Statistics and Probability Letters 129 (2017) 252–259

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Subsampling for nonstationary time series with non-zero mean function

Anna E. Dudek [a,b], Łukasz Lenart [c,*]

[a] Institut de Recherche Mathématique de Rennes, Université Rennes 2, Rennes, France
[b] AGH University of Science and Technology, Department of Applied Mathematics, al. Mickiewicza 30, 30-059 Cracow, Poland
[c] Department of Mathematics, Cracow University of Economics, ul. Rakowicka 27, 31-510 Cracow, Poland

ARTICLE INFO

Article history:
Received 9 November 2016
Received in revised form 29 May 2017
Accepted 1 June 2017
Available online 21 June 2017

MSC 2010:
62G09
62G10
62M10

Keywords:
Almost periodically correlated
Autocovariance function
Fourier coefficients
Periodicity
Testing

ABSTRACT

In this paper a subsampling approach for nonstationary time series with a non-zero mean function is proposed. It is applied for periodically and almost periodically processes. Two statistical tests are constructed. An example with real data is presented.

1. Introduction and problem formulation

In this paper we discuss a subsampling approach for nonstationary time series, which have a non-zero mean function.

First we introduce some notation. Let $\{X_t,\ t \in \mathbb{Z}\}$ be a nonstationary time series and $\mu(t) = \mathrm{E}(X_t)$ be its mean function. We assume that $\mu(t) \not\equiv 0$. Moreover, we consider the following decomposition of the mean function

$$\mu(t) = \mu_1(t) + \mu_2(t).$$

Functions $\mu_1(t)$ and $\mu_2(t)$ can be interpreted differently depending on the application. In economics they can represent e.g., seasonal or business fluctuations or a long-term trend, while in telecommunications these can be two periodic functions e.g., the first one may be a pilot tone and the latter one the mean function of some periodic signal. A pilot tone is a signal transmitted for control, synchronization or reference purposes. In the following we assume that $\mu_1(t)$ is non-zero and we know how to estimate it. On the other hand, $\mu_2(t)$ may be a zero function and moreover we may not have complete knowledge of its form and properties. Such a model describes a real data situation in which only partial information about the mean function is available e.g., we know the nature of some structure contained in the data like seasonality, but other components require further investigation.

* Corresponding author.
E-mail addresses: aedudek@agh.edu.pl (A.E. Dudek), lenartl@uek.krakow.pl (Ł. Lenart).

http://dx.doi.org/10.1016/j.spl.2017.06.002
0167-7152/

We denote the observed sample by $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$. Let $\hat{\mu}_{1,n}(t)$ be the estimator of $\mu_1(t)$ based on $\mathbf{X}_n$. We consider the modified sample $\tilde{\mathbf{X}}_n = (\tilde{X}_{1,n}, \tilde{X}_{2,n}, \ldots, \tilde{X}_{n,n})$, where

$$\tilde{X}_{t,n} = X_t - \hat{\mu}_{1,n}(t).$$

In the sequel, for the sake of simplicity we skip the subscript n and we denote elements of the triangular array $\tilde{\mathbf{X}}_n$ by $\tilde{X}_i$, $i = 1, \ldots, n$. From now we will be using only the sample $\tilde{\mathbf{X}}_n$, i.e., we consider partially or entirely demeaned data. This corresponds to the standard procedures of data analysis in which the data is assumed to be demeaned before any further analysis (e.g. second-order) is performed. In the next section we present how the subsampling idea for nonstationary time series (see chapter 4 in Politis et al. (1999)) can be adapted to our setting. In Section 3 we construct two tests designed for time series with periodic and almost periodic structure. The first one is devoted to the first-order frequency detection, while the second one can be used for testing the second-order periodicity. Finally, Section 4 contains a real data application of our results.

2. Subsampling

In the sequel we use notation introduced by Politis et al. (1999).

Let $\theta = \theta(\mathcal{P}) \in \mathbb{R}$ be the parameter of interest, where $\mathcal{P}$ is the joint probability law governing $\{X_t, t \in \mathbb{Z}\}$. Parameter θ can simply relate to $\mu_2(t)$, but one may consider other characteristics of X_t. Moreover, let $\hat{\theta}_n = \hat{\theta}_n(\tilde{\mathbf{X}}_n)$ be the estimator of θ and $\hat{\theta}_{n,b,t} = \hat{\theta}_{n,b,t}(\tilde{X}_t, \tilde{X}_{t+1}, \ldots, \tilde{X}_{t+b-1})$ be the estimator of θ based on subsample $\tilde{X}_t, \tilde{X}_{t+1}, \ldots, \tilde{X}_{t+b-1}$. In addition, let $\hat{\theta}'_{n,b,t} = \hat{\theta}_{n,b,t}(X'_t, X'_{t+1}, \ldots, X'_{t+b-1})$, where $X'_i = X_i - \mu_1(i)$, $i \in \mathbb{Z}$.

By $J'_{b,t}(P)$ and $J_n(P)$ we denote the sampling distributions of $\tau_b(\hat{\theta}'_{n,b,t} - \theta)$ and $\tau_n(\hat{\theta}_n - \theta)$, where $\tau_{(\cdot)}$ is an appropriate normalizing sequence. The corresponding cumulative distribution functions are defined as

$$J'_{b,t}(x, P) = P\left\{\tau_b(\hat{\theta}'_{n,b,t} - \theta) \le x\right\} \tag{1}$$

and

$$J_n(x, P) = P\left\{\tau_n(\hat{\theta}_n - \theta) \le x\right\}. \tag{2}$$

Finally, following Politis et al. (1999) we assume that X_t is α-mixing.

Conditions below are sufficient conditions for the subsampling consistency.

Assumption 2.1. There exists a limiting law $J(P)$ such that

(i) $J_n(P)$ converges weakly to $J(P)$ as $n \longrightarrow \infty$.

(ii) For any continuity point x of $J(P)$ and for any sequences n, b with $n, b \longrightarrow \infty$ and $b/n \longrightarrow 0$, we have $\frac{1}{n-b+1}\sum_{t=1}^{n-b+1} J'_{b,t}(x, P) \longrightarrow J(x, P)$.
OR
For any continuity point x of $J(\cdot, P)$ and for any index sequence $\{t_b\}$ we have that $J'_{b,t_b}(x, P) \longrightarrow J(x, P)$ as $b \longrightarrow \infty$.

(iii) $\max_{1\le t\le n-b+1} \tau_b \left|\hat{\theta}_{n,b,t} - \hat{\theta}'_{n,b,t}\right| \xrightarrow{p} 0$.

One may notice that comparing to Politis et al. (1999) (see Assumption 4.2.1. in Politis et al. (1999)) we need an additional assumption (iii). It is caused by the fact that in contrast to Politis et al. (1999) we are using partially demeaned data. Finally, condition (i) is identical to the one proposed by Politis et al., while (ii) is adjusted to our problem.

The approximation of $J_n(x, P)$ is defined as

$$\tilde{L}_{n,b}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1} \mathbf{1}\left\{\tau_b\left(\hat{\theta}_{n,b,t} - \hat{\theta}_n\right) \le x\right\},$$

where $\mathbf{1}\{A\}$ is an indicator function of the event A.

Below we state the asymptotic validity of subsampling for general statistics.

Theorem 2.1. *Assume that $\{X_t, t \in \mathbb{Z}\}$ is α-mixing time series. Under Assumption 2.1 and taking $\tau_b/\tau_n \longrightarrow 0$, $b = b(n) \longrightarrow \infty$ such that $b/n \longrightarrow 0$ as $n \longrightarrow \infty$, we have*

(i) *if x is a continuity point of $J(\cdot, P)$, then $\tilde{L}_{n,b}(x) \xrightarrow{p} J(x, P)$,*

(ii) *if $J(\cdot, P)$ is continuous, then $\sup_{x\in\mathbb{R}} |\tilde{L}_{n,b}(x) - J(x, P)| \xrightarrow{p} 0$,*

(iii) *for $\alpha \in (0, 1)$, let $c(1-\alpha) = \inf\{x : J(x, P) \ge 1-\alpha\}$ and $\tilde{c}_{n,b}(1-\alpha) = \inf\{x : \tilde{L}_{n,b}(x) \ge 1-\alpha\}$. If $J(\cdot, P)$ is continuous at point $c(1-\alpha)$, then subsampling confidence intervals are asymptotically consistent, i.e.,*

$$P\left(\tau_n(\hat{\theta}_n - \theta) \le \tilde{c}_{n,b}(1-\alpha)\right) \longrightarrow 1-\alpha. \tag{3}$$