Parametric and Non Parametric Approach to Choice of Model in Descriptive Time Series

Enegesele D. & Ijomah M. A.
Department of Mathematics/Statistics,
University of Port Harcourt,
Port Harcourt, Nigeria

Iwueze I. S.
Department of Statistics,
Federal University of Technology,
Owerri, Nigeria
zubikeijomahs@yahoo.com

Abstract
In this paper, we investigated the choice between additive and multiplicative models using statistical test in time series decomposition. The Buys-Ballot procedure was adopted and two parametric and one non parametric test for constant variance were applied to the column variance of the Buys-Ballot table. The results of the illustrative examples using the monthly demand data of Abeokuta base Danico food limited showed that the parametric (Bartlett and Hartley) and non-parametric (Square Rank) were significant at 5% level indicating that the appropriate model for decomposition is the additive model. Also the application of the same test to quarterly data of the price of square meter of housing in Spain from January 1987 until October 2003 show that the respective statistic were less than their corresponding critical level at 1% significant level indicating that the appropriated model for decomposition of the series is the multiplicative model.

Keywords: *Descriptive time series, trend, seasonality, choice of model, additive model, multiplicative model*

Introduction

Choice of model plays an important role in descriptive time series as it help to facilitate forecasting. Two patterns that may be presented in the identification of time series data are trend and seasonality and the two competing models are the additive and multiplicative models (Iwueze and Nwogu, 2014). Descriptive time series is the separation of the observe time series into four components represented by the trend (T_t), seasonal (S_t), cyclical (C_t) and the irregular (e_t) components. One of the advantages for decomposition in descriptive time series is to estimate seasonal effect that can be used to create and present seasonally adjusted values.

Decomposition models are typically additive or multiplicative, but can also take other forms such as pseudo-additive. The cyclical component is embedded in the trend for short series (Chatfield, 2004). These models are;

Additive Model:

$$X_t = M_t + S_t + e_t \qquad 1$$

Multiplicative Model:

$$X_t = M_t \times S_t \times e_t \qquad 2$$

Pseudo-Additive:

$$X_t = M_t \times S_t + e_t \qquad 3$$

where M_t is the trend-cycle component; S_t is the seasonal component and e_t is the irregular component. For the additive model (1), it is assumed that the error component e_t is the Gaussian white noise $N(0, \sigma_1^2)$ and the sum of the seasonal component over a complete period is zero $\left(\sum_{j=0}^{s} S_j = 0\right)$. While for the multiplicative model (2), e_t is the Gaussian white noise $N(1, \sigma_2^2)$ and the sum of the seasonal component over a complete period is s $\left(\sum_{j=0}^{s} S_j = s\right)$

An important part of the analysis of descriptive time series is the selection of a suitable model for decomposition. The graphical method and non-graphical method has been proposed in the literature to aid the choice of model between additive and multiplicative models. Brockwell and Davis (2002), use the time plot of the entire series to choose a particular model for decomposition. The multiplicative model was adopted when the magnitude of the seasonal pattern in the data depends on the magnitude of the series. In other words, the magnitude of the seasonal pattern increases as the data value increases and decreases as the data value decreases. The additive model was adopted when the magnitude of the seasonal pattern does not change as the series goes up and down. Chatfield (2004) noted that if the seasonal variation stays roughly the same size regardless of the mean level, then it is additive but if it increases in size in direct proportion to the mean level, the appropriate model for decomposition is the multiplicative models. However, there are situations when such plots are not easy to interpret and as such choosing between additive and multiplicative model becomes difficult. Instead of using the seasonal pattern as shown by the time series plot of the entire series, Iwueze et al (2011) used the relationship between the plot of the seasonal means and seasonal standard deviation derived from the Buys-Ballot Table (see Table 1) to choose between additive and multiplicative models.

The aim of this paper is to provide a test that will aid in the choice between additive and multiplicative models in time series decomposition. The rational for this paper is that most existing methods are subjective and tedious. Hence, the need for simpler method and statistical to ascertain the need to choose between both models

Table 1.0: Buys-Ballot Table

Period(i)	Seasons								
	1	2	…	j	…	s	$T_{i.}$	$\bar{X}_{i.}$	$\hat{\sigma}_{i.}$
1	X_1	X_2	…	X_j	…	X_s	$T_{1.}$	$\bar{X}_{1.}$	$\hat{\sigma}_1$
2	X_{s+1}	X_{s+2}	…	X_{s+j}	…	X_{2s}	$T_{2.}$	$\bar{X}_{2.}$	$\hat{\sigma}_2$
3	X_{2s+1}	X_{2s+2}	…	X_{2s+j}	…	X_{3s}	$T_{3.}$	$\bar{X}_{3.}$	$\hat{\sigma}_3$
…	…	…	…	…	…	…	…	…	…
i	$X_{(i-1)s+1}$	$X_{(i-1)s+2}$	…	$X_{(i-1)s+j}$	…	$X_{(i-1)s+s}$	$T_{i.}$	$\bar{X}_{i.}$	$\hat{\sigma}_{i.}$
…	…	…	…	…	…	…	…	…	…
m	$X_{(m-1)s+1}$	$X_{(m-1)s+2}$	…	$X_{(m-1)s+j}$	…	X_{ms}	$T_{m.}$	$\bar{X}_{m.}$	$\hat{\sigma}_{m.}$
$T_{.j}$	$T_{.1}$	$T_{.2}$	…	$T_{.j}$	…	$T_{.s}$	$T_{..}$	-	-

$\bar{X}_{.j}$	$\bar{X}_{.1}$	$\bar{X}_{.2}$	$\cdots$	$\bar{X}_{.j}$	$\cdots$	$\bar{X}_{.s}$	-	$\bar{X}_{..}$	-
$\hat{\sigma}_{.j}$	$\hat{\sigma}_{.1}$	$\hat{\sigma}_{.2}$	$\cdots$	$\hat{\sigma}_{.j}$	$\cdots$	$\hat{\sigma}_{.s}$	-	-	$\hat{\sigma}_{..}$

Source: Iwueze and Nwogu (2014)

where $X_{ij} = X_{(i-1)s+j}$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, s$ is the series, m is the number of periods/years, s is the periodicity, and $n = ms$ is the overall number of observation/sample size.

$T_{.j}$ = Total for jth season, $\bar{X}_{.j}$ = Average of jth season

$\hat{\sigma}_{.j}$ = Standard deviation for jth season.

Column averages and standard deviation is defined as follows:

$$\bar{X}_{.j} = \frac{T_{.j}}{m} = \frac{1}{m}\sum_{i=1}^{m} X_{(i-1)s+j}, \quad j = 1, 2, \ldots, s$$

$$\hat{\sigma}_{.j} = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}\left(X_{(i-1)s+j} - \bar{X}_{.j}\right)^2}, \qquad j = 1, 2, \ldots, s$$

2 Methodology

This paper will adopt the method of comparison of seasonal/column variances of the Buys-Ballot Table proposed by Iwueze and Nwogu (2014) with the inclusion of error component in the derivation of column mean and variances as this was ignored in Iwueze and Nwogu (2014) derivation of the column mean and variance. With the use of the Buys-Ballot Table, the problem of choice between additive and multiplicative models reduces to test for constant variance in the column variances. Table 2 below shows the summary of the result of the derivation of the column variance and mean when trend-cycle component is linear.

Table 2: Row, Column Totals, Averages and Variances of Buys-Ballot for Additive and Multiplicative Models

	Linear trend-cycle component: $M_t = a + bt$, $t = 1, 2, \ldots n$ $\quad n = sm$	
	Additive Model	**Multiplicative Model**
$\bar{X}_{.j}$	$a + \frac{bs}{2}(m-1) + bj + S_j + \bar{e}_{.j}$	$\left[a\bar{e}_{.j} + \frac{bs}{m}\sum_{i=1}^{m} i\, e_{ij} - bs\,\bar{e}_{,j} + bj\,\bar{e}_{.j}\right] * S_j$
$\hat{\sigma}^2_{.j.}$	$b^2\left[\frac{n(n+s)}{12}\right] + \sigma_1^2$	$\left\{\frac{b^2}{12}\left(n^2 - s^2\right) + \left(\alpha + bj\right)^2\right\}\sigma_2^2 S_j^2$

Footnote: $\alpha = a + \frac{b}{2}(n-s)$, σ_1^2 = Error Variance (Additive model),

σ_2^2 = Error Variance (Multiplicative model)

The column variance for the additive model is a constant, while that of the multiplicative model contains seasonal effect.

For the linear trend curve, the column variances are

$$\hat{\sigma}^2_{.j} = \begin{cases} b^2\left(\frac{n(n+s)}{12}\right) + \sigma_1^2, \text{ for additive model} \\ \left[b^2\left(\frac{n^2-s^2}{12}\right) + (\alpha+\beta j)^2\right] * \sigma_2^2 * S_j^2, \text{ for multiplicative model} \end{cases} \quad (4)$$

$$H_0 : \sigma^2_{.1} = \sigma^2_{.2} = \ldots = \sigma^2_{s}$$

against the alternative;

$$H_1 : \sigma^2_{.j} \neq \sigma^2_{.j'}, \text{ for at least one } j \neq j'$$

The parametric tests for constant variances are listed in (a-b) while the nonparametric tests are shown in section (c-d).

a. Bartlett's Test

This test statistic is used to compute homogeneity of variance and the sampling distribution is approximated by the chi-square distribution with k-1 degrees of freedom (Bartlett, 1937; Montgomery, 1997). The test statistic is

$$B = 2.3026 * G \quad (5)$$

where

$$G = \frac{(N-k)\log S_p^{\,2} - \sum_{i=1}^{k}(n_i - 1)\log S_i^2}{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\left(\frac{1}{n_i - 1}\right) - \left(\frac{1}{N-k}\right)\right)} \quad (6)$$

$$S_p^{\,2} = \frac{\sum_{i=1}^{k}(n_i - 1)S_i^{\,2}}{N-k} \quad (7)$$

Using the Buys –Ballot technique where n_i = m, k = s, N = ms, Equation (6) can be written as

$$G = \frac{3s(m-1)^2\left[s\log\left(\frac{\sum_{j=1}^{s} S_j^2}{s}\right) - \sum_{j=1}^{s}\log S_j^2\right]}{s(3m-2)+1} \quad (8)$$

where

$S_j^{\,2}$ is the column sample variances from the jth sample, 2.3026 is a constant value, m is the sample size of jth group, s is the number of groups. The null hypothesis is rejected if the statistic (5) is greater than the tabulated Chi-Square value or not rejected otherwise.

b. Hartley's Test

The Hartley's F_{max} utilizes only the highest and lowest variances of the groups. The statistic makes do with the Hartley F_{max} table with s and m-1 degree of freedom. The test is defined as

$$F = \frac{\max(S_j^2)}{\min(S_j^2)}$$

(9)

where,

S_j^2 = column variance of the jth group, $j = 1, 2, \ldots, s$

$S_j^2(\max) = \max(S_1^2, S_2^2, \ldots, S_s^2)$, $S_j^2(\min) = \min(S_1^2, S_2^2, \ldots, S_s^2)$

s = number of groups, m = sample size of jth group

The null hypothesis is rejected if the calculated statistic (9) is greater than the F_{max} table critical value with s and m-1 degree of freedom, (Hartley, 1950).

Square Rank Test

The Square rank test also known as Conover test is a non-parametric test for either the two sample case or the case of k group, (Conover, 1999).

$$X_{ij} = \left|Y_{ij} - \overline{Y}_{.j.}\right|$$

The test statistic is define as

$$T = \frac{1}{P^2}\left(\sum_{i=1}^{k}\frac{s_i^{\,2}}{n_i} - \frac{\left(\sum R_{ij}^2\right)^2}{N}\right)$$

(10)

where,

s_i is sum of square ranks in subsample i, n_i is number of observations in subsample I Y_{ij} is the original series, $\overline{Y}_{.j}$ is the column mean and X_{ij} is the new ranked series

$$P^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N} R_i^{\,4} - \frac{\left(\sum R_{ij}^2\right)^2}{N}\right) \tag{11}$$

R_i = rank of observation i

If there are no ties in the Rank like the Buys-Ballot Table, then the test statistic is given by

$$T = \frac{5\left[36\sum_{j=1}^{s} S_j^2 - mn(n+1)^2(2n+1)^2\right]}{mn(n+1)(2n+1)(8n+11)} \tag{12}$$

Where n is the total sample size of the series, m is the sample size of the jth subgroup and S_j is the sum of square rank of the jth group. If the assumptions are met, the distribution of this test statistic (12) follows approximately the chi-square distribution with k-1 degree of freedom.

3. Empirical Examples

The Results of the monthly demand of Abeokuta base Danico foods limited from (2002-2009) for the computed column variances of the Buys-Ballot table for the parametric and nonparametric test for constant variance are shown in Table 3

Table 3: Test Statistic for parametric and non-parametric test of Abeokuta Danico food limited

Level of Significance	Test Statistics		
	Bartlett	Hartley	Square Rank
Calculated Statistic	11.09	5.791	6.78
Critical value at 5%	19.675	15.80	19.675

$$\text{Bartlett}, \text{Square Rank}\ (\chi^2_{11}(0.05)) = 19.675 \quad , \text{Hartley}(F_{\max(12,7)}) = 15.8$$

Remark: The results in Table 3 show that all calculated test statistic lies within the respective critical values indicating that the null hypothesis is accepted and the model for decomposition of the monthly demand of Abeokuta base Danico foods limited is the additive model.

Table 4: Test Statistic for parametric and non-parametric test of Average price of square meter of housing in Spain

Level of Significance	Test Statistics		
	Bartlett	Hartley	Square Rank
Calculated Statistic	10.09	15.79	16.78
Critical value at 5%	7.81	4.98	7.81

$$\text{Bartlett}, \text{Square Rank}\ (\chi^2_{3}(0.05)) = 7.81 \quad , \text{Hartley}(F_{\max(4,17)}) = 4.98$$

Remark: The results in Table 4 show that all calculated test statistic were greater than their respective critical values indicating that the null hypothesis is accepted and the model for decomposition of the price of housing in Spain is the multiplicative

4. Conclusion

It is important to check the features of the data before appropriate model is chosen for decomposition. The selection of an adequate model is very important as it shows the underlying structure of the series because the fitted model will be used for future forecasting. The illustrative examples shows that the additive model should be used to decompose the monthly sales of Abeokuta Danico food limited and the multiplicative model should be used to decompose the average price housing of Spain. the parametric (Bartlett and Hartley) and the Square Rank non parametric test can be used to justify the choice of model using the Buys-Ballot procedure.

REFRENCES

Bartlett, M.S. (1937). Properties of sufficiency of statistical tests, Proc. Roy. Soc, Basingstoke. 160, 268-287.

Brockwell P.J and Davis R.A (2002). "Introduction to Time Series and Forecasting 2nd Edition, Springer New York.

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

Conover, W.J (1999), Pratical Nonparametric Statistics. 3rd edition Wiley, New York

Dagum, E. B. (1983), "The X-12 Seasonal Adjustment Method", Technical Report 12-564E, Statistics Canada.

Findley D.F., Monsell B.C., Bell W.R., Otto M.C., and Chen B.C (1998). New Capabilities and Methods of the X-12 ARIMA Seasonal Adjustment Program. Journal of Business & Economic Statistics, 16(2), 127-152.

Hartley, H.O., (1950). The maximum F-ratio as a short-cut test of heterogeneity of variance, Biometrika, 37, 308-312.

Iwueze I.S, Nwogu E.C, Ohakwe J. and Ajaraogu J.C (2011), "Uses of the Buys-Ballot Table in Time Series Analysis", Applied Mathematics, pp 633-645.

Iwueze I.S and Nwogu E.C, (2014) "Framework For Choice of Model and Detection of Seasonal Effect in Time Series" Far East Journal of Theoretical Statistics, Vol. 48 No. 1, pp 45-66.

Justo P. and Rivera M.A.,(2010) "Descriptive analysis of Time Series applied to housing prices in Spain", Management Mathematics for European Schools, 94342-CP-1-2001-DE-COMENIUS-C21

Montgomery, D.C. (1997). Design and analysis of experiments. 4th ed. John Wiley and Sons New York.

Buys-Ballot Table for Monthly Demand of Abeokuta Base Danico Foods Limited

Year	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
2002	11200	14900	25102	16210	14592	10000	9920	11200	12401	22105	27820	25780
2003	9000	14900	21000	25600	29200	25200	15600	15600	16000	21200	29100	27600
2004	12000	22900	28000	32600	41600	30200	23600	19600	26000	31200	41700	36400
2005	21600	22600	17700	27700	32250	22000	15000	15000	24000	30000	40000	42121
2006	28500	29250	30500	31350	30230	29990	20500	21300	29300	31320	30200	50200
2007	27321	29440	30120	32450	35000	25420	26770	25827	30720	32350	40102	52720
2008	34060	35610	36111	30109	27105	26850	25910	25103	25965	45700	51100	57102
2009	57102	27805	35965	51111	34905	36724	55103	38344	45251	41205	52750	50320

Buys-Ballot Table of Housing Price in Spain

Year	Q1	Q2	Q3	Q4
1987	289.89	308.64	324.99	345.55
1988	369.13	389.79	404.39	423.12
1989	456.58	480.17	502.72	516.43
1990	550.4	559.73	570.77	580.6
1991	613.42	637.9	652.8	681.23
1992	650.49	635.7	633.78	630.72
1993	625.44	634.83	639.69	640.61
1994	634.72	636.69	644.36	642.63
1995	652.92	661.06	665.46	667.47
1996	669.98	674.79	675.18	676.45
1997	677.74	683.06	686.68	691.78
1998	694.34	709.66	723.95	738.58
1999	755.21	780.25	803.89	829.81
2000	857.25	891.76	926.36	953.42
2001	994.5	1030.77	1065.78	1096.57
2002	1148.23	1193.66	1254.09	1287.73
2003	1349.11	1402.57	1450.6	1467.34