



Evaluation of Some Heteroscedastic Models of Generator Plants Noise Production

A. J. SAKA^{1*}, I. AKEYEDE² AND A. BABAWUYA³

ABSTRACT

Regression analysis requires the homoscedasticity condition to fit and predict time series observations. Basically, most real life data have a tendency to exhibit changing in variances which violates the homoscedasticity assumption. As such, it is appropriate to consider outlines that allow the variance to be subjected to the history of the data. This study therefore compare some heteroscedastic models in fitting and forecasting generator noise from one location to another. A stationarity and heteroscedasticity procedures are considered on the data across the locations using Augmented Dickey Fuller (ADF) and Breusch-Pagan (BP) tests respectively. Also, the four heteroscedastic models of different orders are used to model the data. Thereafter, the inadequacies of the models selected are determined for future forecasting in which all generating plants noise data were found to be stationary and heteroscedastic. GARCH(1,1) and GARCH(1,2) are used to fit and forecast the data. It was observed that forecasted values are stationary over time across the locations.

Received: 08/07/2020, Accepted: 01/10/2020, Revised: 02/11/2020. * Corresponding author.
2015 *Mathematics Subject Classification*. 62M10 & 62G08.

Keywords and phrases. Heteroscedasticity, Homoscedasticity, stationarity, GARCH, & Generating noise

¹Department of Mathematics, Obafemi Awolowo University, Ile-Ife 220005, Nigeria.

²Department of Statistics, Federal University Lafia, PMB 146 Lafia, Nigeria

³Department of Mechatronics Engineering, Federal University of Technology, Minna, Nigeria.

E-mails: ajsaka@oauife.edu.ng; imamakeyede@gmail.com; babawuyal@futminna.edu.ng

ORCID of the corresponding author: 0000-0003-0276-5579

1. INTRODUCTION

According to [13], noise pollution is one of the major problems facing people across the globe especially in the city where the number of industries, machinery, vehicles use, of electrical generating plants, use of vibrating and pressurizing equipment during road and building constructions and so on over time. This has led to increase in the noise. Meanwhile, rigorous sound upsets the serenity of the atmosphere of a location and can result to a negative impact on society, climate as well as human fitness. Electrical generating plant noise, vehicle traffic noise and pressure, industrial noise, machinery noise, and construction noise and are joint causes of noise pollution that directly undermine climate variation [11]. The quantification of health effects for other noise sources including wind turbine, neighbour, industrial, and combined noise remain research priority [5].

The long-term noise monitoring data using artificial neural networks (ANN) and autoregressive integrated moving averages (ARIMA) models was studied by [7] in which the ANN models observed to be performed better than the ARIMA. Also, the configuration of ARIMA forecasting model is steering and as such the time-series predictive model utilizing ANN method. In [1], the generators noise was evaluated and its impact on human and the atmospheric condition was examined by taking samples from noise generation across different locations in a city. Thus, the investigation revealed that noise level at the sample locations beyond the recommended level for residential and industrial areas by World Health Organization (WHO). The effective models was developed which predict the levels of noise in an area using a hybrid model [4]. The model is based on two different approach that is Time Series Analysis (TSA) and Artificial Neural Network (ANN). The hybrid of the two models revealed a significant variation in prediction of the steps ahead. An IoT-based noise monitoring system was set up by [14] to capture the environmental noise data, and a two-layer short-term memory (TLSTM) network was suggested for the prediction of noise for the large data. The proposed model outperformed the other existing typical approaches. It was concluded that the TLSTM could disclose variation in noise levels for a day and also has better forecasting precision.

The time series data may possess properties like heteroscedasticity, volatility clustering, skewed tails, absence of autocorrelations and outliers in the response variables. The Autoregressive Conditional Heteroscedasticity (ARCH) model that contains normal errors which was initially introduced by [6] captured some of conventional features of time series data. Also, [3] further introduced additional parameter(s) of the ARCH model to generalized ARCH model (GARCH) to capture the phenomenon earlier mentioned in time series data. However, traditionally, time series were modeled with normal innovations. Unfortunately, such models still failed to sufficiently capture the main conventional features of some time series data, such as the skewed tails and leptokurtic. Several researches

have been examined on the efficiency of GARCH models when the distribution of error term is violated. For example, GARCH models with skewed generalized error distribution (SGED) was suggested by [10] and [2]. The robustness of volatility modeling of GARCH(1,1) was investigated by [8] with first order Exponential GARCH [EGARCH(1,1)] model using the monthly stock market returns of seven emerging countries. It was found that the GARCH(1,1) model formed better than the EGARCH model.

Regression analysis requires the homoscedasticity condition to fit and predict the time series observations. Basically, most real life data have a tendency to exhibit changing in variances which violates the homoscedasticity assumption. As such, it is appropriate to consider outlines that allow the variance to be subjected to the history of the data. Indeed, a robust estimation can be studied when there is a sturdy notion of heteroscedasticity. In homoscedastic phenomenon, it is expected that the discrepancy of the error term is constant for all values of x . Heteroscedasticity permits the variance to be dependent on x , which is more common for many real scenarios. For example, the variance of the generator noise is often different from one location to another. In this study, electrical generating plant is studied, it is the major contributor of noise pollution in the selected locations across the city of Kwara State of Nigeria. The study further evaluated some heteroscedastic models in fitting and forecasting generator noise from one location to another.

2. MATERIALS AND METHODS

This study was conducted at four different locations in Ilorin township of Kwara State. This is an urban area predominately with residential development, shopping complexes, offices and industries. Actual noise levels in these selected areas were measured. Four readings each from the four locations were taken with the aid of the sound level meter, at an interval of thirty seconds for a period of one hour and the average for each location was recorded. The readings were taken from 8am to 9am in the morning, 1pm to 2pm in the afternoon, 4pm to 5pm in the evening, and 9pm to 10pm at night respectively. The plot that displays the observed values on Y-axis and time intervals on the X-axis that was used to assess the behaviour of the data over time are presented in the Figure 1 for the four locations. The data obtained were analyzed to examine whether stationary or not Augmented Dickey Fuller (ADF) statistic. Also, heteroscedasticity phenomenon were tested for various data collected across locations using Breusch-Pagan (BP) test before the models of different orders were used to fit the data with the aim of selecting the best model. Thereafter, the inadequacies of the models selected were determined for future forecasting. Thus, the GARCH model that best fit and forecast the noise data is determined using AIC, BIC and HQIC criteria.

2.1. Analytical Framework of ARCH/GARCH Family Models. The specifications of ARCH and GARCH are the mean and variance equations. The conditional heteroscedasticity in time series denoted by y_t can be modeled using ARCH by expressing the mean equation as follows:

$$(1) \quad y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \epsilon_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \epsilon_t$$

where $\epsilon_t \sim NIID(0, \sigma^2)$. Equation (1) is an autoregressive model of order p and the mean equation is constant for all ARCH and GARCH models.

2.2. The Autoregressive Conditional Heteroscedastic (ARCH) Models. The conditional variance equation of the ARCH (q) model is specified by the series σ_t^2 in the form:

$$(2) \quad \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \dots + \alpha_q \epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2$$

The parameter of ARCH (q) model can be obtained by ordinary least square (OLS) method of estimation. The coefficient of the ARCH terms (α_i) can be tested for, whether is statistically significant or not to determine the presence of ARCH component. The best fitting ARCH (q) model is estimated and the squares of the error (ϵ^2) is obtained. Thus, (ϵ^2) is regressed on a constant (α_0) and the lagged values of the error term

$$\hat{\epsilon}_t^2 = \alpha_0 + \sum_{i=1}^q \hat{\alpha}_i \epsilon_{t-i}^2$$

where q is the lag order. The significance of ARCH parameter are tested as follows:

$$(3) \quad H_0 : \alpha_i = 0, \quad i = 1, \dots, q \quad Vs \quad H_1 : \alpha_i \neq 0$$

The test statistic:

$$(4) \quad \frac{SSR_0 - SSR_1}{SSR_1(T - 2m - 1)}$$

where

$SSR_1 = \sum_{t=1}^T \epsilon_t^2$, ϵ_t^2 is the residual of least square of the linear regression and $SSR_0 = \sum_{t=1}^T (a_t^2 - w)$, $w = \frac{1}{T} \sum_{t=1}^T a_t^2$ is the sample mean of a_t^2

Decision rule: If p-value of the test statistic is less than 0.05, H_0 is rejected in favour of H_1 and conclude that there is an ARCH effect.

2.3. Generalized Autoregressive Conditional Heteroscedastic Model.

The specification of GARCH (p, q) model is given as

$$(5) \quad \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$$

where α_i are parameter of the ARCH component; β_i are parameter of the GARCH component. The coefficients (α_0, α_i and β_i) are non-negative such that $\alpha_i + \beta_i < 1$ to achieve stationarity. To test for GARCH effect in (4), the best fitting AR(q) model is estimated which is a lag order that has the lowest information criteria and highest log-likelihood ratio, see [3] and [12]. Then, we compute and plot the autocorrelations of $\hat{\epsilon}_t^2$ by

$$\frac{\sum_{t=i+1}^T (\hat{\epsilon}_t^2 - \hat{\sigma}_t^2)(\hat{\epsilon}_{t-1}^2 - \hat{\sigma}_{t-1}^2)}{\sum_{t=1}^T (\hat{\epsilon}_t^2 - \hat{\sigma}_t^2)^2}$$

3. DATA ANALYSIS AND RESULTS

This section consists of detail description of the analysis of the data on the generating plants noise which consists of four locations in Nigeria namely, Geri-Alimi, Unity, Taiwo and Fate. The data were collected in seconds sparing from 30sec to 3600sec from the four locations, which creates 120 observations.

3.1. Graphical Display of Data. For the purpose of the flow of the analysis, the data on noise measured per second from generating plants at different locations are shown in figure 1 below.

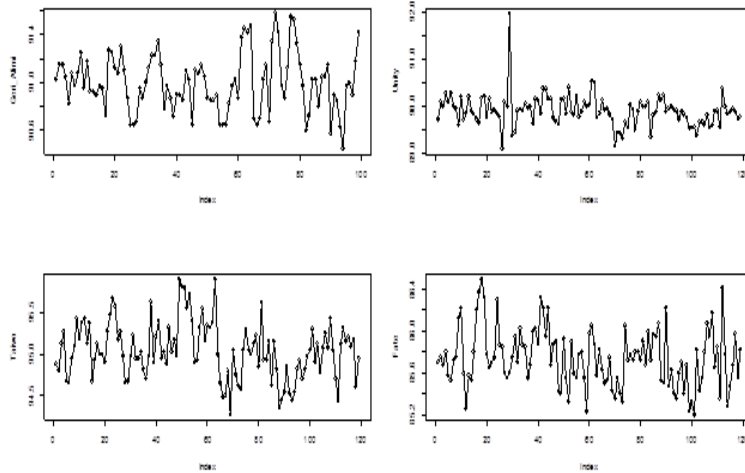


Figure 1: Time Series Plot of Generating Plants Noise from Different Locations

The Figure 1 above indicated clearly that the generating plants noise appeared to be constant over time, at different locations, with systematically visible pattern, no structural breaks and no outlier components found with no constant increasing or decreasing. This is indeed an indication that the data is stationary. Also, a

further test is performed to justify the stationary status using autocorrelation function, partial autocorrelation function and ADF test.

3.2. Model Identification process for the Displayed Data. The Figure 2 and 3 below displayed the ACF and the PACF for consecutive lags in a specified range of lags.

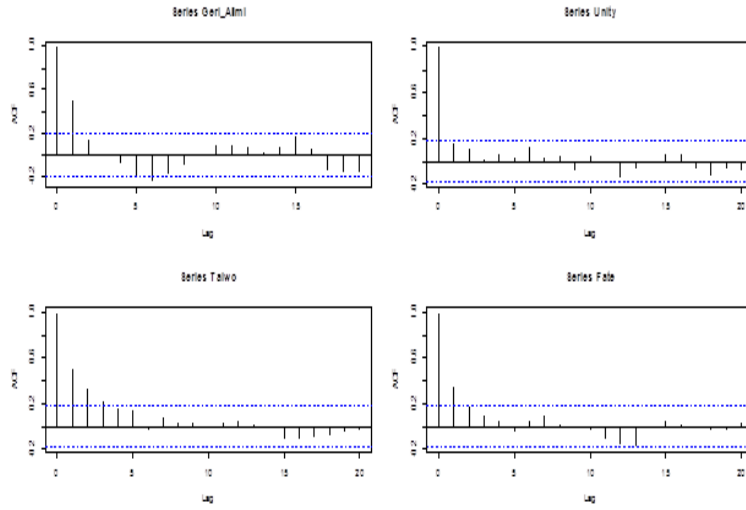


Figure 2: ACF plots of the Generating plants Noise from Different Locations

The ACFs tend or approximate to zero as lag increases, indicating a typical case of stationary series for a stationary process, the main characteristics of the ACF plot suggests that the autocorrelations of the series tend to zero when the lag increases [9].

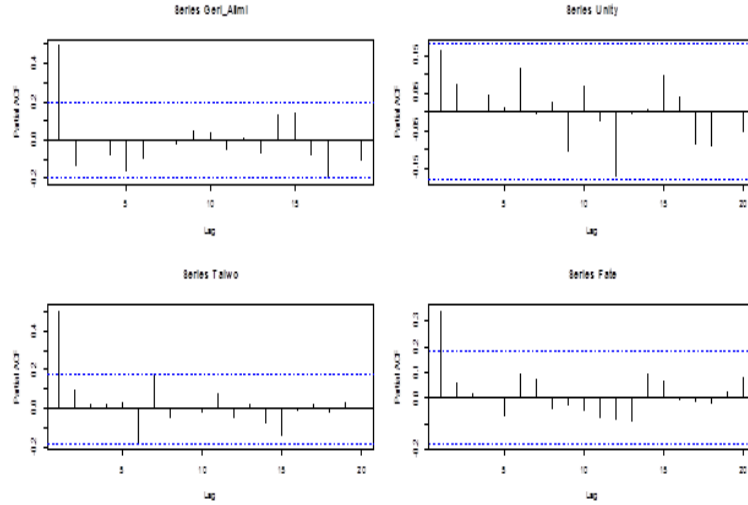


Figure 3: PACF plots of the Generating plants Noise from Different Locations

Again, the partial autocorrelations are examined in the same vein, just like that of the ACF with the partial autocorrelations approximating to zero as the lags increases thereby making the series to be stationary.

3.3. Testing for Unit Root / Stationarity in Generating Plant Noises.

The ADF statistic tests the null hypothesis that the data series have a unit root with the alternative that the data series is stationary. The results obtained are presented in Table 1 below.

Table 1: ADF Test for unit root with respect to the locations

Location	Test Values	Lag order (H_0)	P-value	Hypothesis	Decision	Remark
Geri Alimi	-4.7048	4	0.01	Unit root	Reject (H_0)	Stationary
Unity	-4.4855	4	0.012	Unit root	Reject (H_0)	Stationary
Taiwo	-3.6605	4	0.0308	Unit root	Reject (H_0)	Stationary
Fate	-4.7104	4	0.01	Unit root	Reject (H_0)	Stationary

Table 1 above shows the ADF values -4.7048, -4.4855, -3.6605 and -4.1704 with p-values 0.010, 0.012, 0.031 and 0.010 respectively across the locations, which are less than the critical value of 0.05. We therefore reject null hypothesis of having a unit root in favour of alternative of being a stationary series. Indeed, the tests confirmed that the data series is stationary.

3.4. Testing for Heteroscedasticity in the Data. The test of heteroscedasticity for the data across the locations at different time lags were carried out using Breusch-Pagan test. The Breusch-Pagan tests the null hypothesis of no

heteroscedacity against its alternative of there is heteroscedacity. The Table 2 below shows the detail of the analysis.

Table 2: Testing for hetroscedaticity in the data

Location	Test Values (H_0)	DF	P-value	Hypothesis	Decision	Remark
Geri Aimi	9.668	4	0.0018	No Heteroscedacity	Reject H_0	Data is hetroscedastic
Unity	15.72	4	0.0034	No Heteroscedacity	Reject H_0	Data is hetroscedastic
Taiwo	10.831	4	0.0285	No Heteroscedacity	Reject H_0	Data is hetroscedastic
Fate	11.177	4	0.0246	No Heteroscedacity	Reject H_0	Data is hetroscedastic

It was observed that the p-values of the statistic from the four locations of the experimentation were less than the critical value of 0.05 and we therefore reject the null hypothesis of data being homoscedastic in favour of alternative of being heteroscedastic. Indeed, the test confirmed that the data series of the four locations is heteroscedastic.

3.5. Fitting the GARCH (p,q)Models. Different orders of GARCH models were fitted to the stationary data of the generating plant. The parameters, AIC, BIC and HQIC of each order of the model are displayed in Table 3 below. The model with minimum criteria is considered as the best to capture the generating noise rate.

Table 3: Estimated Values of the Parameter and Information Criteria of the Fitted the GARCHs

Place	Model	μ	ω	α_1	β_1	α_2	β_2	AIC	BIC	HQIC
Geri Alimi	(1,1)	-0.2362	-0.3034	-0.2606	0.7324	-	-	0.0096	0.1145	0.052
	(1,2)	9.10E+01	3.06E-02	5.03E-01	1.00E-08	-	1.00E-08	-0.0275	0.1035	0.0255
	(2,1)	9.10E+01	6.02E-03	1.59E-01	7.49E-01	1.00E-08	-	0.0342	0.1653	0.0872
	(2,2)	9.10E+01	3.06E-02	5.03E-01	1.00E-08	1.00E-08	1.00E-08	-0.0073	0.1499	0.0563
Unity	(1,1)	8.9921+01	0.036803	0.111574	0.561008	-	-	0.6534	0.7468	0.6912
	(1,2)	8.99E+01	2.34E-02	4.08E-01	1.00E-08	-	0.4491	0.577	0.6937	0.6244
	(2,1)	8.99E+01	2.34E-02	4.08E-01	1.00E-08	-	4.49E-01	0.6456	0.7624	0.693
	(2,2)	8.99E+01	2.34E-02	4.08E-01	1.00E-08	1.00E-08	4.49E-01	0.5938	0.7339	0.6507
Taiwo	(1,1)	95.0489	0.0544	0.2429	0.2766	-	-	0.6746	0.768	0.7125
	(1,2)	95.0507	0.05701	0.2563	0.1229	-	0.1169	0.6915	0.8083	0.7389
	(2,1)	9.51E+01	5.53E-02	2.42E-01	1.00E-08	2.70E-01	-	0.6926	0.8093	0.74
	(2,2)	9.51E+01	6.41E-02	2.52E-01	3.88E-02	1.00E-08	1.42E-01	0.7079	0.848	0.7649
Fate	(1,1)	85.737	0.04317	0.24162	0.2029	-	-	0.3044	0.3977	0.3422
	(1,2)	8.57E+01	4.47E-02	2.30E-01	1.92E-01	-	1.00E-08	0.324	0.4407	0.3714
	(2,1)	8.58E+01	5.31E-02	2.23E-01	8.96E-02	1.00E-08	-	0.3208	0.4376	0.3682
	(2,2)	9.51E+01	6.41E-02	2.52E-01	3.88E-02	1.00E-08	1.42E-01	0.708	0.8481	0.7648

With regard to the parameters reported in Table 3 above, The estimated coefficient values of all GARCH (p,q) strictly conforms to the bounds of parameter, between -1 and 1. This has made the models to be stationary. Additionally, comparing the GARCH models above in terms of the AIC, BIC and HQIC of (-0.0275, 0.1035 and 0.0255), (0.5770, 0.6937 and 0.6244), (0.6746, 0.7680, and 0.7125) and (0.3044, 0.3977 and 0.3422) respectively, indicates that GARCH(1,2), GARCH(1,1), GARCH(1,1) and GARCH(1,1) are the best for Geri-Alimi, Unity, Taiwo and Fate respectively since their estimated AIC, BIC and HQIC are smaller as compared to other models. Based on the parameter estimates and the criteria, GARCH(1,1) and GARCH(1,2) is chosen as the best model to capture the generating noise from different locations.

3.6. Model Adequacy (Diagnostic) checking of estimated models (Standardized Residuals Tests). Having chosen the GARCH (1,1) and GARCH (1,2) as the best or tentative models as opposed to others, based on the conclusion in Table 3 above, the adequacy of the chosen model is further tested to draw empirical conclusions regarding the model as good fit. These tests carried out are Ljung-Box, normality test of the residuals using Shapiro-Wilk normality test statistic and LM Arch Test. The results are reported in Table 4.

Table 4: Selected Model Diagnostic with Respect to Locations

Model		GARCH(1,1)		GARCH(1,2)	
Place	Test statistic	Values	P-value	Values	P-value
Geri Alimi	Jarque-Bera Test	0.4696	0.79069	1.19568	0.5499
	Shapiro-Wilk Test	0.9921	0.835	0.9874	0.4781
	Ljung-Box Test	12.4541	0.2558	16.3196	0.0908
	LM Arch Test	15.2688	0.22706	15.0771	0.2372
Unity	Jarque-Bera Test	672.9237	0	147.7828	0
	Shapiro-Wilk Test	0.8753	1.45E-08	0.92897	8.99E-06
	Ljung-Box Test	4.0887	0.9433	11.8877	0.2926
	LM Arch Test	3.8847	0.9854	13.3307	0.3455
Taiwo	Jarque-Bera Test	1.1971	0.54961	1.318	0.5174
	Shapiro-Wilk Test	0.9911	0.6509	0.9904	0.5721
	Ljung-Box Test	3.0403	0.9804	3.0778	0.9795
	LM Arch Test	2.8552	0.9964	3.0347	0.9952
Fate	Jarque-Bera Test	1.9554	0.3762	1.972	0.3731
	Shapiro-Wilk Test	0.9869	0.308	0.9866	0.2899
	Ljung-Box Test	6.3663	0.7836	6.423	0.7786
	LM Arch Test	7.3128	0.8362	7.3744	0.8319

Table 4 shows Ljung-Box and Jarque-Bera tests for the noise data with chi-square statistics that give corresponding p-values. The tests are not significant,

therefore, the residuals appeared to be uncorrelated. This indicates that the residuals of the fitted GARCH (1,1) and GARCH (1,2) model are white noise, as such, the model fits the series quite well (the parameters of the model are significantly different from zero), so we can use this model to make forecasts.

Furthermore, the normality is not significant; hence, the Shapiro-Wilk test suggests that the standardized residuals are normal. This also supports the fact that the residuals of the fitted GARCH models are white noise, and the model fits the series quite well (since one of the assumptions of the residual being white noise is normality). Hence, the model is stationary due to the presence of white noise. The p-values of LM Arch Test also indicated that the fitted model capture the heteroscedasticity of the data over time.

3.7. Forecasting for Future Generating Noise Using the Best Fitted model [GARCH(1,1) and GARCH(1,2)]. A 10-step ahead sample forecast was conducted on the data of the generating plant noise and the forecast is visually displayed in Figure 2 and 3. The 10 horizons were forecasted based on data from the preceding time intervals. The forecast was obtained by using data from the previous periods to estimate the future occurrence period of the noise using the GARCH (1,1) and GARCH (1,2) models.

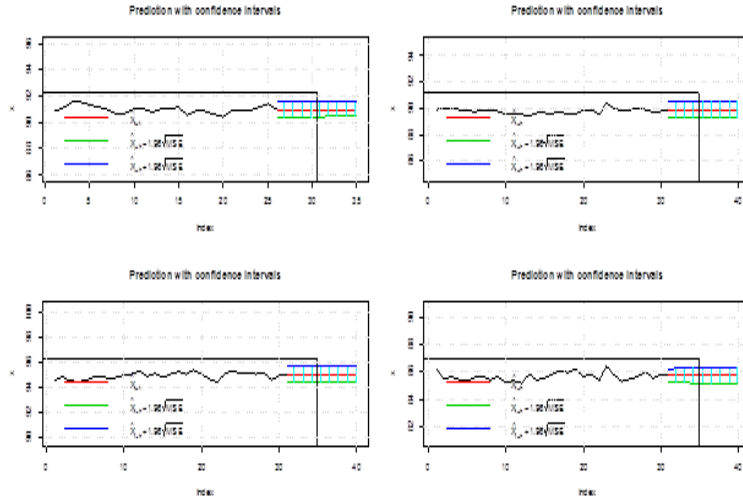


Figure 4: Forecast for Future Occurrence of Noise Generating Plants Across the Locations using GARCH (1,1)

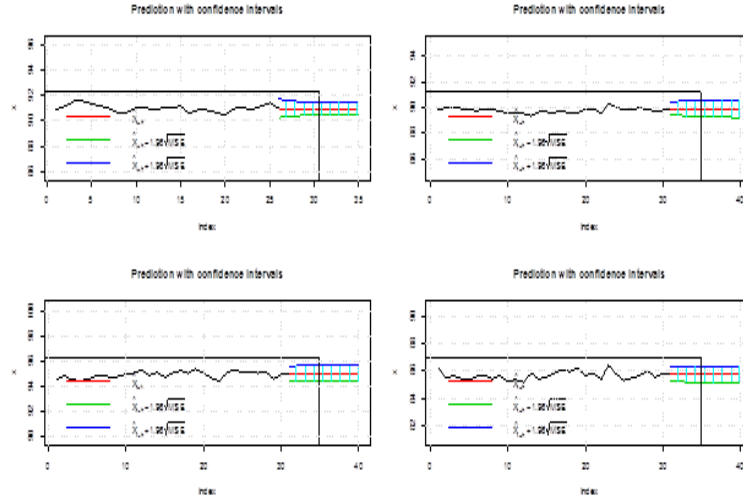


Figure 5: Forecast for Future Occurrence of Noise Generating Plants Across the Locations using GARCH (1,2)

It can be seen from Figure 4 and 5 that the forecast is quite accurate. It seems that GARCH (1,1) and GARCH (1,2) have respectively efficient in capturing the dynamic nature of the data and forecasting. The generating plant noise seems to be stationary over time and their estimates are within the confidence limits.

Conclusions: A class of GARCH models that appropriately describe the generating plant noise over time across four locations in Kwara State of Nigeria are identified. The estimated coefficient values of all GARCH (p,q) strictly conform to the bounds of parameter, between -1 and 1. This has made the models to be stationary. Meanwhile, comparing the GARCH models using the selected criteria, GARCH (1,1) and GARCH(1,2) are obviously preferred to be the best model that captured the generating plant noises from different locations. Importantly, the generating noises are said to be heteroscedastic and stationary over time at considered locations. Finally, the forecasting values seems to be stationary and their estimates are within the confidence limits.

Acknowledgement: The authors are grateful to Obafemi Awolowo University, Federal University Lafia and Federal University of Technology, Minna for the supports they received during the compilation of this work.

Competing interests: The manuscript was read and approved by all the authors. They therefore declare that there is no conflicts of interest.

Funding: The Authors received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

- [1] BABAWUYA A., BORI I., BAKO M. D. S., YUSUF A., JIBRIN A., ELKANAH A. J. AND MOHAMMED A. (2016). Effects of generating plant noise on humans and environment. *International Journal of Engineering Trends and Technology (IJETT)*. **35**(9), 417-422.
- [2] BALI T. G. (2007). Modeling the dynamics of interest rate volatility with skewed fat-tailed distribution. *Annals of Operations Research*. **151**, 151-178.
- [3] BOLLERSLEV T. (1986). Generalised autoregressive conditional heteroscedasticity. *Journal of Econometrics*. **31**, 307-327.
- [4] CLAUDIO G., JOSEPH Q. AND CARMINE T. (2017). A hybrid predictive model for acoustic noise in urban areas based on time series analysis and artificial neural network. *AIP Conference Proceedings*. 1-8.
- [5] CLARK C., CRUMPLER C. AND NOTLEY H. (2020). Evidence for environmental noise effects on health for the United Kingdom policy context: A systematic review of the effects of environmental noise on mental health, wellbeing, quality of life, cancer, dementia, birth, reproductive outcomes and cognition. *Int. J. Environ. Res. Public Health*. **17**, 393-426.
- [6] ENGLE R. F. (1982). Autogressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrics*. **50**, 987-1000.
- [7] GARG N., SHARMA M. K., PARMAR K. S., SONI K., SINGH R. K. AND MAJI S. (2016). Comparison of ARIMA and ANN approaches in time-series predictions of traffic noise. *Noise Control Engr. J.* **64**(4), 522-531.
- [8] GOKCAN S. (2000). Forecasting volatility of emerging stock market: linear versus nonlinear GARCH models. *Journal of Forecasting*. **19**, 499-504.
- [9] HANSEN B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*. **35**, 730-750.
- [10] LIU H. C. AND HUNG J. C. (2010). Forecasting S&P-100 stock index volatility: The role of volatility asymmetry and distributional assumption in GARCH models. *Expert Systems with Applications*. **37**, 4928-4934.
- [11] OLOKOBA S. M., IBRAHIM I. AND ABDULRAHEEM-MUSTAPHA M. A. (2005). Noise pollution: A major catalyst climate change and human health. *Journal of Social Sciences Unilorin*.
- [12] TAYLOR S. (1986). *Modelling financial time series*. Wiley, Chichester.
- [13] YILMAZ H. AND OZER S. (2005). Evaluation and analysis of environmental noise pollution in the city of Erzurum, Turkey. *Int J Environ Pollut*. **23**, 438-448.
- [14] ZHANG X., ZHAO M. AND DONG R. (2020). Time-series prediction of environmental noise for urban IoT based on long short-term memory recurrent neural network. *Appl. Sci.* **10**, 1-18.