Journal of the Nigerian Association of Mathematical Physics
Volume 47 (*July, 2018 Issue*), **pp141 – 146**
© *J. of NAMP*

THE CHOICE OF OPTIMAL MODEL IN A DESCRIPTIVE TIME SERIES ANALYSIS

[1]David Adiele and [2]Kingsley Uchendu

Department of Statistics Michael Okpara University Of Agriculture Umudike

Abstract

This paper examines optimal choice in model selection among the two competing models (additive and multiplicative) commonly used in descriptive time series analysis. To investigate this, the components of time series were isolated, the choice of model using mean square error, the choice of model using coefficient of variation were all examined and a comparative analysis of domination between mean square error method and coefficient variation method was carried out using data on CO_2 measured in ppm units from 1995-2015. Other methods employed in order to forester more understanding of the study are the autocorrelation function (ACF), Partial autocorrelation function (PACF),root mean square error(RMSE),mean absolute deviation (MAD), mean absolute percentage error (MAPE). After Toro examination it is found that the method of coefficient of variation dominates the mean square error method. Therefore, this investigation shows that in order to make an optimal choice in the mist of competing models in descriptive time-series the coefficient of variation approach is the most preferred.

***Keywords*;** coefficient of variation, (MAPE), (MAD), (ACF), (PACF), (RMSE)

1.0 Introduction

In descriptive time series, a suitable model is fitted to a given time series data and the corresponding parameters are estimated using the sample observations. Time series analysis comprises methods that attempts to uncover the underlying mechanism of the nature of the series and is often useful for forecast. It enables us develop an appropriate mathematical model which captures the underlying data generating process for the series. Three common techniques for time series analysis are: Descriptive techniques, probability model techniques (ARMA) model and spectra-density analysis. In descriptive techniques, time series is commonly represented in three functional forms or model. However, the commonly used models are additive, multiplicative, and mixed model [1].

CHOICE OF APPROPRIATE MODELS

There is always the problem of which model to use in analyzing a given series using descriptive techniques. The argument is whether to use the multiplicative or additive model.

Traditionally, the time plot of the entire series is used to make appropriate choice between additive and multiplicative models. In some time series, the amplitude of both the seasonal and irregular variations do not change as the level of the trend rises or falls. In such case, an additive model is appropriate. In many time series, the amplitude of both the seasonal and irregular variation increases as trend rises, in this case a multiplicative model is usually appropriate ,however, another method of choosing an appropriate model using Buys-Ballot table in making appropriate choice of model by considering the relationship that exist between the seasonal means and seasonal standard deviation [2]. Hence, their relationship gives indication of the desired model; they found that the additive model is appropriate when the seasonal standard deviation shows no appreciable increase or decrease relative to any increase or decrease in seasonal means. On the other hand, a multiplicative model is usually appropriate when the seasonal standard deviation shows appreciable increase or decrease in the seasonal means, which implies that the choice of model according to [2] is based on seasonal standard deviation and means using Buys-Ballot table approach. Descriptive techniques which is our main focus in this work assumes that a series is a combination of four components which could be represented thus:

$$X_t = F(T_t, S_t, C_t, e_t) \qquad (1)$$

Corresponding Author**:** David A., Email: fridayd59@gmail.com, Tel: +2348039535155

Where
$X_{t=}$Original value of the series at time t
$T_{t=}$Trend component
S_t=Seasonal variation
C_t=Cyclical component
e_t=Irregular component
The representation of the model in equation (1) depends on its various functional forms. They include:
Additive model:

$$X_t = T_t + S_t + C_t + I_t \tag{2}$$

Multiplicative model:

$$X_t = T_t * S_t * C_t * I_t \tag{3}$$

Mixed or Pseudo Additive model

$$X_t = T_t * S_t * C_t + I_t \tag{4}$$

One of the greatest challenge is how to determine which of the above functional forms will be more appropriate for a giving set of series, that is what this paper intend to address in this study. Therefore, to avoid specification bias or error in the choice of an appropriate model there is a need to make an appropriate choice in descriptive time series .The two competitive models in time series are additive and multiplicative model. [3] In this work we propose coefficient of variation of the seasonal differences and seasonal quotient as a means of making an optimal choice in the selection of an appropriate model in a given set of data among the competing models. That is, this paper is billed to investigate the choice of appropriate model in descriptive time series analysis using coefficient of variation method . Hopefully, this study will be helpful in choice of appropriate functional form or model in the decomposition of time series data using descriptive technique .It will ease us the ambiguity of choosing an appropriate model using long computation as those found in literature. This work considers only descriptive time series analysis techniques and choice between additive and multiplicative model.

METHODOLOGY

Given a series X_t and assuming all the component available (Trend, Seasonal, Cyclical and Irregular),the functional form of the series X_t could be represented thus:

$$X_t = T_t + S_t + C_t + I_t \tag{5}$$

Additive Model

$$X_t = T_t * S_t * C_t * I_t \tag{6}$$

Multiplicative Model

$$X_t = T_t * S_t * C_t + I_t \tag{7}$$

Mixed Model

The additive in (5) and multiplicative in (6) are the major competitive model ,while mixed rarely occur in nature and are seldomly applied [4] . Hence, our choice will be between additive and multiplicative.

To choose the appropriate model for X_t, we adopt a coefficient of variation approach developed by [3]. The procedure is given thus:

(i) Calculate the seasonal difference of the series X_t given by

$$d_{t,i} = X_{t,i} - X_{t-1,i} \tag{8}$$

And its corresponding seasonal quotient given by

$$q_{t,i} = {X_{t,i}}\big/{X_{t-1,i}} \tag{9}$$

Where

$$d_{t,i} = i^{th} \text{ seasonal difference at time } t$$

$$q_{t,i} = i^{th} \text{ seasonal quotient at time } t$$

$$X_{t,i} = i^{th} \text{ observed series at time } t$$

Where seasonal difference is the difference ($d_{t,i}$) between a certain season of a year and same season from the year before and seasonal quotient ($q_{t,i}$) is the quotient of certain season of a year and the same season from the year before.

(ii) Calculate the coefficient of variation CV of the seasonal difference and seasonal quotient given

$$CV(d_{t,i}) = \frac{Standard\ Deviation\ of\ d_{t,i}}{Mean\ of\ d_{t,i}} \qquad (10)$$

$$CV(q_{t,i}) = \frac{Standard\ Deviation\ of\ q_{t,i}}{Mean\ of\ q_{t,i}} \qquad (11)$$

(iii Apply the following decision rule:

If $CV(q_{t,i}) > CV(d_{t,i})$ Additive model is appropriate . (12)

If $CV(q_{t,i}) \leq CV(d_{t,i})$ Multiplicative model is appropriate. (13)

Having determined the appropriate model ,we therefore estimate and isolate the components of the series. Assuming that the series is for a short period , the trend and cyclical component are merged together .Hence the model is giving by:

$$X_t = T_t + S_t + I_t \qquad (14)$$

Additive Model.

$$X_t = T_t * S_t * I_t \qquad (15)$$

Multiplicative Model.

FORECAST ACCURACY MEASURES

Let the l-step ahead error be:

$$e_t(l) = X_{t+l}(l) - \hat{X}_t(l) \qquad (16)$$

Forecast error could be evaluated to determine forecast performance. Among the two common statistics for measures are:

(a) Measures based on error

(i) Root mean square error: this is the square root of the mean square error and is given by

$$RMSE = \sqrt{\frac{1}{m}\sum_{l=1}^{m} e_l^2}$$

(ii) Mean Absolute deviation:this is the mean of absolute value of the error forecast . it is given by :

$$MAD = \frac{1}{m}\sum_{l}^{m}|e_l|$$

(b) Measures based on percentage Error

Mean Absolute percentage error given by :

$$MAPE = \left(\frac{1}{m}\sum_{i=1}^{m}\left|\frac{e_l}{X_{n+1}}\right|\right)\times 100$$

TREND ANALYSIS AND ESTIMATION

The result of application of least squares approach of trend analysis and parameter estimation to the study data (concentration of atmospheric CO2) are shown in table 1 and table 2 for linear, quadratic and exponential trend model respectively .The result is on trend parameter estimate for linear, quadratic and exponential trend model respectively .

TABLE 1: TREND ANALYSIS AND ESTIMATION

Model / **Parameter Estimate**	**Linear Model**	**Quadratic Model**	**Exponential Model**
β_0	358.918 (0.000)	360.0180 (0.000)	359.31 (0.000)
β_1	0.166 (0.000)	0.1400 (0.000)	1.00 (0.000)
β_2	-	0.00003 (0.0001)	-
R^2	96.6%	96.8%	96.7%
MAD	0.51	0.50	0.50
MAPE	1.92	1.90	1.91
RMSE	2.26	2.21	2.23

The result in table 2 indicates that both linear, quadratic and exponential trend model fit the data with the parameter estimates been significantly different from zero p<0.01. The coefficient of determination for regression line (R^2) for linear ,quadratic and exponential trend model are approximately equal to 97% for all the models .The minimum absolute deviation (MAD) , root mean square error (RMSE) and mean absolute percentage error (MAPE) for three competing trend model were found to be approximately equal again .Hence based on parsimony the linear trend model is choosen .Therefore ,subsequent discussions will be based on linear trend. The fitted trend for the data is given by

$$\widehat{T} = 358.918 + 0.166t \qquad (17)$$

CHOICE OF APPROPRIATE MODEL

The result of application of coefficient of variation approach in choosing appropriate model as discussed and shown in Equations 4, 5,.6 and 7 are shown in table 4. From table 4 ,the mean for seasonal difference and seasonal quotient are 2.000375 and 1.005272112 respectively, while the standard deviations are 0.632320116 and 0.001653783 for seasonal difference and seasonal variation respectively. The corresponding coefficient of variation for seasonal difference and seasonal quotient are 0.316100789 and 0.00164511 respectively , therefore applying the decision rule in section 3 in Equation 4 and 5 the multiplicative model is more appropriate, since the coefficient of variation for seasonal quotient is less than coefficient of variation for seasonal difference .The multiplicative model will therefore be used for subsequent decomposition and component isolation .However at the end ,a comparison will be made using minimum error method.

Table 2

Statistic	$d_{t,i}$	$q_{t,i}$
Mean	2.000375	1.005272112
Std	0.632320116	0.001653783
Cv	0.316100789	0.00164511

DIAGNOSIS CHECK OF THE RESIDUAL

The remainder when the trend , seasonal and cyclical component are isolated from the series is the residual and this residual was checked for normality ,constant mean (unit mean for multiplicative model) and autocorrelation .The plot of the residual , ACF of the residuals as well as the graphical summary of the residuals were also examined result shows that the series is adequately fitted using multiplicative model .

Table 3. The result of the application forecast using multiplicative model additive model and there forecast performance measures is below.

$t+l$	$\hat{T}_{t+l}$	$\hat{S}_{t+l}$		Forecast		Forecast Error ($\lvert e_t \rvert$)		Error Square (e_t^2)		$\lvert e_t / X_{t+1} \rvert$	
		Multiplicative	Additive	Multiplicative	Additive	Multiplicative	Additive	Multiplicative	Additive	Multiplicative	Additive
253	400.916	1.000	0.154	401.096	401.070	1.554	1.580	2.415	2.495	0.386	0.392
254	401.082	1.002	0.860	402.017	401.942	2.143	2.218	4.591	4.920	0.530	0.549
255	401.248	1.004	1.402	402.673	402.650	2.187	2.210	4.781	4.884	0.540	0.546
256	401.414	1.007	2.777	404.381	404.191	3.189	3.379	10.171	11.419	0.782	0.829
257	401.58	1.008	3.124	404.919	404.704	2.731	2.946	7.456	8.678	0.670	0.723
258	401.746	1.006	2.342	404.207	404.088	2.793	2.912	7.799	8.477	0.686	0.715
259	401.912	1.002	0.702	402.640	402.614	1.860	1.886	3.460	3.556	0.460	0.466
260	402.078	0.996	-1.576	400.396	400.502	1.844	1.738	3.401	3.019	0.458	0.432
261	402.244	0.991	-3.358	398.626	398.886	2.384	2.124	5.682	4.512	0.594	0.530
262	402.41	0.991	-3.441	398.771	398.969	2.729	2.531	7.448	6.405	0.680	0.630
263	402.576	0.994	-2.172	400.300	400.404	3.340	3.236	11.158	10.473	0.828	0.802
264	402.742	0.998	-0.815	401.895	401.927	2.655	2.623	7.049	6.881	0.656	0.648
Mean						**2.4510**	**2.449**	**6.284**	**6.3010**	**0.606**	**0.605**

TABLE 4 COMPARISON OF THE METHODS USING THE RESIDUAL AND ERROR OF THE FORECAST

	MODEL COMPARISON using the residual		FORECAST COMPARISON using the error of the forecast	
MODEL	**MULTIPLICATIVE**	**ADDITIVE**	**MULTIPLICATIVE**	**ADDITIVE**
MEAN	1.0000	0.00018	2.45066	2.54850
MEDIAN	1.0000	0.01863	2.519	2.374
VARIANCE	0.000000047	0.06738	0.303741	0.39344
MAD	0.0000063	0.20063	1.45066	2.44850
RMSE	0.003484	0.25903	1.543643	2.511966
MAPE	0.00264	0.20066	0.60591	0.805208

SUMMARY:

Often the choice of appropriate model poses a great task in descriptive time series analysis , consequently the minimum mean square error approach has always been used in choosing between these competitive models ,this approach involve fitting the two competitive models and then choose the model with the minimum mean square error .However ,an alternative to minimum mean square error approach was adopted in this work . This approach (coefficient of variation approach) was employed in choice of model for time series analysis using the descriptive technique . The coefficient of variation approach entails comparing the coefficient of variation of the seasonal difference and seasonal quotient as illustrated and the decision rule leads to the choice of appropriate model .We illustrated this using monthly record of atmospheric CO2 and the result showed that multiplicative model is appropriate for the series under study .Comparatively the minimum mean square error was also used for the choice of appropriate model and the result was in agreement with coefficient of variation approach .Multiplicative model was used for further analysis. We estimated the trend using the least squares method and the series was detrended. The estimated trend, seasonal and cyclical components were also examined, before conclusion. The remainder after the isolation of all the component which is the residual was checked for normality using Anderson test , constant mean (unit mean for multiplicative model) and autocorrelation. The time series plot of the residual showed that the residual is stationary and the ACF plot showed no significant spikes indicating that the residual are not auto correlated hence a white noise .The result of the normality test and autocorlation shows that the model adequately fits the series.

This work therefore conclude that the use of coefficient of variation approach in choosing an appropriate model in descriptive time series analysis technique is optimal. .Therefore this approach should be applied as better alternative to minimum mean square error approach which seems to be ambiguous and involve extensive calculation.

REFERENCE

[1] Chatfield C. (2004). "The Analysis of Time Series: and Introduction" 6th edition, Chapman and Hall/CRC.

[2] wueze I.S, Nwogu, E.C, Ohakwe J., Ajaraogu J.C (2004). "Use of Buys-Ballot Table in time series Analysis" Journal of Applied Mathematics.

[3] Puerto. J. and Rivera M.P (2001). "Descriptive Analysis of Time Series Applied to housing Prices in Spain. 94342-CP-l-2001-De-Comentus-C21.

[4] Steve A.E. (20014). "Time Series Analysis: Univariate and Multivariate Methods in the 21st century. Journal of Global Sciences Gueeni York