World Applied Sciences Journal 21 (Mathematical Applications in Engineering): 159-165, 2013
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2013.21.mae.99944

Analysis of Time Series by Re-Sampling

B.I. Mohamed, F.A.M. Elfaki, J.I. Daoud and M. Azram

Department of Science in Engineering, Faculty of Engineering,
IIUM, P.O. Box: 10, 50728 Kuala Lumpur, Malaysia

Abstract: The Box-Jenkins methodology is very often used in financier when the time series are analyzed. The estimations of the parameters of the selected models are one of the first tasks of the analysis. The important problem that emerges in connection with the parameters estimation is the problem of their accuracy. This accuracy is often characterized by the bias and standard deviation. When we want to determine these characteristics by the exact methods some problems often emerge. One possibility of the solution of these problems is the bootstrap methods application. Three different approaches of the application of these methods in the autoregressive model are demonstrated in this paper. Simulation studies are conducted to evaluate the methods.

Key words: Bootstrap · Block Bootstrap · Least Square Methods · and Standard deviation

INTRODUCTION

The bootstrap is a form of a larger class of methods that resample from the original data set and thus are called resampling procedures. The term 'bootstrapping,' due to [1], is an allusion to the expression 'pulling oneself up by one's bootstraps' - in this case, using the sample data as a population from which repeated samples are drawn. It is a method for estimating the distribution of an estimator or test statistic by resampling one's data or a model estimated from the data. The methods that are available for implementing the bootstrap and the accuracy of bootstrap estimates depend on whether the data are an Independently and Identically Distributed (IID) random sample or a time series. If the data are IID, the bootstrap can be implemented by sampling the data randomly with replacement or by sampling a parametric model of the distribution of the data. The distribution of a statistic is estimated by its empirical distribution under sampling from the data or parametric model. [2, 3, 4, 5] provide detailed discussions of bootstrap methods and their properties for data that are sampled randomly from a distribution. Some results are available on the practical application of the bootstrap to time series models. These results apply to stationary Auto-Regressive (AR) processes, which is a subset of the stationary Auto-Regressive-Moving Average (ARMA) models, will be discussed in section 2.1, with illustration of how the bootstrap can be applied to an autoregressive model.

The block bootstrap is the best-known method for implementing the bootstrap with time-series data. It consists of dividing the data into blocks of observations and sampling the blocks randomly with replacement. The blocks may be non-overlapping ([6], [7]) or overlapping ([6], [8], [9]).

To describe blocking methods more precisely, let the data consist of observations $\{X_i;\ i = 1,...,l\}$. In general, the parameter of interest may depend on current and lagged values of X up to order $q \geq 0$. An example is $\theta = E[G(X_1,....,X_{1+q}$, where G is a known function. We assume that $q < \infty$. [9] discuss the case of $q = \infty$ (e.g, spectral density estimation). See, also, [10]. For $q < \infty$, define $Y_i = \{X_i,....,X_{i-q}\}$. With non-overlapping blocks of length l, block 1 is observations $\{Y_j: j = 1,...,l\}$, block 2 is observations $\{Y_{1+j} : j = 1,...,l\}$ and so on. With overlapping blocks of length l, block 1 is observations $\{Y_j:\ j = 1,....,l\}$ block 2 is observations $\{Y_{1+j}:\ j = 1,...,l\}$ and so forth. The bootstrap sample is obtained by sampling blocks randomly with replacement and laying them end-to-end in the order sampled. The procedure of sampling blocks of Yi's instead of Xi's is called blocks-of-blocks bootstrap. Monte Carlo experiments have shown that the blocks-of-blocks bootstrap methods provides greater finite-sample accuracy than does sampling blocks of Xi's, but the two methods have the same (higher-order) asymptotic properties.

Corresponding Author: F.A.M. Elfaki, Department of Science in Engineering, Faculty of Engineering, IIUM, P.O. Box: 10, 50728 Kuala Lumpur, Malaysia.

MATERIALS AND METHODS

Auto-regressive Scheme of Order ρ AR (ρ): In time series we have two main models, Autoregressive, AR (p) model and Moving Average, MA (q) model, where p and q are the order of the AR and MA. According to the value of p and q, we can get $AR(1), AR(2),...,AR(p)$ and $MR(1)$, MR(2),...,$MR(p)$. In this paper we deal with $AR(\rho)$.

Let $..., X_{-2}, X_{-1}, X_0, X_1, X_2,...$, be a time series. When for their single elements hold true $E(Y_t) = \mu$ and $\text{cov}(X_t, X_{t+r}) = kr$ or every integer t,r then this time series is weakly stationary stochastic process. Then the X_t will be obtained from the following equation:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + E_t \quad (1)$$

For every t and certain $\rho < t$, where E_t are independent identically distributed random variables with $\mu = 0$ and standard deviation σ, then we say, that the process is ρ order autoregressive scheme $AR(\rho)$. Random variables E_t are called the residuals or the white noise. The estimate of the coefficients $\alpha_1, \alpha_2,...,\alpha_\rho$ is one of the fundamental problems of the $AR(\rho)$ model analysis.

The estimate of the parameters coefficients can be obtained by different ways from the empirical data. Let $x_1,x_2,...,x_\rho,...,x_{n+\rho}$ be the observed data of the time series. Vector of estimates of coefficients $\alpha_1, \alpha_2,...,\alpha_n$ is indicated as $\alpha^* = (\alpha_1^*, \alpha_2^*,...,\alpha_\rho^*)'$. When the least squares method is used, then

$$\alpha^* = (X'X)^{-1} X'x \quad (2)$$

where $X = \begin{pmatrix} x_\rho & x_{\rho-1} & x_1 \\ x_{\rho+1} & x_\rho & x_2 \\ x_{\rho+n-1} & x_{\rho+n-2} & x_n \end{pmatrix}$

and $x = \begin{pmatrix} x_{\rho+1} \\ x_{\rho+2} \\ x_{\rho+n} \end{pmatrix}$

The logically consequential question that must be solved in connection with the estimate of the $\hat{\alpha}$ is the question of its accuracy. The accuracy of the estimate is usually evaluated by its bias and standard error. Whereas the distribution function is unknown, it is problematical to state these values. The bootstrap method can help in this case.

The First-order Autoregressive Process AR (1): AR (1) model only have one past value and current error that is not correlated to the past values. This AR (1) process is sometimes called the MARCOV process where any value depending on the immediate value. For the first-order auto-regression (AR (1) model) the model is given by

$$X_t = \alpha X_{t-1} + E_t \quad (3)$$

where X_t is the observation at time t (possible centred to have zero mean) and E_t is mentioned in equation (1).

Generating the Data and Calculating the Bias and the Standard Error: We generate a random sample from the Normal distribution N (μ,σ^2), to represent the random disturbances $e_t = (e_1, e_2,....e_n)$, then we generate the time series X_t through AR (1) model which given by:

$$X_t = \alpha X_{t-1} + e_t \quad (4)$$

where in this case an unknown parameter α is a real number and $|\alpha| \le 1$.

The generated random disturbances $e_t = (e_1, e_2,...e_n)$ are resampled with replacement to obtain $e_t^* = (e_1^*, e_2^*,...e_3^*)$ which consider to be a random sample also, the bootstrap samples $(x_1^*, x_2^*, x_3^*,...x_n^*)$ are generated with nonparametric bootstrap. The first step is to determine the value x_1^*. Let $x_1^* = x_1$ the initial value of this time series. Then,

$$x_2^* = \alpha x_1^* + e_2^*$$
$$x_3^* = \alpha x_2^* + e_3^*$$
$$\cdots\cdots\cdots\cdots$$
$$x_{n+1}^* = \alpha x_n^* + e_{n+1}^*$$

We get the bootstrap replications $\hat{\alpha}^*$ of the estimate α by applying the least square estimation procedure through the formula,

$$\hat{\alpha}^* = \frac{\sum_{t=2}^{n+1} x_t . x_{t-1}}{\sum_{t=2}^{n+1} (x_{t-1})^2} \quad (5)$$

Then we calculate the bootstrap estimates of the bias by the empirical bias,

$$B(\theta^*) = \frac{1}{R} \sum_{i=1}^{R} (\hat{\theta}_i^* - \theta^*) \quad (6)$$

where R is the number of generated replications. Then, the standard error is estimated by the standard deviation:

$$SE(\theta^*) = \sqrt{\frac{1}{R-1}\sum_{i=1}^{R}\left(\theta_i^* - \bar{\theta}^*\right)^2},$$

where $\bar{\theta}^* = \frac{1}{R}\sum_{i=1}^{R}\hat{\theta}_i^*$ (7)

RESULTS AND DISCUSSION

Simulation Studies

Simulation 1 (Choice of Replications (R))

Model-Based Bootstrapping Method: We generate the time series $X_t = 0.2\ X_{t-1} + e_t$. The random disturbances e_t were generated from $N(0,1)$. In this part we fixed the sample size and the value of the parameter α [n = 100, α = 0.2] and based on 100 to 800 simulations we obtained the bias and standard error of the estimate to compare the accuracy of the three methods that we mentioned above. In the next part, the sample size [n = 100] and the numbers of simulation [100], where α takes the values [0.8, 0.6, 0.4, 0.2, 0.1, -0.1, -0.2, -0.4 and -0.6].

Figure 1 shows the plot of the generated values of the time series. Since pattern of the first two moments are constant over time, then we can say that the data is weakly stationary stochastic process.

Table 1 shows the Summary statistics for model-based bootstrapping estimates of Mean, Bias and Standard Error (SE) for α, the parameter α is known in advance [0.2], the mean values almost the same and approach to 0.2 our chosen value, which indicates high accuracy, the result can be supported with the bias which turn to be zero when we round the values to 2 d.p and the standard errors to 0.01.

Figure 2 demonstrates the development of bias and standard error of the estimate α when 100 to 800 replications obtained. The bias's line doesn't show any convergence after approximate 700 bootstrap replications, where the standard error line after approximate 300 bootstrap replications. The values of bias converge to 0.002 and the values of the standard error to 0.005 when AR (1) and model-based bootstrap applied.

Table 1: Summary statistics for model-based bootstrapping estimates of Mean, Bias and Standard Error (SE) for α

	Model-based Bootstrapping Method		
R	Mean	Bias	SE
100	0.202	0.002	0.005
200	0.204	0.004	0.004
300	0.198	-0.002	0.005
400	0.200	0.000	0.005
500	0.199	-0.001	0.005
600	0.201	0.001	0.005
700	0.202	0.002	0.005
800	0.202	0.002	0.005

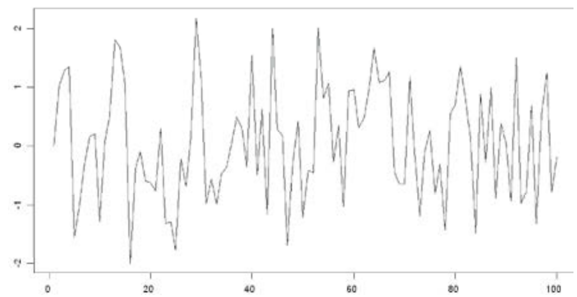

Fig. 1: The graph of the generated time series from $X_t = 0.2\ X_{t-1} + e_t$

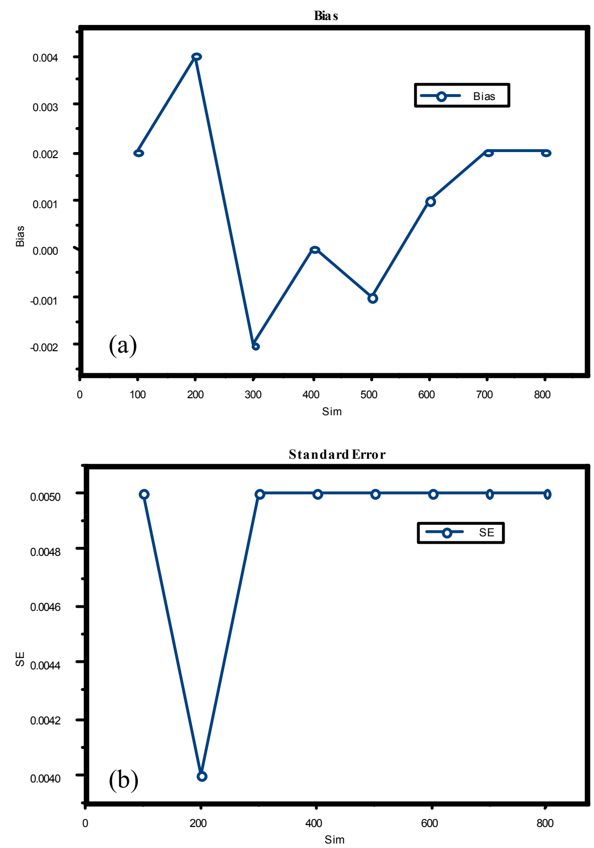

Fig. 2a, b: The demonstration of the development of bias and standard error of the estimate α when 100 to 800 replications obtained and AR (1) applied for model-based bootstrap

Block-Bootstrapping Method: The method described in section 3.1. fall under the category of model-based resampling methods, because the residuals are generated

Table 2: Summary statistics for block- bootstrapping estimates of Mean, Bias and Standard Error (SE) for α

	Block- Bootstrapping Method Overlapping blocks [size = 5]		
R	Mean	Bias	SE
100	0.204	0.004	0.006
200	0.199	-0.001	0.005
300	0.195	-0.005	0.006
400	0.196	-0.004	0.005
500	0.197	-0.003	0.005
600	0.197	-0.003	0.005
700	0.196	-0.004	0.005
800	0.196	-0.004	0.005

Table 3: Summary statistics for least square estimates of Mean, Bias and Standard Error (SE) for α

	Least Square Method		
R	Mean	Bias	SE
100	0.194	-0.006	0.005
200	0.194	-0.006	0.005
300	0.194	-0.006	0.005
400	0.195	-0.005	0.005
500	0.195	-0.005	0.005
600	0.195	-0.005	0.005
700	0.195	-0.005	0.005
800	0.195	-0.005	0.005

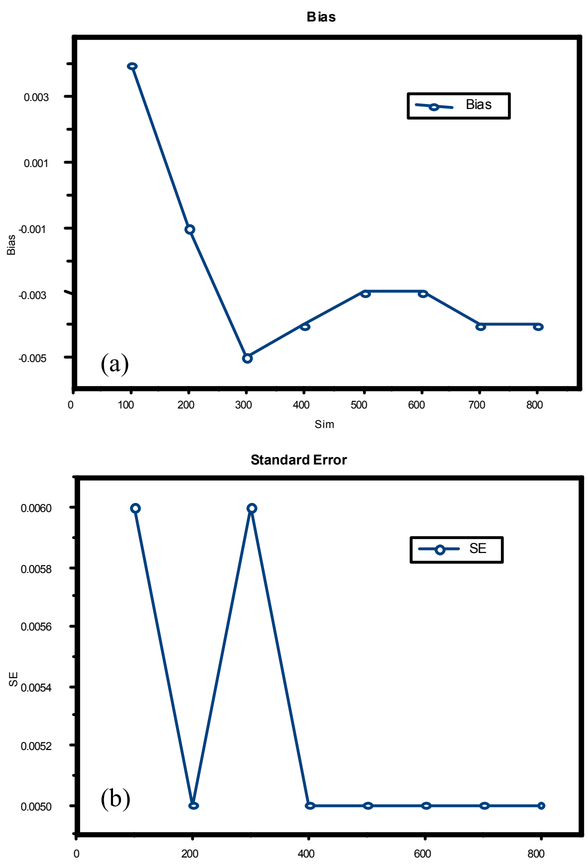

Fig. 3a,b: The demonstration of the development of bias and standard error of the estimate α when 100 to 800 replications obtained and AR (1) applied for block- bootstrap method

and resampled based on a time series model [i.e., the AR (1) model]. In this second method we bootstrap the original data instead of the residuals. But if the individual data are bootstrapped, the dependences that the original data are bound with are lost. This shortage can be reduced the way that we don't bootstrap individual data but the whole blocks of the neighbouring original data [i.e., the resampling or bootstrap scheme here is to resample with replacement from the set of b blocks]. We bootstrap the overlapping blocks [size 5] that we obtained from the original set.

Table 2 shows the Summary statistics for block- bootstrapping estimates of Mean, Bias and Standard Error (SE) for α, the mean values we obtained here approach also to 0.2 correct to 1 d.p as the first method with small differences in the bias and standard error values which indicates less accuracy compare to the model-based bootstrapping.

Figure 3 shows the development of bias and standard error of the estimate α when 100 to 800 replications obtained. The bias's line level of after approximate 700 bootstrap replications, where the standard error line after approximate 400 bootstrap replications. The values of the biasness converge to -0.004, whereas, the values of the standard error to 0.005 when AR (1) and block- bootstrap method applied.

Least Square Method: In this section, we generate a random sample from the normal distribution $N(\mu,\sigma^2)$ to represent the random disturbances $e_t = (e_1, e_2, ..., e_n)$ then we generate the time series X_t through AR (1) model which given in equation (4) by choosing $|\alpha| \leq 1$. Then, the estimation of $\hat{\alpha}^*$ is obtained by applying the least square procedure that given in formula (5).

Table 3 shows the Summary statistics for least Square estimates of the Mean, Bias and Standard Error (SE) for α, all the mean values are less than 0.2 our chosen value for α, all the values of the bias are negative and slightly bigger (in their absolute values) than the values in the other two method and the standard error is constant

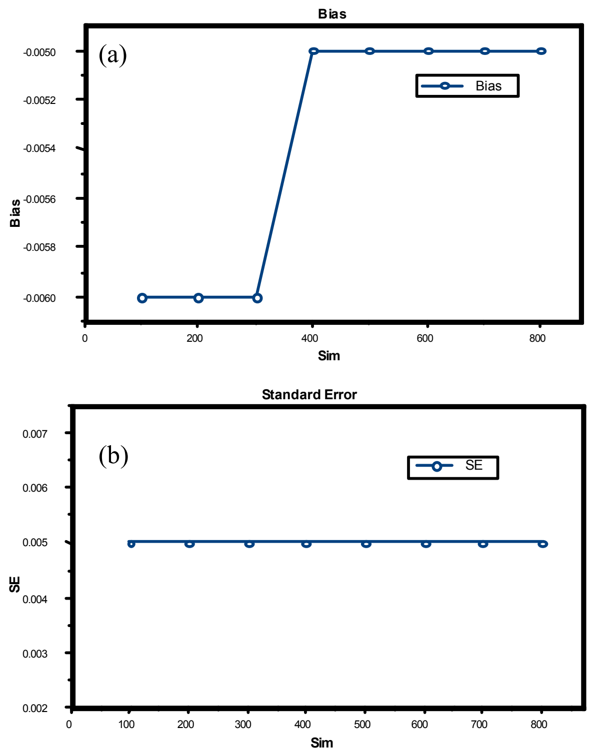

Fig. 4a,b: The demonstration of the development of bias and standard error of the estimate α when 100 to 800 replications obtained and AR (1) applied for least square method

over all values of (R) which reflect the meaningless of varying the replications and showing less accuracy compare to the previous two methods.

Figure 4 demonstrates the development of bias and standard error of the estimate α when 100 to 800 replications obtained. The bias's line implies small values of (R) tend to be more bias than big values, where the standard error line showing the same error for all bootstrap replications.

Simulation 2 (Choice of α): In the first part of this paper, we fixed the value of the parameter [α = 0.2] and the sample size [n = 100], then we look to different numbers of replications. In this section, we consider a sample of size [n = 100] and we will consider different values of the parameter α [0.8, 0.6, 0.4, 0.2, 0.1, -0.1, -0.2, -0.4 and -0.6].

Table 4 shows the Summary statistics for model-based bootstrapping estimates of Mean, Bias and Standard Error (SE) for α, the parameter α is known in advance. From the result, we can that positive and biggest values of α are tend to be more accurate than the negative and small values with model-based bootstrapping method.

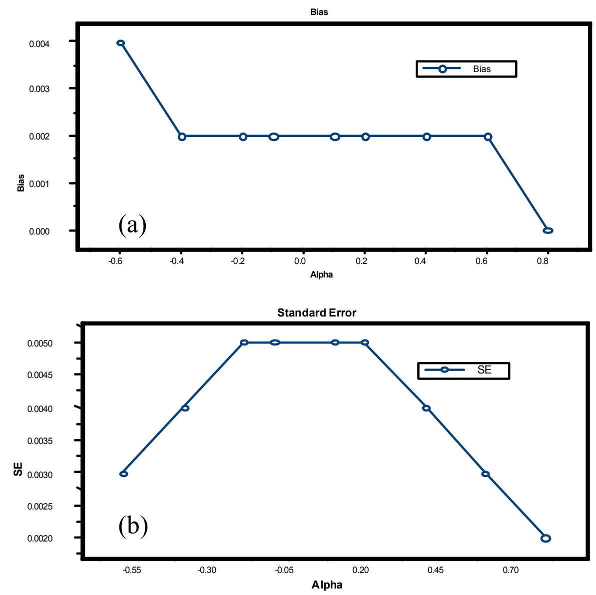

Fig. 5a,b: The demonstration of the development of bias and standard error of the estimate α when 100 replications obtained and AR (1) applied for model-based bootstrapping method with different values of α

Table 4: Summary statistics for model-based bootstrapping estimates of Mean, Bias and Standard Error (SE) for α with different values and 100 replications

	Model-based Bootstrapping Method		
α	Mean	Bias	SE
0.8	0.800	0.000	0.002
0.6	0.602	0.002	0.003
0.4	0.402	0.002	0.004
0.2	0.202	0.002	0.005
0.1	0.102	0.002	0.005
-0.1	-0.098	0.002	0.005
-0.2	-0.198	0.002	0.005
-0.4	-0.398	0.002	0.004
-0.6	-0.596	0.004	0.003

Figure 5 shows the development of bias and standard error of the estimate α when 100 replications obtained and varying α's values. The line of the bias value is almost symmetry to show the balance in biasness for positive and negative values of α and the positive bigger numbers are less bias the standard error line implies any convergence for the small values of α.

Table 5 shows the Summary statistics for block-bootstrapping estimates of Mean, Bias and Standard Error (SE) for α, the parameter α is known in advance, in the column of the mean the values are slightly different with

Table 5: Summary statistics for block- bootstrapping estimates of Mean, Bias and Standard Error (SE) for α with different values of α and 100 replication

	Block- Bootstrapping MethodOverlapping blocks [size = 5]		
α	Mean	Bias	SE
0.8	0.782	-0.018	0.004
0.6	0.592	-0.008	0.005
0.4	0.399	-0.001	0.006
0.2	0.204	0.004	0.006
0.1	0.107	0.007	0.006
-0.1	-0.087	0.012	0.006
-0.2	-0.184	0.016	0.006
-0.4	-0.377	0.022	0.005
-0.6	-0.570	0.029	0.004

Table 6: Summary statistics for least square estimates of Mean, Bias and Standard Error (SE) for α with different values of α and 100 replication

	Least Square Method		
α	Mean	Bias	SE
0.8	0.781	-0.019	0.002
0.6	0.587	-0.013	0.003
0.4	0.391	-0.009	0.004
0.2	0.194	-0.006	0.005
0.1	0.095	-0.005	0.005
-0.1	-0.103	-0.003	0.005
-0.2	-0.202	-0.002	0.005
-0.4	-0.397	0.003	0.005
-0.6	-0.591	0.009	0.004

the chosen value of α which indicates more bias compared to model-based method which support again the preference of the first method in both part 1&2.

Figure 6 demonstrates the development of bias and standard error of the estimate α when 100 replications were preformed. In general, the bigger values of α are more bias than the small. Standard error line implies any convergence for the small values of α. The values of the standard error converge to 0.006 when AR (1) and block- bootstrapping method applied.

Figure 7 demonstrates the development of bias and standard error of the estimate α when 100 replications were preformed. The bias line implies any convergence for the small values of α. The values of the standard error converge to 0.005 when AR (1) and Least Square Method applied.

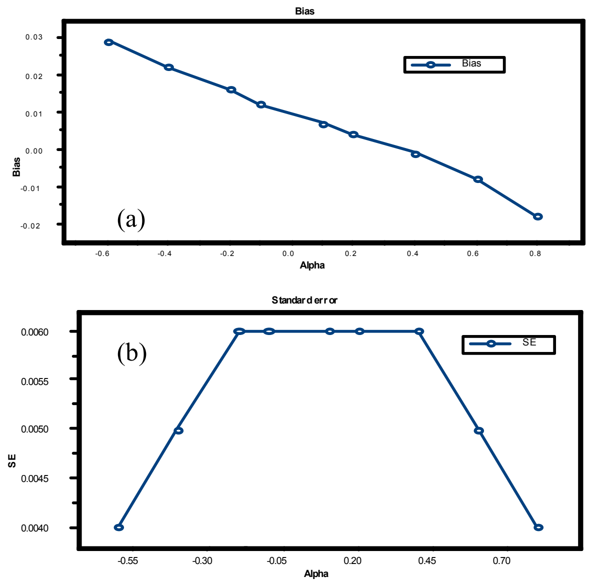

Fig. 6a,b: The demonstration of the development of bias and standard error of the estimate α when 100 replications obtained and AR (1) applied for block- bootstrapping method with different values of α

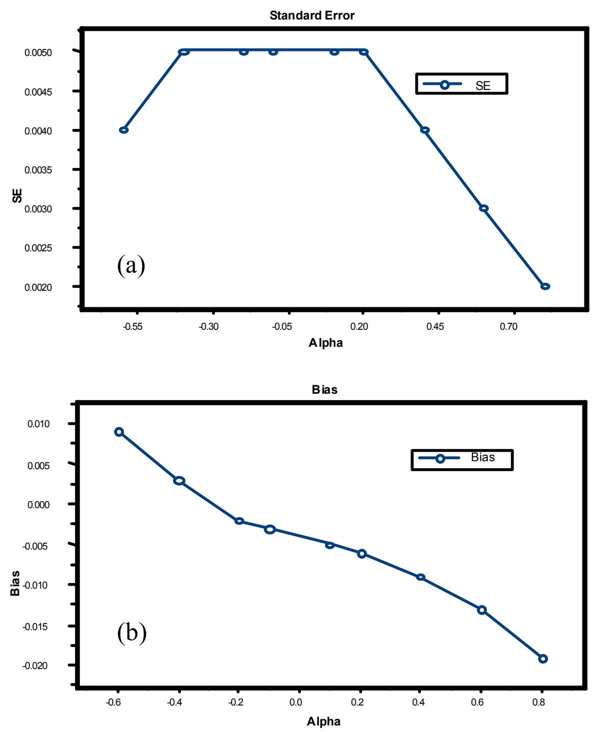

Fig. 7a,b: The demonstration of the development of bias and standard error of the estimate α when 100 replications obtained and AR (1) applied for least square method with different values of α

CONCLUSION

In this paper, three methods are discussed to evaluate the accuracy of the estimated parameter in the autoregressive model through their biasness and standard error. It is possible to say that the bootstrap methods are not absolutely exact but they are the practicable solution in such cases when real situation require using the complicated model because the mathematical complexity of the model is not related to accuracy of the bootstrap analysis. From the results obtained, we can conclude that model-based bootstrapping (re-sampling the residual of the model) perform better compared to the least square method and overlapping-block bootstrap in terms of having small bias and standard error. Biggest values of the parameter tend to be high bias and have big standard error.

A direction for future research is to increase the numbers of the replications for the model base, considering the block size and non-overlapping blocks for block bootstrapping.

ACKNOWLEDGMENTS

I am grateful to my supervisor Associate Professor Dr. Faiz Ahmed Mohamed for his encouragement, helpful comments and suggestions through the writing of this paper.

REFERENCES

1. Efron, B., 1979. Bootstrap methods: another look at the jacknife. Annals of Statistics, 7: 1-26.
2. Beran, R.J., G.R. Ducharme and Universite´ De Montre´al, 1991. Asymptotic theory for bootstrap methods in statistics. Montre´al: Les Publications CRM.
3. Hall, P., 1992. The Bootstrap and Edgeworth Expansion. New York: Springer-Verlag.
4., Efron, B. and R. Tibshirani, 1993. An introduction to the bootstrap. New York: Chapman and Hall.
5. Davison, A.C. and D.V. Hinkley, 1997. Bootstrap methods and their application. Cambridge: Cambridge University Press.
6. Hall, P., 1985. Resampling a Coverage Process. Stochastic Processes and their Applications, 19: 259-269.
7. Carlstein, E., 1986. The use of subseries methods for estimating the variance of ageneral statistic from a stationary time series. Annals of Statistics, 14: 1171-1179.
8. Küunsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. Annals of Statistics, 17: 1217- 1241.
9. Politis, D.N. and J.P. Romano, 1993. On the sample variance of linear statistics derived from mixing sequences. Stoch. Proc. Appl., 45: 155-167.
10. Bühlmann, P., 1995. The blockwise bootstrap for general empirical processes of stationary sequences. Stochastic Processes and their Applications, 58: 247-265.