African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 276-286)
www.abjournals.org

ON DETECTING AN APPROPRIATE MODEL IN TIME SERIES ANALYSIS

Chukwudi Anderson Ugomma[1] and Emmanuel Uchenna Ohaegbulem[2]

[1&2]Department of Statistics, Imo State University, Owerri, Nigeria.

Cite this article:

Chukwudi A. U., Emmanuel U. O. (2024), On Detecting an Appropriate Model in Time Series Analysis. African Journal of Mathematics and Statistics Studies 7(2), 276-286. DOI: 10.52589/AJMSS-16JZMGDJ

Manuscript History

Received: 17 Mar 2024

Accepted: 30 May 2024

Published: 3 Jul 2024

ABSTRACT: *This study sought to present yet another method of decomposition in time series data. The data for this study were of secondary source and obtained from sources which comprised of both the open and the closed stock prices. The two data were firstly tested for randomness and they were confirmed fit for time series analysis. The two data were also subjected to trend curve analysis, and it was observed that both data were of exponential curve since the exponential trend curve exhibited the highest coefficient of determination r^2 (88%), among other trend curves which included linear, quadratic, cubic and logarithmic curves. In the decomposition of the two data series, using the exponential trend, it was revealed that the model, for each data were of multiplicative type since the multiplicative model had the Minimum Mean Squared Error (MSE) of 0.00827 and 0.003665respectively for both Open and Closed Stock Prices of Coca-Cola Data. Hence, in this study, it was recommended that this traditional method of statistics should be applied in the decomposition of any time series data.*

KEYWORDS: Time Series Models, Test of Randomness, Trend Curves, Decomposition of Time Series Models, Mean Squared Error.

INTRODUCTION

A time series is a set of observations taken at a specified times, usually at equal intervals. Times series analysis is one of the descriptive method of analysis that involves four components which include the Secular Trend (T), Cyclical trend (C), Seasonal Component (S) and the Irregular Component (I).

Time series analysis amounts to investigation of the factors, T_t, C_t, S_t and I_t; and it is often referred to as a decomposition of a time series into its basic components. One of the descriptive procedures for time series data is to decompose and build a time series model.

Several authors have used different procedures to decompose the time series models in time series data, which include, Davey and Flores (1993), Chatfield (2004), Iwueze and Nwogu (2004), Iwueze and Akpanta (2009), Iwueze, et al (2011), Nwogu, et al (2019), Dozie and Ijomah (2020), Dozie and Ibebuogu (2020), Dozie and Uwaezuoke (2021). Almost all the studies mentioned above used the Buys-Ballot procedure for the choice of time series models, but in this study, an alternative known statistical procedure to decompose the trends shall be employed, and, hence, choose an appropriate model for the data in this study.

MATERIALS AND METHOD

Time Series Models

Basically, three models exists in time series modeling, and they include the additive model, the multiplicative model and the mixed model.

The Additive Model: This is the simpler to use arithmetically. It assumes that the actual data is the sum of the four effects, $, T_t, C_t$, S_t and I_t . This model assumes that the effects of the seasons, the cycles and the residual components are equal in absolute terms throughout the period of study. This assumption is probably true when short periods are involved or where the rate of growth or decline in the trend is small and no transformation is required.

Mathematically, the additive model is of the form,

$$Y_t = T_t + C_t + S_t + I_t \quad (1)$$

The Multiplicative Model: This model suggests that the actual data are the product of the effects of the trend, the seasonal, and the cyclical and residual components. This model is possible when one considers a long term and a marked growth rate, and a logarithmic transformation is certainly appropriate. Mathematically, the multiplicative model is given by,

$$Y_t = T_t.C_t.S_t.I_t \quad (2)$$

The Mixed Model: This model has multiplicative seasonality and additive errors and it is mathematically given by

$$Y_t = T_t + C_t + S_t.I_t \quad (3)$$

For further details about the models, see, for example, Chatfield (1975), Iwueze (2006), Linde (2005), Gupta (2013), Oladugba, et al (2004)

Article DOI: 10.52589/AJMSS-16JZMGDJ
DOI URL: https://doi.org/10.52589/AJMSS-16JZMGDJ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 276-286)

Testing for Time Dependency (Randomness) in Data Series

Prior to embarking on any time series analysis, it is traditionally most appropriate to test the series/sequence for randomness (otherwise known as test for dependency) especially where the series is a time series. The outcome of the test showing randomness would imply that the series is not time – dependent; and consequently, it would be most inappropriate to run a time series analysis on the data set. However, if the test shows non-randomness, then the series is time-dependent and the times analysis would be an appropriate analysis applied on the data set (see, for example, Nwobi and Nduka 2003).

One of the customary statistical tests for randomness in a sequence/series is the one-sample Runs Test. This test is basically used to test the null hypothesis the sequence of observations is random when the number of positive observations, (r_+), and the number of negative observations, (r_-), are equal to the total number of observations being less than 20; that is $r_+ + r_- = r < 20$ (Nwobi and Nduka, 2003)

In most real-life situations, the sequence encountered are usually large series. In this situation, the One-Sample Runs Test becomes impracticable; since it is limited to the number of observations should not be more than 20 observations. For a large sequence, the test statistic can be calculated using an approximation to the normal distribution. The test statistic for the test of randomness of a large sample sequence, according to Ross (2009), is given by

$$Z_R = \frac{R - \mu_R}{\sigma_R} \sim N(0,1) \qquad (4)$$

where,

R is the number of runs in the sequence;

μ_R is the expected number of runs in the sequence, and it is given by

$$\mu_R = \frac{2nm}{n+m} + 1 \qquad (5)$$

σ_R is the standard deviation of the number of runs in the sequence given by

$$\sigma_R = \sqrt{\frac{2nm(2nm - n - m) - \mu_R}{(n+m)^2 (n+m-1)}} \qquad (6)$$

n is the positive runs in the sequence; and

m is the negative runs in the sequence.

In this test the null hypothesis (that the series is random) is to be rejected if the test statistic, Z_R, is greater than or equal to critical value (Z_α) at 5% level of significance. Otherwise, the null hypothesis is not to be rejected.

Article DOI: 10.52589/AJMSS-16JZMGDJ
DOI URL: https://doi.org/10.52589/AJMSS-16JZMGDJ

Decomposition of the Time Series Data

The general method of decomposition of time series data has included the four possible components which make up a time series, but, it is not a constant rule that all the four components must be present. If an annual data are confronted, there can be no seasonal component. Similarly, if short periods are involved, the cyclical component can be ignored (Iwueze, 2006). In decomposition procedure, either of the additive or multiplicative may be used to effect the decomposition of the time series data. The first step to undertake before the decomposition of the time series data will usually be to estimate the trend and then to eliminate the trend for each time period from the actual data either by subtraction or division. If additive or multiplicative respectively, given a de-trended series that expresses the effect of the seasons, the cycles and the residual components. In this study, we shall de-trend the residual component from the data in order to detect whether the data is additive, multiplicative or mixed in nature.

The de-trended residual component for the additive model will be mathematically given as

$$Y_t - T_t - C_t - S_t = I_t \qquad (8)$$

For the Multiplicative model, it is given as

$$\frac{Y_t}{T_t.C_t.S_t} = I_t \qquad (9)$$

and for the Mixed model, it is given as

$$Y_t - T_t.C_t.S_t = I_t \qquad (10)$$

Method of Data Analysis

Two approaches shall be employed in analyzing the data used in this study for the purpose of illustration. The two approaches are the Time plot Approach and the Mathematical Trend Curves Approach.

The Time Plot Approach: This approach will be used here in order to indicate the appropriate movement (curve) of the data series and to enable the choosing of the appropriate mathematical trend to estimate the data series and to obtain the trend curves given by

$$T = \hat{Y}_t, \quad t = [-\infty, \infty] \qquad (11)$$

where, $\hat{Y}_t$ depends on the exact mathematical curve to be used in the estimation of the trend curves.

Mathematical Trend Curves Approach: Fitting a mathematical trend curve to time series data has obvious applications of forecasting due to the fact that the curve may easily be projected forwards as the basis of the forecast, and also that the trend values for the whole series will be made available, therefore, permitting a complete analysis of historical data to be undertaking (Iwueze, 2006). The mathematical methods used to deal with non-seasonal data that contains a trend is to fit a single function such as:

(a) **The Linear Trend**: In this method, the least squares method is applied with little modification and some simplifications. In this case, if the pattern of the trend is linear, a line shall be fit; and it is given by

Article DOI: 10.52589/AJMSS-16JZMGDJ
DOI URL: https://doi.org/10.52589/AJMSS-16JZMGDJ

$$Y_t = \alpha + \beta t \tag{12}$$

Using the least squares method, the normal equations will be reduced to

$$\sum_{t=1}^{n} Y_t = n\alpha \tag{13}$$

and

$$\sum_{t=1}^{n} tY_t = \beta \sum_{t=1}^{n} t^2 \tag{14}$$

so that $\hat{\alpha}$ and $\hat{\beta}$ can be obtained as

$$\hat{\alpha} = \frac{\sum_{t=1}^{n} Y_t}{n} \tag{15}$$

and

$$\hat{\beta} = \frac{\sum_{t=1}^{n} tY_t}{\sum_{t=1}^{n} t^2} \tag{16}$$

(b) Polynomial Curves: The general expression for the polynomial curve is given by

$$Y_t = \sum_{t=0}^{q} \alpha_i t^2 \tag{17}$$

where,

$\alpha_0, \alpha_1, \ldots, \alpha_q$ are constants and Y_t is the observation at time t

Generally, fitting a polynomial curve is of the form

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_n t^n \tag{18}$$

The polynomial curve can be in the form of, Quadratic curve, cubic curve, exponential curve, etc, but in this study, emphasis shall be laid on the linear, Exponential, Quadratic, Cubic and Logarithmic curves.

The Quadratic Curve: The quadratic curve is of the form

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + e_t \tag{19}$$

where, $\alpha_0, \alpha_1, \alpha_2$ can be determined by the method of least squares by minimizing

$$S = \sum_{t=1}^{n} \left(y_t - \alpha_0 - \alpha_1 t - \alpha_2 t^2 \right)^2 \tag{20}$$

such that

$$\sum_{t=1}^{n} y_t = n\hat{\alpha}_0 + \hat{\alpha}_2 \sum_{t=1}^{n} t^2 \tag{21}$$

and

Article DOI: 10.52589/AJMSS-16JZMGDJ
DOI URL: https://doi.org/10.52589/AJMSS-16JZMGDJ

$$\sum_{t=1}^{n} t^2 y_t = \hat{\alpha}_0 \sum_{t=1}^{n} t^2 + \hat{\alpha}_2 \sum_{t=1}^{n} t^4 \quad (22)$$

So that

$$\hat{\alpha}_2 = \frac{n\sum_{t=1}^{n} t^2 y_t - \left(\sum_{t=1}^{n} y_t\right)\left(\sum_{t=1}^{n} t^2\right)}{n\sum_{t=1}^{n} t^4 - \left(\sum_{t=1}^{n} t^2\right)^2} \quad (23)$$

and

$$\hat{\alpha}_0 = \frac{\sum_{t=1}^{n} y_t - \hat{\alpha}_1 \sum_{t=1}^{n} t - \hat{\alpha}_2 \sum_{t=1}^{n} t^2}{n} \quad (24)$$

Exponential Curve: The Exponential trend can be estimated with the formula given by:

$$Y_t = \alpha\beta^t \quad (25)$$

where α and β are constants. The exponential curve can be linearized by taking the logarithms as

$$\log_e Y_t = \log_e \alpha + t \log_e \beta = a + bt \quad (26)$$

Where $a = \log_e \alpha$ and $b = \log_e \beta$

As a matter of fact, any of the mathematical trend curves fittings would be used if the data series exhibits any of them.

Time Series Model Selection Criteria

Trend Curve Selection

To select the best trend curve among the mathematical trends, we shall employ the coefficient of determination, r^2, such that the fit with highest r^2 becomes the best.

The coefficient of determination r^2, given by Hamburg (1970) is given as

$$r^2 = 1 - \frac{S_{y_t t}^2}{S_{y_t}^2} \quad (27)$$

Equation (27) which is the case of two variables can also be written as

$$r^2 = 1 - \frac{\alpha\sum_{t=1}^{n} Y_t + \beta\sum_{t=1}^{n} tY_t - n\bar{Y}_t^2}{\sum_{t=1}^{n} Y_t^2 - n\bar{Y}_t^2} \quad (28)$$

For more than two variables (multiple linear regression), the coefficient of determination is given by

$$r^2 = 1 - \frac{S_{y_t.1.2...(k-1)}^2}{S_{y_t}^2} \quad (29)$$

where,

Article DOI: 10.52589/AJMSS-16JZMGDJ
DOI URL: https://doi.org/10.52589/AJMSS-16JZMGDJ

$$S^2_{y_t.1.2...(k-1)} = \frac{\sum (Y - Y_c)^2}{n-k} \tag{30}$$

Note that the coefficient of determination is expressed in percentage and it helps in finding the percentage of total variability in the dependent variable that is explained by the regression of the dependent variable on the independent variable(s).

Model Selection

Here, to select the appropriate times series model, the Mean Squared Error (MSE) shall be made use of; such that the model with minimum Mean Squared Error would suggest the appropriate model.

The Mean Squared Error (MSE) is given by

$$MSE = \frac{1}{n} \frac{\sum_{t=1}^{n} (e_t - \bar{e}_t)^2}{\sum_{t-1}^{n} e_t^2} \tag{31}$$

where,

$$e_t = Y_t - \hat{Y}_t \tag{32}$$

e_t is the error term (residual);

Y_t is the actual value;

$\hat{Y}_t$ is the trend value that is dependent on the appropriate mathematical curve; and

$\bar{e}_t$ is the mean of the residual.

DATA AND ANALYSIS

In this section, the Coca-Cola Open and Closed Stock Prices Data, obtained from *https://ng.www.investing.com/equities/cocacola-bottle-historical-data*, are used to illustrate the procedures of selecting an appropriate model for time series data described in this study.

Testing for Randomness in the Coca-Cola Data

(a) Open Stock Price

The Test Hypotheses for the data are stated as:

H_0: The Coca-Cola stock price is random

H_1: The Coca-Cola stock price is not random

The level of significance of the test is chosen at 5%

The test statistic expressed in (5) is used, and the following quantities were obtained:

The Median for the open stock price of the Coca-Cola data is 201.78

The runs for the data series were obtained as:

Article DOI: 10.52589/AJMSS-16JZMGDJ
DOI URL: https://doi.org/10.52589/AJMSS-16JZMGDJ

Runs: 47(+), -, +, 9(-), 0, 10(+), 16(-), +, 34(-).

Thus, the number of runs in this sequence is obtained as $R = 8$, while the number of positive and negative runs in the sequence are, respectively, obtained as: $n = 4$ and $m = 4$.

The computation of the expected number of runs and standard deviation in the sequence using (5), (6) and (7) are summarized in Table 1.

Table 1: Computation of Runs Properties of Coca-Cola Open Stock Price

$\mu_{R(open)}$	$\sigma_{R(open)}$	$Z_{R(open)}$	$Z_{\alpha(open)}$	Decision
3	2.61	1.915	1.645	Reject

The results in Table 1 showed that the computed test statistic (1.915) is greater than the critical value (1.645). Therefore, the null hypothesis (that the open coca-cola stock price is random) is rejected, and the conclusion is that the time series analysis can be applied on the Coca-Cola Stock Price data series.

(b) Closed Stock Price Series

The Test Hypotheses for the data are stated as:

H_0 : The Coca-Cola stock price is random

H_1 : The Coca-Cola stock price is not random

The level of significance of the test is chosen at 5%.

The test statistic expressed (5) is used, and the following quantities were obtained:

The Median for the open stock price of the Coca-Cola data is 199.45

The runs for the data series were obtained as:

Runs: 34(+), -, 13(+), -, +, 9(-), 11(+), 16(-). +, 33(-)

Thus, the number of runs in this sequence is obtained as $R = 10$ while the number of positive and negative runs in the sequence are, respectively, obtained as: $n = 5$ and $m = 5$

The computation of the expected number of runs and standard deviation in the sequence using (5), (6) and (7) are summarized in Table 2.

Table 2: Computation of Runs Properties of Coca-Cola Closed Stock Price

$\mu_{R(closed)}$	$\sigma_{R(closed)}$	$Z_{R(closed)}$	$Z_{\alpha(closed)}$	Decision
6	1.49	2.685	1.645	H_0 Reject

The result in Table 2 showed that the computed test statistic (2.685) is greater than the critical value (1.645). Therefore, the null hypothesis (that the closed coca-cola stock price is random)

Article DOI: 10.52589/AJMSS-16JZMGDJ
DOI URL: https://doi.org/10.52589/AJMSS-16JZMGDJ

is rejected; and the conclusion is that time series analysis can be applied on Coca-Cola Closed Stock Price data series.

The Time Series Plots for the Coca-Cola Open and Closed Stock Prices

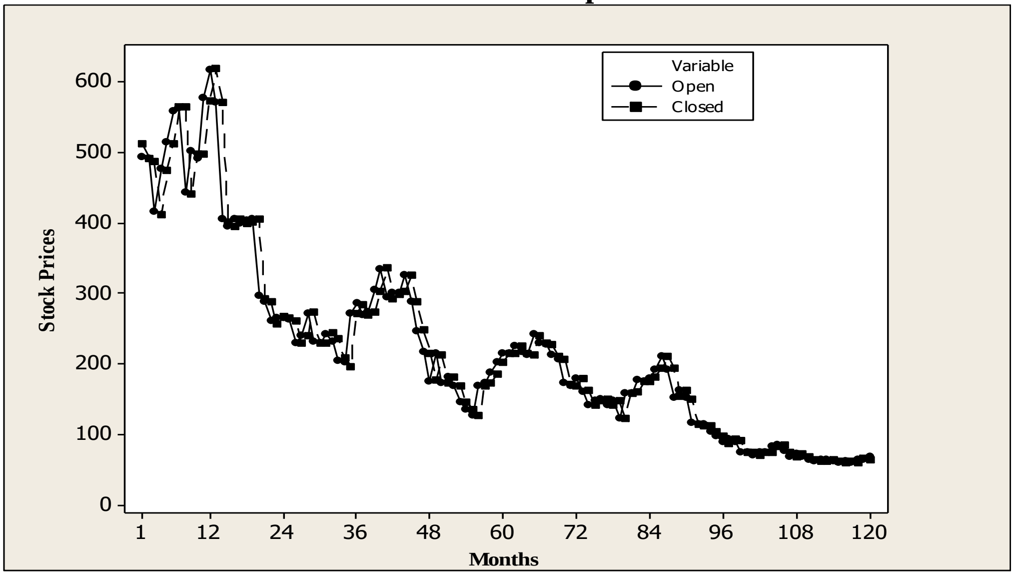

Figure 1: Plot of Open and Closed Coca-Cola Stock Prices

Model Selection

The results of the coefficients of determination for the different trends stated in this study for both Open and Closed Stock prices are summarized in Table 3

Table 3: Model Selection for both Coca-Cola Open and Closed Stock Prices

Trend	r^2	
	Open price	**Closed price**
Linear	0.788	0.790
Exponential	0.879	0.878
Quadratic	0.839	0.841
Cubic	0.869	0.871
Logarithmic	0.808	0.807

From Table 3, it is observed that both open and closed prices of Coca-cola stock exhibited exponential curve as the best since it revealed the highest coefficient of multiple determination, $r^2(88\%)$ which is higher than those of other curves. Therefore, the exponential model of the Time Series is to be employed to decomposition of the Coca-Coca Open and Closed Stock Prices data. The process of decomposition of the two data series are presented in Table 4

Article DOI: 10.52589/AJMSS-16JZMGDJ
DOI URL: https://doi.org/10.52589/AJMSS-16JZMGDJ

Table 4: Computation of Exponential Model Properties

Stock Price	Parameters		Model	MSE		Choice of Model
	A	b		Additive	Multiplicative	
Open	2.264429	-0.0037	$\hat{Y}_o = 2.2644 - 0.0037t$	0.00833	0.008265	Multiplicative
Closed	2.271202	-0.0037	$\hat{Y}_c = 2.271202 - 0.0037t$	0.008402	0.003665	Multiplicative

From Table 4, it is observed that both open and closed stock prices have almost the same intercept and exhibit the same values of slope. The minimum Mean Squared Errors of both open and the closed stock prices show the multiplicative model as the choice model for both Coca-Cola Open and Closed Stock Prices data series.

CONCLUSION

This study has presented another method of detecting an appropriate model in time series data. Based on the selection criterion in Section 3.2 of the study, it revealed that both Open and Closed Stock Prices data exhibited multiplicative model as the best choices of models.

REFERENCES

Chatfield, C (1975): The Analysis of Time Series: Theory and Practice, Monographs on Statistics and Applied Probability (MSAP).

Chatfield, C (2004): The Analysis of Time Series: An Introduction. Chapman and Hall/CRC Press, Boca Raton.

Davey, A.M and Flores, B.E (1993): Identification of Seasonality in Time Series: A Note, Maths.Comput. Modeling, 18(6), 73 – 81.

Dozie, K.C.N and Ibebuogu, C.C (2020): Statistical Analysis of Time Series Model: Additive Case. International Journal of Research and Innovation in Applied Science (IJRIAS), 8(3) 130 – 136.

Dozie, K.C.N and Uwaezuoke. M.U (2021): Procedure for Estimation of Additive Time Series Model. International Journal of Research and Scientific Innovation (IJRSI), 8(2), 251 – 256.

Dozie, KCN and Ijomah, M.A (2020): A Comparative Study on Additive and Mixed Models in Descriptive Time Series. American Journal of Mathematical and Computer Modeling, 5(1), 12 – 17.

Gupta, S.C (2013): Fundamentals of Statistics, Himalaya Publishing House PVT LTD, Mumbai-400004.

Hamburg, M (1970): Statistical Analysis for Decision Making. Harcourt, Brace and World Inc, New York, USA.

Article DOI: 10.52589/AJMSS-16JZMGDJ
DOI URL: https://doi.org/10.52589/AJMSS-16JZMGDJ

Iwueze, I.S (2006): Time Series Analysis in the Time Domain, Unpublished Note Book, Department of Statistics, Abia State University, Uturu, Nigeria.
Iwueze, I.S and Akpanta, A.C (2009): On Applying the Bartlett Transformation Method to Time Series Data. Journal of Mathematical Sciences, 20(3), 227 – 243.
Iwueze, I.S and Nwogu, E.C (2004): Buys-Ballot Estimates for Time Series Decomposition. Global Journal of Mathematics, 3(2), 83 – 98.
Iwueze, I.S, Nwogu, E.C, Ohakwe, J and Ajaraogu, J.C (2011): Uses of Buys-Ballot Table in Time Series Analysis. Applied Mathematics, 2, 633 – 645.
Linde, P (2005): Seasonal Adjustment, Statistics, Denmark, www.dst.dk/median/konrover/13-forecasting-org/seasonal/001pdf
Nwobi, F.N and Nduka, E.C (2003): Statistical Notes and Tables for Research (2nd Edition). Alphabet Nigeria Publishers.
Nwogu, E.C, Iwueze, I.S, Dozie, K.C.N and Mbachu, H.I (2019): Choice between Mixed and Multiplicative Models in Time Series Decomposition. International Journal of Statistics and Applications, 9(5), 153 – 159.
Oladugba, A.V, Ukaegbu, E.C, Udom, A.U, Madukaije, M.S, Ugah, T.E and Sanni, S.S (2014): Principles of Applied Statistics, University of Nigeria Press Limited, Nsukka, Enugu, Nigeria.
Ross, S.M (2009): Nonparametric Hypothesis Tests: An Introduction to Probability and Statistics for Engineers and Scientists (4th Edition). Elsevier , New York, USA.

Article DOI: 10.52589/AJMSS-16JZMGDJ
DOI URL: https://doi.org/10.52589/AJMSS-16JZMGDJ