

Application of Bootstrap Data Generating Processes on Nigerian Capital Market Data

C. K. Acha¹; I. A. Acha²

¹Department of Statistics,
Michael Okpara University of Agriculture,
Umudike, Nigeria.
e-mail: acha.kelechi@mouau.edu.ng

²Department of Banking and Finance,
University of Uyo,
Uyo, Nigeria.
e-mail: ikechukwuacha@uniuyo.edu.ng

Abstract — The passion to carry out this research work was ignited by wondering under what conditions the nonparametric bootstrap data generating processes (DGP) produces minimum error when compared with the parametric bootstrap DGP and also to ascertain the performance of the Nigeria stock market, in terms of bias, standard error and other information criteria. Nigeria stock market data sets (1987-2016) were analyzed using parametric and nonparametric bootstrap DGPs techniques for simple linear regression (SLR) models were employed to fit the data set. Evidence showed that the original data Stock Market Capitalization (ODSMC) model without bootstrap resulted to a poor model with high bias. Results also showed that when the sample size was small ≤ 200 , in almost all the bootstrap conditions, the nonparametric bootstrap method performed better than all the parametric counterparts evident from the reported smallest conditional bias. If the Parametric Stock Market Capitalization (PSMC) bootstrap model with the high bootstrap level had not been explored in this study, the nonparametric Stock Market Capitalization Model (NPSMC) bootstrap model would have been exclusively ranked first in producing the smallest bias in the three tests length. But for large sample (≥ 10000), the reverse is the case. The PSMC and NPSMC models show that real income and saving rate are positively significant linear functions of the stock market capitalization and covaries with it. Across all the bootstrap conditions, it was obvious that all the models that worked well have very low HQIC, SBIC, AIC, adjusted R^2 with standard error term ≤ 0.005 and minimum bias confirming that the stock market in Nigeria performance is beyond average. This also provides good models for prediction of stock market performance.

Keywords—*Bootstrap, Stock market, Performance, Information criteria, Prediction, Data generating processes.*

I. INTRODUCTION

The capital market in any country is one of the major pillars of long-term economic growth and development. The market serves a broad range of clientele, including different levels of government, corporate bodies and individuals within and outside the country. Capital markets help in boosting the financial system as well as mobilizing long-term debt and equity finance for investments in long-term assets. This market is a framework through which businesses and government raise long term funds for investment. It has been one of the major means through which foreign funds are injected into most economies and the tendency towards a global economy is more visible there than anywhere else. It is therefore quite valid to state that the growth of the capital market has become one of the most important tools for measuring the overall economic growth of a nation. Many studies have been carried out on the stock market, [1-10]. This paper presents examples of problems in estimation that demonstrate the use and performance of the parametric and nonparametric bootstrap in econometric settings. The examples are illustrated with two empirical applications of bootstrap on Nigeria stock market data sets.

Several research works have been carried out in the area of reviewing empirically the bootstrap methods in econometrics, [11-26], but none has been applied on Nigeria stock market data sets. Some of the key determinants of capital market capitalization are real income, saving rate, financial intermediary development, and stock market liquidity.

In this study, the various bootstrap DGP from parametric and nonparametric approaches were the estimation methods of interest. Therefore, the purpose of this study is to ascertain the performance of the Nigeria

stock market and to compare the nonparametric and parametric bootstrap DGP methods in estimating the bias, standard errors and other information criteria of Nigeria stock market data sets. The sample size of Nigeria stock market data sets available was very small that bootstrap needed to be carried out. However, this study will focus on the bootstrapping regression models from real income and savings rate on stock market capitalization under a variety of assessment conditions. In the context of regression, standard error, bias and other information criteria can be examined at estimation point. It is pertinent to note that finding bias, standard error and other information criteria may lead to selection of accurate statistical inference, test samples, or test procedures. Reporting standard error and bias on available data, such that observations of certain kind may be more likely to be reported and consequently used in this research.

II. MATERIALS AND METHODS

The procedures for nonparametric and parametric bootstrap methods are as follows;

Step 1: Randomly draw a sample of size “N” from the population distribution. This random sample is called Replication *r*. In this study, only the stock market capitalization was considered, thus, a sample point reflects both the real income and savings rate.

Step 2: From Replication *r*, randomly draw a sample of size *N* with replacement. This random sample is named the Bootstrap Sample *b*.

Step 3: Evaluate the bootstrap samples from Step 1 and Step 2. For a specific point is denoted as * () where *b* indicates the *b*th bootstrap sample, and the superscript * emphasizes that these results are from the bootstrap samples (as opposed to the original sample).

Step 4: Repeat Step 1 to Step 3 *B* times to obtain * ().

Step 5. Obtain the estimated bias and standard error at each point.

While, the parametric bootstrap method uses simulation steps that are similar to those in the nonparametric bootstrap method except that:

- i. A parametric model is first fit to the replications drawn from the population.
- ii. Bootstrap samples are drawn from the fitted replication distribution rather than the original replication.

A. Parametric Bootstrap Data Generating Process (DGP)

Given (1), under the assumption that the error terms are normally distributed.

$$y_t = X_t\beta + u_t, \mu_t^* \sim NID(0, \sigma^2) \quad (1)$$

The first step in constructing a parametric bootstrap DGP is to estimate (2) by OLS, yielding the restricted estimates $\hat{\beta}$, and $\hat{\sigma}^2$. Then the bootstrap DGP is given by

$$y_t^* = X_t\hat{\beta} + \mu_t^*, \mu_t^* \sim NID(0, \hat{\sigma}^2), \quad (2)$$

y_t which is just the element of the model (2) characterized by the parameter estimates under the null, with stars to indicate that the data are simulated. In order to draw a bootstrap sample from the bootstrap DGP (2), we first draw an *n*-vector u^* from the $N(0, \hat{\sigma}^2)$ distribution. The rest of the procedure for computing a bootstrap *P* value is identical to the one for computing a bootstrapped *P* value for exact test. For each of the *B* bootstrap samples, θ_t^* a bootstrap test statistic θ_t^* is computed from y_t^* in just the same way as $\hat{\theta}$ was computed from the original data, y_t in (1).

B. Nonparametric Bootstrap Data Generating Process (DGP)

The parametric bootstrap procedure that we have just described, based on the DGP (2), does not allow us to relax the strong assumption that the error terms are normally distributed. How can we construct a satisfactory bootstrap DGP if we extend the models (2) to admit non-normal errors? If we knew the true error distribution, whether or not it was normal, we could always generate the μ^* from it. Since we do not know it, we will have to find some way to estimate this distribution.

The nonparametric method that will be adopted in this study is called “pairs bootstrap”, which was proposed by [22]. This method was applied to regressions with instrumental variables by [27]. The idea is to resample the data instead of the residuals. Thus, in the case of the regression model (1), we resample from the matrix [y X] with typical row [ytXt]. Each observation of the bootstrap sample is, a randomly chosen row from [y X]. This method is called the pairs (or pairwise) bootstrap because the dependent variable and the independent variables are always selected in pairs unlike the parametric methods.

The pairs bootstrap DGP which is

$$y_t^* = X_t^*\hat{\beta} + \mu_t^*, [y_t^*, x_t^*] \sim NID(\bar{x}, s^2) \quad (3)$$

C. Evaluation Criteria

The following statistical evaluation criteria were used to investigate and understand the bootstrap DGP methods and to compare the different types of bootstrap DGP for the simple linear regression (SLR) models under a variety of assessment conditions. Generally, to evaluate a satisfactory degree of performance and validity of the DGP methods for the stock market capitalization and its best model, several assessment conditions were evaluated 200 times to estimate the standard error and bias from the two bootstrap

methods in this paper. In bias test by [28, 29], a difference of 0.1 standard deviation units is generally considered relatively large, whereas a difference of 0.25 is regarded as very large. This style will be adopted in this paper.

These several assessment conditions are described below;

- i. Bootstrap method; as indicated earlier, the parametric bootstrap (PB) and nonparametric bootstrap (NPB) methods were considered.
- ii. The bootstrapped ability levels investigation and evaluation will be described in the same form; N(0,1). Here, we use the standard error to get the group differences.
- iii. Different sample sizes of three tests lengths: $n_1=200$, $n_2=1000$ and $n_3=10000$ were studied. These levels represented typical small, medium, and large sample sizes.
- iv. Bootstrap levels (PB and NPB levels); B-Level will be 99, 499, 1999. These levels also represented typical small, medium, and large sample sizes and satisfies the pivotal conditions.
- v. Other information criteria (model selection criteria): In this section several criteria that will be used to compare and choose among bootstrap DGP. In-sample will be considered since it essentially tells us how the chosen model fits the data in a given sample.

The formulas for other information criteria used to generate the values in Table 3 (see Appendix), are shown as follows;

a. Adjusted R² criterion

As a penalty for adding regressors to increase the R² value. Formula for adjusted R denoted by \bar{R}^2 is

$$\bar{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)} = 1 - (1 - R^2) \frac{n-1}{n-k} \quad (4)$$

It is pertinent to note that $\bar{R}^2 \leq R^2$, unlike R², the adjusted R² will increase only if the absolute t value of the added variable is greater than 1. For comparative purposes, therefore, R² adjusted is a better measure than R².

b. Akaike Information Criterion (AIC):

The idea of imposing a penalty for adding regressors to the model has been carried further in the AIC criterion, which is defined as:

$$AIC = e^{2k/n} \frac{\sum \hat{u}_i^2}{n} = e^{2k/n} \frac{RSS}{n} \quad (5)$$

Where k is the number of regressors (including the intercept) and n is the number of observations. For mathematical convenience, equation A, is written as

$$AIC = \left(\frac{2k}{n}\right) + \ln\left(\frac{RSS}{n}\right) \quad (6)$$

Where \ln AIC = natural log of AIC and $2k/n$ = penalty factor. In comparing two or more models, the model with the lowest value of AIC is preferred. One advantage of AIC is that it is useful for not only in-sample but also out-of-sample performance of a regression model, [30].

c. Schwartz Bayesian Information Criterion (SBIC):

Similar in spirit to the AIC, [31], also known as Schwart Information Criterion (SIC) or Bayesian Information Criterion (BIC) is also consistent, unbiased and sufficient. The only difference between AIC and SBIC is that SBIC imposes a harsher penalty than AIC. The SBIC criterion is defined as:

$$SBIC = n^{k/n} \frac{\sum \hat{u}_i^2}{n} = n^{k/n} \frac{RSS}{n} \quad (7)$$

Or in log-form

$$\ln SBIC = \frac{k}{n} \ln n + \ln\left(\frac{RSS}{n}\right) \quad (8)$$

Where $[(k/n)\ln n]$ is the penalty factor. Like AIC, the lower the value of SBIC, the better the model. Again, like AIC, SBIC can be used to compare in-sample or out-of-sample performance of a model.

d. Hannan-Quinn information criterion (HQIC)

In statistics, the Hannan-Quinn Information Criterion (HQIC) is a criterion for model selection. It is an alternative to Akaike Information Criterion (AIC) and Schwartz Bayesian Information Criteria (SBIC) or Bayesian Information Criterion (BIC). And it is given as

$$HQIC = n \log\left(\frac{RSS}{n}\right) + 2k \log \log n \quad (9)$$

where k is the number of parameters, n is the number of observations, and RSS is the residual sum of squares that results from linear regression or other statistical

Table 1: Comparison of Real Income Bias and Standard Error of the SLR for Parametric and Nonparametric Bootstrap Models in a Stock Market Capitalization.

Bootstrap Ability Level	Sample Size	Bias for the Bootstrap Models		Standard error for Bootstrap Models	
		PSMCM	NPSMCM	PSMCM	NPSMCM
B=99	N(0,1) 200	0.0025	0.0017	0.0809	0.0790
		0.0008		0.0019	
	0.0317	0.0312	0.0282	0.0275	
	0.0005		0.0007		
	1000	0.0018	0.0015	0.0121	0.0117
	10000	0.0003		0.0004	
B=499	N(0,1) 200	0.1224	0.1220	0.1189	0.1181
		0.0004		0.0008	
	0.0319	0.0311	0.0498	0.0500	-
	0.0008	0.0158	0.0002		

10000	0.0160 0.0002	-	0.03370.03370.0000
B=1999 N(0,1) 200	0.1220 0.0001	0.1219 0.0590	0.1197 0.0006 0.0282 0.0336
1000	0.0572 0.0018		0.0022 0.0022 0.0022 0.0022
10000	0.0009 0.0000	0.0009	0.0000 0.0000

Note. Bold values signify smallest bias and standard error in the parametric and nonparametric bootstrap models in table 1, including their differences.

Table 2: Comparison of Saving Rate Bias and Standard error of the SLR for Parametric and Nonparametric Bootstrap Models in a Stock Market Capitalization.

BootstrapAbility sample Level Size	Level	Bias for the Bootstrap Models		Standard error for Bootstrap Models	
		PSMCM	NPSMCM Diff	PSMCM Diff	NPSMCM
B=99	N(0,1)	0.0024	0.0015	0.0806	0.0790
	200	0.0009		0.0016	
	1000	0.0316	0.0312	0.0282	0.0275
	10000	0.0004	0.0015	0.0121	0.0116
B=499	N(0,1)	0.1223	0.1220	0.1129	0.1111
	200	0.0003		0.0008	
	1000	0.0318	0.0312	0.0498	0.0500 -
	10000	0.0006	0.0162	0.0002	0.0032
B=1999	N(0,1)	0.1222	0.1219	0.1197	0.1128
	200	0.0003		0.0006	
	1000	0.0591	0.0572	0.0282	0.0336
	10000	0.0019	0.0007	0.0000	0.0021 0.0021 0.0000

Note. Bold values signify smallest bias and standard error in the parametric and nonparametric bootstrap models in table 1, including their differences.

IV. DISCUSSIONS

In this study, three models were selected to represent the real and bootstrap data sets after more than 200 trials were carried out within each bootstrap level (B). The selection was based on the fact that as n (number of trials) increase, the models maintain the same pattern, and unless there is change in the pattern another model will not be selected. The three equations (11, 13, 15), represent each of the groups of models selected; results presented in Tables 1, 2 & 3 will be discussed. Also, to present the results more clearly, three subsections included as stated above will be further explained.

Table 1, shows that across all the bootstrap conditions, it was obvious that all the models that worked well have very low HQIC, SBIC, AIC, adjusted R²>50%. In order to highlight the differences among the models, evaluate and ascertain the performance of the bootstrap DGP methods on Nigerian capital market data sets, two separate groups are laid out, with (a) for model -ODSMC from the real data set (b) for models PSMC and NPSMC from the bootstrap data sets. This will enable determine the effects of the factors of sample size and bootstrap level on a stock market data sets. Extreme values in the ranges stated above were truncated and very low estimates were also observed, results in these ranges are presented in order to demonstrate the trends and the performance at the lower ends of the distributions for each bootstrap model. The bootstrap models provide information to evaluate the relative effects of sample size, and group proficiency level on the bias, standard error of the stock market.

It can be seen from Tables 1 & 2, that sample size and test length of bootstrap level had large effects on bias of the SLR, ability level had relatively small or mixed effects under some conditions bias was smaller for a larger sample size and a shorter test length. Given the same test length, a smaller ratio normally yielded slightly larger bias, especially for the parametric model with the smaller estimate. Table 2&3 show that NPSMC model performed very well in all the various assessment conditions, expect at B=499, N(0,1), n=10000 in Table 2, that is, the bias and standard error values of PSMC and NPSMC models are (0.0158 and 0.0337) and (0.0160 and 0.0337) respectively. At this point PSMC and NPSMC models have the same standard error. Also, B=1999, N(0,1), n=10000, both models have approximately the same bias and standard error in Tables 2 & 3. Although the effect of group proficiency level on bias of the simple linear regression

(SLR) was quite small, it seemed there was an interaction effect between real income and savings rate.

For the nonparametric bootstrap DGP method, as the group differences became larger, the bias of the SLR became somewhat smaller; however, for the parametric methods with a longer test, the bias of the SLR became slightly larger as the groups were more different. There was no evidence showing any effect of the group proficiency level on the parametric method with a short test. Across all the conditions considered, models PSMC model yielded much larger bias than NPSMC model in most of the estimates. Therefore, for the bootstrap models considered, the pattern was clear that at lower sample sizes and bootstrap levels NPSMC model behave better with respect to bias and standard error. However, it does not mean the higher sample sizes were always associated with the smaller bias along all the estimated values [14], [15], [19], [16] and [32]. As a result, the real income sector can exert a positive influence on the stock market through the volume of savings. Therefore, stock market and economic growth positively influence each other in the process of development. In reality, real income, saving rate interact during all stages of development. In other words, there exist a positive correlation between real income and saving rate on the stock market.

A general observation is that across the group ability level, as the sample size and bootstrap level increased, the bias reduced, meanwhile, the differences among the parametric and nonparametric bootstrap models were becoming more similar. Also, as the sample size, bootstrap level increased, the standard error generally decreased. However, the differences between the results from PSMCM and NPSMC were small (≤ 0.005) both bootstrap models can be recommended for predictions of stock market performance, since the determinants are highly significant.

V. CONCLUSION

When the sample size was moderate (≤ 200) in almost all the bootstrap conditions, the nonparametric bootstrap method performed better than all of the parametric bootstrap models by showing the smallest bias and standard error. Based on the information criteria, across all the bootstrap conditions, it was obvious that two models worked well, having very low HQIC, SBIC, AIC and adjusted R^2 and confirming that the models are good model for further studies and predictions in the economic sectors especially in the stock market. Based on the evidence from

this study, the paper concludes that bootstrap data generating processes can be used to infer the statistical properties of the estimators and constructed inference based on bootstrapped data sets as done with the stock market capitalization data set.

REFERENCES

- [1] Adebisi, M. A. (2005). Capital Market Performance and the Nigerian Economic Growth. In Oluwatayo O.F. and Olasupo, A. (eds) Issues in Money, Finance and Economic Management in Nigeria. Lagos: University Press.
- [2] Alite, H. I. (1986). The Nigerian Stock Exchange Market in Operation. Ibadan: Jeromeliabo and Associates Nigeria.
- [3] Al-Faki, M. (2006). The Nigerian Capital Market and Socioeconomic Development. Paper presented at the 4th Distinguished Faculty of Social Science Public Lecture, University of Benin, 26 July, pp. 9-16.
- [4] Al-Faki, M. (2007). Capital Market Reforms and Real Sector Financing: What Option? SEC News, Vol.4 No. 1, October.
- [5] Anao O. E. (1994). Essence of Capital Market. SEC quarterly reports, April 1994 Vol. 3 No. 2, Abuja.
- [6] CBN Publication (2008). Statistical Bulletin. Vol. 18 December, Abuja.
- [7] CBN Publication (2004 - 2008). Annual Reports and Statement of Accounts. Abuja.
- [8] Dare, A (2003). Principles of Economics (MACRO). Isheri-Osun, Lagos: Darkol Press and Publishers.
- [9] Ekezie, E. S. (2002). The Elements of Banking: Money, Financial Institutes and Markets. Onitsha: Africana - Feb Publishers Limited, Nigeria.
- [10] Ekineh, S. D. (1996). The Securities and Exchange Commission and Investor Protection in the Capital Market. Security Market Journal, Special Edition, (9);1-86.
- [11] Acha, C. K. (2010). Impact of Housing and Age of Birds on Egg Production. *International Journal of Numerical Mathematics*, 5(2):275 – 298.
- [12] Acha, C. K. (2012a). Regression and Principal Component Analyses: a Comparison Using Few Regressors. *American Journal of Mathematics and Statistics*, 2(1): 1-5 DOI: 10.5923/j.ajms.20120201.01.
- [13] Acha, C. K. (2012b) On two Methods of Analysis Balanced Incomplete Block Designs. *Pakistan Journal of Statistics and Operation Research*, 8(4), 749 –757.

- [14] Acha, C. K. (2014). Parametric Bootstrap Methods for Parameter Estimation in SLR Models. *International Journal of Econometrics and Financial Management*, 2014, 2(5), 175-179. Available online at <http://pubs.sciepub.com/ijefm/2/5/2> © Science and Education Publishing DOI: 10.12691/ijefm-2-5-2.
- [15] Acha, C. K. (2014b) Bootstrapping Normal and Binomial Distributions. *International Journal of Econometrics and Financial Management*, 2(6), 253– 256. Doi:10.12691/ijefm-2-6-2.
- [16] Cover, T. M.; Thomas, J. A. (2006). *Elements of Information Theory*. John Wiley and Sons. p. 254.
- [17] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- [18] Chernick, M. R. (1999). *Bootstrap Methods: A Practitioners Guide*. Wiley, New York.
- [19] Acha, C.K. and Acha I.A. (2015) Smooth Bootstrap Methods on External Sector Statistics. *International Journal of Econometrics and Financial Management*, 3(3), 115–120. Doi:10.12691/ijefm-3-3-2.
- [20] Chernick, M. R. and LaBudde, R. (2011). *An Introduction to the Bootstrap with Applications in R* Wiley, Hoboken.
- [21] Acha, C. K. and Omekara, C. O. (2016) Towards Efficiency in the Residual and Parametric Bootstrap Techniques. *American Journal of Theoretical and Applied Statistics*. 5(5) 285-289. doi: 10.11648/j.ajtas.20160505.16.
- [22] Freedman, D.A. (1981) Bootstrapping regression models, *Ann. Statist.* 6, 1218- 1228.
- [23] Lehikoinen, A., Saurola, P., Byholm, P., Linden, A. and Valkama, J. (2010). Life history event of eurasian sparrow hawk in a changing climate. *Journal of Avian Biology*, 41; 627-636.
- [24] Acha, I. A. and Acha, C. K. (2011). Interest Rates in Nigeria: An Analytical Perspective. *Research Journal of Finance and Accounting*, 2(3); 71-81 www.iiste.org ISSN 2222-1697 (Paper) ISSN 2222-2847 (Online).
- [25] Chatfield, C. (2004). *The Analysis of Time Series.- An Introduction*. Chapman and Hall, London, UK. Sixth Edition.
- [25] Chatfield, C. (2004). *The Analysis of Time Series.- An Introduction*. Chapman and Hall, London, UK. Sixth Edition.
- [26] MacKay, D. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press; First Edition. ISBN 978-0521642989.
- [27] Freedman , D , A. , and Peters , S. C. (1984). Bootstrapping a regression equation: Some empirical results *J. Am. Statist. Assoc.* **79** , 97 – 106.
- [28] Wang, T., Lee, W., Brennan, R.L. and Kolen, M.J. (2008). A comparison of frequency estimation and chained equipercetile methods under the common-item nonequivalent groups design. *Applied Psychological Measurement*, 32, 632-651.
- [29] Wang, T. and Brennan, R.L. (2009). A modified frequency estimation equating method for
- [30] Akaike, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, AC-19: 716-723.
- [31] Schwarz, G. (1978) Estimating the Dimension of a Model. *The Annals of Statistics*, 6, (2), 461-464.
- [32] Acha, C. K. (2016). Rescaling Residual Bootstrap and Wild Bootstrap. *International Journal of Data Science and Analysis*. 2(1): 7-14. doi: 10.11648/j.ijdsa.20160201.12

APPENDIX:

Table 3: Comparison of PSMC and NPSMC Extract from the Model Output Selection

Lag	HQIC _{NPSMC}	SBIC _{NPSMC}	AIC _{NPSMC}	R ² _(adjusted) _{NPSMC}	HQIC _{PSMC}	SBIC _{PSMC}	AIC _{PSMC}	R ² _(adjusted) _{PSMC}
0	74.7422	69.5277	84.7114	74.7174	78.5476	66.8854	70.1124	73.9982
1	68.1163	66.3343	69.6323	68.6613	62.4412	70.1242	63.3225	62.1863
2	78.4265	65.6843	89.9901	65.9923	61.3321	65.4113	88.0111	64.5213
3	67.7632	77.0111	70.8976	77.1142	79.0001	63.5326	69.5542	69.4475
4	75.0141	64.5376*	66.5221*	70.7701	60.9771*	76.8890	60.5313*	74.7621
5	61.4312*	70.8742	72.0042	63.7236*	71.4002	61.2287*	71.5590	71.2474*