PARAMETER ESTIMATION OF LOG-ARIMAX MODEL VIA GENERALIZED LINEAR METHOD OF EXPONENTIAL FORM

By

K. I. EKERIKEVWE* and T.O. OLATAYO**

*Department of Statistics, Delta State Polytechnic, Otefe-Oghara, Delta State, Nigeria

**Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye

Email: kennedyekerikevwe@gmail.com.

ABSTRACT

To improve forecast accuracy, an autoregressive integrated moving average model, ARIMAX (p, d, q, b), was developed for short-memory observational time series data with exogenous covariate(s) .However, for long-memory frequency observations, a modification will be necessary to neutralize the model for a better and improved prediction of the system. This study, therefore, is designed to propose and formulate a logarithmic autoregressive integrated moving average (LOG-ARIMAX)modelwhose distributional form would be robust and sufficient in capturing and accommodating both the external covariate (s) and the heavy-tailed properties of long-memory frequencyobservational time series data. The parameter estimation of the LOG-ARIMAX model will be carried out via Generalized Linear Method (GLM)of exponential form.The comparison of the model performance indexes will bedone with the traditional ARIMAX model under in-sample forecasts conditions.

Keyword: ARIMAX, Time Series, Accuracy, Frequency, Model, Estimation

1.0 INTRODUCTION

Statistical methods and models are either linear or non-linear based on some assumptions theoretically and analytically. These assumptions led to the splitting of approach of dealing with time varying observations (time series) models into two approaches; Time domain (otherwise known as probabilistic approach) and frequency domain (spectral function) analyzes. The linear probabilistic models range from Autoregressive (AR), Moving Average (MA), Double Moving Average (DMA),Markov-Autoregressive, Autoregressive Moving Average(ARMA), Vector Autoregressive (VAR), Vector Autoregressive Moving Average (VARMA), and other purely random processes (white processes) etc. However, these mentioned linear probabilistic models are applied only in univariate series and do not depict turning points, volatility and cycle traits in time events. These led to the propounded of some non-linear models, like Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models by Eagle (1982) and Bollerslev (1986) for time-varying volatility of asset series returns or otherwise regard as non-linear models for variance random processes. The models are affected by the ignorance of only taken cognizance of constrained positive values (not fully a reflection of fluctuation (risk) in two sides' scenarios) (Cepeda*et al.*, 2014). In addition, some other fluctuation models like Threshold Autoregressive (TAR), Self-Exciting Threshold

Autoregressive (SETAR), Smooth Transition Autoregressive (STAR), other GARCH variants' models (EGARCH, APARCH, etc.), Simple Exponential Smoothing (SES), Double Exponential Smoothing (DES), Triple Exponential Smoothing (TES), and the Bayesian SETAR etc. failed to incorporate external exogenous variable(s).

This study aimed to specify a model for a long-memory frequency Autoregressive Integrated Moving Average (ARIMAX) model that is coupled with an external time-varying covariate(s) with heavy tailed distributional lognormal form of a residual structure. This research is significant as it would lead to a full specification of log-ARIMAX model with standardized lognormal distribution as its residual structure. The lognormal is expected to capture and accommodate both the external covariate (s) and the heavy-tailed properties of long-memory (high frequency) observational time series data to achieve better and improved prediction. Also, this study will be useful to oil companies; climate change and environmental researchers who have interest in employing time series and Log-ARIMAX approach in modeling high or long-memory series for efficient, sufficient and reliable precision.

Some of the literatures reviewed considered ARIMAX model for short-memory frequency data. In view of this, this study, therefore, proposes a hybrid ARIMAX model to capture and accommodate both the external covariate(s) and the heavy-tailed properties of observational time series events using secondary datasets of the long memory types of oil spillage and temperatures. A hybridization of Logarithm and ARIMAX would be propounded to model time series data with heavy-tailed trait.

2.0 THE LITERATURE

Silvestrini and Veredas (2008),Groger*et etal* (2012), and Wadud(2014) highlighted the usefulness and essence of propounded Autoregressive Moving Average Exogenous Variables (ARMAX (p, q, b) or Autoregressive Integrated Moving Average Exogenous Variables (ARIMAX (p, d, q, b) models by explicitlyrelating it to a regression model with a univariate lagged dependent variable uniform time-varying observations resting on the shoulder of the exogenous independent lagged timely variables. ARIMAX (p, d, q, b) works more fine and perfect because of its integrating part or otherwise regarded as differencing to ascertain stationarity (constant rate of slope change) and constant trend (constant mean). However, it works perfectly well for short-memory frequencies data (like interest rate, number of stock sold on a daily/weekly/monthly/yearly basis, indexes, changes in monthly prices of commodity etc.) and gives near perfect forecast of systems (Chi &Baek 2012; Doktoringenieurs, 2010; Lee, *et al.*, 2010).Long-memory frequencies' observations,like the climate-measured daily/weekly/average monthly temperatures recorded, changes in daily recorded climate; currencies exchange rates; Consumer Price Index (CPI) and GDP,a modification will be necessarily needed to neutralize or log-linearize the ARIMAX (p, d, q, b) model for the betterment and improvement of the (constant rate of slope change), constant trend (constant mean), transfer function, residual process and prediction of the system. Having ascertained the power of logarithm in the process of differencing or transformation to help stabilize, eliminate (or reducing) trends, mean of the time series and seasonality signal if any when characterized with high or long memory series. So, a log function would be added to the ARIMAX(p, d, q, b) to make it Log-ARIMAX (p, d, q, b)to neutralized the threat posed by long-memory traits that might likely affect not only the parameterization (over-parameterization or under-parameterization) and the end product of given a reliable system forecasts. Forecast generally usually emanated from the generalization of the residual processes attached to the model. However, Log-ARIMAX (p, d, q, b) would not be an exception but additional components of in-sample and out-sample forecast would be incorporated and tested via forecast indexes via AR, MA and the exogenous residual processes.The Log-ARIMAX abstraction of reality would not only make room for merging linear/non-linear regression with ARMA model for better broadening of the applicability of non-linear time series models but also going to serve as a platform of introducing Generalized Non-linear/linear time series for transfer function (otherwise called mean function and impulse weight function in Generalized Linear Model (GLM). This technique of solving residual of different residual structures of different distributional forms and different variable types will be employed to treat the white noise of the Log-ARIMAX model. Additionally, appropriate formulations for the autocorrelation structure (Partial Autocorrelation Function (PACF)) of the error term from the regression equation function (transfer function) of the long memory (highly frequency data) would be sufficiently identify and compare to ARMA and ARMAX models. Another meritorious trait of both the Log-ARIMAX and ARIMAX models would be the attachment of degree of fit (contributions of the exogenous variables) as measured by the coefficient of R-squared and its variants to fitted models; and capturing of the dynamics of seasonal variation change patterns over time. The long memory (high frequency) associated to economic, environmental, climate change, wave data (sea and ocean wavy pattern record) etc. will be a typical examples of most long memory data due to their fluctuations, higher values, dependency, and switching circular traits. However, economic index like inflation, which is persistent and appreciable rise in the general level of prices. In other words, general price level might be response or predictive variable (Frimpong & Oteng-Abayie, 2010). Another example of long memory and conditionally covariates series exchange rate (which is the value of the domestic currency in terms of foreign currency). Exchange rate changes can affect the relative prices, thereby the competitiveness of domestic and foreign producers. Theoretically, exchange rate will have a negative or positive relationship with economic growth. This is because currency depreciation will foster a country export that will lead to an increase in Gross Domestic Product (GDP) while currency depreciation will also discourage a country import, thus leading to decrease in GDP of that country. It counts out that appreciation of exchange rate exerts positive influence on GDP and real economic growth (Aliyu, 2011). Therefore, exchange rate and GDP might be dependency or predictor for an ARIMAX or ARIMA model depending on their context. Among other congenital examples of realization that best fit the Log-ARIMAX conceptualization are environmental and climate changes of oil spillage and temperature.

According to Alves *et al.* (2014) and Alves *et al.* (2015), continuous demand of energy by human has led to rapid development and increment of exploitation of crude oil and gas in oil source regions. This has led to high-risk level of oil spill accidents to its source environs (e.g. seas, rivers, streams, farms, poultry etc.). Frequent occurrence of oil spill pollutions has been a as result of rapid and infrastructural development, ship grounding accidents, collisions maritime transportation and capsized tankers. The movement of oil spillage in water (spreading and drift of spilled oil on the sea) and land (evaporation and adsorption of spilled oil by farms, ponds, poultry etc. are usually influenced by dynamic factors, non-dynamic factors, and variable oil properties (These influencer/ factors are regarded as the covariates, supportive, explanatory variables needed by ARIMAX.) (Dietrich *et al*, 2012; Alves *et al*, 2016). Similarly, water level of seas, oceans, streams, riverine areas, evaporation level affects, influence and dictate the temperature (in degree Celsius) level to be measured in particular region. These influencers can be directly or indirectly influence or affected plants' growth, level of raw materials from land (like charcoal, crude oil, limestone etc.) and influence harvest time via temperature (Ragulina and Reitan, 2017). Like that of the oil spillage factors, these influencers or factors played a major role in recorded temperature.
This study aimed at conceptualizing the long memory associated to oil spillage and temperature (environmental and climate) influencers or explanatory respectively via log-ARIMAX model. The burning issues of environmental and climate factors in the world and in Nigeria led to diverting and channeling modification and fine tune ARIMA to the mentioned datasets. No doubt, some external, internal, randomness (natural influence or human influence), and other climate factors would have been affecting or changing oil spillage and temperature for a direct or indirect effect on environmental and climate factors respectively.

3.0 MATERIALS AND METHODS

For long-memory (highly frequency) observational series, ARMAX or ARIMAX might unable to dissolve the characterized long-memory in both the exogenous covariates and series or in any of the two observational.
In category of such long-memory data are high valued financial returns, climate indexes recorded data, sea wave measurements, evolutional data, inflation, GDP etc. However, the stochastic disturbance would be less in power to dissolve high constant flexibility.
An ideal distributional form of the stochastic disturbance (ε_t)via the lognormal variate shall be introduced.
The distributional form of $(\varepsilon_t)$is then given as

$$f(y_t) = \frac{1}{y_t\sigma\sqrt{2\pi}} exp\left[-\left(\frac{(ln(y_t))^2}{2\sigma^2}\right)\right] y_t > 0 \tag{3.1}$$

or

$$f(\varepsilon_t) = \frac{1}{\varepsilon_t\sigma\sqrt{2\pi}} exp\left[-\left(\frac{(ln(\varepsilon_t))^2}{2\sigma^2}\right)\right] \varepsilon_t > 0 \tag{3.2}$$

Because, the error term and the observational series share the same distributional form

$$\text{With } y_t \sim \varepsilon_t \sim N\left[exp\left(\frac{\sigma^2}{2}\right), exp(\,2\sigma^2) - exp(\,\sigma^2)\right] \tag{3.3}$$

$$Y_t = \sum_{j=1}^{n} \frac{\eta_h(B)B^b}{\lambda_r(B)} X_t + \frac{\theta_q(B)\varepsilon_t}{\varphi_p(B)} \sim N\left[exp\left(\frac{\sigma^2}{2}\right), exp(2\sigma^2) - exp(\sigma^2)\right]$$

(3.4)

For log-ARMAX

$$Y_t = \sum_{j=1}^{n} \frac{\eta_h(B)B^b}{\lambda_r(B)} X_t + \frac{\theta_q(B)\varepsilon_t}{\nabla^d \varphi_p(B)} \sim N\left[exp\left(\frac{\sigma^2}{2}\right), exp(2\sigma^2) - exp(\sigma^2)\right]$$

(3.5)

where,

$x_t \Rightarrow x_t$ is the independent exogenous input series at time "t"

$\frac{\eta_h(B)B^b}{\lambda_r(B)}$is the transfer function of the input series.

The Probability Density Function (PDF) of lognormal distribution is given by

$$f(y) = \frac{1}{y\sigma\sqrt{2\pi}} exp\left(-\frac{(\log(y)-\mu)^2}{2\sigma^2}\right) \; y\epsilon(0,\infty)$$

(3.6

The lognormal is said to belong to the exponential family if its (PDF) can be express as

$$f(y;\mu,\sigma^2) = exp\left[\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i,\phi)\right] \qquad (3.27)$$

Where, the $\theta_i$is the location parameter, e.g., mean, ϕ scale parameter, e.g., variance

$$a_i(\phi) = \frac{\phi}{P_i} \Longrightarrow a_i(\phi) \approx 1$$

$$E(Y_i) = b^T(\theta_i)$$

$$Var(Y_i) = b^{TT}(\theta_i)a_i(\phi)$$

$$f(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2}(\log(y) - \mu)^2\right) \quad (3.28)$$

Taking logarithms,

$$\log f(y) = log\left[\frac{1}{y\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2}(\log(y) - \mu)^2\right)\right]$$

$$= \log\left[\frac{1}{y}(2\pi\sigma^2)^{-1/2} exp\left(-\frac{1}{2\sigma^2}(\log(y) - \mu)^2\right)\right] \quad (3.29)$$

$$= -\frac{1}{2}\log(2\pi\sigma^2) - \log y - \frac{1}{2\sigma^2}[\log(y)^2 - 2\mu logy + \mu^2] \quad (3.30)$$

$$= -\frac{1}{2}\log(2\pi\sigma^2) - \log y - \frac{log(y)^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\log(y) - \frac{\mu^2}{2\sigma^2} \quad (3.31)$$

Taking exponential such that it neutralizes the logarithm to be able to go back to the distributional form

$$f(y) = exp\left[-\frac{1}{2}\log(2\pi\sigma^2) - \log y - \frac{log(y)^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\log(y) - \frac{\mu^2}{2\sigma^2}\right] \quad (3.33)$$

$\theta_{\mu_i} = \frac{\mu}{\sigma^2} b_\mu(\theta_i) = \frac{\mu^2}{2\sigma^2}$ $\quad \theta_0 = -\frac{1}{2\sigma^2}$ $\quad b_\sigma(\theta_i) = \frac{\log(\sigma^2)}{2} c(y_i, \theta) = -\frac{1}{2}\log(2\pi) - \log(y)$

This show that the log-normal distribution belongs to the exponential family due to explicit form in a canonical form.

From (3.16), $\varphi(B)x_t = \frac{\eta_h(B)B^b}{\lambda_r(B)} X_t$

$$Y_i = \sum_{j=1}^{n} \frac{\eta_h(B)B^b}{\lambda_r(B)} X_t + \frac{\theta_q(B)}{\varphi_p(B)}\varepsilon_t \quad (3.34)$$

$$= \sum_{j=1}^{n}(\varphi_i B^i)B^b + \frac{\theta_q(B)}{\nabla^d \varphi_p(B)}\varepsilon_t \quad (3.35)$$

Inserting equation (3.35) into equation (3.33)

$$f(y) = exp\left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{\log\left[\sum_{j=1}^{n}(\varphi_i B^i)B^b + \frac{\theta_q(B)}{\nabla^d\varphi_p(B)}\varepsilon_t\right]^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\log\left[\sum_{j=1}^{n}(\varphi_i B^i)B^b + \frac{\theta_q(B)}{\nabla^d\varphi_p(B)}\varepsilon_t\right] - \frac{\mu^2}{2\sigma^2}\log\left[\sum_{j=1}^{n}(\varphi_i B^i)B^b + \frac{\theta_q(B)}{\nabla^d\varphi_p(B)}\varepsilon_t\right]\right] \quad (3.36)$$

The log-likelihood function is

$$\ln L(y) = \sum_{i=1}^{n}\left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{\log\left[\sum_{j=1}^{n}(\varphi_i B^i)B^b + \frac{\theta_q(B)}{\nabla^d\varphi_p(B)}\varepsilon_t\right]^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\log\left[\sum_{j=1}^{n}(\varphi_i B^i)B^b + \frac{\theta_q(B)}{\nabla^d\varphi_p(B)}\varepsilon_t\right] - \frac{\mu^2}{2\sigma^2}\log\left[\sum_{j=1}^{n}(\varphi_i B^i)B^b + \frac{\theta_q(B)}{\nabla^d\varphi_p(B)}\varepsilon_t\right]\right] \quad (3.37)$$

For simplicity, working at AR=2, MA=2 at B_p of p-coefficients and X_p-covariates, that is, Log-ARIMAX $(2, \nabla^d, 2)$

$$\ln L(y) = \sum_{i=1}^{n}\left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{\log\left[\sum_{j=1}^{n}(\varphi_i B^i)B^b + \frac{\theta_q(B)}{\nabla^d\varphi_p(B)}\varepsilon_t\right]^2}{2\sigma^2} - \frac{\mu}{\sigma^2}\log\left[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2}) + \frac{\theta_1 y_{t-1} + \theta_2 y_{t-2}}{\nabla^d(\theta_1 y_{t-1} + \theta_2 y_{t-2})}\varepsilon_t\right] - \log\left[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2}) + \frac{\theta_1 y_{t-1} + \theta_2 y_{t-2}}{\nabla^d(\theta_1 y_{t-1} + \theta_2 y_{t-2})}\varepsilon_t\right]\frac{\mu^2}{2\sigma^2}\right] \quad (3.38)$$

$$= \sum_{i=1}^{n}\left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{2}{\sigma^2}\log\left[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2}) + \frac{\theta_1 y_{t-1} + \theta_2 y_{t-2}}{\nabla^d(\theta_1 y_{t-1} + \theta_2 y_{t-2})}\varepsilon_t\right] - \frac{\mu^2}{2\sigma^2} - \frac{\mu}{\sigma^2}\log\left[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2}) + \frac{\theta_1 y_{t-1} + \theta_2 y_{t-2}}{\nabla^d(\theta_1 y_{t-1} + \theta_2 y_{t-2})}\varepsilon_t\right]\right]\sum_{i=1}^{n}\left[-\frac{1}{2}\log(2\pi\sigma^2) - \left(\frac{2}{\sigma^2} + \frac{\mu}{2}\right)\log\left[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2}) + \frac{\theta_1 y_{t-1} + \theta_2 y_{t-2}}{\nabla^d(\theta_1 y_{t-1} + \theta_2 y_{t-2})}\varepsilon_t\right]^2 - \frac{\mu^2}{2\sigma^2}\right] \quad (3.39)$$

$$= \sum_{i=1}^{n}\left[-\frac{1}{2}\log(2\pi\sigma^2) - \left(\frac{4+\mu\sigma^2}{2\sigma^2}\right)\log\left[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2}) + \frac{\theta_1 y_{t-1} + \theta_2 y_{t-2}}{\nabla^d(\theta_1 y_{t-1} + \theta_2 y_{t-2})}\varepsilon_t\right]^2 - \frac{\mu^2}{2\sigma^2}\right] \quad (3.40)$$

$$= \sum_{i=1}^{n}\left[-\frac{1}{2}\log(2\pi\sigma^2) - \left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\log\left[\left[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2}) + \frac{\theta_1 y_{t-1} + \theta_2 y_{t-2}}{\nabla^d(\theta_1 y_{t-1} + \theta_2 y_{t-2})}\varepsilon_t\right]\right] - \frac{\mu^2}{2\sigma^2}\right] \quad (3.41)$$

$$\frac{\delta\ln(L)}{\delta\theta_1} = \sum_{i=1}^{n}\left[-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})] \times \frac{y_{t-1}}{\nabla^d\varphi_1 y_{t-1} + \varphi_2 y_{t-2}} \times \frac{\nabla^d\varphi_1 y_{t-1} + \varphi_2 y_{t-2}}{\theta_1 y_{t-1} + \theta_2 y_{t-2}}\right] \quad (3.42)$$

$$= \sum_{i=1}^{n}\left[-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})] \times \frac{y_{t-1}}{\theta_1 y_{t-1}+\theta_2 y_{t-2}}\right] \quad (3.43)$$

$$= \sum_{i=1}^{n}\left[-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})]y_{t-1}(\theta_1 y_{t-1} + \theta_2 y_{t-2})^{-1}\right] \quad (3.44)$$

But $(\theta_1 y_{t-1} + \theta_2 y_{t-2})^{-1}$ can be expanded via Negative Binomial expansion

$$(\theta_1 y_{t-1} + \theta_2 y_{t-2})^{-1} = (\theta_1 y_{t-1})^{-1}[(-1)(\theta_2 y_{t-1})^{-2}\theta_2 y_{t-1}]\left[\frac{(-1)\times(-2)}{1\times 2}(\theta_1 y_{t-1})^{-3}(\theta_2 y_{t-2})\right] \times \left[\frac{(-1)\times(-2)(-3)}{1\times 2\times 3}(\theta_1 y_{t-1})^{-4}(\theta_2 y_{t-2})^{3}\right] \quad (3.45)$$

$$= \frac{1}{\theta_1 y_{t-1}} - \frac{\theta_2 y_{t-2}}{(\theta_1 y_{t-1})^2} + \frac{(\theta_2 y_{t-2})^2}{(\theta_1 y_{t-2})^3} \quad (3.46)$$

$$= \frac{(\theta_1 y_{t-1})^2 - (\theta_2 y_{t-2})(\theta_1 y_{t-1}) + 1}{(\theta_1 y_{t-1})^3} \quad (3.47)$$

Therefore,

$$\Rightarrow \sum_{i=1}^{n}\left[-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})]y_{t-1}(\theta_1 y_{t-1} + \theta_2 y_{t-2})^{-1}\right] \quad (3.48)$$

$$\sum_{i=1}^{n}\left[-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})]y_{t-1}\frac{(\theta_1 y_{t-1})^2 - (\theta_2 y_{t-2})(\theta_1 y_{t-1}) + 1}{(\theta_1 y_{t-1})^3}\right] \quad (3.49)$$

$$\sum_{i=1}^{n}\left[-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})]\frac{(\theta_1 y_{t-1})^2 - (\theta_2 y_{t-2})(\theta_1 y_{t-1}) + 1}{(\theta_1)^3(y_{t-1})^2}\right] \quad (3.50)$$

$$\frac{\delta \ln(L)}{\delta\theta_1} = 0$$

And dividing through by $(\theta_1)^3(y_{t-1})^2$ gives

$$\sum_{i=1}^{n}\left[-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})][(\theta_1 y_{t-1})^2 - (\theta_2 y_{t-2})(\theta_1 y_{t-1}) + 1]\right] = 0 \quad (3.51)$$

$$-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\sum_{i=1}^{n}[\log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})]\theta_1[\theta_1 y_{t-1}^2 - \theta_2 y_{t-2} y_{t-1}] + 1] = 0 \quad (3.52)$$

$$-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\sum_{i=1}^{n}[\log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})][\theta_1(\theta_1 y_{t-1}^2 - \theta_2 y_{t-2} y_{t-1}) + 1]] = 0 \quad (3.53)$$

$$-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\sum_{i=1}^{n}[\log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})][\theta_1(\theta_1 y_{t-1}^2 - \theta_2 y_{t-2} y_{t-1}) + 1]] = 0 \quad (3.54)$$

$$-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\sum_{i=1}^{n}[\log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})][\theta_1(\theta_1 y_{t-1}^2 - \theta_2 y_{t-2} y_{t-1})] + \log[(\varphi_0 B^b + \varphi_1 B^{b+1} + \varphi_2 B^{b+2})]] = 0 \quad (3.55)$$

$$-\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\sum_{i=1}^{n}\left[\log[(\varphi_0 B^b+\varphi_1 B^{b+1}+\varphi_2 B^{b+2})](\theta_1\sum_{i=1}^{n}(\theta_1 y_{t-1}^2-\theta_2 y_{t-2}y_{t-1}))=\left(\frac{4+\mu\sigma^2}{\sigma^2}\right)\log[(\varphi_0 B^b+\varphi_1 B^{b+1}+\varphi_2 B^{b+2})]\right] \quad (3.56)$$

$$\sum_{i=1}^{n}\theta_1(\theta_1 y_{t-1}^2-\theta_2 y_{t-2}y_{t-1})=\frac{\log[(\varphi_0 B^b+\varphi_1 B^{b+1}+\varphi_2 B^{b+2})]}{\log[(\varphi_0 B^b+\varphi_1 B^{b+1}+\varphi_2 B^{b+2})]} \quad (3.57)$$

$$\theta_1^2\sum_{i=1}^{n}y_{t-1}^2-\theta_2\sum_{i=1}^{n}y_{t-2}y_{t-1}=1 \quad (3.58)$$

$$\theta_1^2\sum_{i=1}^{n}y_{t-1}^2=1+\theta_2\sum_{i=1}^{n}y_{t-2}y_{t-1} \quad (3.59)$$

$$\theta_i^2=\frac{1+\theta_2\sum_{i=1}^{n}y_{t-2}y_{t-1}}{\sum_{i=1}^{n}y_{t-1}^2} \quad (3.60)$$

Taking the square root of (3.60)

$$\theta_i=\sqrt{\frac{1+\theta_2\sum_{i=1}^{n}y_{t-2}y_{t-1}}{\sum_{i=1}^{n}y_{t-1}^2}} \quad (3.61)$$

4.0 RESULTS AND DISCUSSION

Table 1: Results for ARIMAX and LOG-ARIMAX Models Selection

Ticker	Model Type	Selected Model	AIC
DBG	ARIMAX	(0,1,2)	781.65
DBG	LOG-ARIMAX	(0,1,2)	765.72
DBG*	ARIMAX	(0,1,2)	533.38
DBG*	LOG-ARIMAX	(0,1,2)	525.53

Table 2: Estimation of Model Parameters for ARIMAX and LOG-ARIMAX

ESTIMATES	ARIMAX	LOG-ARIMAX	ARIMAX*	LOG-ARIMAX*
b	-	0.6785	-	0.6366
AR (1)	-	-	-	-
AR (2)	-	-	-	-
AR (3)	-	-	-	-
MA (1)	-0.0199	-0.0297	-0.022	-0.033

MA (2)	0.2916	0.3322	0.3103	0.3367
MA (3)	-	-	-	-

Table 4.5: Forecast Accuracy Measures

	Test Type		
Ticker	MAE	RMSE	MSE
ARIMAX	44.7725	56.5525	3198.1830
LOG-ARIMAX	36.80373	49.8227	2482.3040
ARIMAX*	53.4383	65.2898	4262.7570
LOG-ARIMAX*	42.6022	54.0129	2917.3950

Data used was monthly adjusted data recorded by four Oil and Gas companies from 2005 - 2020 with a total of 72 observations. The Mean Square Error (MSE), Mean absolute Error (MAE) and Root Mean Square Error (RMSE) serve as the error matrices in evaluating the forecastability of the models. The effect of Akaike Information Criterion (AIC) and the linear correlation on candidate models among the considered oil spill data tested. Table 1 show that the Log-ARIMAX model has the least AIC in the two time horizon as compared to the classical ARIMAX model. Results for ARIMAX and LOG-ARIMAX Models Selection with respect to AIC show ARIMAX (0,1,2) with AIC 781.65, LOG-ARIMAX (0,1,2) with AIC 765.72, ARIMAX (0,1,2) with AIC 533.38, LOG-ARIMAX (0,1,2) with AIC 525.53. Also, with respect to error metrics ((Forecast Accuracy Measures), the results show ARIMAX (MAE = 44.7725, RMSE= 56.5525, MSE= 3198.1830) and LOG-ARIMAX (MAE = 36.80373, RMSE= 49.8227, MSE= 2482.3040). The results also show that the data used for the analysis are not significantly correlated. None of the random walk test of all the considered data was significant both with homoskedastic and heteroskedastic errors. This implies that the LOG-ARIMAX model has a better forecasting strength and accuracy as compare to that of ARIMAX model. Tables 2 &3 shows estimation of model parameters and error metrics (forecast accuracy measures) respectively. The values of the error metrics, in terms of

MAE, RMSE, and MSE, shows that the LOG-ARIMAX model gives better forecasting accuracy than the traditional ARIMAX model.

CONCLUSION

This study proposes a hybrid ARIMAX model to capture and accommodate both the external covariate(s) and the heavy-tailed properties of observational time series events using secondary datasets of the long memory types of oil spillage and temperatures. The results of the analysis show that the hybridization of Logarithm and ARIMAX (LOG-ARIMAX) as propounded in this work is more robust, efficient, sufficient and reliable in forecasting long-memory data characterized by heavy tailed traits.

REFERENCES

Alves, T.M., Kokinou, E., Zodiatis, G. Lardner, C.P. (2016). Multidisciplinary oil spill modeling to protect coastal communities and the environmentof the Eastern Mediterranean Sea, Scientific Reports, 6(36882).

Alves, T.M., Kokinou, E., Zodiatis, G. Lardner, C.P., Radhakrishnan, H. (2015). Modelling of oil spills in confined maritime basins: the case for early response in the Eastern MediterraneanSea, *Environmental Pollution*, 206(8069): 390–399.

Alves, T.M., Kokinou, E., Zodiatis, G. (2014). A three-step modelto assess shoreline and offshore susceptibility to oil spills: the southAegean (crete) as an analogue for confined marine basins, *Marine Pollution Bulletin*, 86(2): 443–457.

Büyükşahin, Ü. Ç., & Ertekin, Ş. (2019). Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing*, *361*, 151-163.

Chrobok, R.; Kaumann, O.; Wahle, J.; Schreckenberg, M. (2004). Different methods of traffic forecast based on real data.*European Journal of Operational Research,* 155 (3): 558-568.

Dietrich, J.C., Trahan, C. J., Howard.,M. T. (2012). “Surfacetrajectories of oil transport along the Northern Coastline of theGulf of Mexico,” *Continental Shelf Research*, 41: 17–47.

Doktoringenieurs,G. (2010).Short-Term Freeway Traffic Flow Forecasting with ARIMAX Modeling.

Farhana I. And Monzur A. I. (2020) Use Of Teleconnections To Predict Western Australian Seasonal Rainfall Using ARIMAX Model 2020. Hydrology 7(3):52

Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications*, *103*, 25-37.

Lee, M. H., Suhartono, D.,Hamzah, N.A. (2010). Calendar variation model based on ARIMAX for forecasting sales data with Ramadhan effect.*Proc. RCSS'10, Malaysia*, 349-361.

Ling, A. S. C., Darmesah, G. , Chong, K. P. , Ho, C. M. (2019). Application of ARIMAX Model to Forecast Weekly CocoaBlack Pod Disease Incidence. Mathematics and Statistics 7(4A):29-40

Ragulina, G., Reitan, T. (2017).Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the Bayesian approach.*Hydrological Sciences Journal*, 4(1):2150-3435.doi: 10.1080/02626667.2016.1260134.

Wadud, Z. (2014) Simultaneous modeling of passenger and cargo demand at an airport.Transportation Research Record: *Journal of the Transportation Research Board*, 2336: 63- 74.ISSN 0361-1981.https://doi.org/10.3141/2336-08.