ComparingThree Methods on Bootstrap Approach to Time Series

David Peter Ilesanmi, Okafor Joseph, Morrison Mudiaga

Abstracts

This research reviews the concept of Bootstrapping, and suggests the bootstrap method that performs best for Time Series Analysis. Various types of bootstrap schemes have been developed. This research work compare only three methods: Stationary bootstrap, Block bootstrap and Sieve bootstrap methods using Monte Carlo simulation and a real life data from Nigeria stock index data. We consider the following models: ARIMA(1,0,0), ARIMA(1,1,0), ARIMA(1,1,1), ARIMA(2,1,1) and ARIMA(1,1,3). Our results are based on 2000 simulations; the number of bootstrap replicates is 500. We only report the bootstrap estimates for the Bias and the Mean Square error (MSE). Nigerian Stock Index (NSI) Data from 5/1/2006 to 10/9/2015.The Nigerian Stock Index data was fitted and it generated an ARIMA (5,2,1). For the specific applications, the simulation results suggested that the sieve bootstrap performed better than both the block and the stationary bootstrap methods. The analysis of the real life data (NSI data) results supports our simulation findings. Hence, we draw the final conclusion that based on our study; the Sieve bootstrap is generally superior over both the Stationary bootstrap and the Block bootstrap.

Keywords: Bootstrap, ARIMA, Simulation, Mean Square error, Bias, Asymptotic, Time series.

INTRODUCTION

The bootstrap 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 carrying the bootstrap and the improvements in accuracy that it achieves relative to first-order asymptotic approximations depend on whether the data are a random sample from a distribution or a time series. If the data are random samples, then,the bootstrap can be carried out 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. (Beran and Ducharme, 1991; Hall,1992;Efron and Tibshirani,1993; and Davison and Hinkley, 1997) provided detailed discussions of bootstrap methods and their properties for data that are sampled randomly from a distribution.

- David Peter Ilesanmi, MSc Statistics, University of Lagos, Nigeria. PH-08064864045. E-mail: davidilesanmi@yahoo.com
- Okafor Joseph, Mathematics and Statistics Lecturer, Delta State Polytechnic, Ogwashi-Uku, Delta State, Nigeria. PH: 08032275807. E-mail: josephiokafor@yahoo.com
- Morrison Mudiaga, Mathematics and Statistics Lecturer,Delta State Polytechnic, Ogwashi-Uku, Delta State, Nigeria . PH: 08037614769. E-mail: morrismudiaga@yahoo.com

The situation becomes more complicated when the data are a time series data, because bootstrap sampling must be carried out in a way that suitably captures the dependence structure of the data generation process (DGP). This is not difficult if one has a finite-dimensional parametric model (e.g., a finite-order ARMA model) that reduces the DGP to independent random sampling. In this case and under suitable regularity conditions, the bootstrap has properties that are essentially the same as they are when the data are a random sample from a distribution. (See, for example: Andrews, 1999; Bose, 1990)

This research work is concerned with the situation in which one does not have a finite-dimensional parametric model that reduces the DGP to independent random sampling. We reviewthe three methods that have been proposed for carrying out the bootstrap in this situation and discuss the ability of these methods to achieve asymptotic transformation. We note that methods for carrying out the bootstrap with time-series data are not as well understood as methods for data that are sampled randomly from a distribution. Moreover, the performance of the bootstrap as measured by the order of the asymptotic transformations that are available from known methods tends to be poorer with time series than with random samples. This is an important problem for applied research because first-order asymptotic approximations are often inaccurate and misleading with time-series data and samples of the sizes encountered in applications. We shall see that there is a need for further research in the application of the bootstrap to time series.

STATEMENT OF PROBLEM

Formally, the bootstrap consists of a methodology for estimating standard errors by repeatedly re-sampling with replacement from the original finite sample, which is believed to be sample of independent and identically distributed (i.i.d.) observations from an unknown probability distribution. The re-samples obtained (known as bootstrap samples) are used to estimate the statistic of interest. However, it was notpossible to use this bootstrap procedure with Time Series data. The reason lies in the assumption of i.i.d. random variables which is violated when observations are serially correlated. The original bootstrap becomes inconsistent in time series data situation because the assumption of i.i.d. random variables is violated when observations are serially correlated and in some cases when the error becomes conditionally volatile. The performance of these bootstrap methods using real life data(NSI) in context of time-series merits further research.

OBJECTIVES OF THE STUDY

Various bootstrap methods have been developed.In this project,

1. We shall investigate the performance of three (3) competing bootstrap methods for time-series analysis.
2. The performance of the bootstrap methods was found using simulation study
3. We shall apply the bootstrap methods to a real life Data.

METHODOLOGY

We discuss the methods used in this research work.

Monte Carlo Simulation: Monte Carlo simulations design is used to investigate the performance of various bootstrap methods used in the present study. The models for the design are ARIMA (p,d,q) structured

Set Seed: We specify the random seed and the number of simulations. And set the seed to specify the seed for generating the next forecasts.

ARIMA(p,d,q):ARIMA models are, in theory, the most general class of models for forecasting a time series which can be stationarized by transformations such as differencing and logging. In fact, the easiest way to think of ARIMA models is as fine-tuned versions of random-walk and random-trend models: the fine-tuning consists of adding lags of the differenced series and/or lags of the forecast errors to the prediction equation, as needed to remove any last traces of autocorrelation from the forecast errors.

The acronym ARIMA stands for "Auto-Regressive Integrated Moving Average." Lags of the differenced series appearing in the forecasting equation are called "auto-regressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series. Random-walk and random-trend models, autoregressive models, and exponential smoothing models (i.e., exponential weighted moving averages) are all special cases of ARIMA models.

A non-seasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of non-seasonal differences, and
- **q**is the number of lagged forecast errors in the prediction equation.

To identify the appropriate ARIMA model for a time series, you begin by identifying the order(s) of differencing needing to stationarize the series and remove the gross features of seasonality, perhaps in conjunction with a variance-stabilizing transformation such as logging or deflating. If you stop at this point and predict that the differenced series is constant, you have merely fitted a random walk or random trend model. (Recall that the random walk model predicts the first difference of the series to be constant, the seasonal random walk model predicts the seasonal difference to be constant, and the seasonal random trend model predicts the first difference of the seasonal difference to be constant--usually zero.) However, the best random walk or random trend model may still have auto-correlated errors, suggesting that additional factors of some kind are needed in the prediction equation.

The ARIMA models used in this project are:

I. ARIMA (1,0,0)

An ARIMA (1,0,0) model with constant would have the prediction equation:

$$\hat{Y}(t) = \mu + \phi(Y(t-1) - Y(t-2))$$

II. ARIMA(1,1,0) = differenced first-order autoregressive model:

This would yield the following prediction equation:

$$\hat{Y}(t) - Y(t-1) = \mu + \phi(Y(t-1) - Y(t-2))$$

which can be rearranged to

$$\hat{Y}(t) = \mu + Y(t-1) + \phi(Y(t-1) - Y(t-2))$$

This is a first-order autoregressive, or "AR(1)", model with one order of nonseasonal differencing and a constant term--i.e., an "ARIMA(1,1,0) model with constant." Here, the constant term is denoted by "mu (μ)" and the autoregressive coefficient is denoted by "phi (ϕ)", in keeping with the terminology for ARIMA models popularized by Box and Jenkins.

III. A "mixed" model—ARIMA(1,1,1):

An ARIMA(1,1,1) model with constant would have the prediction equation:

$$\hat{Y}(t) = \mu + Y(t-1) + \phi\big(Y(t-1) - Y(t-2)\big) - \theta e(t-1)$$

IV. ARIMA (2,1,1):

An ARIMA (2,1,1) model with constant would have the prediction equation:

$$\hat{Y}(t) = \mu + Y(t-1) + \phi_1\big(Y(t-1) - Y(t-2)\big) + \phi_2(Y(t-1) - Y(t-2) - Y(t-3)) - \theta e(t-1)$$

v. ARIMA (1,1,3):

An ARIMA (1,1,3) model with constant would have the prediction equation:

$$\hat{Y}(t) = \mu + Y(t-1) + \phi\big(Y(t-1) - Y(t-2)\big) - \theta_1 e(t-1) - \theta_2 e(t-2) - \theta_3 e(t-3)$$

BOOTSTRAP METHOD PERFORMANCE EVALUATION

To assess the performance of the methods we used the following performance techniques.

1. Bias: Let $\emptyset$ denote the thing that we are trying toestimate.

Let $\hat{\emptyset}$ denote the result of an estimation based onone data set.

Bias, $b(\hat{\emptyset}) = \emptyset - E[\hat{\emptyset}]$ = difference between the truevalue and the average of all possible estimates.

2. Mean Square Error: The average squared residual (MSE) is a measure of how closely the forecasts track the actual data. The statistic is popular because it shows up in analysis of variance tables. However, because of the squaring, it tends to exaggerate the influence of outliers (points that do not follow the regular pattern).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

3 Bootstrap method Performance Evaluation Table

The bootstrap performance evaluation table shows the bootstrap methods Comparison Report for this study.

RESULTS AND DISCUSSIONS

In this section, we study and compare the performance ofStationary, Block and Sieve Bootstrap methods using

A. Simulation Data (ARIMA).

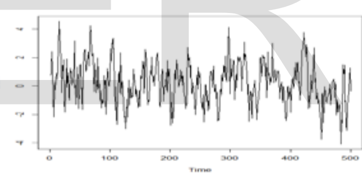
We consider the following models: ARIMA(1,0,0), ARIMA(1,1,0), ARIMA(1,1,1), ARIMA(2,1,1) and ARIMA(1,1,3).Our results are based on 2000 simulations; the number of bootstrap replicates is 500. We only report the bootstrap estimates for the Bias and the Mean Square error (MSE).

B. Nigerian Stock Index (NSI) Data from 5/1/2000 to 10/9/2009.

The Nigerian Stock Index data was fitted and it generated an ARIMA (5,2,1).Our results are based on 2000 simulations; the number of bootstrap replicates is 500.We only report the bootstrap estimates for the Bias and the Mean Square error.

A. Results from Simulation Data (ARIMA).

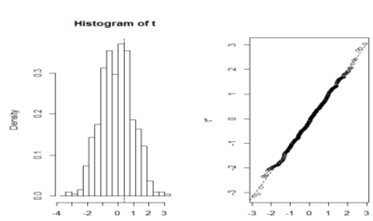
R Output of ARIMA(1,0,0)

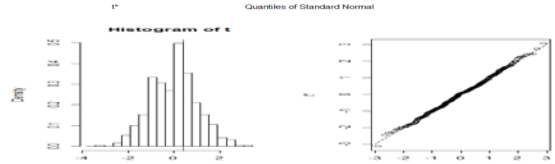


1. **BIAS:**

Stationary Bootstrap

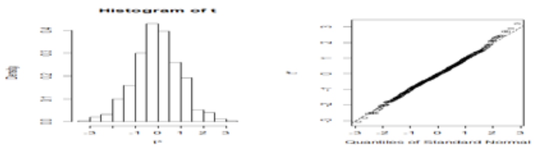
Block Bootstrap





Sieve Bootstrap

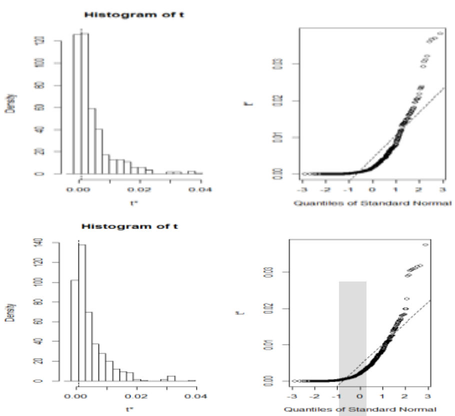
ISSN 2229-5518



2. MSE:

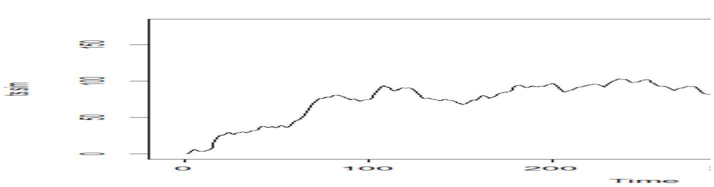
Stationary Bootstrap

Block Bootstrap



Sieve Bootstrap

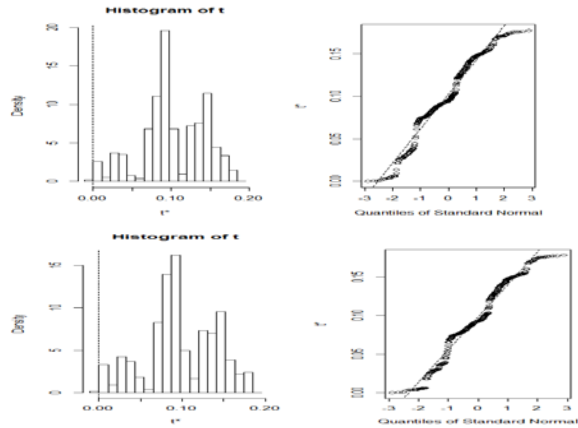
R Output of ARIMA(1,1,0)



1. BIAS:

Stationary Bootstrap

Block Bootstrap

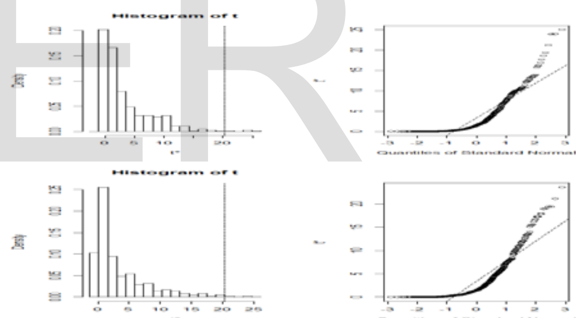


Sieve Bootstrap

2. MSE:

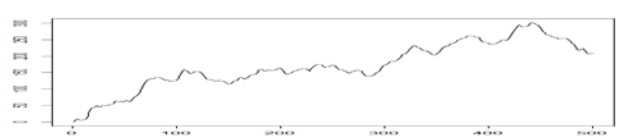
Stationary Bootstrap

Block Bootstrap



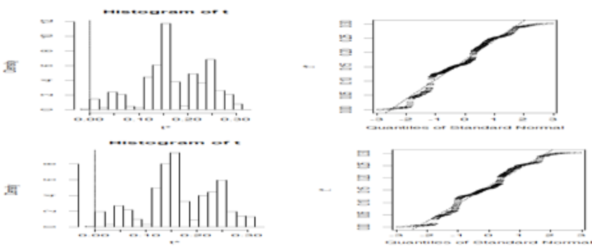
Sieve Bootstrap

R Output of ARIMA(1,1,1)



1. Bias:

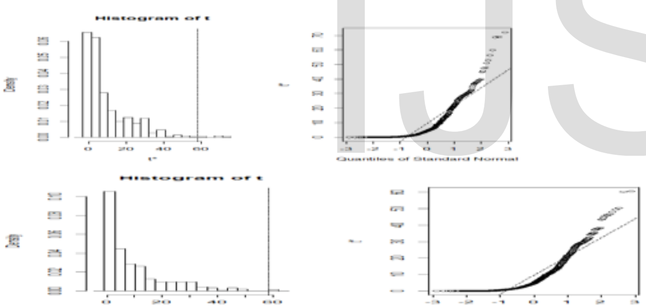
Stationary Bootstrap
Block Bootstrap



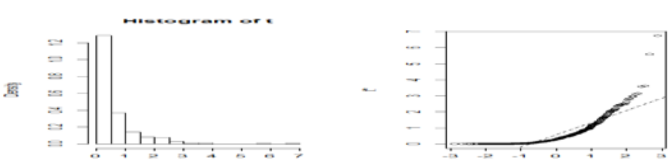
Sieve Bootstrap

2. MSE

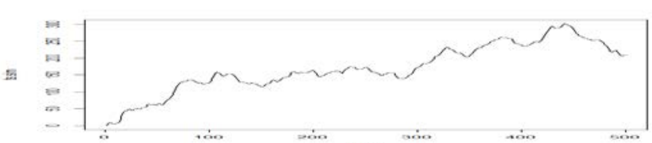
Stationary Bootstrap
Block Bootstrap



Sieve Bootstrap

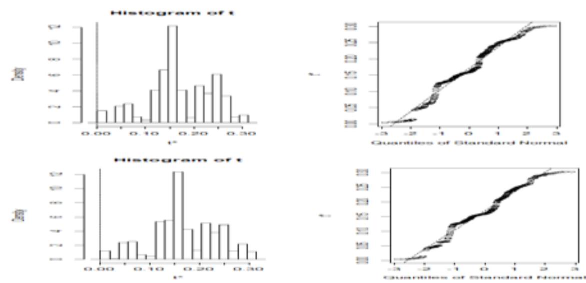


R Output of ARIMA (2,1,1)

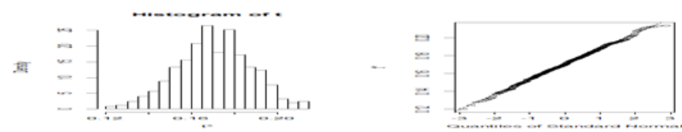


1. Bias:

Stationary Bootstrap
Block Bootstrap

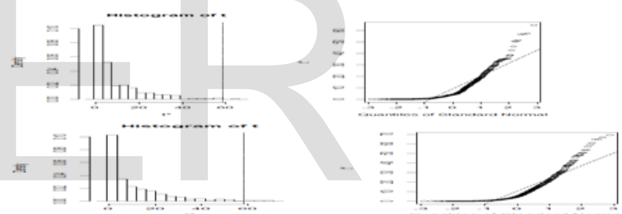


Sieve Bootstrap



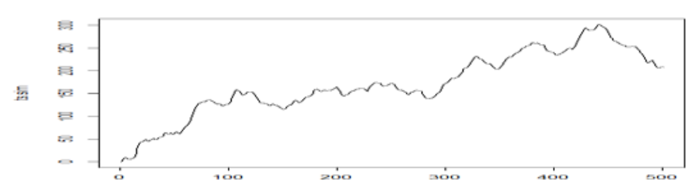
2. MSE

Stationary Bootstrap
Block Bootstrap



Sieve Bootstrap

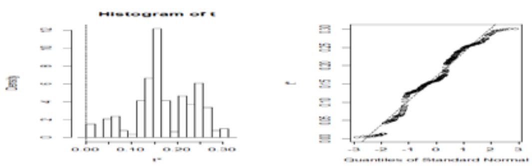
R Output of ARIMA (1,1,3)

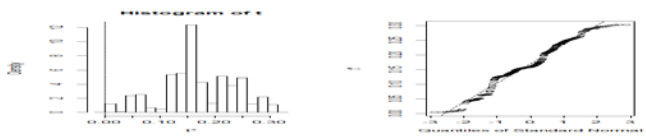


1. Bias:

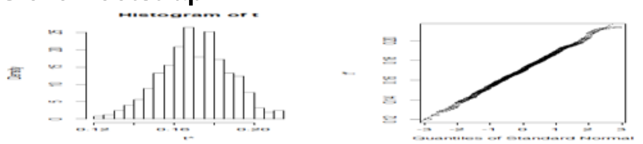
Stationary Bootstrap

Block Bootstrap



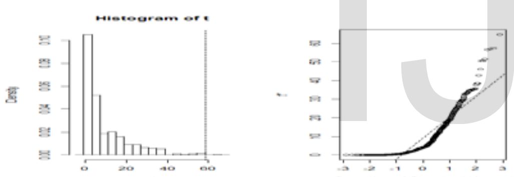


Sieve Bootstrap

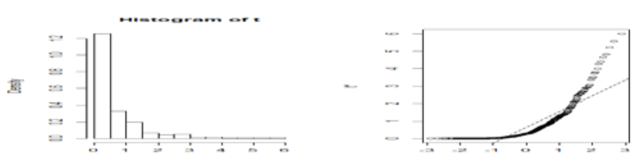


2. MSE

Stationary Bootstrap
Block Bootstrap



Sieve Bootstrap



Bootstrap method Performance Evaluation Table

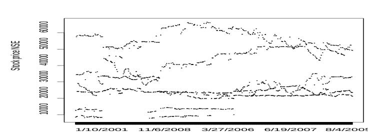
MET		Arima	Arima(	Arima	Arima	Arima

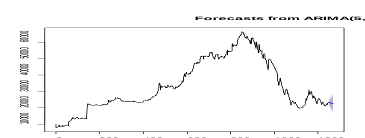
HOD		(1,0,0)	1,1,0)	(1,1,1)	(2,1,1)	(1,1,3)
Simulation Results	Bias	-0.001044584	0.0699193	0.07155598	0.06580487	0.06795338
	MSE	0.004060544	3.365487	9.657499	9.657499	9.657499
Stationary Bootstrap	Bias	5.980708e-05	-0.003942727	-0.00652402	-0.001935742	-0.002188269
	MSE	0.004053825	3.319714	9.447134	9.53052	9.429147
Block Bootstrap	Bias	8.533645e-07	-0.0003281766	0.002577846	6.310652e-05	-0.00421015
	MSE	0.004052018	3.265242	9.457113	9.377295	9.283563
Sieve Bootstrap	Bias	-1.471237e-05	-1.471237e-05	-2.898846e-05	1.124808e-06	6.003037e-05
	MSE	0.004059374	0.2447629	0.5616879	0.563979	0.5681795

From the above table, we observe that all the three methods are good bootstrap procedures for the Arima Time series. They perform better than those from the simulation results. The most surprising is the Sieve Bootstrap, as it performs better than both the Stationary bootstrap and the Block bootstrap methods across the two ways simulation design in this study. This result suggests that the Sieve bootstrap is superior over both the Stationary bootstrap and the Block bootstrap.

B. Results obtained bootstrapping the NSE Data from 5/1/2006 to 10/9/2015.

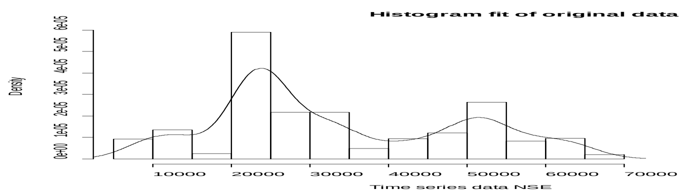
R Output fromNSE Data
NSE data:





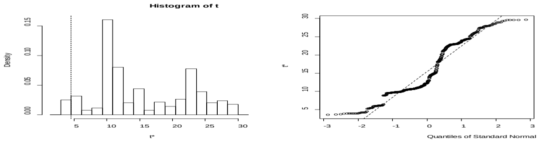
Series: NSE

ISSN 2229-5518

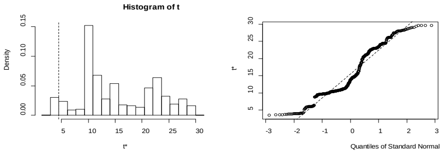


1. Bias:

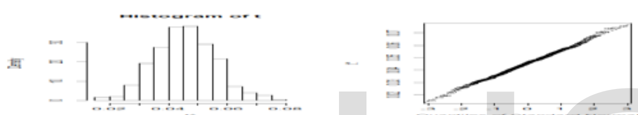
Stationary Bootstrap



Block Bootstrap

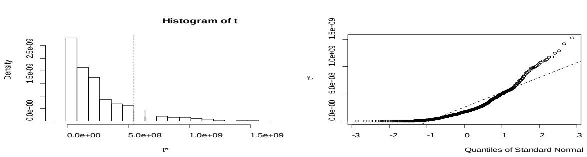


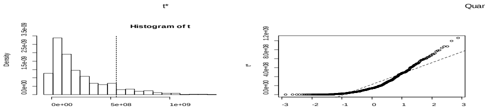
Sieve Bootstrap



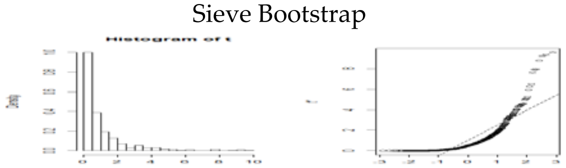
2. MSE:

Block Bootstrap





Sieve Bootstrap



Bootstrap Method Performance Evaluation Table

METHOD		Arima(5,2,1)
NSI Data	Bias	-0.02949217
	MSE	4.750791
Stationary Bootstrap	Bias	-0.001442936
	MSE	4.31575
Block Bootstrap	Bias	-0.002401189
	MSE	4.504353
Sieve Bootstrap	Bias	-0.000614193
	MSE	0.7466976

From the results tabulated above, we observe that the Sieve Bootstrap performed better than both the Stationary bootstrap and the Block bootstrap methods. Hence,analysis of the Nigerian Stock Index data results supports our simulation findings.We draw the final conclusion that in the framework of our study, the Sievebootstrap is generally superior over both the Stationary bootstrap and the Block bootstrap.

Conclusions

Formally, the bootstrap consists of a methodology for estimating standard errors by repeatedly re-sampling with replacement from the original finite sample, which is believed to be sample of independent and identically distributed (i.i.d.) observations from an unknown probability distribution. The re-samples obtained are used to estimate the statistic of interest. However, it was notpossible to use this bootstrap with Time Series data. The reason lies in the assumption of i.i.d. random variables which is violated when observations are serially correlated. Few approaches to this problem were considered in this project.

For the specific applications, the simulation results suggested that the sieve bootstrap performed better than both the block and the stationary bootstrap methods. The analysis of the real life data (NSI data) results supports our simulation findings.

Hence, we draw the final conclusion that in the framework of our study, the Sievebootstrap is generally superior over both the Stationary bootstrap and the Block bootstrap.

REFERENCES

[1] Bickel, P. and D. Freedman (1981). Some Asymptotic Theory for the Bootstrap, Annals of Statistics, 9, 1196-1217.
[2] Bose, A. (1988). Edgeworth Correction by Bootstrap in Autoregressions, Annals of Statistics, 16, 1709-1722.
[3] Bose, A. (1990). Bootstrap in Moving Average Models, Annals of the Institute of Statistical Mathematics, 42, 753-768.
[4] Bühlmann, P. (1998). Sieve Bootstrap for Smoothing Nonstationary Time Series, Annals of Statistics, 26, 48-83.
[5] Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series, Annals of Statistics, 14, 1171-1179.
[6] Carlstein, E. K.-A. Do, P. Hall, T. Hesterberg, and H.R. Künsch (1998). Matched-block bootstrap for dependent data, Bernoulli, 4, 305-328.
[7] Dahlhaus, R. and D, Janas (1996). A Frequency Domain Bootstrap for Ratio Statistics in Time Series, Annals of Statistics, 24, 1934-1963.
[8] Datta, S. and W.P. McCormick (1995). Some Continuous Edgeworth Expansions for Markov Chains with Applications to Bootstrap, Journal of Multivariate Analysis, 52, 83-106.
[9] Davison, A.C. and P. Hall (1993). On Studentizing and Blocking Methods for Implementing the Bootstrap with Dependent Data, Australian Journal of Statistics, 35, 215-224.
[10] Efron B (1979). Bootstrap methods: Another look at the jackknife. Ann. Statist. 7 1–26
[11] Efron, B. and R.J. Tibshirani (1993). An Introduction to the Bootstrap. (Chapman & Hall, New York).
[12] Franke, J., J.-P. Kreiss, E. Mammen, and M.H. Neumann (2000). Properties of the Nonparametric Autoregressive Bootstrap, working paper, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.
[13] Götze, F. and C. Hipp (1994). Asymptotic distribution of statistics in time series, Annals of Statistics, 22, 2062-2088.
[14] Hafner, C. (1996). Nonlinear Time Series Analysis with Applications to foreign Exchange Rate Volatility (Physica-Verlag Heidelberg).
[15] Hall, P. (1992). The Bootstrap and Edgeworth Expansion (Springer-Verlag,New York).
[16] Hall, P. and J.L. Horowitz (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators, Econometrica, 64:891-916.
[17] Hall, P. and B.-Y. Jing (1996). On sample reuse methods for dependent data, Journal of the Royal Statistical Society, Series B, 58, 727-737.
[18] Hamadu, D., "A bootstrap Approach to Bias-Reduction of Nonlinear Parameter in Regression Analysis", Journal of Scientific Research and Development, Vol 12, 110-127.
[19] Hamadu, D., R. Okafor and O. Abass (2004), "A Model based bootstrap Method for Heteroscedastic Regression Model", Journal of Scientific Research and Development, Vol.9, 9-22.
[20] Hardle, W., J.L.Horowitz, and J.-P. Kreiss (2001). Bootstrap Methods for Times Series, working paper,Institute for Statistics and Economics Humboldt Universitatzu Berlin, Berlin, Germany.
[21] Hansen, B. (1999). "Non-Parametric Dependent Data Bootstrap for Conditional Moment Models" working paper, Department of Economics, University of Wisconsin, Madison, WI.
[22] Horowitz, J.L. (2001a). The bootstrap, in J.J. Heckman and E.E. Leamer (eds.), Handbook of Econometrics, Vol. 5 (North-Holland Publishing Co., Amsterdam).
[23] Horowitz, J.L. (2001b). Bootstrap Methods for Markov Processes, working paper, Department of Economics, Northwestern University, Evanston, IL.
[24] Kreiss, J.-P. (2000). Residual and Wild Bootstrap for Infinite Order Autoregressions, working paper, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.
[25] Kreiss, J.-P., and M.H. Neumann, and Q. Yao (1998). Bootstrap Tests for Simple Structure in Nonparametric Time Series Regression, working paper, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.
[26] Kreiss, J.-P. and E. Paparoditis (2000). Autoregressive Aided Periodogram Bootstrap for Time Series, working paper, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.
[27] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations, Annals of Statistics, 17, 1217-1241.
[28] Inoue A. and M. Shintani (2001). Bootstrapping GMM estimators for time series, working paper, Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC.

[29] Lahiri, S.N. (1991). Second order optimality of stationary bootstrap, Statistics and Probability Letters, 11, 335-341.

[30] Lahiri, S.N. (1992). Edgeworth correction by 'moving block' bootstrap for stationary and nonstationary data, in R. LePage and L. Billard (eds.), Exploring the Limits of Bootstrap (Wiley, New York).

[31] Lahiri, S.N. (1996a). On Edgeworth expansion and moving block bootstrap for Studentized M-estimators in multiple linear regression models, Journal of Multivariate Analysis, 56, 42-59.

[32] Lahiri, S.N. (1996b). Asymptotic expansions for sums of random vectors under polynomial mixing rates, Sankhya, Series A, 58, Pt. 2, 206-224.

[33] Lahiri, S.N. (1999). Theoretical comparisons of block bootstrap methods, Annals of Statistics, 27, 386-404.

[34] Nordgaard, A. (1992). Resampling a Stochastic Process Using a Bootstrap Approach, in: Bootstrapping and Related Techniques in K.H. Jöckel, G. Rothe , W. Sendler (eds.), Lecture Notes in Economics and Mathematical Systems 376 (Springer-Verlag, Berlin).

[35] Paparoditis, E. (1996). Bootstrapping Autoregresive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes, Journal of Multivariate Analysis, 57, 277-296.

[36] Paparoditis, E. and D.N. Politis (2000). The Local Bootstrap for Markov Processes, Journal of Statistical Planning and Inference, forthcoming.

[37] Paparoditis, E. and D.N. Politis (2001). Tapered block bootstrap, Journal of Statistical Planning and Inference, forthcoming.

[38] Politis, D.N. and J.P. Romano (1993). The stationary bootstrap, Journal of the American Statistical Association, 89, 1303-1313.

[39] Politis, D.N., and J.P. Romano, and M. Wolf (1999). Subsampling (Springer-Verlag, New York).

[40] Rajarshi, M.B. (1990): Bootstrap in Markov-Sequences Based on Estimates of Transition Density," Annals of the Institute of Statistical Mathematics, 42, 253-268.

[41] Theiler, J., and L.S. Paul, and D.M. Rubin (1994). Detecting Nonlinearity in Data with Long Coherence Time, in: Time Series Prediction, A. Weigend and N. Gershenfeld (eds.) (Addison-Welsey, Reading, MA).

[42] Varian, H.(2005). "Bootstrap Tutorial". *Mathematica Journal*, 9, 768-775.

[43] Zvingelis, J. (2001). On Bootstrap Coverage Probability with Dependent Data, in Computer-Aided Econometrics, D. Giles (ed.) (Marcel Dekker).