.

Second, very important, question which need to be considered is how to make the selection of the block length b in the class of periodic models adequate and effective?

Third, how will behave resampling methods in non-stationary case, if we replace the $\alpha-$mixing assumption by weak dependence?
For the results in stationary case, for selected bootstrap methods in some econometrics models, see Ango Nze, Doukhan [2].

It is important to develop statistical tools for models with periodic structure. The reason is very simple: many of the phenomena that we observe in real life is characterized by seasonality. If we also consider long memory and heavy tails it is even better. Because long memory is an often occurring phenomenon and heavy-tails are everywhere therefor they are more "normal" than the Gaussian.

Our results provide consistent statistical procedures for mean function and confidence intervals.

9 Bibliography

References

[1] Andrews D. W. K., (1984), *Non-strong mixing autoregressive processes*, J. Appl. Probab. 21, 930-934

[2] Ango Nze P., Doukhan P. (2004), *Weak Dependence: Models and Applications to Econometrics*, Econometric Theory Vol. 20, No. 6 , p. 995-1045

[3] Antoni J. (2009), *Cyclostationarity by examples*, Mech. Syst. Signal Process. 23, 987-1036

[4] Arcones M. (1994), *Limit theorems for nonlinear functionals of a stationary sequence of vectors*, Ann. Prob., 22:2242-2274

[5] Athreya K. B. (1987), *Bootstrap of the mean in the infinite variance case*, Ann. Statist. 15, 724-731

[6] Bardet J-M., Doukhan P., Lang G., Ragache N. (2008), *Dependent Lindeberg central limit theorem and some applications*, ESAIM: Probability and Statistics, Volume: 12, p. 154-172

[7] Beran J., Feng T., Ghosh S., Kulik R. (2013), *Long-Memory Processes. Probabilistic Properties and Statistical Methods*, Springer

[8] Bertail P. (2011), *Some comments on subsampling weakly dependent time series and application to extremes*, TEST, Volume 20, p. 487-490

[9] Bertail P., Haefke C., Politis D. N., White W. (2004), *Subsampling the distribution of diverging statistics with applications to finance*, Journal of Econometrics 120, p. 295-326

[10] Bertail P., Politis D. N., Rhomari N. (2000), *Subsampling continuous parameter random fields and a Bernstein inequality*, Statistics, Vol. 33, No.4, p. 367-392

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

[12] Besicovitch A.S. (1932), *Almost periodic functions*, University Press, Cambridge

[13] Bickel, P., Bühlmann, P. (1999), *A new mixing notion and functional central limit theorems for a sieve bootstrap in time series*, Bernoulli 5-3, p. 413-446

[14] Billingsley P. (1995), *Probability and measure*, John Wiley and Sons Inc., New York

[15] Bradley R. (2005), *Basic properties of strong mixing conditions*, Probab. Surv., 2:107-144

[16] Brockwell P., Davis R. (1991), *Time Series: Theory and Methods*, Springer-Verlag, New York

[17] Chan V., Lahiri S. N., Meeker W. (2004), *Block bootstrap estimation of the distribution of cumulative outdoor degradation*, Technometrics, 46:215-224

[18] Dedecker J., P. Doukhan, Lang G., León J. R., Louhichi S., Prieur C. (2008), *Weak Dependence: With Examples and Applications* Lecture Notes 190 in Statistics, Springer-Verlag

[19] Dehay D., Dudek A.E., Leśkow J. (2014), *Subsampling for continuous-time almost periodically correlated processes*, Journal of Statistical Planning and Inference 150:142-158

[20] Dehay D., Hurd H. L. (1994), *Representation and estimation for periodically and almost periodically correlated random processes*, In Cyclostationarity in communications and signal processing., p. 295-326, New York, IEEE

[21] Dehay D., Leśkow J. (1996), *Functional Limit Theory for the Spectral Covariance Estimator*, Journal of Applied Probability, Vol. 33, No. 4, p. 1077-1092

[22] Doukhan P., Prohl S., Robert C. Y. (2011), *Subsampling weakly dependent times series and application to extremes*, TEST, Volume 20, Issue 3, p. 487-490

[23] Doukhan P., Lang G. (2002), *Rates in the empirical central limit theorem for stationary weakly dependent random fields*, Statis. Inf. for Stoch. Porc. 5, p. 199-228. MR1917292

[24] Doukhan P. (1994), *Mixing: properties and examples*, Lecture Notes in Statistics 85, Springer Verlag

[25] Doukhan P., Louhichi S. (1999), *A new weak dependence condition and applications to moment inequalities*, Stoch. Proc. Appl. 84, p. 313-342

[26] Doukhan P., Wintenberger O. (2006), *Invariance principle for new weakly dependent stationary models under sharp moment assumptions*, Preprint available in http://samos.univ-paris1.fr/ppub2005.html.

[27] Dudek A.E., Leśkow J., Politis D., Paparoditis E. (2014), *A generalized block bootstrap for seasonal time series*, J. Time Ser. Anal., 35, p. 89-114

[28] Dudek A.E. (2014), *Circular block bootstrap for coefficients of autocovariance function of almost periodically correlated time series*, Metrika, Published online: 09 August 2014

[29] Durrett R. (1991), *Probability: Theory and Examples*, Wadsworth and Brooks/Cole, Pacific Grove, California

[30] Efron B. (1979), *Bootstrap methods: Another look at the jackknife*, Ann. Statist., 7:1-26.

[31] Feller W. (1971), *An Introduction to Probability Theory and Its Applications, vol. II*, John Wiley and Sons Inc., New York

[32] Ferrara L., Guégan D. (2001), *Forecasting with k-factor Gegenbauer processes: theory and applications*, Journal of Forecasting, Volume 20, Issue 8, p. 581-601

[33] Fitzenberger B. (1997), *The moving blocks bootstrap and robust inference for linear last squares and quantile regressions*, J. Econometrics, 82:235-287

[34] Gajecka-Mirek E. (2014), *Subsampling for weakly dependent and periodically correlated sequences*, Lecture Notes in Mechanical Engineering, Cyclostationarity: Theory and Methods, Springer

[35] Gardner W,. Napolitano A., Paura L. (2006), *Cyclostationarity: half a century of research*, Signal Process., 86:639-697

[36] Gladyshev E. G. (1961), *Periodically correlated random sequences*, Sov. Math., 2

[37] Gladyshev E. G. (1963), *Periodically and almost PC random processes with continuous time parameter*, Theory Probab. Appl., 8

[38] Granger, C. W. J. and Joyeux, J. (1980), *An introduction to long memory time series models and fractional differencing*, J. Time Series Anal., 1, p. 15-30

[39] Gray H. L., Zhang N.-F., Woodward W. A. (1989), *On generalized fractional processes*, Journal of Time Series Analysis, 10, p. 233-257

[40] Gray H. L., Zhang N.-F., Woodward W. A. (1994) *On generalized fractional processes a correction*, Journal of Time Series Analysis, 15(5), p. 561-562

[41] Guégan D., Ladoucette S. (2001), *Non-mixing properties of long memory sequences*, C.R. Acad. Sci. Paris, 333, p. 373-376

[42] Hall P., Jing B.-Y. (1998), *Comparison of bootstrap and asymptotic approximations to the distribution of a heavy-tailed mean*, Statistica Sinica 8, p. 887-906

[43] Hall P., Jing B.-Y., Lahiri S. (1998), *On the sampling window method for long-range dependent data*, Statistica Sinica 8, p. 1189-1204

[44] Hosking J. R. M. (1981), *Fractional differencing*, Biometrika, 68, p. 165-176

[45] Hosking J. R. M. (1982), *Some models of persistence in time series*, in (Anderson, O. D. (Ed.)) Time Series Analysis: Theory and Practice, vol. 1, North - Holland

[46] Hui Y. V., Li W. K.(1995), *On Fractionally differenced periodic processes*, The Indian Journal of Statistics, Volume 57, Series B, Pt. 1, p. 19-31

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

[48] Hurd H., Makagon A., Miamee A. G. (2002), *On AR(1) models with periodic and almost periodic coefficients*, Stochastic Process. Appl., 100:167-185

[49] Hurd H., Leśkow J. (1992), *Strongly consistent and asymptotically normal estimation of the covariance for almost periodically correlated processes*, Stat Decis 10:201-225

[50] Hurd H., Leśkow J. (1992), *Estimation of the Fourier coefficient functions and their spectral densities for $\varphi-$mixing almost periodically correlated processes*, Statist. Probab. Lett., 14:299-306

[51] Ibragimov I. A., Rozanov Y. A. (1978), *Gaussian Random Processes*, Springer-Verlag, New York

[52] Ibragimov I. A., (1962), *Some limit theorems for stationary processes*, Teor. Veroyatn. Primen. 7 361392 (in Russian), [English translation: Theory Probab. Appl. 7, 349382]

[53] Jach A., McElroy T., Politis D. N. (2012), *Subsampling inference for the mean of heavy-tailed long memory time series*, J. Time Ser. Anal., vol. 33, no. 1, p. 96-111

[54] Jakubowski J., Sztencel R. (2004), *Wstp do teorii prawdopodobieństwa*, Wydanie III, Script, Warszawa

[55] Janicki A., Weron A., (1994), *Simulation and Chaotic Behavior of a-Stable Stochastic Processes*, Marcel Dekker, New York

[56] Javorskyj I., Leśkow J., Kravets I., Isayev I., Gajecka E. (2012), *Linear filtration methods for statistical analysis of periodically correlated random processesPart I: Coherent and component methods and their generalization*, Signal Processing 92:7, p.1559-1566

[57] Javorskyj I., Leśkow J., Kravets I., Isayev I., Gajecka E. (2011), *Linear filtration methods for statistical analysis of periodically correlated random processesPart II: Harmonic series representation*, Signal Processing 91:2506-2519

[58] Javorskyj I., Isayev I., Majewski J., Yuzefovych R. (2010), *Component covariance analysis for periodically correlated random processes*, Signal Processing 90:1083-1102

[59] Javorskyj I., Isaev I., Zakrzewski Z., Brooks S. P. (2007), *Coherent covariance analysis of periodically correlated random processes*, Signal Processing 87(1):13-32

[60] Kolmogorov A. N., Rozanov Y. A. (1960), *On strong mixing conditions for stationary Gaussian processes*, Theory Prob. Appl. 5, p. 204-208

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

[62] Lahiri S. (2003), *Resampling Methods for Dependent Data*, Springer-Verlag Inc., New York

[63] Lahiri S. (1999), *Theoretical comparison of block bootstrap methods*, Ann. Statist., 27:386-404

[64] Lahiri S. (1993), *On the moving block bootstrap under long-range dependence*, Statistics and Probability Letters 18, p. 405-413

[65] Lawrance A. J., Kottegoda N. T. (1977), *Stochastic modeling of river flow time series*, J. Royal Stast. Soc., Ser. B., 140, p. 1-47

[66] Leśkow J., Synowiecki R. (2010), *On bootstrapping periodic random arrays with increasing period*, Metrika, 71:253-279

[67] Leśkow, J., Synowiecki, R. (2006), *Asymptotic distribution and subsampling for the estimator of autocovariance coefficient for APC time series*, Preprint No. 1/2006, Faculty of Applied Mathematics, AGH University of Science and Technology

[68] Leśkow, J., Weron A. (1992), *Ergodic behavior and estimation for periodically correlated processes*, Statistics and Probability Letters 15 (4), 299-304

[69] Lii K. S., Rosenblatt M. (2002), *Spectral analysis for harmonizable processes*, Ann. Statist. 30, 258-297

[70] Lii K. S., Rosenblatt M. (2006), *Estimation for almost periodic processes*, Ann. Statist. 34, 1115-1139

[71] Liu R., Singh K. (1992), *Moving blocks jacknife and bootstrap capture weak dependence*, In LePage R. and Billard L., editors, Exploring the Limits of Bootstrap, p. 225-248, Wiley, New York.

[72] Lunetta G. (1963), *Di una generalizzazione dello schema della curva normale*, Annali della Facoltá di Economia e Commercio di Palermo 17, 237-244

[73] Marinucci D. (2005), *The empirical process for bivariate sequences with long memory*, Statist. Inf. for Stoch. Proc., 8(2):205-223

[74] McLeod A. I., Hipel K. W. (1978), *Preservation of the rescaled adjusted range, I. A reassessment of the Hurst phenomenon*, Water Resources Research, 14, p. 491-508

[75] McElroy T., Politis D. N. (2007), *Self-Normalization for Heavy-Tailed Time Series with Long Memory,* Statist. Sinica, vol. 17, no. 1, p. 199-220

[76] Philippe A., Surgailis D., Viano M. C. (2006), *Almost periodically correlated processes with long-memory*, Lecture Notes in Statistics 187, Springer

[77] Politis D. (2001), *Resampling time series with seasonal components*, Proceedings of the 33rd Symposium on the Interface of Computing Science and Statistics

[78] Politis D. N., Romano J. P. (1994), *A general theory for large sample confidence regions based on subsamples under minimal assumptions*, Ann. Statist., 22, p. 2031-2050.

[79] Politis D. N., Romano J. P. (1992), *A circular block-resampling procedure for stationary data. Exploring the limits of bootstrap.*, Wiley series in probability and mathematical statistics, Wiley, New York

[80] Politis D. N., Romano J. P., Wolf M. (1999), *Subsampling*, Springer-Verlag, New York

[81] Priestley M. B. (1981), *Spectral Analysis and Time Series*, Academic Press, London

[82] Radulović D. (1996), *The bootstrap of the mean for strong mixing sequences under minimal conditions*, Statist. Probab. Lett., 28:65-72

[83] Rosenblatt M., (1956), *A central limit theorem and a strong mixing condition*, Proceedings of the National Academy of Sciences USA 42, 43-47

[84] Silverman R. L. (1957), *Locally stationary random processes*, IEEE Transactions on Information Theory 3(9):182-187

[85] Singh K. (1981), *On the asymptotic accuracy of the Efrons bootstrap*, Ann. Statist., 9:1187-1195

[86] Subbotin M. Th. (1923), *On the law of frequency of error*, Mat. Sb., 31:2, 296-301

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

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

[89] Taqqu M. S., Samorodntsky G. (1994), *Stable non-Gaussian random processes, Stochastic models with infinite variance*, Chapman and Hall

[90] Varanasi M. K., Aazhang B. (1989), *Parametric Generalized Gaussian Density Estimations*, Journal of the Acoustical Society of America., vol. 86, no. 4, p. 1404-1415

[91] Volkonski V. A., Rozanov Y. A., (1959), *Some limit theorems for random functions*, Part I. Teor. Veroyatn. Primen. 4 186207 (in Russian), [English translation: Theory Probab. Appl. 4, 178-197]

[92] Weron A. (1984), *Stable processes and measures: A survey. Probability Theory on Vector Spaces III*, in Szynal D. and Weron A., Lecture Notes in Mathematics, Vol. 1080, Springer-Verlag, Berlin, p. 306-364