

On Statistical Analysis and Modeling of Rare Events with Autoregressive Integrated Moving Average (Arima) Model.

¹T.O. Olatayo, ²Mayor Andrew, ³O.O. Alabi and ⁴R.B. Afolayan

¹Department of Mathematical Sciences, Olabisi Onabanjo University,
Ago-Iwoye, Ogun State, Nigeria.

²Department of Statistics, Delta State Polytechnic, Otefe-Oghara, Nigeria.

³Department of Statistics, Federal University of Technology, Akure, Nigeria.

⁴Department of Statistics, University of Ilorin, Ilorin, Nigeria

Abstract

The study was carried out to analyse and model a rare event and forecast future occurrences. Rare events are those event that follows a Poisson distribution and are mostly non stationary in nature.

The Box-Jenkins approach for Autoregressive Integrated Moving Average (ARIMA) modeling and the Adaptive Response Rate Single Exponential Smoothing (ARRSES) method were employed in comparison for forecasting. The parameters estimated were Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and Mean Square Error (MSE) respectively.

The results was obtained for three models, ARIMA (1,1,1), ARIMA (0,1,3) and ARIMA (2,0,0) and it showed that ARIMA (1,1,1) is the best fit. We compared ARIMA (1,1,1) with the Adaptive Response Rate Single Exponential Smoothing (ARRSES) method, which showed that forecast made by ARIMA models are more accurate and a reliable model to be fitted with minimum AIC, BIC and MSE.

In conclusion, this research provides a framework for the investigation of other rare events in future. The best model to be fitted for a rare event that follow Poisson process and are normally and independently distributed with a non stationary nature is ARIMA (1,1,1).

Keywords: Rare event, ARIMA, ARRSES, MSE, Poisson Distribution.

1.0 Introduction

Forecasting methods to produce numerical estimates range from relative simple techniques to complex and sophisticated techniques are discussed in this paper. Among forecasting methods were extrapolative or projective techniques i.e. moving averages and exponential smoothing. A moving average is a trend method wherein each point of a moving average of a time series is the arithmetic mean of a number of consecutive observations of the series. The number of observations in the moving average computation is chosen to minimize the effect of seasonality or other disturbances in the series [1-4].

Exponential smoothing is a flexible trend whereby past data observations are given different weights in computing the forecast. It has the advantages of providing a simple up-to-date forecast where the new forecast is equal to the previous one plus some stated proportion of the previous periods of a forecasting error. Exponential smoothing methods are adaptable to adjustment to include trend and seasonal projections with adaptive types of optimum weighting procedures. Exponential smoothing methods are used to forecast large numbers of items. Time series decomposition methods are widely used to identify the systematic components of a time series, trend cycle and seasonal pattern and the nonsystematic or random component. The seasonal Pattern is identified by first determining the seasonal indexes for each month or quarter of year and in turn these patterns are projected ahead. The cyclical forecast may be prepared by other systematic projection or by economic judgment. Nonsystematic or irregular variation is usually assumed zero in a forecast but irregular adjustments may be needed for an anticipated stoppage in production or some other casual factors in the time period of the forecast. A highly analytical method for measuring seasonal fluctuations is called census I or X-11 variant. Regression models express the past relationships among the item being forecasted. These models are useful when adequate history of data are available on the major factors associated with variations in the item being forecasted [5,6].

Corresponding author: T.O. Olatayo, E-mail: otimtoy@yahoo.com, Tel.: +2348034052195.

All methods of forecasting do not work equally well in the same circumstances, they need diversified conditions. The moving averages need the past n sample observations and if a large number of items are to be forecasted, these data can require considerable storage space. Equal weights are given to all past observation and no weight is given to observation earlier than period t_{n+1} . Recent observation may contain more information than the older ones in forecasting future movements. We give more weight in forecasting to the recent observations than to the older ones [7].

Exponential smoothing produces self-correcting forecasts with built-in adjustment that regulate forecast values by changing them in the opposite direction of errors. Forecasts generated by exponentials smoothing are sensitive to the specification of the smoothing constant. This technique results in forecasts that lag behind the turning point in actual time series data [8].

Adaptive response rate single exponential smoothing does not require specification of α and is attractive when a many items have to be forecasted. When trend is evident in the time series, single exponential smoothing like simple moving average becomes less useful. Both single and double exponentially smoothed values have the same property of lagging the actual values.

Brown's linear exponential smoothing method provides an additional correction similar to that of linear moving averages. In this method the difference between the single and double smoothed values is added to the single smoothed values and adjusted for a pattern in the data. Holt's double parameter linear exponential smoothing is an extension of Brown's method whereby a growth factor is added to the smoothing equation. This method smoothes the trend values directly. Quadratic exponential smoothing method of forecasting is appropriate when the underlying pattern of the observed time series is quadratic, cubic or of a higher order. Winter's triple parameter linear and seasonal exponential smoothing model of forecasting is an extension of Holt's double parameter method [9].

In this research we have suggested three ARIMA models in comparison with Adaptive response rate single exponential smoothing for forecasting road accidents which is one of the rare events that occur in nature. These models were proposed by Box and Jenkins. The Box-Jenkins method is a statistically sophisticated way of analyzing and building a forecasting model which best represent a time series. This method has a number of advantages over other methods of time series analysis. First this method of analyzing time series is logical and statistically accurate. Second the methods extract a great deal of information from the historical time series data. This method results in an increase in forecast accuracy while keeping the number of parameters to a minimum in comparison with other similar modeling processes. Box-Jenkins methodology assumes no particular pattern in the historical data of the time series to be forecasted. By iterative approach, the procedure identifies a possible useful model from a general class of ARIMA models. After the estimation, the chosen model is then diagnostically checked against the historical time series data. If the mean square error (MSE) is small and the residuals are also small, then the chosen ARIMA model is said to be a good fit. If the chosen model is not satisfactory, the Box-Jenkins process is repeated using another model designated to improve the original one. The process is repeated until a satisfactory model is found [10].

This methodology can be enlarged to include seasonal data by building a multiplicative ARIMA model. The multiplication model requires the identification of both a non-seasonal and seasonal models that is non stationary in nature. Box-Jenkins approach for ARIMA modeling provides for forecasting with a very powerful and flexible tool. It is an excellent method for forecasting a time series from its own current and historical values. Naturally, most of the rare events, such as road accidents are non stationary in nature. Therefore, the problem of non stationarity should be taken care of in any forecasting methodology [11,12].

A great deal of research has been and is being committed to this topic all over the world. Road accident in developing countries has increased over the last ten years. This could be due to increasing vehicle occupancy over time or to the fact that a specific type of accident is much more than it was some years back.

2.0 Poisson Distribution

Let μ be a positive real number. A discrete variable with probability function

$$f(x) = \mu^x e^{-\mu} / x! \quad x = 0, 1, 2, \dots \quad (1.1)$$

is said to have Poisson distribution. The parameter μ is called the mean of distribution. If we interpret probabilities as long run relative frequencies, then μ is the average or mean of the X -values that we would obtain in a very large number of repetitions of the experiment. The mean is the central value of the probability distribution in the same sense that the centre of mass is the central point of a mass distribution. Suppose that at point $X_1, X_2, \dots$ along a rigid horizontal axis, we place weights $f(X_1), f(X_2), \dots$ where $\sum f(X_i) = 1$, then the centre of gravity or balance point of the system is located at the point $\sum x_i f(x_i) = \mu_x$.

If x has mean μ_t .

Then

$$E(x - \mu) = E(x) - \mu = \mu - \mu = 0$$

The total probability in (1.1) is

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} \mu^x e^{-\mu} / x! = e^{-\mu} \left[1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \dots \right] \quad (1.2)$$

The series in square bracket is the Taylors expansion of $e^{-\mu}$ and hence the total probability is $e^{-\mu}e^{\mu} = 1$ for all μ .

One of the possible applications of the Poisson distribution is to describe the number of accident per driver. A sample of n drivers of a particular age group is examined and it driver has X_i accidents for $j = 1, 2, \dots, n$. The number of accidents per driver is assumed to be a Poisson variable except that some of the drivers have no accidents because they are vigilant and careful drivers and not merely because of the chance variation allowed for by the Poisson distribution. Let $Y=1$ or 0 depending on whether a driver is careless in driving or not (vigilant). Let the overall proportion of drivers which are vigilant and careful be ϕ , ie $Pr(Y = 0) = \phi$ which is unknown and a nuisance parameter. When $\phi = 0$, estimation of θ reduces to the standard problem and $\sum_{i=1}^n X_i$ is the minimal sufficient statistics for θ [13].

3.0 Methodology

3.1 Adaptive Response Rate Single Exponential Smoothing (ARRSES)

Adaptive response rate single exponential smoothing (ARRSES) method is also used to forecast the monthly univariate time series on road accidents. This is an attractive method when a great many items have to be forecasted. The ARRSES forecasting is

$$F_{t+1} = a_t X_t + (1 - a_t) F_t \quad (2.0)$$

where $a = |E_t/M_t|$

Smoothed error $E_t = B e_t + (1 - B) E_{t-1}$

Absolute smoothed error $M_t = B |e_t| + (1 - B) M_{t-1}$

Error $e_t = X_t - F_t$

$B = 0.2$

In this method we change the value of unspecified α on an ongoing basis. ARRSES changes the value of α when there is a change in the basic pattern requiring different smoothing constant. The forecast F_{t-1} is a linear combination of the previous period's observed value and forecast. The smoothing constant is the absolute value E_t divided by M_t . E_t is a linear combination of the forecast error, e , and the smoothed E . M_t is absolute smoothed error and is a linear combination of the forecast error and the absolute smoothed error in the previous period. By assigning $B = 0.2$ means that the first term in smoothed error and absolute smoothed error gives much smaller weights than the second. With $B=0.2$ the second term is given four times more importance [5].

3.2 Univariate Time Series

When $k=1$, we shall write $Z_1 = Z_1$

$Z_t = (Z_{1t}, \dots, Z_{kt})^t$

An important class of models for discrete univariate series originally proposed in [14] is stochastic difference equation of the form

$$\phi_l(B) Z_t = \theta_q(B) a_t \quad (2.1)$$

where $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ and $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$

In (2.1) the a_t 's are independently identically and normally distributed random shocks (or white noise) with zero mean and variance σ_a^2 . B is the back-shift operator such that $BZ_1 = Z_{1-1}$ and $Z_1 = Z_1 - \eta$ is the deviation of the observation z_T from some convenient location η .

Relationships between K series $\{Z_n\}, \dots, (Z_n)$ can sometimes be represented by linear transfer function models of the form

$$z_{iu} = \sum_{h \in S_i} [\omega_{shi}(B) B^{b_{shu}} / \delta_{chi}(B)] Z_{it} + [\theta_{qb}(B) / \varphi_{ph}(B)] a_{nt} \quad (2.2)$$

where $Z_{0t} \equiv 0$, $k(h)$ is the set $(1, \dots, h-1)$, $\omega_{shi}(B) \delta_{shi} \varphi_{ph}(h = 1, 2, \dots, k)(B)$, and $\theta_{qb}(B)$; are polynomials in B the b_{tu} 's are nonnegative integers, and $\{a_{if}\}, \dots, \{a_{ki}\}$ are k independent Gaussian white-noise processes with zero mean and variance $\sigma_1^2, \dots, \sigma_k^2$ in particular intervention models of this form has proved useful [4].

Transfer function models of the form (2.2), however assume that the series, when suitable arranged possess triangular relationship, implying for example that z_1 depends only on its own past, z_2 depends on its own past and on the present and past of z_1 , z_1 , on its own past and on the present and past of z_2 and z_1 ; and so on. On the other hand if z_1 depends on the past of z_2 and also z_2 depends on the past of z_2 and also z_2 depends on the past of z_1 then we must have a model that allows for this feedback.

3.3 Univariate ARIMA Forecasts

Let a time series of a set of ordered observations is

$$Y_1, Y_2, Y_3, \dots, Y_{t-1}, Y_t, Y_{t+1}, \dots \quad (2.3)$$

We define trend as any systematic change in the level of time series. A time series that is trendless can be represented by the model.

$$Y_t = b_0 + N_t \quad (2.4)$$

where the parameter b_0 is interpreted as the level of the time series N_t is a "noise" component or as stochastic process. Assuming white noise process, we can write the model for a trendless time series process as

$$Y_t = b_0 + a_t \quad (2.5)$$

where $a_t \approx NID(0, \sigma_a^2)$ i.e. white noise consists of a series of random shocks each distributed and independently about a zero mean with constant variance. σ_a^2 . As the mean of white noise is zero the expected value of a trendless process is $EY_1 = b_0 + Ea_1 = b_0$ where $a_1 \approx NID(0, \sigma_a^2)$ that is, white noise consist of a series of random shocks, each distributed normally and independently with a zero mean and constant variance, σ_a^2 . As the mean of a white noise is zero, the expected value of a trendless process is

$$EY_1 = b_0 + Ea_1 = b_0 \quad (2.6)$$

The parameter b_0 is the level about which the realized time series fluctuates. In case of trend less time series process b_0 is the arithmetic mean define as

$$\bar{b}_0 = \bar{Y} = 1/N \sum_{t=1}^N Y_t \quad (2.7)$$

A time series following a linear trend, requires, a model

$$Y_1 = b_0 + b_1 1a_1$$

Taking expectation

$$E(Y_1) = b_0 + b_1$$

A regression of Y_t on the time series $t(t=0,1,\dots,N)$. The level of the time series process is changing systematically throughout its course. Starting at $t=0$

$$E_1 Y_0 = b_0$$

$$E(Y_1) = b_0 + b_1$$

$$E(Y_2) = b_0 + 2b_1$$

$\vdots$

$$E(Y_N) = b_0 + Nb_1$$

The parameter b_0 is the intercept of the model, that is the expected level of the process when $t = 0$. The parameter b_1 is the slope of the model that is the expected change in level from one observation to the next [15-17].

3.4 The General ARIMA(p, d, q) Model

Let the ARIMA (1,0,1) process as

$$y_t = \phi_1 y_{t-1} + a_t - \theta_1 a_{t-1} \quad (2.8)$$

Or

$$(1 - \phi_1 B)y_t = (1 - \theta_1 B)a_t$$

and solving for y_t

$$y_t = (1 - \phi_1 B)^{-1} (1 - \theta_1 B)a_t$$

$$\text{Since } (1 - \phi_1 B)^{-1} = 1 + \phi_1 B + \phi_1^2 B^2 + \dots + \phi_1^n B^n + \dots$$

So

$$(1 - \phi_1 B) + \phi_1^2 B^2 + \dots + \phi_1^n B^n + \dots)(1 - \theta_1 B)a_t$$

When $\phi_t = \theta_1$ both parameters are completely redundant substituting ϕ_1 for θ_1 in the infinite series form of the ARIMA (1,0,1) process

$$y_t = (1 + \phi_1 B + \phi_1^2 B^2 + \dots + \phi_1^n B^n + \dots)(1 - \phi_1 B)a_t \\ = (1 - \phi_1 B + \phi_1^2 B^2 - \dots - \phi_1^n B^n + \phi_1^n B^n \dots)a_t \quad (2.9)$$

when $\phi_t = \theta_1$ an (ARIMA (1,0,1) model reduces to an ARIMA (0,0,0) model

Now suppose that $\phi \neq \theta$ but that the parameter are close to each other in value, if $\theta = \phi_1$ where c is a very small number the infinite series form of the ARIMA (1,0,1) models is

$$y_t = (1 + \phi_1 B + \phi_1^2 B^2 + \dots + \phi_1^n B^n + \dots)(1 - \phi_1 B - cB)a_t \\ = (1 - cB - c\phi_1 B + \phi_1^2 B^2 - \dots - \phi_1^n B^n + \phi_1^n B^n \dots)a_t \quad (2.10)$$

Now when both c and ϕ_1 are small $c\phi_1$ will be approximately zero and the ARIMA (1,0,1) model reduces approximately to an ARIMA (0,0,1) model

$$y_t = (1 + \theta_1 B)a_t \quad (2.11)$$

where $c=0$ and if $c\phi_1$ is not approximately zero, then the ARIMA (1,0,1) model reduces to approximately an ARIMA (0,0,2) model

$$y_t = (1 - \theta_1 B - \theta_2 B^2)a_t \quad (2.12)$$

where $c\phi_1 = \theta_2$ [11].

4.0 Results Of Analysis

The monthly univariate time series on rare events (road accidents) of South- Western Nigeria during 2011-2013 is used to illustrate the Box-Jenkins method of forecasting investigated earlier. We compare ARIMA models for the series to assess the forecasting performance of the models to see whether any benefit can be gained from the improved fit. We choose to forecasts the 12 observations (year 2014) with each of models (fitted to the whole of the available data) at various forecasting horizons and to calculate the sum of squared forecasting errors for these forecasts. Formal data analysis suggests that the data are non-stationary and differencing is indispensable [18].

The three ARIMA models and their corresponding forecasts are as follows:

Model 1: ARIMA (1,1,1). The model fitted is;

$$(1 - B)(1 + 0.0052B)Y_t = (1 - 0.7473B)e_t$$

The forecast for ARIMA (1,1,1) fitted above is given below

Table 1: Forecast for the ARIMA (1,1,1)

Period	Forecast
157	89
158	89
159	88
160	88
161	88
162	87
163	87
164	87
165	86
166	86
167	86
168	86

Model 2: ARIMA (0,1,3). The model fitted is;

$$(1 - B)Y_t = (1 - 0.7545B + 0.0394B^2 - 0.0470B^3)e_t$$

Table 2: Forecast for the ARIMA 0,1,3)

Period	Forecast
157	89
158	88
159	88
160	88
161	87
162	87
163	87
164	86
165	86
166	86
167	85
168	85

Model 3: ARIMA (0,0,2). The model fitted is;

$$Y_t = 0.4135Y_{t-1} + 0.4070Y_{t-2} + e_t$$

Table 3: Forecast for the ARIMA (0,0,2)

Period	Forecast
157	97
158	96
159	99
160	100
161	101
162	102
163	103
164	104
165	105
166	105
167	106
168	106

The ARIMA (1,1,1) models yields minimum $MSE = 224.00$. The estimate of the models are $AR(1)=0.7473$ with the constant and their standard deviations 0.1076, 0.0715 and 0.3061 respectively. The t-ratio of the estimates are -0.05, 10.46, -0.98. Table 1 and table 2 shows that the forecasts produced by ARIMA (1,1,1) and ARIMA (0,1,3) are very similar.

The ARRSEA model is given as

$$F_{t+1} = 0.0764X_t + (1 - 0.6567)F_t$$

Table 4: The ARRSES Forecast

Obsd. Value X_t	Forecast F	Error e_t	Smoothed error E_t	Absolute M_t	Error α_t value
116					
119	116	3	0.60	0.60	1.00
138	119	19	4.28	4.28	1.00
150	138	12	5.82	5.82	1.00
165	160	15	7.66	7.66	1.00
162	165	-3	5.53	6.73	0.82
139	163	-24	0.38	10.18	0.04
146	162	16	3.50	11.34	0.31
111	157	-46	1.20	18.27	0.66
174	127	-47	4.24	24.02	0.01
149	127	22	8.78	23.61	0.18
158	131	27	7.23	24.29	0.36

Three ARIMA models best fit the time series and performed better than the ARRSES, the forecasts made by ARIMA model are more accurate and showing downward trend in road accidents. The forecasts have incorporated the effect of various measures exercised to improve the worsening situation of road accidents. These measures are highlighted which include human environmental and engineering factors responsible for road accident.

The model selection criteria are given below:

Table 5: Model Selection Criteria

Model	MSE	AIC	BIC
ARIMA (1,1,1)	224.0	9.81672	5.014325
ARIMA (0,1,3)	234.8	9.890158	5.087763
ARIMA (2,0,0)	253.9	9.881705	5.07931

ARRSES forecasts are given in Table 4 which are compared with ARIMA (1,1,1) models forecasts given in Table 1. The MSE for ARRSES is 275.49. while the MSE for ARIMA (1,1,1) model is 224.0 which is smaller than the MSE for ARRSES as seen in Table 5.

5.0 Conclusion

In conclusion, in forecasting rare event, the Autoregressive Integrated Moving Average (ARIMA) model is better in forecasting future values. The assumption was reached after it was compared with the Adaptive Response Rate Single Exponential Smoothing (ARRSES). The selection criteria used the AIC, BIC and MSE all gave a minimum value quite less than the value of MSE of ARRSES. This research provide a framework for the investigation of other rare events in the future.

The best model to be fitted for a rare event that follow a Poisson process and are normally and independently distributed with a non stationary nature is ARIMA (1,1,1).

6.0 References

- [1] BOX.G.E.P. JENKINS. G.M. (2006), "Time series Analysis forecasting and control ," Revision Edition. San Francisco. Holden. Day.
- [2] BOX, G.E.P, JENKINS, G.M, REINSEL, G.G.(2004)," Time series Analysis. Forecasting and control." 3rd. Edition. Englewood Cliffs. Prentice Hall.
- [3] DUFUR, H (2004), "The Comparison and Evaluation of Forecasters," The Statistician V, 32, pp, 12 – 22.
- [4] BOX. G.E.P. TIAO, G.C.(2005), " Change in Level of a Nonstationary Time series." Biometrical 52, pp. 181-200.
- [5] BROWN, R.G. AND NEYMAN, J . (2002). "smoothing, forecasting and precision of Discrete Time series, "Englewood cliff, NJ: prentice Hall.
- [6] KENNY, P.B., DURBIN,J (2002), " Local Trend Estimation and Seasonal Adjustment of Economic and Serial Time Series." Journal of Royal Statistical Society. A, V. 145, pp.1-41
- [7] PIERCE, D.A. (2000), 'Data Revisions with Moving Average Seasonal Adjustment Procedures'. *Journal of Econometrics*, V.14, pp. 320-325.
- [8] WINKLER, R.L. AND MAKRIDAKIS, S. (2003)," the Contribution of forecast." Journal of Royal Statistical Society. A, V. 146 pp. 150-157.
- [9] WALLIS, K.F. (2002), "Seasonal Adjustment and Revision of Current Data: Linear Filters for the X-11 Method". Journal of Royal Statistical Society. V. 145, pp. 76-85.
- [10] WEIS, A.A., ANDERSEN, A.P. (2004). Estimating Time Series Models Using The Relevant forecast Evaluation criterion: Journal of Royal statistical society, A,V, 147. 481-488.
- [11] RAO, C.R. (2008), " Prediction of Future Observation with Special Reference to Linear Models in Multivariate Analysis.' V.4 Edition, P.R. Krishnaiah, Amsterdam: North Holland pp. 200-208.
- [12] ALI, M.M. (2009), Tests for Autocorrelation and Randomness in Multiple Time Series." Journal of American Statistical Association V.B4. No 406, pp. 633 – 540.
- [13] KUNIROMO,N.L AND VAMAMOTO,S. (2005), " Discrete Distribution Boston." Houghton Niffin
- [14] YULE G.U (2007). ' ON A method of investigating periodicities in Disturbed series with special Reference to wafer's sunspot Numbers.' Philosophical transactions of Royal society London, V226,pp 267-289.

- [15] PERRON,P .(2001), " A Continuous Time Approximation to the Unstable First-Order Auto-Regressive Process." The Case without an Intercept Econometrica. V 59, pp. 211-239.
- [16] TSAY, R.S. TIAO, G.C. (2004), "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Non Stationary ARMA Models."Journal of American Statistical Association. V. 79 No 385, pp. 84-96.
- [17] TSAY, R.S, (2006), "Time Series Model Specification. V.81 No 393, pp 132-141.
- [18] MAYOR Andrew (2012), Statistical Analysis and Modelling of Rare Events; An Unpublished M.SC. Dissertation, in the Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye Ogun State Nigeria.