Bootstrapping the Seasonal Means and the Overall Mean in Rainfall Time Series

Lorena Margo Zeqo

Eljona Milo Tasho

Denisa Kafazi

Department of Mathematics and Physics,
Fan S. Noli University,
Korca, Albania

Received: 12 January 2023 / Accepted: 16 March 2023 / Published: 20 March 2023
© 2023 Zeqo et al.

Doi: 10.56345/ijrdv10n108

Abstract

Several bootstrap methods have been studied nowadays for estimating the parameters in periodic data and it is important to consider the periodicity present in choosing the right bootstrap method. In this study we conducted a comparison of the performance of several block bootstrap methods designed for dependent data in the case of a real data time series with periodic structure such as the rainfall time series that is a time series with meteorological data collected monthly in a region in Albania. R programming language is used to perform the bootstrap and to obtain the results. We proposed a block bootstrap procedure in the case of estimating the seasonal means and the overall mean in the time series studied and based on the results obtained we notice a good performance of our proposed bootstrap procedure compared with other procedures considered.

Keywords: *block bootstrap, periodicity, time series, parameter estimation*

1. Introduction

Bootstrap proposed by Efron (1979) resulted to be an important method in the case of estimating the distribution of an estimator or test statistic by applying the resampling of the data or to a model estimated from the data ([4], [6]). This method gives good results in the case of independent and identically distributed (i.i.d) observations, but in the case of dependent observations, like time series, the i.i.d bootstrap gives incorrect answers (Singh (1981) [13], Babu and Singh (1983) [1]). Because of independence requirement the application of the bootstrap method is limited. For this reason, block bootstrap was introduced and resulted to be an important method for implementing the bootstrap with the time series data.

Block bootstrap consists in dividing the data into blocks of observations and then resampling these blocks with replacement. Several block bootstrap methods were presented from researchers with intention to estimate the parameters of interest such as Moving Block Bootstrap (MBB) proposed by Künsch (1989) [7] and Liu and Singh (1992) [8] and is considered a good method designed especially for stationary time series.

When we study periodic data, these methods must be able to considerate the periodicity present. For the time series with a seasonal component, Politis (2001) [12] proposed a resampling algorithm that takes in consideration the periodicity present in the data. The Seasonal Block Bootstrap (SBB) proposed by Politis is a version of Kunsch's (1989) block bootstrap with blocks whose size and starting points are integer multiples of the period. Although the good results of SBB presented by Politis, the method possess a restriction on the relative size of the period and the block size, because the block size must be of the order of the period (See [3]).

Dudek et. al [3] proposed a modification of the block bootstrap, the Generalized Seasonal Block Bootstrap (GSBB) that is suitable for periodic time series with fixed length periodicities of arbitrary size as related to block size and also to the sample size. The series of observations is divided in blocks of desired length independent from periodicity and then these blocks are resampled in a way that retains periodicity. Because the block size can be chosen independently from the period, the usual asymptotic considerations for block size choice avoid the inconsistency problems of other methods and the problems that associate the Seasonal Block Bootstrap of Politis. The GSBB preserve the periodic structure of the data not requiring any particular relationship between the block length and the period length. This method is consistent for the seasonal means and the overall mean of a periodically correlated time series (See [3] for more).

The goal of this study is to realize a comparison of the performance of the aforementioned bootstrap procedures and another procedure, Block bootstrap based in residuals, BBR, proposed for seasonal time series in the case of real data with periodic structure such as Rainfall time series (see [9], [10]). The Rainfall time series is a series of rainfall readings for a period of 40 years in a region in Albania.

In this study the seasonal means and the overall mean estimations are obtained using these three different block bootstrap procedures: MBB, GSBB and BBR. There are good results obtained in the case of using BBR with a similar performance as GSBB and this is a good indication to use this block bootstrap procedure in the case of time series with real data and periodic structure.

2. Objectives

In this study we consider a time series with data collected from monthly rainfall readings in a region in Albania for a period of 40 years. We used elements of periodic time series analysis to confirm the periodic structure of the series and we also realized a comparison of the performance of three different block bootstrap procedures in estimating the seasonal means and the overall mean.

We proposed a block bootstrap procedure in the case of periodic time series such as Block Bootstrap based in Residuals and the goal is to compare its performance with a well-known block bootstrap procedure such as Generalized Seasonal Block Bootstrap and also with Moving Block Bootstrap.

R programming language is used in performing bootstrap replications and obtaining and presenting the results.

3. Methodology

3.1 Data in this study

In this study we used meteorological data collected monthly over a period of 40 years in a region in Albania. The dataset length is 480 data. Several studies involve meteorological time series and they are more focused in forecasting, but we are interested in comparing the performance of block bootstrap procedures in the case of parameter estimating in the series considered.

3.2 Bootstrap methodology

The bootstrap method was introduced by Efron in 1979 and is a computer-intensive method for approximating the sampling distribution of any statistic derived from a random sample by independently sampling with replacement from an existing data sample with the same sample size. The main idea behind the bootstrap is that in the conditions of the lack of any information about the population distribution, the sample contains all the required information and using this information correctly can lead to good results (See [2], [4], [5] for more).

This method gives good results in the case when the observations are independent and identically distributed (i.i.d), but in the case of dependent observations, like time series, the i.i.d bootstrap gives incorrect answers (Singh (1981) [13], Babu and Singh (1983) [1]). Moving Block Bootstrap (MBB) proposed by Künsch (1989) [7] and Liu and Singh (1992) [8], is considered a good method for bootstrapping the data designed especially for stationary time series.

For the time series with a seasonal component, Politis (2001) [12] proposed the Seasonal Block Bootstrap (SBB) that is a version of Kunsch's (1989) block bootstrap with blocks whose size and starting points are integer multiples of the period. Also Dudek et. al [3] proposed a modification of the block bootstrap, the Generalized Seasonal Block Bootstrap (GSBB) that is suitable for periodic time series with fixed length periodicities of arbitrary size as related to block size and also to the sample size.

We propose to use a block bootstrap procedure, Block bootstrap based on residuals (BBR) that we adapted in the case of periodic models based on the idea from the Residual-based block bootstrap of Paparoditis and Politis ([11]).

3.3 R programming language

In this study R programming language is used. R is a free software environment for statistical computing and graphics (see [14] for more). We use this programming language to implement the algorithms proposed and to obtain and present the results.

4. Results and Discussion

We used R programming language in order to apply the bootstrap methods for the point estimation of the seasonal means and the overall mean in the Rainfall time series, as well as for the calculation of the Bias, Standard Deviation and the Mean Squared Error.

1000 bootstrap replications as well as 1000 Monte Carlo trials were used in the case of finding the bootstrap estimations of the seasonal means and the overall mean using the three procedures, GSBB, BBR and MBB (their circular version) and performing the comparison of the results with the point estimation. The chosen block length is b=5.

Figure 1 presents the Rainfall time series plot and also the ACF (Autocorrelation Function) plot:

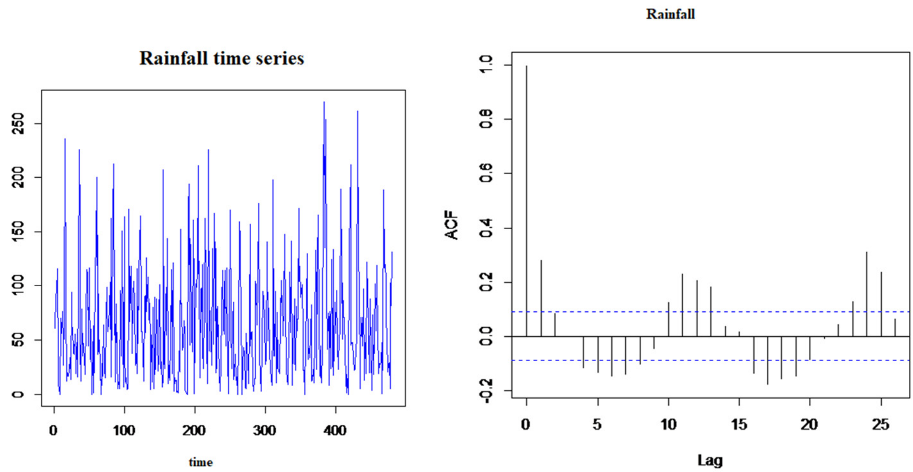

Figure 1. Rainfall time series plot of monthly rainfall values over a period of 40 years (left) and for the ACF (autocorrelation function) plot (right).

As expected in this time series, a periodic structure is observed. The period is thought to be T=12, but this needs to be confirmed by further analysis. For this, the periodogram function from the TSA package is used (see the figure below). The periodogram of the Rainfall time series is constructed and the period is found. As seen in the Figure 2 the period will be T=12.

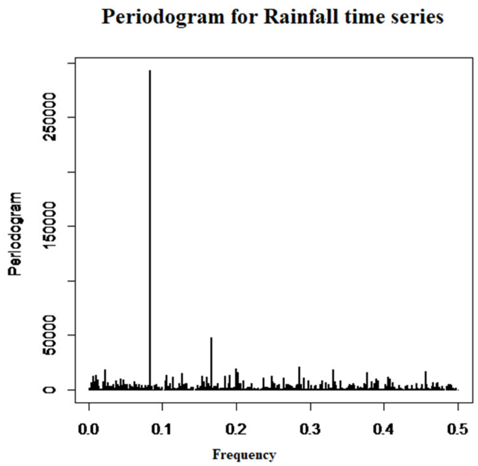

Figure 2. Periodogram for the Rainfall time series, obtained from the periodogram function, TSA package

We construct the graph of the squared coherence statistic, which is used to determine the periodic structure of the time series, as well as the strength of the periodic structure on the various support lines (See the figure below).

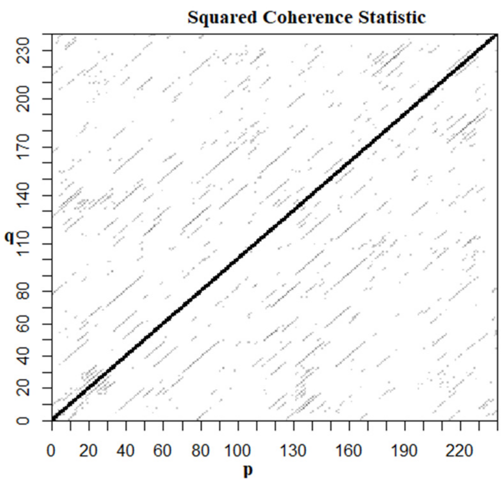

Figure 3. Squared coherence statistic, for the Rainfall time series (values passing the threshold $\alpha = 0.05$, applied to the time series after removing the periodic mean using M=16).

Knowing the period gives us the opportunity to calculate the main characteristics of the time series such as: periodic means and periodic standard deviations. For this purpose, we make the graphical presentation as a function of the season with a 95% confidence interval based on non-missing values for both of these characteristics.

In Figure 4, we graphically present the seasonal (periodic) means together with the 95% confidence intervals, while in Figure 5, we present the estimation of seasonal (periodic) standard deviations together with the 95% confidence intervals.

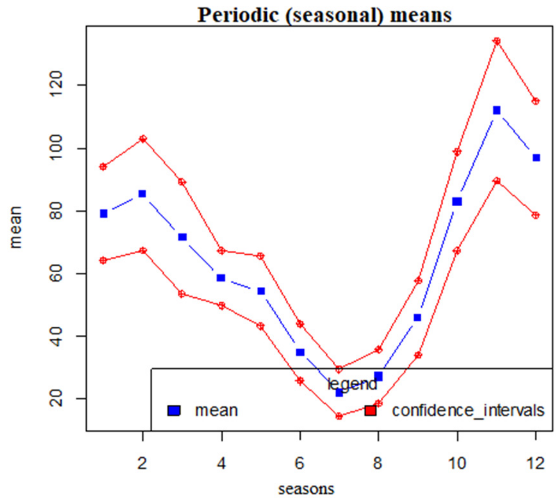

Figure 4. Point estimations of seasonal means along with 95% confidence intervals for the Rainfall time series

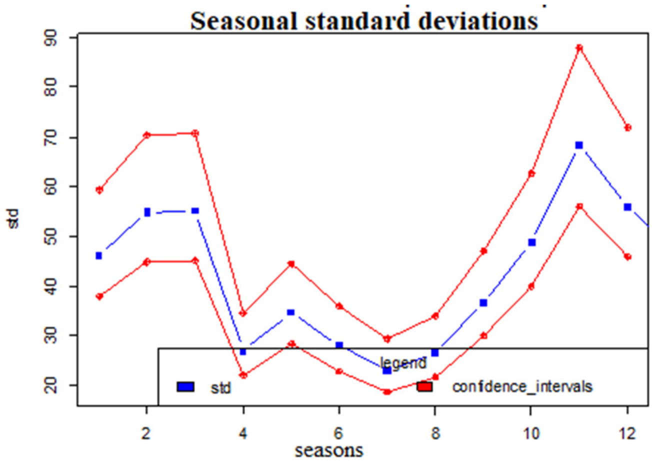

Figure 5. Point and interval estimations for periodic standard deviations for the Rainfall time series.

In the figures above it can be observed the lack of stationarity of the time series and a periodic structure based on the seasonal means and seasonal standard deviations. The applied test is One-Way ANOVA for equality of the seasonal means and Bartlett's test is applied for Homogeneity of variances. The p-value obtained from the test for homogeneity of variances is zero, which indicates that the hypothesis of equality of variances is rejected. This also applies to the case of seasonal means.

We are interested in the application of bootstrap methods in periodic time series and from the results above we confirmed the periodic structure of Rainfall time series and we found the period. In this time series we apply bootstrap methods in the case of point estimation of seasonal means and the overall mean. The methods used to obtain bootstrap estimations are GSBB, BBR and MBB (their circular version). The results are presented as follows:

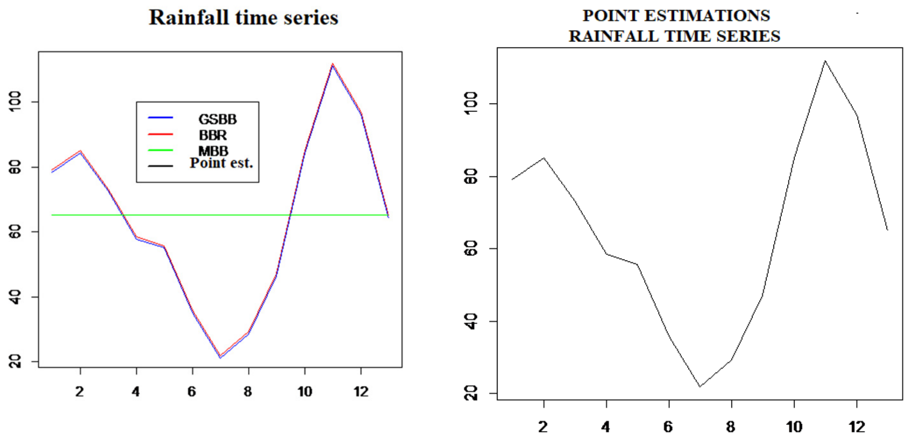

Figure 6. Estimations of the seasonal means and the overall mean for the Rainfall time series using GSBB, BBR, MBB (left) and the point estimations (right).

From Figure 6, we can see that the estimations obtained by the GSBB and BBR methods are close to the point estimations proposed for the seasonal means and the overall mean and are far from the estimations obtained using MBB. The estimations obtained using the MBB method for the seasonal means are very close to the estimations of the overall mean obtained by the three bootstrap methods and close to the classical point estimation of the overall mean.

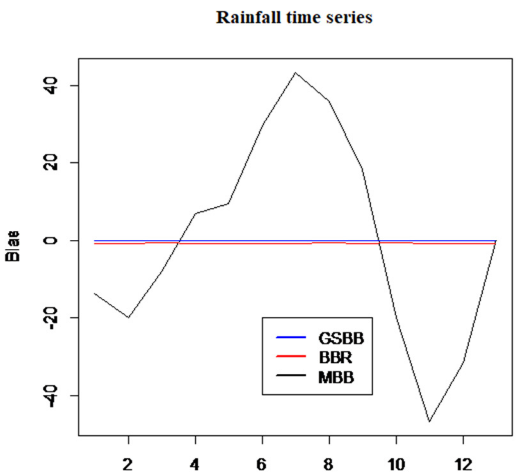

Figure 7. Bias for estimations of seasonal means and the overall mean when using the GSBB, BBR, MBB methods for the Rainfall time series.

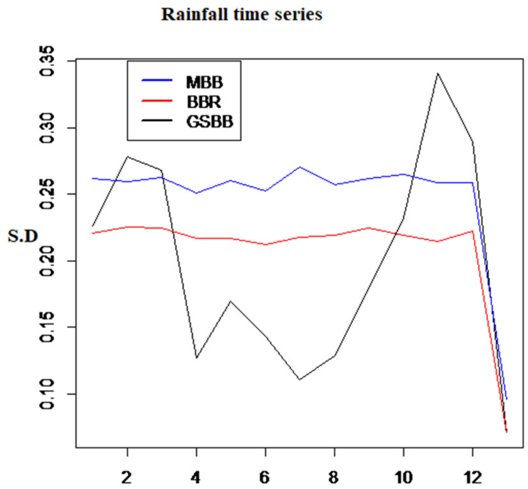

Figure 8. Standard deviation (S.D) for estimations of the seasonal means and the overall mean using the GSBB, BBR and MBB methods.

We can notice a similar and a good performance in the case of Bias for the estimations of seasonal means and the overall mean using the GSBB and BBR methods, but not a good performance of the MBB method (See Figure 7).

In the figure below we present the Mean Squared Error (MSE) of the estimations of the seasonal means and the overall mean, when using the three methods considered for the Rainfall time series.

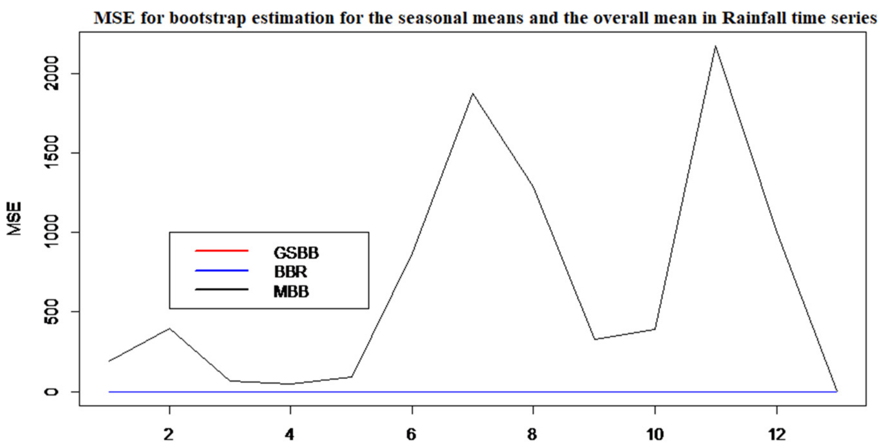

Figure 9. Mean squared error (MSE) for the estimations of seasonal means and the overall mean using the GSBB, BBR and MBB methods for the Rainfall time series.

A similar and quite good performance of the GSBB and BBR methods is observed, which are accompanied by small MSE values (there is an overlap of the graphs in the figure and they are indistinguishable from each other, see Figure 9). The

MBB method gives us fairly large values of MSE, since all seasonal means estimations are close to the estimation of the overall mean. Only in the case of the estimation of the overall mean we have an almost identical performance for the three methods.

5. Conclusions

From the results obtained when studying the performance of the three block bootstrap procedures in the case of Rainfall time series, the estimations obtained by the GSBB and BBR methods are very close to each other and also close to the classical point estimations used for periodically time series. The performance of the GSBB and BBR methods is quite good in this situation and this happens because we take in consideration the periodicity of the time series during the application of these methods. Also GSBB and BBR estimations are characterized by very small Bias and MSE values and we have a similar performance of the two methods.

The MBB method does not take into account the presence of periodicity and the results for the seasonal means are quite far from the results of the point estimations and the bootstrap estimations from the other two methods. Also MBB estimations of the seasonal means are characterized by large values of Bias and MSE.

Only in the case of estimating the overall mean we have a similar performance of the three methods.

References

Babu G. J., Singh K. (1983), Inference on means using the bootstrap. Ann. Statist. 11, 999-1003.
Davidson, C., Hinkley, D.V. (1997), Bootstrap Methods and their Application. Cambridge: Cambridge University Press.
Dudek A. E., Leskow J., Paparoditis E., Politis D. N. (2014), A generalized block bootstrap for seasonal time series, Journal of Time Series Analysis.
Efron B. (1979), Bootstrap methods: Another look at the jackknife, Ann. Statis. 7, 1-26.
Efron B., Tibshirani R. J. (1993), An Introduction to the Bootstrap. New York: Chapman and Hall.
Horowitz J. L. (2019), Bootstrap Methods in Econometrics, Annual Review of Economics 2019 11:1, 193-224
https://www.r-project.org/
Kunsch H. (1989), The Jackknife and the Bootstrap for general stationary observations, Ann. Stat.17, 1217-1241.
Liu R.Y., and Singh K., (1992), Moving Blocks Jackknife and Bootstrap Capture Weak Convergence. In: Lepage, R. and Billard, L., Eds., Exploring the Limits of Bootstrap, Wiley, New York, 225-248.
Margo L., Ekonomi L. (2017), Some empirical results using bootstrap procedures in periodically correlated time series models, Applied Mathematical sciences, Vol. 11, 2017, no. 11, 539-547.
Margo Zeqo L. (2023), Twenty-Sixth International Conference on: "Social and Natural Sciences-Global Challenge 2023" (ICSNS XXVI-2023), Berlin, Germany, Comparison of bootstrap procedures in estimating seasonal means in AROSA time series, pp. 132, ISBN: 978-9928-178-21-3.
Paparoditis E., Politis D. N. (2003), Residual-based block bootstrap for unit root testing, Econometrica, Vol. 71, No. 3, May 2003, 813-855.
Politis (2001), D. N., Resampling time series with seasonal components, in Frontiers in Data Mining and Bioinformatics. In: Proceedings of the 33rd Symposium on the Interface of Computing Science and Statistics, Orange County, California, pp. 619–621.
Singh K. (1981), On the asymptotic accuracy of Efron's bootstrap. Ann. Statist. **9** 1187–1195.