IJRP.ORG
International Journal of Research Publications
ISSN: 2708-3578 (Online)

ON TIME SERIES MODEL ORDER SELECTION OF NON-STATIONARY AND NON-NORMAL DATA STRUCTURE

H. A. Chamalwa[a*], H. R. Bakari[b] and I. Akeyede[c]

[a,b]Departmemnt of Mathematical Sciences, University of Maiduguri

[c] Department of Statistics, Federal University Lafia, PMB 146, Lafia Nigeria

Corresponding Author: chamalwa@gmail.com,

Abstract

The study is aimed at identifying the orders of Time Series Models in Non-Stationarity Non-normal data structure from Uniform distributions with a view to determining the best Autoregressive/ Moving Average orders from time series models (ARMA and ARIMA) under different underline distributions when Non-Stationarity assumption is assumed. The data is generated from Autoregressive (AR) linear of second orders of general classes of Autoregressive functions. The generation of the data used for this simulation study is non-stationary cases and non-normal. The data were simulated for both response variables and error terms from non-normal Distribution. The result shows that the values of the penalty function: AIC, BIC, HQIC and FPE of the order selection increases with increase in sample size but decreases with increasing order. It was observed that at both lower (20, 40) and larger (160,180 and 200) sample sizes, models with smallest orders. Similarly, the selection process is tied to the principle of parsimony i.e smaller orders are selected, but vary with the variation in the distribution of the series. The study recommends the need to develop a methodology for model selection combining objective and subjective techniques.

Keywords: Data, Generation, Linear, Normal, Results, Simulation

1.0: Introduction

Time series analysis in its entirety serves two broad purposes which are, to model a successive observation of a given variable as well as to provide a better understanding of the stochastic mechanism that characterized any observed series and secondly to make predictions of the future values of the observed series on the basis of the historical observation and other possibly relevant factors. Therefore, the need to have a proper and better description of the series cannot be over emphasis. Since, the primary objective of the model selection process is to assess competing models and select the best that describes the data Ongbali, Igboanugo, Afolalu, Udo and Okokpujie (2018). Modeling and forecasting are the main stay of any time series analysis or investigation; in modeling aspect, order identification is very crucial, as the parameters of the fitted model depends largely on the correct order selected and subsequent estimation and performance, as there should be method that is suitable for fitting particular model. Model identification is still a ‘thorny issue’ in robust time series analysis (Martin and Yohai 1986). As Chang (1982) remarks: ‘We need to protect not only the parameter estimation process against the adverse effect of exogenous interventions but also the model identification process so that appropriate model forms for the underlying time series can be specified in the very first place.’ Knowledge of the data generating mechanism also plays a great role in the identification of the tentative model (Cryer and Chan, 2008). Hastie, Tibshirani and Friedman (2009) opined that cardinal reason of model identification is to assess the performance of different models and select the best fit among the several models for particular data. the use final prediction Error (FPE) for model order selection being the expected variance of prediction error when an Autoregressive(AR) model was fitted was suggested and objective of model selection include finding a good predictor that describes a system and Akaike Information Criterion (AIC) is a principal model selection method (Burnham , Anderson and Huyvaert , 2011)

Numerous literatures (Beguin, 1980, Davies 1984, Tsay and Tiao 1985, Choi, 1992, Chan, 1999, Eija, 2015, Norhayati, 2016) abound on order selection since the pioneer work of Box and Jenkins in (1975) but not yet extensive. All of them failed to consider different distributions of error terms and responses in respect to both stationary and non-stationary data structures. More so, the ACF and PACF frequently used in the literature for identification of models' orders depend on the size of the series and very sensitive to outliers or violation of the ideal normal assumption (Stadnytskaet'al, 2008). This study therefore intends to identify a suitable order for the different time series models under different distributions and sample size in fitting and forecasting stationary and non-stationary data structures using different criteria.

2.0 Methodology

For non-stationary cases, the data will be simulated for response variables from Uniform distribution with the parametres of 200, 32 and 38 while that of error terms from normal distribution with mean 1000 and variance 10, so as to violate the white noise assumption of zero mean and difference means and variances of the responses and error terms, thereby violate the Stationarity assumptions i.e;

$Y_{ti} \sim N(200,20) and e_{ti} \sim N(200,10)$For non Stationarity cases

$t = 1,2, \ldots ,20, 40, 60, 80, 100, 120, 140, 160, 180\ and\ 200.\ i = 1,2, \ldots , 1000$

The generation of the data used for this simulation study is non-stationary cases.

x <w- runif(120,32, 38)

x[t] <- 0.7*x[t-1]+0.9*x[t-2]+w[t]

A Non stationary data from a Non-Normal data structures were simulated at various sample sizes of 20, 40, 60, 80, 100, 120, 140, 160, 180 and 200 respectively. Several ARMA (p, q) were fitted on the simulated data; the four different models considered are: ARMA (1, 1), ARMA (1, 2), ARMA (2, 1) and ARMA (2, 2), ARIMA (1, 1, 1), ARIMA (1, 1, 2), ARIMA (2, 1, 1) and ARIMA (2, 1, 2) respectively.

The effect of different levels of ORDER (0.3, 0.6, and -0.3, -0.6) at the sample size of 20, 40, 60, 80, 100, 120, 140, 160, 180 and 200 which represent small, moderate and large sample sizes respectively on the simulated data from the non-stationary and non-normal data . The simulation study was carried out with 1000 iteration on each case in R Statistical software. respectively.

2.1: Model Assessment Criteria

The goodness of fit for each model was assessed using the criteria: of AIC, BIC, HQIC and FPE. The model with lowest criteria value is considered the best among the models for the simulated data. Penalty function are equally widely used for model identification: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Hannan-Quin Information Criterion (HQIC) and Final Prediction Error (FPE)

$$\text{AIC}= \log[\left(\hat{\sigma}^2{}_{p,q}\right) + \frac{(p+q)2}{n} \quad 1.0$$

$$\text{BIC}=\log[\left(\hat{\sigma}^2{}_{p,q}\right) + \frac{(p+q)\log(n)}{n} \quad 2.0$$

$$\text{HQIC}= \log[\left(\hat{\sigma}^2{}_{p,q}\right) + \frac{(p+q)2\log(\log(n))}{n} \quad 3.0$$

$$\text{FPE(p)} = \sigma_p^2\left(1+\frac{p}{N}\right), where \sigma_p^2 = \frac{N}{N-p}\hat{\sigma}^2 \quad 4.0$$

$$\text{FPE(p)} = \frac{N}{N-p}\left(1+\frac{p}{N}\right)\hat{\sigma}^2 = \frac{N+p}{N-p}\hat{\sigma}_p^2 \quad 5.0$$

3. Analysis, Results and Discussion

The orders of each model family considered are investigated and the results were presented in tables according to the model families at various sample sizes of 20, 40, 60, 80, 100, 120, 140, 160, 180 and 200 respectively and the criteria for the assessment, at different sample sizes. Here, the data were simulated from normal of both observed values and error terms to achieve the normality assumption. The criteria like AIC, BIC, HQIC and FPE are used to determine the models' orders such that a model's order with the least criteria is chosen for the model's family.

3.1: Order Determination for ARMA on Stationary Data Structure from Non-Normal Distribution

From each iteration simulated, the values of the criteria for the assessment (AIC and BIC) were computed and their average values were recorded according to sample sizes as shown in table 3.1. The values from the tables were plotted in figures 3.1a and3.1b respectively. The model with lowest criteria value is considered as the best. For each of the iteration the values of the criteria for the assessment (AIC, BIC, HQIC and FPE) were computed and their average values were recorded according to sample sizes as shown in tables above

Table 3.1: AIC and BIC Values of ARMA (p) Model for UNIFORM Data

	AIC				**BIC**			
Sample Sizes	**ARMA(1,1)**	**ARMA (1,2)**	**ARMA (2,1)**	**ARMA (2,2)**	**ARMA (1,1)**	**ARMA (1,2)**	**ARMA (2,1)**	**ARMA (2,2)**
20	110.295	115.601	112.926	105.838	114.2785	120.5797	117.905	111.8131
40	194.483	187.526	190.616	201.834	201.2392	195.9707	199.061	211.9681
60	282.374	285.84	280.839	280.850	290.7514	296.3118	291.310	292.7765

80	368.496	346.248	390.522	370.153	378.0245	358.1585	402.432	384.4456
100	453.962	443.504	452.556	454.149	464.383	456.5299	465.581	469.7808
120	447.750	582.033	524.198	541.950	458.1709	595.9712	538.136	558.6753
140	638.439	604.282	592.576	590.849	650.2057	618.9902	607.284	605.4997
160	738.885	672.108	681.703	670.410	751.1865	687.4843	697.079	688.8611
180	764.501	766.719	777.199	775.848	777.2734	782.6839	793.164	795.0064
200	912.324	851.023	846.967	841.241	925.5177	867.515	863.458	861.0312

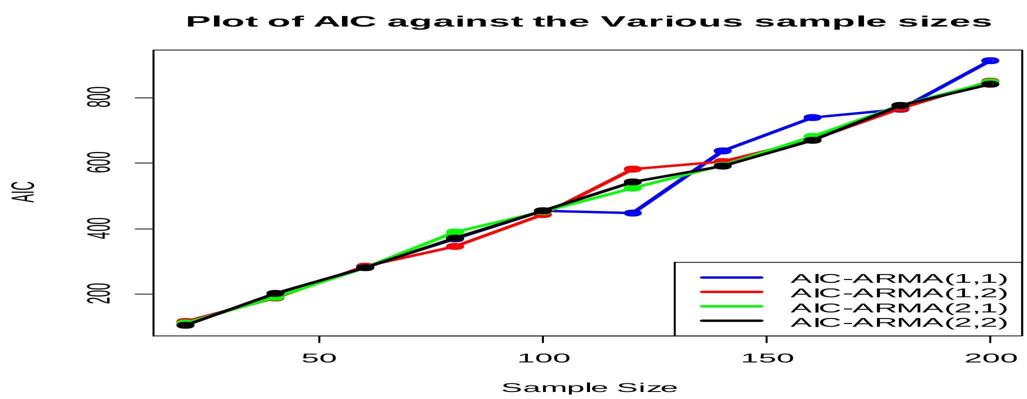

Fig.3.1a: AIC values for ARIMA (p, d, q) Models form a Non-normal Data

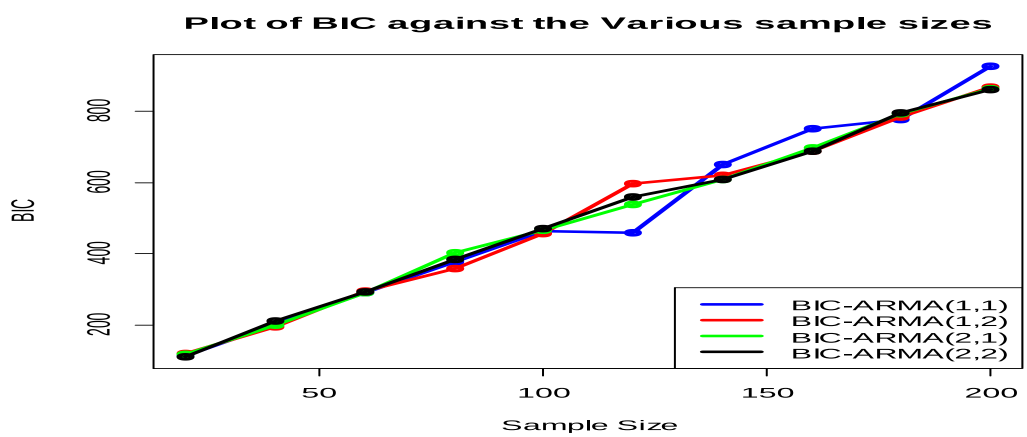

Fig.3.1b: BIC values for ARIMA (p, d, q) Models form a Non-normal Data

IJRP.ORG
International Journal of Research Publications
ISSN: 2708-3578 (Online)

Table 3.1 shows the relative performance of the four fitted models on the simulated data at various sample sizes. The average values of AIC and BIC for each of the models are recorded in the table above and then plotted in figures 3.1a and 3.1b respectively. At sample sizes 20,140,160 and 200 ARMA (2, 2) was selected as the best fit. ARMA (2, 1) was the best fit at sample sizes 60 and 120. While ARMA (1, 2) was the best fit at sample sizes 40, 80 and 100 respectively.

Table 3.2: HQIC and FPE Values of ARMA (p) Model for Stationary Non-normal Data(UNIFORM)

	HQIC				**FPE**			
Sample Sizes	**ARMA(1,1)**	**ARMA (1,2)**	**ARMA (2,1)**	**ARMA (2,2)**	**ARMA (1,1)**	**ARMA (1,2)**	**ARMA (2,1)**	**ARMA (2,2)**
20	107.143	115.717	107.077	106.511	68.02588	72.34176	66.4970	64.63
40	197.331	190.742	192.882	194.093	110.1149	104.7689	106.012	105.1989
60	281.677	271.916	270.576	293.316	150.99	144.0379	143.297	154.3058
80	385.265	355.420	380.140	361.475	202.8649	185.187	198.510	186.8578
100	463.163	424.137	441.477	452.271	241.0412	218.6998	227.906	232.0155
120	514.856	513.568	528.608	528.600	265.6905	263.3672	271.272	269.6223
140	614.786	595.521	610.501	609.037	315.8526	304.1307	311.948	309.5167
160	698.926	688.554	704.754	671.923	357.759	350.6888	359.098	340.3689
180	796.383	771.138	822.278	787.773	406.5805	391.8268	418.263	398.7229
200	924.504	850.939	870.839	865.413	471.1764	431.5553	441.808	437.295

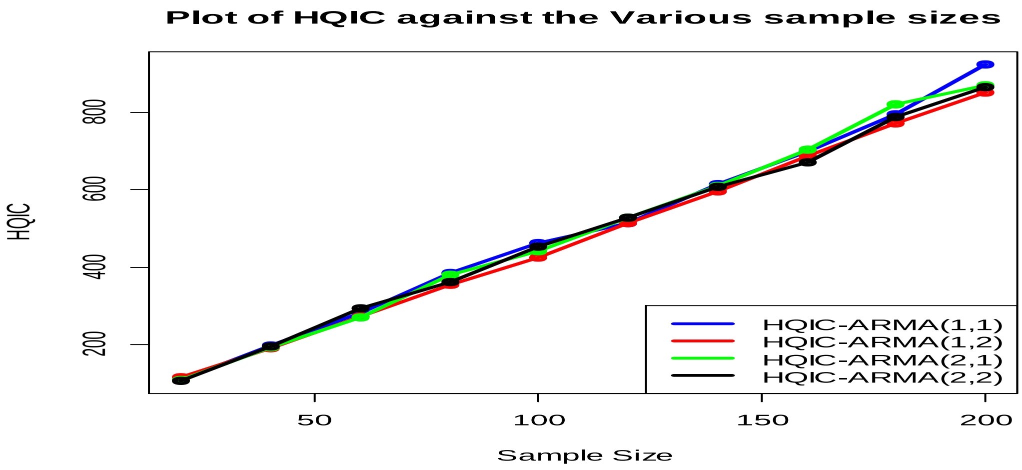

Fig.3.2a: AIC values for ARIMA (p, d, q) Models form a Non-normal Data

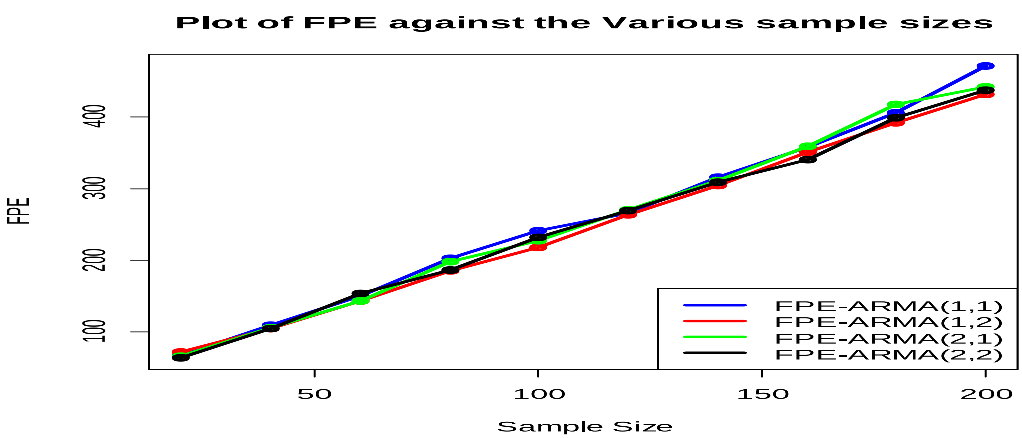

Fig.3.2b: AIC values for ARIMA (p, d, q) Models form a Non-normal Data

Similarly, table 3.2 above shows the relative performances of the four fitted models on the simulated data at various sample sizes are presented. The average values of HQIC and SIC for each of the models are recorded in the table above and plotted in figures 3.2a and 3.2b respectively. It was observed that the values of both AIC and BIC increases with increase in sample sizes of the simulated data. At sample sizes 20,140,160 and 200 ARMA (2, 2) was selected as the best fit. ARMA (2, 1) was the best fit at sample sizes 60 and 120. While ARMA (1, 2) was the best fit at sample sizes 40, 80 and 100 respectively.

3.2: Determination of Best order of ARIMA Based on AIC and BIC Criteria

From each iteration simulated, the values of the criteria for the assessment (AIC and BIC) were computed and their average values were recorded according to sample sizes as shown in table 3.3. The values from the tables were plotted in figures 3.3 and 3.3b respectively. The model with lowest criteria value is considered as the best.

IJRP.ORG International Journal of Research Publications ISSN: 2708-3578 (Online)

Table 3.3 AIC and BIC Values of ARIMA (p) Model for UNIFORM Data

	AIC				BIC			
Sample Sizes	**ARIMA(1,1,1)**	**ARIMA (1,1,2)**	**ARIMA (2,1,1)**	**ARIMA (2,1,2)**	**ARIMA (1,1,1)**	**ARIMA (1,1,2)**	**ARIMA (2,1,1)**	**ARIMA (2,1,2)**
20	102.756	108.535	110.195	108.164	105.59	108.4151	113.973	112.887
40	202.937	188.239	212.130	187.926	207.9285	194.8939	218.785	196.2447
60	292.051	278.122	290.181	274.508	298.2836	286.4323	298.491	284.8965
80	373.930	360.644	358.135	357.878	381.0387	370.1222	367.613	369.7259
100	463.040	458.664	448.306	441.796	470.8255	469.0448	458.686	454.7721
120	570.121	521.541	543.579	537.352	578.4584	532.6576	554.696	551.2484
140	614.511	597.187	624.306	584.610	623.3148	608.9254	636.044	599.2827
160	726.655	712.008	712.051	681.150	735.8625	724.2845	724.326	696.4951
180	822.195	765.903	782.957	729.186	831.758	778.6526	795.707	745.1232
200	898.535	850.923	861.888	853.599	908.4151	864.097	875.061	870.0656

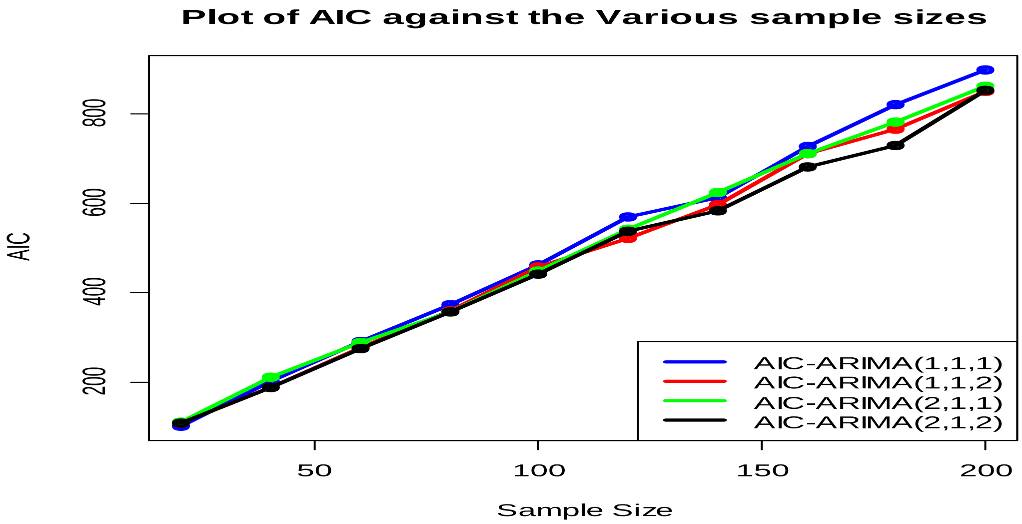

Fig.3.3a: AIC values for ARIMA (p, d, q) Models from a Non-normal Data

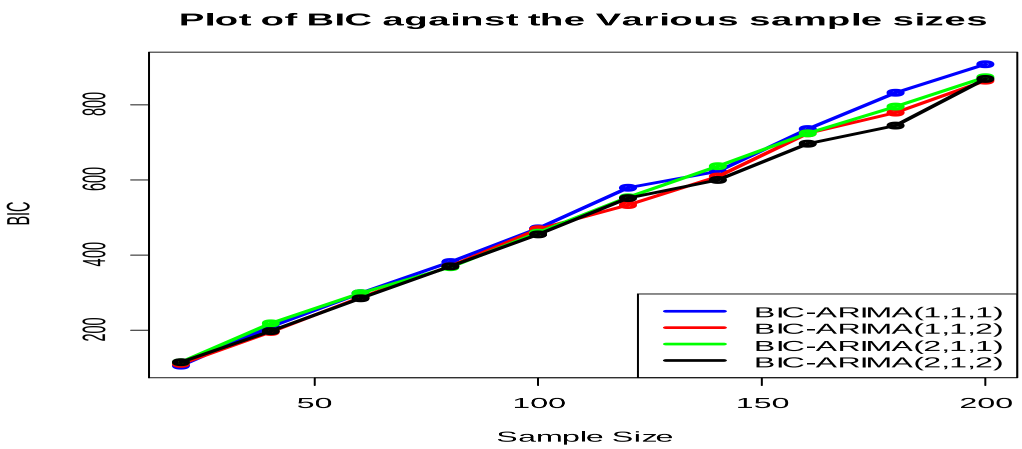

Fig.3.3b: BIC values for ARIMA (p, d, q) Models from a Non-normal Data

Table 3.3 above shows the relative performance of the four fitted models on the simulated data with 1000 iterations at various sample sizes. The average values of AIC and BIC for each of the model are recorded in the table above and then plotted in figures 3.3a and 3.3b respectively. At sample sizes 20, ARIMA (1,1,1) was selected as the best fit. ARMA (2, 1, 2) was the best fit at sample sizes 40, 60, 80,140, 160, 180 and 200 respectively. While ARIMA (1, 1, 2) was the best fit at 120

Table 3.4 HQIC and FPE Values of ARIMA (p) Model for Stationary UNIFORM Data

	HQIC				FPE			
Sample Sizes	**ARIMA(1,1,1)**	**ARIMA (1,1,2)**	**ARIMA (2,1,1)**	**ARIMA (2,1,2)**	**ARIMA (1,1,1)**	**ARIMA (1,1,2)**	**ARIMA (2,1,1)**	**ARIMA (2,1,2)**
20	98.4687	102.163	109.063	92.5975	53.64235	64.65706	69.3247	56.70176

40	196.121	192.831	192.431	190.682	126.1213	107.5	107.267	102.3827
60	267.178	277.857	287.237	273.896	144.5353	148.8789	154.062	145.1321
80	391.59	357.765	352.845	366.100	207.8665	188.0435	185.391	190.9431
100	478.188	452.143	455.803	463.097	250.6404	235.1904	237.133	239.3847
120	556.144	532.496	575.216	541.568	289.0395	274.9628	297.418	287.5723
140	640.450	616.526	616.186	620.641	330.9145	316.7607	316.583	317.2408
160	723.697	696.066	702.806	697.454	372.3045	356.2744	359.773	355.3089
180	806.649	767.843	772.523	780.898	413.5903	391.8268	394.246	396.8722
200	888.189	854.864	857.964	815.899	454.1842	435.2959	436.893	413.5017

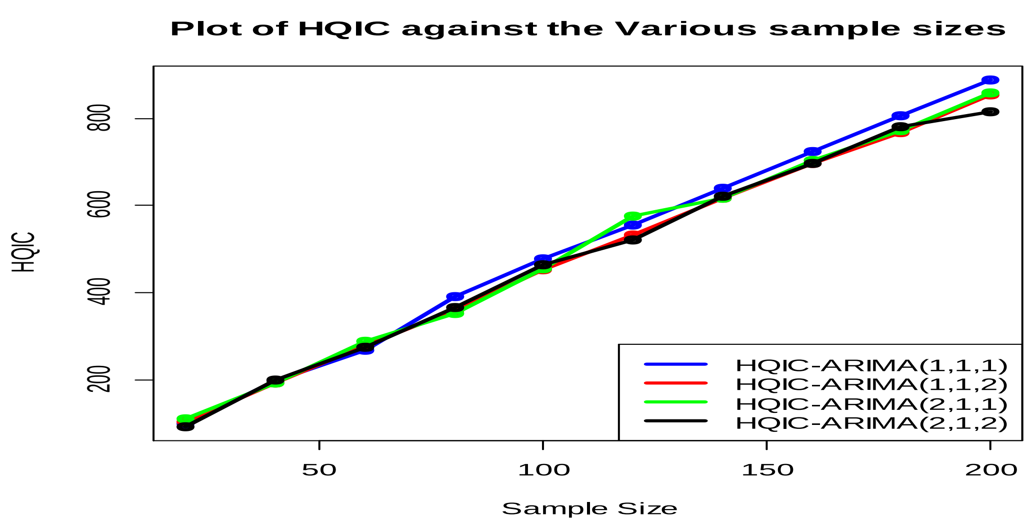

Fig.3.4a: HQIC values for ARIMA (p, d, q) Models from a Non-normal Data

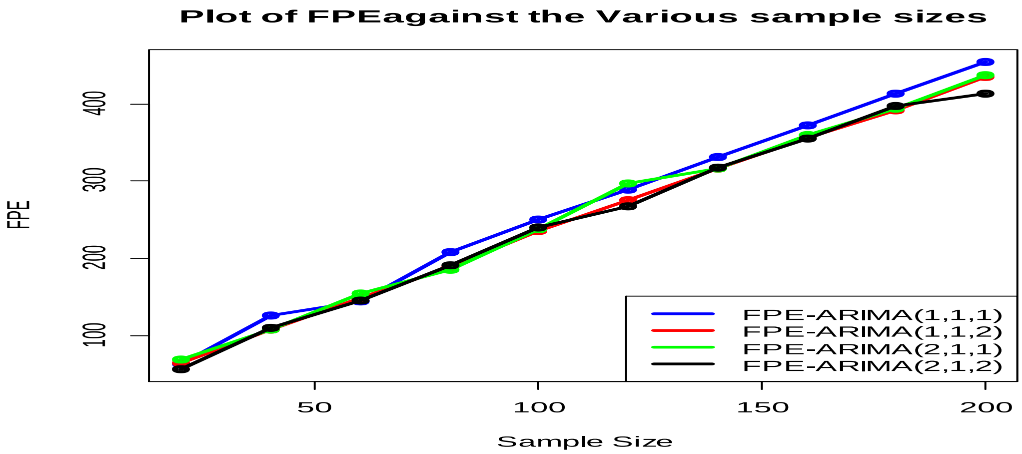

Fig.3.4b: FPE values for ARIMA (p, d, q) Models from a Non-normal Data

Similarly, table 3.4 above shows the relative performance of the four fitted models on the simulated data with 1000 iterations at various sample sizes. The average values of HQIC and FPE are recorded and then plotted in figures 3.4a and 3.4b. At sample sizes 20, ARIMA (1, 1, 1) was selected as the best fit. ARMA (2, 1, 2) was the best fit at sample sizes 40, 60, 80,140, 160, 180 and 200 respectively. While ARIMA (1, 1, 2) was the best fit at 120

4.0: Conclusion and Recommendation

The general conclusion is that for Non-stationary Non-normal data, ARMA smaller orders were picked at almost all the sample sizes, for ARMA and ARIMA respectively. That at both lower (20, 40) and larger (160,180 and 200) sample sizes, models with smallest orders [ARMA (1, 2) and ARMA (2, 1)] were picked on the average while at medium sample sizes, models with larger orders were picked. Similarly, for ARIMA (p, d, q) models, from sample sizes 0f 20-60, models with smaller orders were picked [ARIMA (1, 1, 2) and ARIMA (2, 1, 1)] and at medium sample sizes from80-120 larger orders were picked ARIMA (2, 1, 2). The selections are almost identical in both d non-normal data structures, but vary with the variation in the distribution of the series.

Based on the results of the present study there is the need to develop a methodology for model selection combining objective and subjective techniques. Since, majority of researchers are unlikely to know the type of ARIMA process underlying the data understudy, so it to avoid relying on precarious procedures.

REFERENCES

Akaike, H., (1976). *Canonical correlation analysis of Time Series and the use of an Information Criterion In Mehra*, R.K., Lainiotis, D.G. (Eds.), Systems Identification: Advances and Case Studies. Academic Press, New York, pp. 27—96

Burnham K. P., Anderson D. R. and Huyvaert K. P. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons [J]. Behav Ecol Sociobiol, 2011, 65:23–35

Chan, W. S. (1999). A comparison of some of Pattern Identification Methods for order Determination of Mixed ARMA models *Statistics& Probability Letters*, 42, 69-79.

Chang, I. (1982) Outliers in Time Series**;** Ph.D Dissertation, Department of Statistics, University of Wisconsin, Madison, USA

Choi, B., (1992), *ARMA Model Identification,* Springer Verlag, New York

Cryer J. D. & Chan K.S (2008) *Time Series Analysis with Applications in R Second Edition*, Springer Science Business Media, LLC

Davies, N., Petruccelli, J.D., (1984). On the use of the general partial autocorrelation function for order determination in ARMA (p; q) processes. *J. Amer. Statist. Assoc.* 79, 374{377

Eija Ferreira.(2015) Model Selection in Time Series Machine Learning Applications [D]. Academic dissertation of Technology and Natural Sciences of the University of Oulu, Linnanmaa,.

Hastie T., Tibshirani R. and Friedman J.(2009) The Elements of Statistical Learning: Data Mining [J]. Inference and Prediction. Springer Series in Statistics. Springer-Verlag, New York, NY, 2nd edition,.

Martin, R.D. – Yohai, V.J. (1986) Influence Functional for Time Series (with discussion). The *Annals of Statistics 14*, 781–818.

Norhayati Y.(2016) SURE-Autometrtcs Algorithm for Model Selection in Multiple Equations [D]. PhD Thesis, Universiti, Utara, Malaysia,

Ongbali S. O. , Igboanugo A. C., Afolalu S. A. , Udo M. O. and Okokpujie I. P.(2018) Model Selection Process in Time Series Analysis of Production System with Random Output: *Materials Science and Engineering Journal*