Journal of Data Science 14(2016), 641-656

AN EMPIRICAL COMPARISON OF BLOCK BOOTSTRAP METHODS: TRADITIONAL AND NEWER ONES

Beste H. Beyaztas[a,b,] Esin Firuzan[b*]

[a]*Department of Statistics, Istanbul Medeniyet University, Istanbul, Turkey*
[b]*Department of Statistics, Dokuz Eylul University, Izmir, Turkey*

Abstract: In this study, we compared various block bootstrap methods in terms of parameter estimation, biases and mean squared errors (MSE) of the bootstrap estimators. Comparison is based on four real-world examples and an extensive simulation study with various sample sizes, parameters and block lengths. Our results reveal that ordered and sufficient ordered non-overlapping block bootstrap methods proposed by Beyaztas et al. (2016) provide better results in terms of parameter estimation and its MSE compared to conventional methods. Also, sufficient non-overlapping block bootstrap method and its ordered version have the smallest MSE for the sample mean among the others.

Keywords: Block bootstrap, bootstrap, estimation, linear time series, sufficient bootstrap

1. Introduction

The original non-parametric bootstrap method proposed by Efron (1979) does not work for dependent data, since the assumption of independently and identically distributed data is violated. Block bootstrap method is one of the techniques to extend this method to serially correlated data. In this technique, the series of size n is divided into blocks consisting of l consecutive observations. These blocks are then resampled with equal probability, and then pasted end-to-end to form the bootstrap series, in which the dependency structure of the original data can be preserved, at least within the adjacent observations in each block. This method is used in a wide variety of fields. For example, Suda et al. (2009) developed an automatic real-time monitoring system for deep nonvolcanic tremors in southwest Japan. Srinivas and Srinivasan (2005) presented a method for resampling multiseason hydrologic time series. Amiri and Zwanzig (2011) proposed a family of tests based on the bootstrap method and applied it to chemical experiments. Also, Sherman et al. (2010) analyzed tidal data using blockwise bootstrap in regression setting.

In blocking technique, the blocks may be comprised of non-overlapping or overlapping subsets from the original series. The non-overlapping block bootstrap (NBB) approach was proposed by Carlstein (1986), and overlapping blocks known as moving block bootstrap (MBB) were independently proposed by Künsch (1989) and Liu and Singh (1992). In these techniques, let b and l denote the number of blocks and block length, respectively. In MBB, the first and last l – 1 observations appear less frequently than the rest. To fix this disadvantage, the circular block bootstrap method (CBB) was suggested by Politis and Romano (1992) and Shao and Yu (1993) by wrapping the data around a circle so that each observation in the original data set has an equal probability. In addition to these methods, Politis and Romano (1994) proposed a stationary bootstrap method (SB) with random block lengths which have a geometric distribution. Recently, Beyaztas et al. (2016) proposed sufficient and ordered versions of the NBB methods to provide better inference in linear time series analysis. For sufficient NBB, they proposed replacing sufficient bootstrapping within the NBB to reduce the computing time and obtain more efficient results for sample mean. Since the sufficient bootstrap proposed by Singh and Sedory (2011) uses only distinct units in resamples, this method can only be extended to the NBB method. Because, for other block bootstrap methods, although the selected blocks are distinct, the observations in blocks may be the same since the consecutive blocks have the same observations. In more detail, let our time series data be $Y = (y_1, y_2, \dots, y_{12})$ and let *l=3*, then the data is represented with blocks as $b_1 = (y_1, y_2, y_3)$, $b_2 = (y_4, y_5, y_6)$, $b_3 = (y_7, y_8, y_9)$ and $b_4 = (y_{10}, y_{11}, y_{12})$. Suppose the bootstrapped blocks are $b_1^* = (y_1, y_2, y_3)$, $b_2^* = (y_1, y_2, y_3)$, $b_3^* = (y_{10}, y_{11}, y_{12})$ and $b_4^* = (y_7, y_8, y_9)$. Then, the new data is obtained by NBB as $Y_1^* = (y_1, y_2, y_3, y_1, y_2, y_3, y_{10}, y_{11}, y_{12}, y_7, y_8, y_9)$. However, it is obtained as $Y_{\upsilon_1}^* = (y_1, y_2, y_3, y_{10}, y_{11}, y_{12}, y_7, y_8, y_9)$, where ι denotes the expected number of distinct observations, when sufficient NBB (SNBB) is used. Since the first and second blocks are the same, sufficient block bootstrap discards one of them. For ordered NBB (ONBB) and its sufficient version (SONBB), Beyaztas et al. (2016) proposed ordering the bootstrapped blocks according to given labels to each original block for capturing more dependence structure compared to the conventional NBB method. For the same example mentioned above, the labels are determined as $b_1 = 1$, $b_2 = 2$, $b_3 = 3$ and $b_4 = 4$, and then each bootstrapped block is sorted according to these labels $b_1^* = 1$, $b_2^* = 2$, $b_4^* = 3$ and $b_3^* = 4$, and pasted end-to-end to obtain new data. Again, to obtain SONBB data, the repeated blocks are discarded. Hence, the new data set is obtained by the ONBB as $Y_{O,n}^* = (y_1, y_2, y_3, y_1, y_2, y_3, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}$. However, it is obtained as $Y_{O,\upsilon}^* = (y_1, y_2, y_3, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}$when the SONBB method is used.

Bootstrap methods have been involved in a large variety of applications with different nature, ranging from natural phenomena to biomedical researches. That's why, to assess the performances of the block bootstrap methods mentioned in this study, four real-world data sets are chosen from different fields and linear time series models. The rest of the paper is organized as follows. In Sections 2 and 3, we present the results of the simulation and case studies, respectively. Some concluding remarks are given in Section 4.

2. Simulation Study

To assess the performances of the aforementioned block bootstrap methods, we conduct a simulation study under first order autoregressive (AR(1)) and moving average (MA(1)) models with $\phi_1 = \theta_1 = -0.9, -0.5, -0.1, 0.1, 0.5, 0.9$ autoregressive and moving average parameter values. The sample sizes and block lengths are chosen as n=100, 250 and l=5, 10, respectively. The comparisons are made in terms of biases and MSEs of the block bootstrap estimators of ϕ_1 andθ_1. The number of Monte Carlo simulations MC are set at MC=1000, and for each simulation, the bootstrap replicates B=1000 are used. It should be noted that all the calculations were carried out using R 3.1.1. The simulation results are presented in Figures 1-4. As presented in these figures, the SONBB estimators of autoregressive and moving average parameters have the smallest bias and MSE values among all the block bootstrap estimators. Also, another prominent method is the ONBB with smaller biases and MSEs compared to other block bootstrap methods. Moreover, the estimators obtained by the ONBB and SONBB methods are more robust to the parameters close to unity which means that the process is close to nonstationarity. We also performed a simulation study for the first order autoregressive and first order moving average model (ARMA(1,1)) with various parameters and sample sizes. Since our conclusions do not vary significantly, therefore to save space, the numerical details are omitted for the mixed model. As a consequence, considering the simulation results, the ordered versions of the NBB methods produce more stable and reliable estimators than those obtained using the conventional methods in the context of linear time series models.

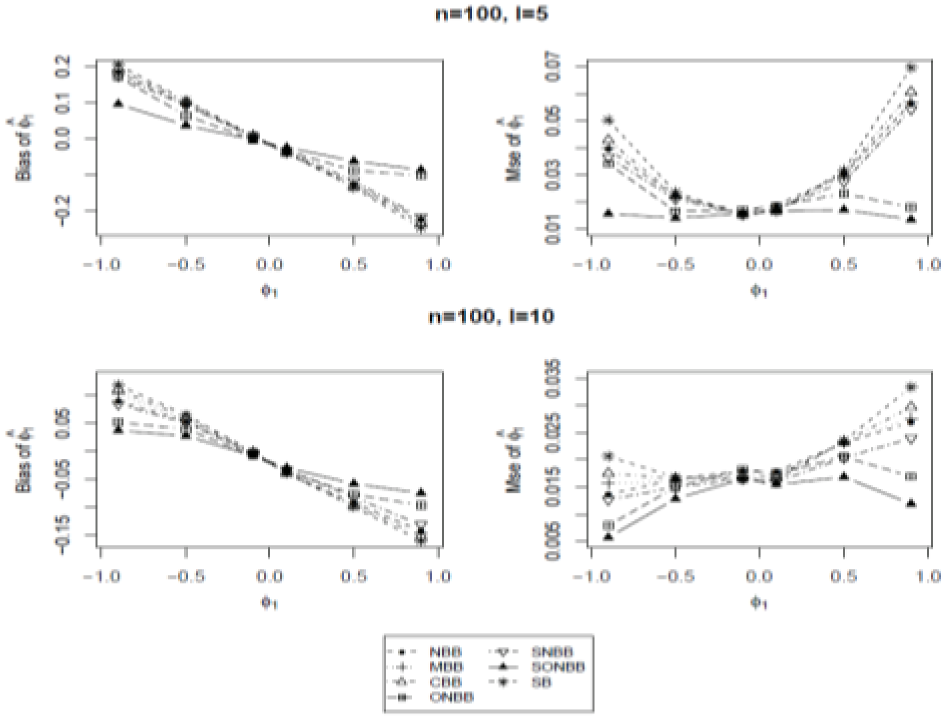

Figure 1. Plots of block bootstrap methods for AR(1) model with sample size n=100, and block lengths l=5, 10

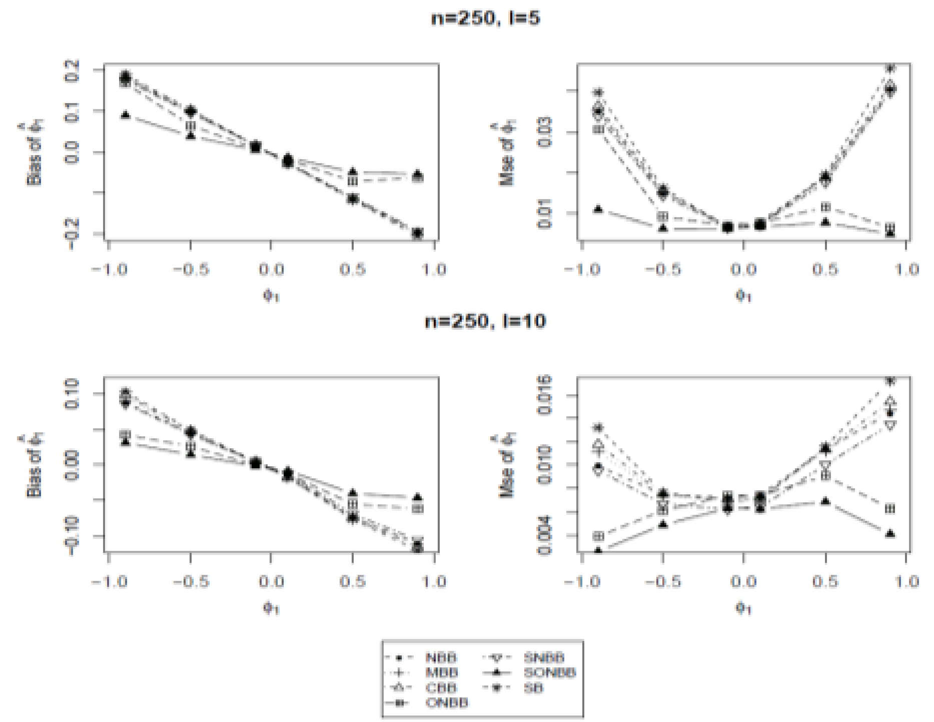

Figure 2. Plots of block bootstrap methods for AR(1) model with sample size n=250, and block lengths l=5, 10

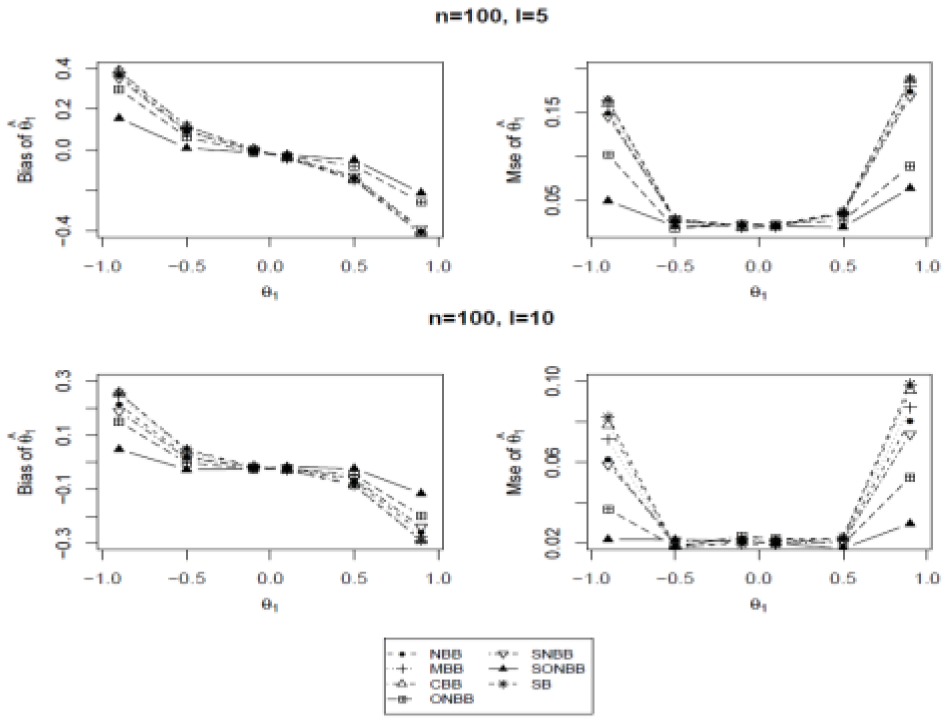

Figure 3. Plots of block bootstrap methods for MA(1) model with sample size n=100, and block lengths l=5, 10

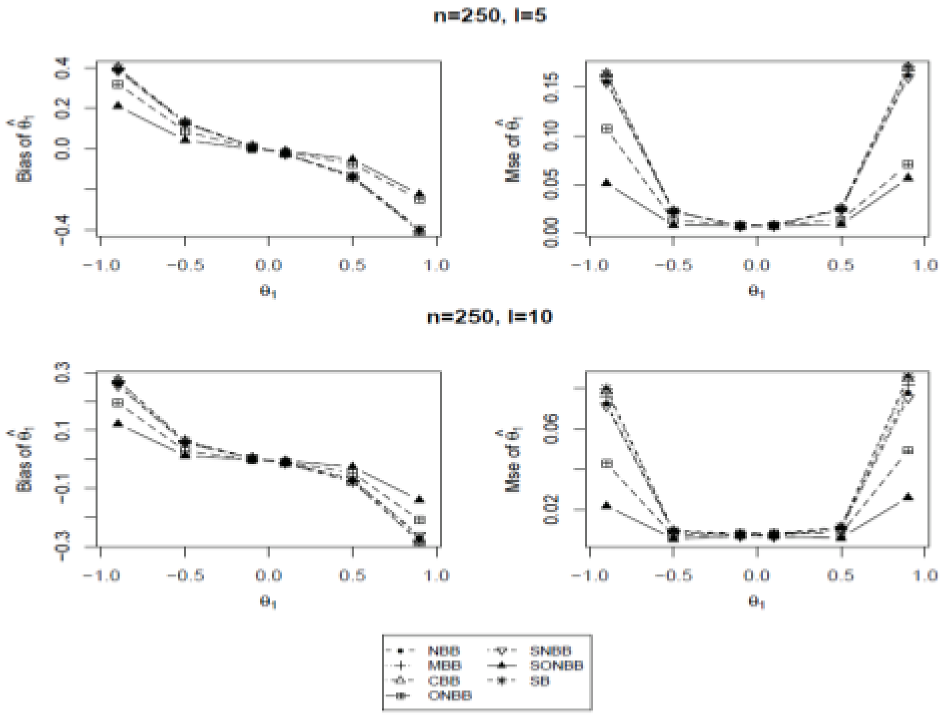

Figure 4. Plots of block bootstrap methods for MA(1) model with sample size n=250, and block lengths l=5, 10

3. Case Study

Next subsections serve to compare of the performances of traditional block bootstrap and the newer ones in the application ground. The performances of the methods were compared in terms of parameter estimation of univariate linear time series models and MSEs of the relevant block bootstrap estimators. The data sets used in this study were chosen with different sample sizes and block lengths. The block bootstrap estimates and their MSE values are calculated based on B=1000 bootstrap replicates. The descriptive statistics of the four real-world data are given in Table 1. Also, the descriptive statistics and Anderson-Darling normality test results for the residuals of the fitted models are given in Table 2.

Table 1. The descriptive statistics of the real-world data

	Operation time	Number of earthquake	Chemical concentration	Unemployment rate
Sample size	70	71	180	241
Mean	488.360	20.141	17.064	4.388
Std. dev.	158.750	7.374	0.395	2.776
Skewness	0.147	0.568	0.229	0.911
Kurtosis	-0.413	0.446	-0.128	-0.440
Block lengths	2, 5, 7, 10	2, 5, 7, 10	3, 5, 9, 12, 15	4, 6, 8, 10, 12, 16

Table 2. The descriptive statistics and normality test results for the residuals of the fitted models

Real-world data	Fitted model	Mean	Std. deviation	Skewness	Kurtosis	Anderson-Darling normality test
Liver operation time	AR(1)	-0.336	151.556	0.266	-0.214	0.231 (p-value=0.797)
Number of earthquakes	IMA(1,1)	-0.075	6.088	0.382	0.630	0.557 (p-value=0.144)
Chemical concentration	ARMA(1,1)	-0.003	0.309	0.356	0.895	0.531 (p-value=0.172)
Unemployment rate	ARI(1,1)	0.000	0.136	0.600	4.070	2.934 (p-value=0.000)

3.1 Liver transplantation operation time

The operation time of liver transplantation patients who had surgery in an university hospital between January 2003 and December 2013 is examined in detail. The liver transplantation operation time is modeled as AR(1) process with Maximum Likelihood estimate (MLE) as in Equation (1).

$$Y_t = 344.06 + 0.298 Y_{t-1} + \varepsilon_t , \quad t \in Z \tag{1}$$

where ε_t follows a N(0,1) distribution.

As reported in Table 3, the calculated p-value < 0.05 of t-tests indicated that both constant term δ and ϕ_1 are statistically significant at %5 level of significance. Beside, Ljung-Box Q-statistics (please see Table 4) to test for autocorrelations in residuals shows that the residuals are not auto-correlated. Also, the p-value > 0.05 of Anderson-Darling test (see Table 2) indicates that the residuals are normally distributed. All of exploratory analysis indicates that AR(1) model is a suitable choice to model operation time data.

The parameter estimates and their MSEs obtained by the block bootstrap methods are presented in Figure 5. As it is seen from this figure, the true parameter value is captured by the SONBB method in a better way. Although the results of the SNBB are not as good as the results of the SONBB for parameter estimation, it has almost the same MSE values with the SONBB since SNBB provides smallest variances for $\hat{\phi}_1$ compared to others. It means that the variance of parameter which is obtained by the SNBB is smaller than the one obtained by the SONBB, but more biased for parameter estimation in comparison with the SONBB. The results show that using SONBB method gets less biasedness and less error for AR(1) parameter.

The worst estimation performance belongs to the MBB, additionally CBB and SB have the similar performances which are not satisfactory. When the block length is 7, all the non-overlapping block bootstrap methods' (i.e. NBB, SNBB, ONBB, SONBB) performances are in the same manner.

Table 3. AR(1) model estimates for Liver operation time data

Parameter	Standard error (s.e.)	t-stat.	p-value
δ	17.981	19.231	0.000
ϕ_1	0.115	2.589	0.009

Table 4. Ljung-Box Q-Statistics for the residuals of the estimated AR(1) model for Liver operation time data

Considered lag	Ljung-Box test	Degrees of freedom	Significance level
12	5.285	10	0.871
24	10.324	22	0.983
36	20.515	34	0.967

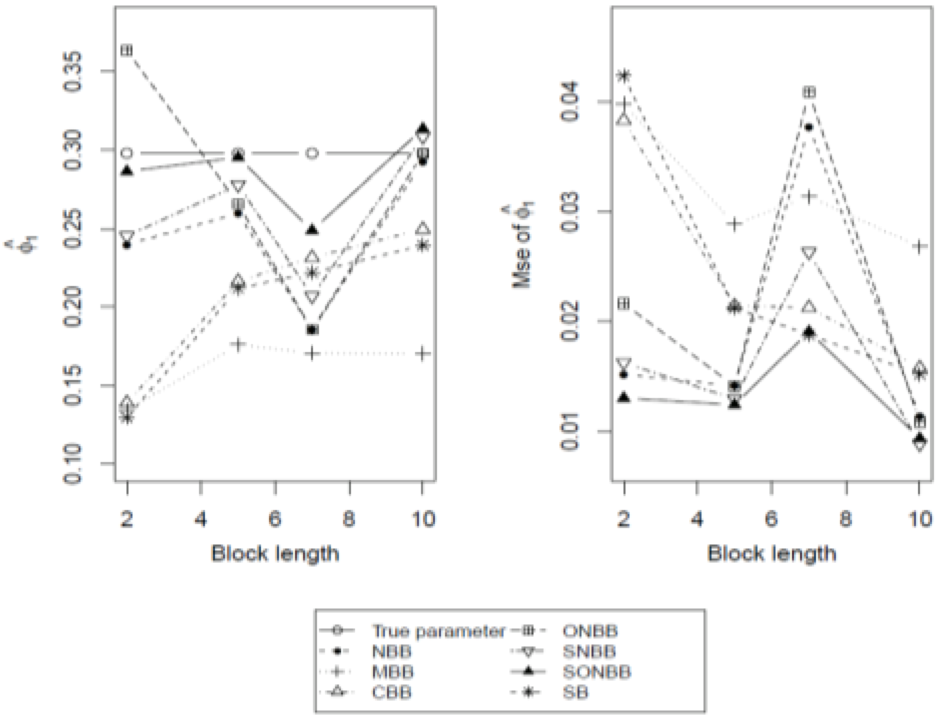

Figure 5. Plots of block bootstrap methods for $\hat{\phi}_1$ and its MSE for Liver operation time data

3.2 Number of earthquakes

The number of earthquakes per year which has occurred greater than magnitude 7.0 over the 1928-1998 period is analyzed. The data is obtained from http://datamarket.com/data/list. For this data set, by using Box-Jenkins' methodology the integrated moving average model of order 1 (IMA(1,1)) is fitted as in Equation (2) and the MLE of θ_1 is statistically significant at the 0.05 level (see Table 5).

$$\Delta Y_t = \varepsilon_t - 0.623\varepsilon_{t-1},\ t \in Z \tag{2}$$

where $\{\varepsilon_t\}$ is a sequence of normally distributed random variables, and $\Delta^1 Y_t = (1-B)^1 Y_t - Y_{t-1}$ show the first-order difference of {Yt} where B and Δ denote the backshift and difference operators, respectively.

The results of Ljung-Box autocorrelation test for this data set is given in Table 6, and it shows that the residuals do not have statistically significant autocorrelation. Moreover, descriptive statistics and normality test results of the residuals given in Table 2 show that the residuals follow the normal distribution.

In Figure 6, we plot the estimated parameters and MSE values of the block bootstrap estimators. As in the first example, SONBB provides better results among the others even in small block lengths which is the point of our study. The other methods tend to have the same trend for both estimation of $\hat{\theta}_1$and its MSE. The results of all the bootstrap methods become more consistent with increasing block length.

Table 5. IMA(1,1) model estimates for Number of earthquakes data

Parameter	Standard error (s.e.)	t-stat.	p-value
θ_1	0.088	7.046	0.000

Table 6. Ljung-Box Q-Statistics for the residuals of the estimated IMA(1,1) model for Number of earthquakes data

Considered lag	Ljung-Box test	Degrees of freedom	Significance level
12	10.404	11	0.494
24	25.024	23	0.349
36	41.329	35	0.214

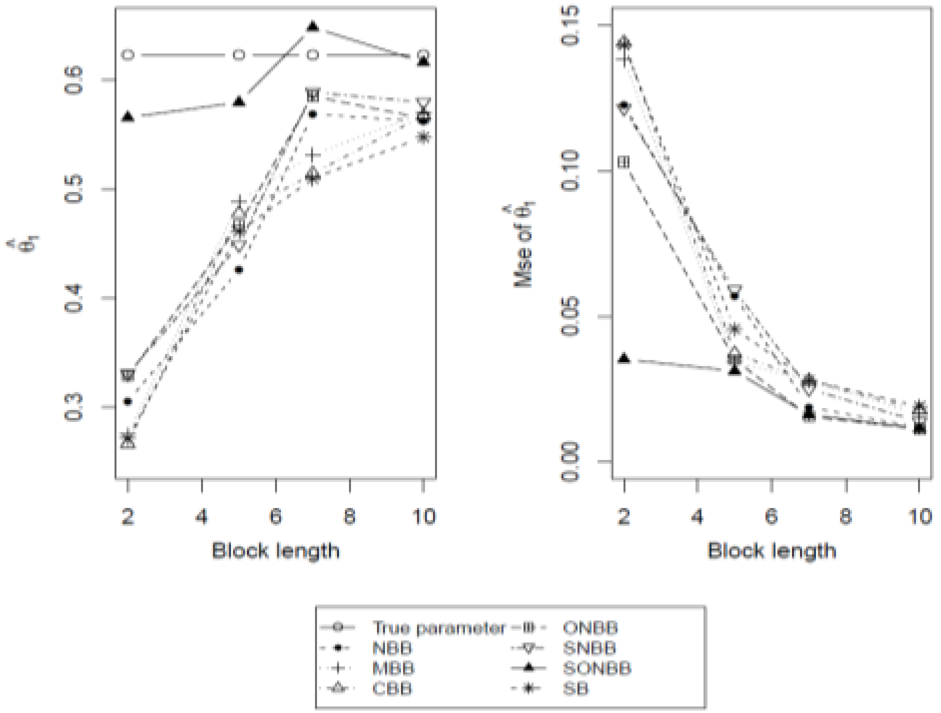

Figure 6. Plots of block bootstrap methods for $\hat{\theta}_1$ and its MSE for Number of earthquakes data

3.3 Chemical concentration

Chemical concentration data is chosen because of being an example from ARMA(1,1) model. The original data is given as Series A in Box-Jenkins (1976). The analysis is implemented for between 14th - 193th observations. The estimated ARMA(1,1) model based on Box-Jenkins methodology is represented as in Equation (3). The results of residual analysis given in Table 8 and Table 2 indicate that the residuals of the fitted model are normally distributed and not serially correlated. Therefore, the p-value < 0.05 of t-tests presented in Table 7 shows that ARMA(1,1) model appears adequate at the significance level under 0.05.

$$Y_t - 0.927\,Y_{t-1} = 1.247 + \varepsilon_t - 0.612\,\varepsilon_{t-1},\ t \in Z \qquad (3)$$

where $\{\varepsilon_t\}$ is a sequence of normally distributed random variables.

Figure 7 and 8, respectively, show the results of the block bootstrap estimators of ϕ_1 and θ_1. It is clear that the ONBB and SONBB produce more preferable results for both parameter estimations and their MSEs while the other methods behave in the same manner and their estimated values get close to the results of the ONBB and SONBB as l increases. The point of interest in this example is that the results of the ONBB and SONBB grow away from the true parameter value as block length increases. The interpretation of this could be that the ONBB and SONBB may be preferred for small block lengths unlike the other methods.

Table 7. ARMA(1,1) model estimates for Chemical concentration data

Parameter	Standard error (s.e.)	t-stat.	p-value

δ	0.048	19.154	0.000
ϕ_1	0.008	143.681	0.000
θ_1	0.113	5.406	0.000

Table 8. Ljung-Box Q-Statistics for the residuals of the estimated ARMA(1,1) model for Chemical concentration data

Considered lag	Ljung-Box test	Degrees of freedom	Significance level
12	13.951	9	0.124
24	23.397	21	0.323
36	45.228	33	0.076

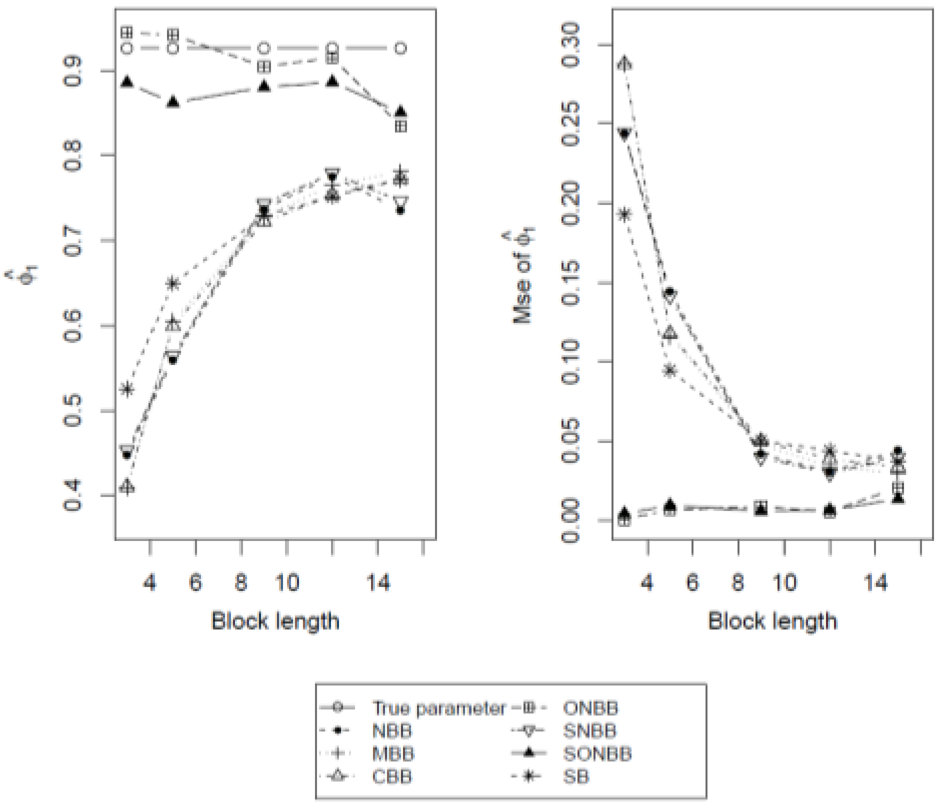

Figure 7. Plots of block bootstrap methods for $\hat{\phi}_1$ and its MSE for Chemical concentration data

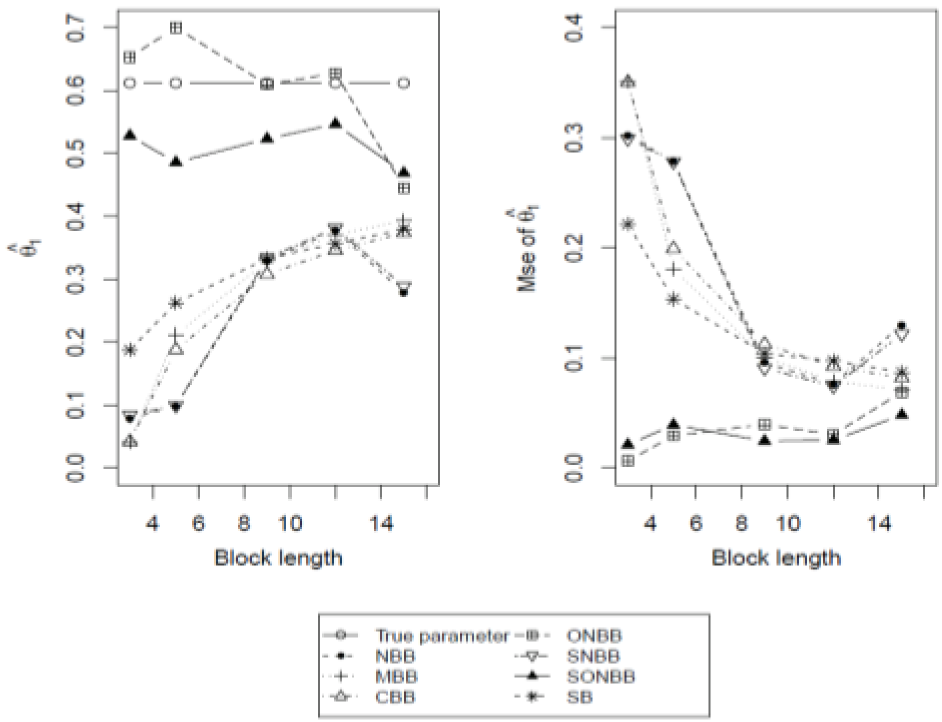

Figure 8. Plots of block bootstrap methods for $\hat{\theta}_1$ and its MSE for Chemical concentration data

3.4 Unemployment rate

The seasonally adjusted registered unemployment rate of United Kingdom is discussed as an example of labour market statistics. The quarterly data (01/01/1955 – 01/01/2015 time period) were obtained from https://research.stlouisfed.org/fred2/series.rate. The autoregressive integrated model (ARI(1,1)) is fitted as in Equation (4), and Table 9 shows that the estimated coefficient is statistically significant at %5 level of significance. Also, Ljung-Box Q-statistics reported in Table 10 indicate that there is no evidence of autocorrelation in the residuals.

$$\Delta Y_t = 0.842\,\Delta Y_{t-1} + \varepsilon_t, \quad t \in Z \tag{4}$$

where $\{\varepsilon_t\}$ is a sequence of random variables which has normal distribution, and $\Delta^1 Y_t = Y_t - Y_{t-1}$ is the first difference of {Yt} process.

Table 2 indicates that the residuals are not normally distributed. All of empirical analysis lead us to consider an ARI(1,1) model as a suitable choice in the context of linear time series models although the residuals do not follow normal distribution.

In this example, the results of the SONBB are better than the others especially in small l values as shown in Figure 9. But the difference between the performances is not statistically significant for large block lengths.

As it is seen from the Figure 10, sufficient versions of the NBB method have the smallest MSE values for the sample mean compared with the conventional block bootstrap methods. The reason of this is based on using distinct observations in the resamples. For more information see Pathak (1961).

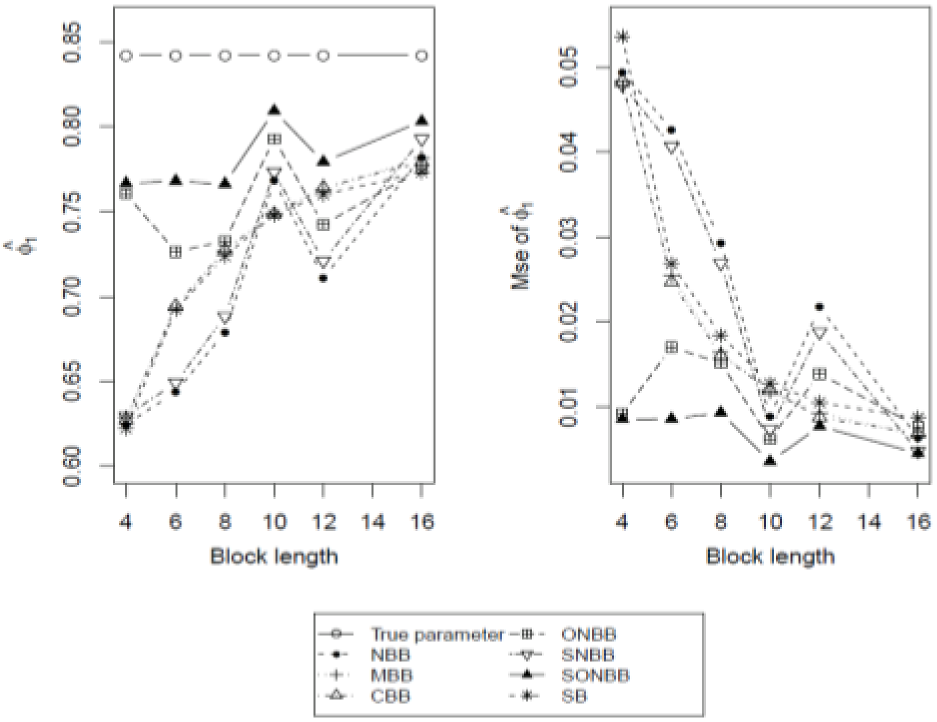

Figure 9. Plots of block bootstrap methods for $\hat{\phi}_1$ and its MSE for Unemployment rate data

Table 9. ARI(1,1) model estimates for Unemployment rate data

Parameter	Standard error (s.e.)	t-stat.	p-value
ϕ_1	0.034	24.548	0.000

Table 10. Ljung-Box Q-Statistics for the residuals of the estimated ARI(1,1) model for Unemployment rate data

Considered lag	Ljung-Box test	Degrees of freedom	Significance level
12	18.684	11	0.067
24	34.537	23	0.058
36	47.971	35	0.071

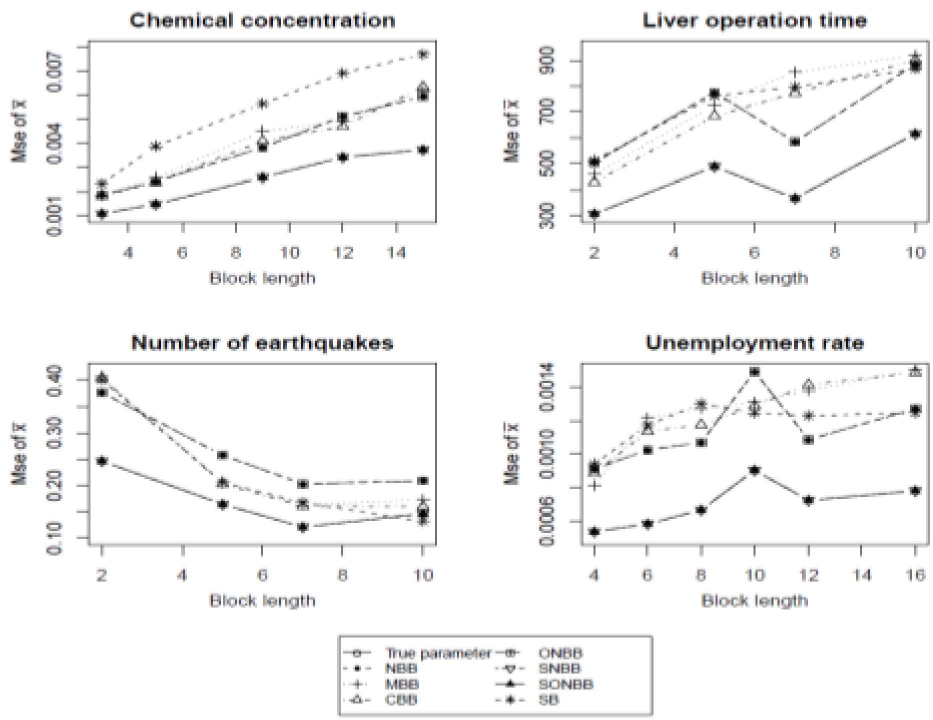

Figure 10. Plots of block bootstrap methods for MSE of the sample means

4. Conclusion

In this study, the performances of four different conventional block bootstrap methods are compared to the newer ones' performances. The evaluation of the performances have been done in terms of estimations of the parameters and their MSEs with a simulation study and four real-world data sets which come from different linear time series models and different fields. The empirical findings for the parameter estimation under considered data sets in this study are consistent with the simulation results presented by Beyaztas et al. (2016).

One of the important results observed from the simulation study is that the estimators obtained by the SONBB have the significantly smaller biases and MSEs compared to other block bootstrap methods. Moreover, the SONBB and ONBB methods have more stable estimators under various parameter values considered in this paper.

For the real-world examples, SONBB method provides better results for the data sets under consideration even for small block sizes in general. For Liver operation time and Registered unemployment rate of UK data sets, the non-overlapping block bootstrap method and its sufficient and ordered versions fluctuate in an unstable manner for different block lengths while these fluctuations do not occur for other linear time series models. MSE values of the sample means for the sufficient versions of the non-overlapping block bootstrap methods are smaller than the other block bootstrap methods' for each data because of using only distinct observations in the resamples.

As a future research, the performances of the newer methods can also be examined under unit root processes or can be studied in other areas such as testing mixture models as studied by Yang et. al. (2010). Finally, it should be noted that all the calculations in this study are performed under the assumption that the considered model forms are correct. Also, changing the block length affects overall results, but it does not make any different decision about the comparison among performances of the bootstrap methods.

References

[1] Amiri, S. and Zwanzig, S. (2011). Assessing the coefficient of variations of chemical data using bootstrap method. Journal of Chemometrics 25, 295-300.

[2] Beyaztas, B. H., Firuzan, E. and Beyaztas, U. (2016). New block bootstrap methods: Sufficient and/or Ordered. Communications in Statistics: Simulation and Computation. DOI: 10.1080/03610918.2015.106680

[3] Box, E. P. and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control, Revised Edition. Holden Day, San Francisco.

[4] Carlstein, E. (1986). The use of subseries values for estimating the variance of general statistics from a stationary sequence. Annals of Statistics 14, 1171-1194.

[5] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Annals of Statistics 7, 1-26.

[6] Künsch, C. H. (1989). The jackknife and the bootstrap for general stationary observations. Annals of Statistics 17, 1217-1241.

[7] Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. Exploring the limits of bootstrap, R. LePage and L. Billard. Wiley, New York.

[8] Politis, D. N. and Romano, J. P. (1992). A circular block resampling procedure for stationary data. Exploring the limits of bootstrap, R. LePage and L. Billard. Wiley, New York.

[9] Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. Journal of American Statistical Association 89, 303-1313.

[10] Shao, J. and Yu, H. (1993). Bootstrapping the sample means for stationary mixing sequences. Stochastic Processes and their Applications 48, 175-190.

[11] Sherman, M., Speed, Jr F. M. and Speed, F. M. (2010). Analysis of tidal data via the blockwise bootstrap. Journal of Applied Statistics 25, 333-340.

[12] Singh, S. and Sedory, S. A. (2011). Sufficient bootstrapping. Computational Statistics and Data Analysis 55, 1629-1637.

[13] Srinivas. V. V. and Srinivasan, K. (2005). Matched block bootstrap for resampling multiseason hydrologic time series. Hydrological Process 19, 3659-3682.

[14] Suda, N., Nakata, R. and Kusumi, T. (2009). An automatic monitoring system for nonvolcanic tremors in southwest Japan. Journal of Geophysical Research-Solid Earth 114, 1-15.

[15] Yang, Y., Fan, J. and May, S. (2010). Bimodality of plasma glucose distributions in whites: A bootstrap approach to testing mixture models. Journal of Data Science 8, 483-493.

Esin Firuzan
Department of Statistics, Dokuz Eylul University
35160 Buca-Izmir/TURKEY
Tel: +90 232 301 85 57
Fax: +90 453 41 88
E-mail: esin.firuzan@deu.edu.tr

Beste Hamiye Beyaztaş
Department of Statistics, Istanbul Medeniyet University
34000 Kadikoy-Istanbul/TURKEY
E-mail: beste.sertdemir@medeniyet.edu.tr