Development of Robust MA(q) Model with Asymmetric Error Innovation

O. I. Shittu[1]; F. E. Awosanya[2]

[1,2]Department of Statistics,
University of Ibadan,
Ibadan, Nigeria.
e-mail: oi.shittu@hotmail.com[1]; jesufunmi45@gmail.com[2]

***Abstract* — Moving Average process of order q (MA(q)) is a dynamic time series model useful in modelling economic and financial series, where normality assumption of the error term is taken for granted. However, most real data are skewed and contains outliers which violate this assumption. This work developed the theoretical framework that modified the conventional Moving Average model to assume asymmetric error innovation, capable of characterizing both normal and non-normal time series data. Both simulated and real data is used to validate the proposed model. The proposed model was compared with conventional MA models using model order determination and forecast evaluation tools. The results showed that MA(2) with exponential power error innovation had lower AIC value (-157.76) when compared with the conventional normal error innovation with AIC= 5.73 and MAE value of 0.2031and 0.6898 for MA with exponential power error innovation and conventional normal error innovation respectively. It is established in that the moving average model with exponential power error innovation performed better in terms of modelling and forecast performance that the moving average model with normal error innovation.**

Keywords- *Moving Average model, Outliers, Normal error innovation, and Asymmetric error innovation*

.

I. Introduction

Moving average of order q is a linear process with finite time domain and its shocks which occur in past periods down to lag q are each assigned values. The process does not assign value to the shock that is coming from the present time. The present shock is a random shock that could attain any unpredicted value due to inadequate information presented. Therefore, an expectation of a current shock should be formed based on the information available along with the knowledge of previous shocks as determined by lag *q* in order to specify the model properly.

Moving average process is a time series model that has long found it uses in modelling returns series of economic series in the financial world. Analyst and investment advisors have recommended it as a tool that could either furnish predictive probabilities for security price movements or aid in minimizing losses. Since Moving average is based on past shocks, it smoothed the price data to form a trend, predict price direction. Thus moving average process helps smooth financial returns and filter out the noise [1].

Estimating the parameters of MA model is usually more difficult especially if the zeros are located close to the unit circle, past works have considered ways to solving this problem and four methods reviewed were Durbin's Method (DM) [2], Inverse Covariance (or Correlation) Method (ICM) [3], the Vocariance Recursion Method (VRM) and Vocariance ESPIRIT Method (VEM) [4,5].

The prominence of developing robust techniques for empirical analysis is of importance since the recent global financial crises in 2008 which has placed economic and finance theories under the spotlight [6]. The classical statistical tests follow normality assumption but significant skewness and kurtosis have indicated that real world data are not normally distributed as a result of the presence of outliers in the data set and the selection of a proper model is extremely important as it reflects the underlying structure of the series. Time series analysis is the procedure of proper modelling of a time series [7].

An outlier defined as an observation that is distant from other observations [8]. [9] was the first to have explicitly considered method of analysis of outliers in time series. An outlier can occur by chance in a series but they often indicate skewness in the data set. A data set exhibiting significant skewness or kurtosis has values of data that occur at irregular frequencies, the mean, median and mode will occur at different points. For a data set having significant skewness and kurtosis, Box-Cox transformation can try to normalize it if moderate right skewness is observed by taking log or square root of a data set. Another approach is to use techniques based on error innovations other than the normal. For example, in reliability studies, the exponential power error innovation, weibull and log-normal distribution are typically used as a basis for modelling and the mentioned innovation belong to the family of asymmetry error innovation.

Estimators capable of coping with non-normality in data set are said to be robust, therefore it is of utmost significance to find the relevant and robust statistical measures that can consider asymmetries in any given data sets so as to ensure precise and accurate forecast are obtained.

II. Materials and Methods

Linear models have drawn much attention due to their relative simplicity in understanding and implementation of time series process. The two widely used linear time series models considered are Autoregressive (AR) and Moving Average (MA) models [10]. The mixture of these two models gives many other models namely Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) [10, 11]. Autoregressive Fractionally Integrated Moving Average (ARFIMA) model was used to generalize ARMA and ARIMA models [12].

The moving average model used for this research was adapted from [10]. It was represented by the systematic component that is generated as a weighted average of random disturbances of past periods.

Let X_t be a response variable assuming $MA_{(q)}$ process defined as

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

where ε_t i.i.d. $\sim N(0, \sigma^2)$

Where $\theta_1, \theta_2, \ldots\ldots, \theta_q$ are the parameters of the model which can be either positive or negative.

ε_t is a white noise process with $E(X_t) = 0$ and $Var(Y_t) = \sigma^2 \Sigma \theta_i^2$, $\theta_0 = 1$

and $1 + \theta_1^2 + \ldots + \theta_q^2 < \infty$. It is always invertible for all values of $\theta_1, \theta_2, \ldots, \theta_q$. Thus, MA model only require for invertibility condition not stationarity. N this study, we consider the Moving Average of order two (MA(2)).

Parameter Estimation of Moving Average Model with Normal Error Innovation

Moving Average of order two (MA_2) is given as

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \quad (1)$$

$$\varepsilon_t = X_t - \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

$$\text{where } \varepsilon_t \approx N(0, \sigma^2)$$

Normal error innovation has its probability density as

$$f(\varepsilon_t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X_t - \mu)^2} \quad (2)$$

where $X_t - \mu = \varepsilon_t$

$$f(\varepsilon_t) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(X_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})^2} \quad (3)$$

Taking the likelihood function of (3) we have;

$$L = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\Sigma(X_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})^2} \quad (4)$$

Taking the log-likelihood function of (4), differentiating with respect to $\theta_1, \theta_2, \sigma^2$ and equating each to zero to obtain:

$$\frac{dl}{d\theta_1} = -2(\varepsilon_{t-1})\left(-\frac{1}{2\sigma^2}\right)\Sigma(X_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}) = 0, \quad \text{gives}$$

$$\hat{\theta}_1 = \frac{\Sigma X_t \varepsilon_{t-1}}{\sigma^2} \quad (5)$$

Parameter θ_2 gives

$$\frac{dl}{d\theta_2} = -2(\varepsilon_{t-2})\left(-\frac{1}{2\sigma^2}\right)\Sigma(X_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}) = 0, \quad \text{gives}$$

$$\hat{\theta}_2 = \frac{\Sigma X_t \varepsilon_{t-2}}{\sigma^2} \quad (6)$$

To obtain the parameter σ^2 we have

$$\frac{dl}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\Sigma(X_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})^2 = 0$$

$$\hat{\sigma}^2 = \frac{\Sigma(X_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})^2}{n} \quad (7)$$

Parameter Estimation of Moving Average Model with Exponential Power Error Innovation

The most prominent and widely used family of skewed probability distribution is the exponential power distribution. The distribution that was first discussed by [13] as well as the Bayes inference have been used in signal processing field and in image processing [14].

The probability distribution function of exponential power error innovation is given as

$$f(X_t, \mu, \sigma, \beta) = \frac{\beta}{2\sigma\Gamma 1/\beta} e^{-[\frac{x-\mu}{\sigma}]^\beta}, \quad -\infty<x<\infty,\ x>0 \quad (8)$$

This family of distribution allows for tails that are either heavier than normal (when β <2) or lighter than normal (when β >2), where β is a shape parameter. The parameter estimate is done by method of maximum likelihood and the method of moments. The estimates do not have a closed form and must be obtained numerically.

If X_t follow moving average model of order 2, then

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

We will use exponential power error innovation because moving average model is usually white noise, we then have:

$$f(\varepsilon_t) = \frac{\beta}{2\sigma\Gamma 1/\beta} e^{-[\frac{x_t-\mu}{\sigma}]^\beta} \quad (9)$$

By substituting the mean equation, we have

$$f(\varepsilon_t) = \frac{\beta}{2\sigma\Gamma 1/\beta} e^{-[\frac{x_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}}{\sigma}]^\beta} \quad (10)$$

The log-likelihood gives

$$l = n\ln\beta - n\ln(2\sigma\Gamma 1/\beta) - \Sigma[\frac{x_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}}{\sigma}]^\beta \quad (11)$$

Differentiating equation (11) with respect to $\theta_1, \theta_2, \sigma, \beta$ and equating to zero

to have

$$\hat{\theta}_1 = \beta \frac{\sum x_t \varepsilon_{t-1}}{\sigma^2} \quad (12)$$

$$\hat{\theta}_2 = \beta \frac{\sum x_t \varepsilon_{t-2}}{\sigma^2} \quad (13)$$

$$\hat{\sigma}^2 = \frac{\Sigma[x_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}]^\beta}{n} \quad (14)$$

$$\frac{dl}{d\beta} = \frac{n}{\beta} - \frac{ne^{-ax^\beta}}{\beta\sqrt{\pi}} - \Sigma[\frac{x_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}}{\sigma}]^\beta \ln[\frac{x_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}}{\sigma}] = 0 \quad (15)$$

$$\frac{n}{\beta} - \frac{ne^{-ax^\beta}}{\beta\sqrt{\pi}} - \Sigma[\frac{x_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}}{\sigma}]^\beta \ln[\frac{x_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}}{\sigma}] = 0 \quad (16)$$

There is no closed form solution to (16) above therefore, it was solved numerically.

III. Results and Discussion

Data were simulated for samples of sizes 50, 100, 200, 500 and 1000 and Nigeria Stock Exchange data from January, 1996 to May, 2017 were used to validate this model while R-package was used to analyse the data. Two sets of data were simulated, one set was simulated with random error innovation while the other set was contaminated with outliers randomly.

Simulated data (one outlier)

The descriptive statistics of the uncontaminated and contaminated data sets are reported in Table 1, while the order determinant and forecast performance of the MA model are given in Table 2 after modelling the simulated series. The results revealed that uncontaminated data sets are normally distributed but the contaminated data sets have positive skewness and tails and are therefore skewed. Also, AIC of the MA(2) with exponential power error innovation are lower when compared to MA(2) with normal error innovation. Fig. 1 presents the plots of simulated data for uncontaminated and contaminated data set of size 50. The plots confirm that uncontaminated data are normally distributed while the contaminated data is not.

Table 1: Descriptive statistics for uncontaminated and contaminated data (one outlier)

Sample Size	Uncontaminated Data Sets		Contaminated Data Sets	
	Skewness	**Kurtosis**	**Skewness**	**Kurtosis**
N = 50	-0.1085	-0.4577	2.8862	13.76
N = 100	-0.0542	-0.5644	2.5490	15.17
N = 200	0.0356	-0.5127	1.7966	11.64
N = 500	0.1396	-0.0559	1.1562	8.04
N = 1000	-0.0151	0.1070	0.4680	3.75

Table 2: Model Order Determinant and Forecast Performance of MA(2) model (with one outlier)

Sample Size		Uncontaminated Data Sets			Contaminated Data Sets		
		σ^2	AIC	MAE	σ^2	AIC	MAE
N = 50	NEI	1.019	159.17	0.5712	2.637	212.74	0.9893
	EPEI	3.18 e-04	159.17	0.5712	0.2991	32.20	0.1242
N = 100	NEI	1.067	361.08	0.5063	1.944	447.66	0.7619
	EPEI	0.0009	361.08	0.5063	0.0113	192.32	0.2014
N = 200	NEI	1.087	850.16	0.4842	1.629	846.93	0.4537
	EPEI	0.0083	850.16	0.4842	0.0063	500.11	0.2002
N = 500	NEI	0.8796	2661.42	0.5094	1.057	2703.72	0.5314
	EPEI	0.0787	2661.42	0.5094	0.1075	2388.73	0.3787

Note: NEI stands for Normal Error Innovation and EPEI stands for Exponential Power Error Innovation

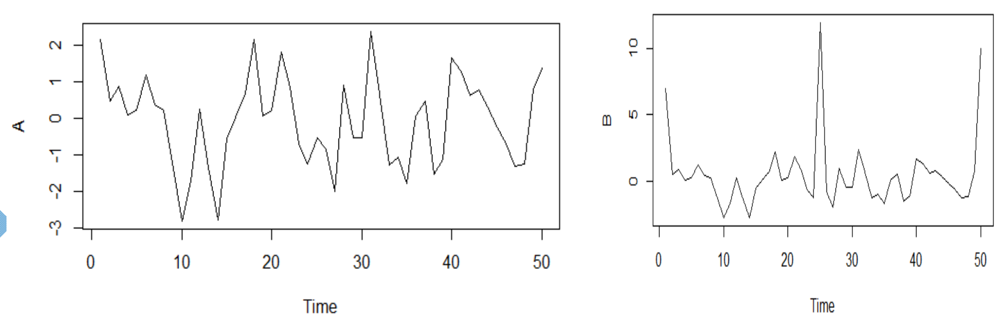

Figure1. Time plots for one outlier when N = 50: a) uncontaminated data set, b) for contaminated data set

Simulated data (two outliers)

The result shown in Table 3 summarizes the statistics after modeling. It revealed that contaminated data sets are highly skewed with high kurtosis. The skewness was reducing as the sample size increases. Table 4 gave lower values of AIC and MAE for contaminated series with exponential error innovation over normal error innovation while both error innovations gave same value of AIC and MAE for uncontaminated for sample sizes 50, 100, 200 and 500 considered. Fig. 2 presents the plots of simulated data for uncontaminated and simulated data set of size 200. The plots confirm that uncontaminated data are normally distributed but contaminated data contain outliers

Table 3: Descriptive statistics for uncontaminated and contaminated data (two outliers)

Sample Size	Uncontaminated Data Sets			Contaminated Data Sets		
	Skewness	**Kurtosis**	**P-Value**	**Skewness**	**Kurtosis**	**P-Value**
N = 50	-0.1085	-0.4577	0.6788	2.6454	9.6673	7.19e-08
N = 100	-0.0543	-0.5644	0.5189	2.5907	12.6299	2.55e-10
N = 200	0.0356	-0.5127	0.4852	2.0076	11.2459	8.57e-12
N = 500	0.1396	-0.0559	0.2776	1.3540	8.5424	2.25e-14
N = 1000	-0.0151	0.1070	0.4053	0.5968	4.2428	4.21e-13

Table 4: Model Order Determinant and Forecast Performance of MA(2) model (two outliers)

Sample Size		Uncontaminated Data Sets			Contaminated Data Sets		
		σ^2	AIC	MAE	σ^2	AIC	MAE
N = 50	NEI	1.019	159.17	0.5712	3.469	198.70	0.8459
	EPEI	0.0058	159.17	0.5712	0.0013	38.51	0.1350
N = 100	NEI	1.067	361.08	0.5063	2.378	402.36	0.5991
	EPEI	0.0128	361.08	0.5063	0.0117	116.56	0.1269
N = 200	NEI	1.087	850.16	0.4842	1.852	877.81	0.4840
	EPEI	0.1000	850.16	0.4842	0.0059	628.72	0.2506
N = 500	NEI	0.8796	2661.42	0.5094	1.138	2717.44	0.5373
	EPEI	0.1131	2661.42	0.5094	0.1000	1137.87	0.1062

Note: NEI stands for Normal Error Innovation and EPEI stands for Exponential Power Error Innovation

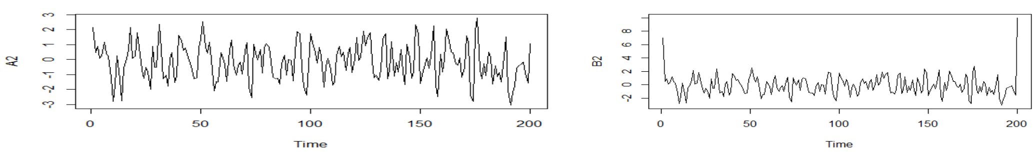

Figure2. Time plots for two outliers when N = 200: a) uncontaminated data set, b) for contaminated data set

Real Data

Table 5 summarizes the descriptive statistics of the Nigerian stock exchange data. It can be seen that the data is highly skewed with heavy tail and leptokurtic. Hence the data is not normally distributed. Table 6 presents the result after modelling the data. The AIC for MA(2) with exponential power error innovation is lower than MA(2) with normal error innovation and this supports the notion that most real life data are not normally distributed. Fig. 3 presents the plots of stock exchange data of size 257. The plots confirm that real life data are contaminated.

Table 5: Descriptive statistics for stock exchange data

Sample Size	**257**
p-value	<2.2e-16
Skewness	3.7891
Kurtosis	23.3558

Table 6: Criteria result of normal and non-normal error innovation for stock exchange data when N = 257

	Normal Error Innovation	Exponential Power Error Innovation
σ^2	1177.79	0.0458
AIC	1519.32	893.34
MAE	0.6898	0.2031

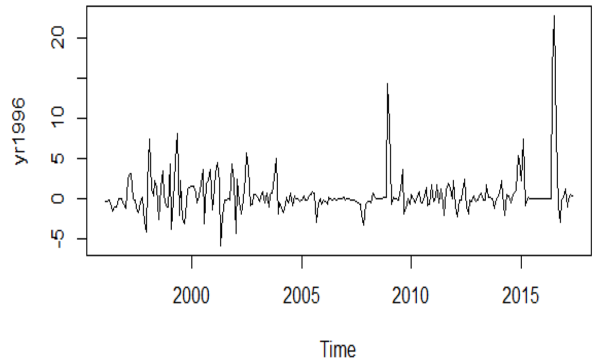

Figure 3: plot of stock exchange data when N = 257

IV. Conclusion

The Moving average model of order 2 with non-normal error innovation performed better in terms of lower model performance tools (AIC) than that with normal error innovation for both simulated and real data sets. Also, the forecast performance is better with power mean absolute error of forecast. It has been established in this study that whether the series is normally distributed or asymmetric, the moving average model with exponential power error innovation performed better in terms of modelling and forecast performance that the moving average model with normal error innovation.

Acknowledgment

The authors are grateful to anonymous reviewers for their valuable comments on the original draft of this manuscript

References

[1] James, F.E. (1968). Monthly Moving Averages- An effective Investment Tool. *Journal of Financial and Quantitative Analysis*, Volume 3, Issue 3, Special Issue: Random Walk Hypothesis, 315-326

[2] Durbin, J. (1959). Efficient estimation of parameters in Moving Average Models. Biometrika, Vol. 46, pp. 306-316

[3] Chatfield, C. (19979). Inverse autocorrelations. J.R. Statist. Soc. A, Vol. 142, pp. 363-377

[4] Kaderli, A. and Kayhan, A.S. (2000). Spectral estimation of ARMA processes using ARMA-cepstrum recursion. IEEE Sign. Proc. Lett., Vol.7, pp. 259 – 261

[5] Kay, S.M., Jackson, L.B., Huang, J. and Djuric, P.M. (1988). An approximate maximum likelihood ARMA estimator based on the power cepstrum. In Proc. Acoust. Speech Sign. Proc. (ICASSP), pp. 2344 – 2347

[6] Brown, S. and Wong, W.K. (2015). Probability and Statistics with Applications in Finance and Economy. *The Scientific World Journal*, vol. 2015 (2015), Article ID 618785, http://dx.doi.org/10.1155/2015/618785

[7] Hipel, K. W. and McLeod, A. I., (1994). Time Series Modelling of Water Resources and Environmental Systems. Elsevier, Amsterdam, http://www.stats.uwo.ca/faculty/aim/1994Book/.

[8] Alkutubi, H.S. and Ali, H.M. (2011). Maximum likelihood estimators with complete and censored data. Eur. J. Sci. Res., 54(3): 407-410.

[9] Fox, J.A. (1972). Outliers in time series. J. Royal Stat. Soc., Series B, 34: 350-363.

[10] Box, G.E.P. and Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day.

[11] Cochrane, J. (1997). Where is the market going? Uncertain facts and novel theories. Economic Perspectives, Issue Nov., Vol. 21. No. 6, pp. 3-37

[12] Galbraith, J.W. and Zinde-Walsh, V. (2001). Autoregression based Estimators for AFRIMA Models. Serie Scientifique, CIRANO; 36 pages

[13] Box, G.E.P. and Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis, Addision-Wesley, Reading, MA.

[14] Gharavi, H. and Tabatabai, A. (1988). Sub-band coding of monochrome and color images, IEEE Transactions on Circuits System, 35, 207-214