Journal of Modern Applied Statistical Methods

Volume 15 | Issue 2 Article 24

11-1-2016

Some Tests for Seasonality in Time Series Data

Eleazar Chukwunenye Nwogu
Federal University of Technology, Owerri, nwoguec@yahoo.com

Iheanyi Sylvester Iwueze
Federal University of Technology, Owerri, isiwueze@yahoo.com

Valentine Uchenna Nlebedim
Federal University of Technology, Owerri, uchenlebedim@yahoo.com

Journal of Modern Applied Statistical Methods
November 2016, Vol. 15, No. 2, 382-399.
doi: 10.22237/jmasm/1478002920

ISSN 1538 − 9472

Some Tests for Seasonality in Time Series Data

Eleazar Chukwunenye Nwogu
Federal University of Technology, Owerri
Owerri, Imo State, Nigeria

Iheanyi Sylvester Iwueze
Federal University of Technology, Owerri
Owerri, Imo State, Nigeria

Valentine Uchenna Nlebedim
Federal University of Technology, Owerri
Owerri, Imo State, Nigeria

This paper presents some tests for seasonality in a time series data which considers the model structure and the nature of trending curve. The tests were applied to the row variances of the Buys Ballot table. The Student *t*-test and Wilcoxon Signed-Ranks test have been recommended for detection of seasonality.

Keywords: Model structure, trending curves and seasonal indices, Buys-Ballot table, row variances, paired sample data

Introduction

Decision making is paramount to any organization. Making a good decision depends largely on predicting future events and conditions. The basic assumption made when forecasting is that there is always an underlying pattern which describes the event and conditions, and that it repeats in the future. A time series is a chronological sequence of observations on a particular variable. Hence, there are two major goals of time series analysis: (1) identifying the nature of the phenomenon represented by the sequence of observations; and (2) forecasting (predicting future values of the time series variable). Identification of the pattern and choice of model in time series data is critical to facilitate forecasting. Thus, both of these goals of time series analysis require that the pattern of observed time series data is identified and described. Two patterns that may be present are trend and seasonality. In order to understand the effectiveness of identification of patterns of observed time series data, it is important to first identify what a time series

Dr. Nwogu is an Associate Professor and Head of the Department of Statistics. Email them at: nwoguec@yahoo.com.

consists of. In time series analysis, it is assumed that the data consists of a systematic pattern (usually a set of identifiable components) and random noise (error). Most time series patterns can be described in terms of four basic classes of components: The systematic pattern includes the trend (denoted as T_t), seasonal (denoted as S_t), and cyclical (denoted as C_t) components. The irregular component is denoted as I_t or e_t, where t stands for the particular point in time. These four classes of time series components may or may not coexist in real-life data.

The two main goals of a time series analysis are better achieved if the correct model is used. The specific functional relationship among these components can assume different forms. However, the possibilities are that they are combined in an additive (additive seasonality) or a multiplicative (multiplicative seasonality) fashion, but can also take other forms such as pseudo-additive/mixed (combining the elements of both the additive and multiplicative models) model.

The additive model (when trend, seasonal and cyclical components are additively combined) is given as:

$$X_t = T_t + S_t + C_t + I_t, \quad t = 1, 2, \ldots, n \tag{1}$$

The multiplicative model (when trend, seasonal and cyclical components are multiplicatively combined) is given as:

$$X_t = T_t \times S_t \times C_t \times I_t, \quad t = 1, 2, \ldots, n \tag{2}$$

and the Pseudo-Additive/Mixed Model (combining the elements of both the additive and multiplicative models) is given as:

$$X_t = T_t \times S_t \times C_t + I_t, \quad t = 1, 2, \ldots, n \tag{3}$$

Cyclical variation refers to the long term oscillation or swings about the trend, and only long period sets of data will show cyclical fluctuation of any appreciable magnitude. If short periods of time are involved (which is true of all examples in this study), the cyclical component is superimposed into the trend (Chatfield, 2004) and then the trend-cycle component is denoted by M_t. In this case, (1), (2), and (3) may, respectively, be written as:

$$X_t = M_t + S_t + I_t, \quad t = 1, 2, \ldots, n \tag{4}$$

$$X_t = M_t \times S_t \times I_t, \quad t = 1, 2, \ldots, n \tag{5}$$

$$X_t = M_t \times S_t + I_t, \quad t = 1, 2, \ldots, n \tag{6}$$

The pseudo-additive model is used when the original time series contains very small or zero values. However, this work will discuss only the additive and multiplicative models.

As long as the trend is monotonous (consistently increasing or decreasing), the identification of the trend component is not very difficult. Tests for trend are given in Kendall and Ord (1990). The cyclical component exhibits variation at periods that may be fixed or not fixed, but which are predictable. Many time series exhibit a variation which repeats itself in systematic intervals over time and this behavior is known as seasonal dependency (seasonality). The seasonal component, S_t, is associated with the property that $S_{(i-j)s+j} = S_j$, $i = 1, 2, \ldots$. The difference between a cyclical and a seasonal component is that the latter occurs at regular (seasonal) intervals, although cyclical factors have usually a longer duration that varies from cycle to cycle.

In some time series data, the presence of a seasonal effect in a series is quite obvious and the seasonal periods are easy to find (e.g., 4 for quarterly data, 12 for monthly data, etc.). Seasonality can be visually identified in the series as a pattern that repeats every k elements. The following graphical techniques can be used to detect seasonality: (1) a run sequence plot (Chambers, Cleveland, Kleiner, & Tukey, 1983); (2) a seasonal sub-series plot (Cleveland, 1993); (3) multiple box plots (Chambers et al., 1983); and (4) the autocorrelation plot (Box, Jenkins, & Reinsel, 1994). Both the seasonal subseries plot and the box plot assume that the seasonal periods are known. If there is significant seasonality, the autocorrelation plot should show spikes at multiples of lags equal to the period, the seasonal lag (Box et al., 1994). For quarterly data, we would expect to see significant spikes at lag 4, 8, 12, 16, and so on. Iwueze, Nwogu, Ohakwe, and Ajaraogu (2011) pointed out that seasonality in time series can be identified from the time plot of the entire series by regularly spaced peaks and troughs which have a consistent direction and approximately the same magnitude every period/year, relative to the trend.

In some cases the presence of a seasonal effect in a series is not quite obvious and, therefore, testing is required in order to confirm the presence of the seasonal effect in a series. Davey and Flores (1993) proposed a method which adds statistical tests of seasonal indexes for the multiplicative model that helps identify seasonality with greater confidence. Tests for seasonality are also given in Kendall and Ord

(1990). Chatfield (2004) suggested the use of the Buys Ballot table for inspecting time series data for the presence of trend and seasonal effects. Fomby (2008) presented various graphs suggested by the Buys Ballot table for inspecting time series data for the presence of seasonal effects. Fomby (2010), in his study of Stable Seasonal Pattern (SSP) models, gave an adaptation of Friedman's two-way analysis of variance by ranks test for seasonality in time series data. Several statistics have also been proposed to test for seasonality. They can be broken down into three groups: the Chi-Square (χ^2) Goodness-of-Fit test and the Kolmogorov-Smirnov type statistic, the Harmonic analyses based on the Edwards' type statistic (Edwards, 1961), and the Nonparametric Tests.

The χ^2 goodness-of-fit test is relatively popular for detecting seasonality because of its simple mathematical theory, which makes it easy to calculate and understand (Hakko, 2000). The test is on whether the empirical data can be a sample of a certain distribution with sampling error as the only source of variability (McLaren, Legler, & Brittenham, 1994). This test requires a sample from a population with an unknown distribution function $F(x)$ and a certain theoretical distribution function $F_0(x)$. Although there is no restriction on the underlying distribution, usually the hypothetical distribution is a uniform distribution.

For seasonality studies, the frequency O_i, $i = 1, 2,\ldots, k$ is the observed value at the i^{th} season, while the frequency E_i, $i = 1, 2,\ldots, k$ is the expected cell frequency at the i^{th} season. Under the null hypothesis that there is no seasonal effect (i.e., $F_0(x)$ is a uniform distribution), then $E_1 = E_2 =\ldots= E_k$ and the statistic

$$T = \sum_{i=1}^{k} \left[\frac{\left(O_i - E_i\right)^2}{E_i} \right] \tag{7}$$

is asymptotically distributed as χ^2 with $v = k - 1$ degrees of freedom (Horn, 1977). The χ^2 goodness-of-fit test for seasonality has been recently used for the analysis of seasonality in suicide, myocardial infarction, diarrhoea, pneumonia, and overall mortality (Flisher, Parry, Bradshaw, & Juritz, 1996; Herring & Hoppa, 1997; Rihmer, Rutz, Pihlgren, & Pestiality, 1998; Sheth, Nair, Muller, & Yusuf, 1999; Underwood, 1991; Villa, Guisecafré, Martinez, & Muñoz, 1999).

In his article, Edwards (1961) explicitly mentions the possibility to estimate cyclic trends by considering the ranking order of the events which are above or below the median number. This idea was used by Hewitt, Milner, Csima, and Pakula (1971) but did not use a binary indicator as suggested by Edwards (1961), instead using all of the ranking information. Rogerson (1996) made an attempt to generalize this test by relaxing the relatively strict assumption of Hewitt et al. (1971) that seasonality is only present if a six-month peak period is followed by a six-month trough period. Rogerson (1996) allowed that the peak period can also last three, four, or five months. Rau (2005) further relaxed these assumptions and allows total flexibility for the basic time duration as well as for the length of the peak period.

The Kolmogorov-Smirnov goodness-of-fit test (KS-Test) is comparable to the χ^2 goodness-of-fit test because both approaches are designed to test if a sample drawn from a population fits a specified distribution. However, the KS-Test does not compare observed and expected frequencies at each season, but rather the cumulative distribution functions between the ordered observed and expected values (Rau, 2005).

For seasonality studies, if $\mathrm{F}_N(t)$, $t = 1, 2,\ldots, s$, is the empirical distribution function based on the observed frequencies at each season and $\mathrm{F}_0(t)$ is the corresponding distribution function under the null hypothesis that there is no seasonal effect, the test-statistic used is:

$$T = V_N \sqrt{N} = \left[\max_{1 \le t \le 12} \left(\mathrm{F}_N(t) - \mathrm{F}_0(t) \right) + \left| \max_{1 \le t \le 12} \left(\mathrm{F}_N(t) - \mathrm{F}_0(t) \right) \right| \right] \tag{8}$$

The statistic T does not follow any of the known distributions (e.g. χ^2, $\mathrm{N}(\mu, \sigma^2)$, etc.). The distribution of T was determined empirically by Freedman (1979) using Monte Carlo simulations and tabulated in Freedman's article. Freedman's modified KS-Type Test has been used for the study of seasonality (Verdoux, Takei, Cassou de Saint-Mathurin, & Bourgeois, 1997).

In all these tests for the presence of seasonal effect in a time series data, the model structure (i.e. whether Additive or Multiplicative models) and nature of the trending curve (Linear, Quadratic, Exponential, etc.) were not taken into consideration. However, Iwueze and Nwogu (2014) have shown that, for precise detection of presence of seasonal effect in a series the model, structure and trending curves are necessary. Some of the questions that come to mind are: "How does the model structure affect the detection of presence of seasonal effect in a time series data?"; "How does the nature of the trending curves affect the test for presence of seasonal effect in a series?" These and other related questions are what this study intends to address.

Therefore, the ultimate objective of this study is to develop tests for seasonality in a series which take into account the nature of the model structure and trending curves for precise detection of the presence of seasonal effect in a series where it exists. The specific objectives are to:

(a) Review the Buys Ballot procedure for selected trending curves,

(b) Construct test(s) for the detection of presence of seasonal effect in a series using the row, column, overall means and variances of the Buys-Ballot table, and

(c) Assess the performance of the developed test statistics in detection of the presence of seasonal effects in a series using empirical examples.

Based on the results, recommendations are made.

The rationale for this study is to fill the gap in the existing tests for seasonality by providing analyst with objective test for the detection of the presence of seasonal effect in a series when it exists.

The Buys-Ballot procedure was developed by Iwueze and Nwogu (2004) for short period data in which trend and cyclical components are jointly estimated; the tests developed in this study are based on this assumption. In their results, on the basis of which the proposition for choice of appropriate model was made, Iwueze and Nwogu (2014) showed that, for the selected trending curves, the column variances depend only on the trend parameters for the additive model and on both trend parameters and seasonal indices for the multiplicative model. Therefore, if the seasonal/column variances are functions of the trend parameters, only then is Additive the appropriate model. However, if the seasonal/column variances are functions of both the trend parameters and seasonal indices, then the appropriate model is Multiplicative. It is the presence of the seasonal effect in the seasonal/column variances that makes the model multiplicative. In other words, once the seasonal/column variances indicate that the appropriate model is Multiplicative, it also indicates that the series contains seasonal effects. Therefore, in this study, tests for detection of the presence of seasonal effect in a time series data are developed for the additive model only.

For the additive model and all trending curves studied, the row variances contain both the trending parameters and the seasonal component, while the column variances do not contain the seasonal component. Therefore, the parameters of the trending curves have been varied in order to see their effects on the powers of the tests. In particular, the slope parameter b of the linear trend has been assigned the values b = 0.02, 0.20, and 2.00 to check its effect on the test(s).

Furthermore, the power of the tests will be measured by considering the percentages of the total simulations in which the test correctly detected the presence of seasonal effect when it exists.

Methodology

The summary of the row variances for the additive model derived by Iwueze and Nwogu (2014) are shown in Table 1 for the selected trending curves, with

$$C_1 = \sum_{j=1}^{s} jS_j, \quad C_2 = \sum_{j=1}^{s} j^2 S_j$$

Tests for seasonality in the Additive model are constructed by applying the tests for the matched pairs of data to the row variances of the Buys-Ballot table.

Table 1. Summary of row variances of the Buys-Ballot table for the additive model and the selected trending curves

Trending Curve	Row Variance (σ_i^2)
Linear: $a + bt$	$b^2\left(\frac{s(s+1)}{12}\right)+\left(\frac{2b}{s-1}\right)\sum_{j=1}^{s} jS_j + \frac{1}{s-1}\sum_{j=1}^{s} S_j^2$
Quadratic: $a + bt + ct^2$	$\frac{s(s+1)}{180}\left\{(2s-1)(8s-11)c^2 - 30(s-1)bc + 15b^2\right\} + \frac{1}{s-1}\left\{\sum_{j=1}^{s} S_j^2 + 2\left[b-2cs\right]C_1 + 2cC_2\right\} + \left\{\frac{s^2(s+1)}{3}\left[bc - c^2(s-1) + \frac{4csC_1}{s-1}\right]\right\}i + \left[\frac{s^3(s+1)c^2}{3}\right]i$
Exponential: be^{ct}	$b^2 e^{2c[(i-1)s+1]}\left[\left(\frac{1-e^{2cs}}{1-e^{2c}}\right) - \frac{1}{s}\left(\frac{1-e^{cs}}{1-e^{c}}\right)\right] + \sum_{j=1}^{s} S_j^2 + 2be^{c(i-1)s}\sum_{j=1}^{s} e^{cj} S_j$

Source: Iwueze and Nwogu (2014).

For the matched pairs of data, (U_i, V_i), $i = 1, 2, \ldots, n$, define

$$d_i = U_i - V_i \tag{9}$$

where, for the i^{th} observation unit, U_i and V_i denote measures on two characteristics. For the variable d_i, any of these test statistics: (a) the Student's t-distribution; (b) the sign test; or (c) the Wilcoxon Signed-Rank; can be used to test the null hypothesis that the two characteristics have the same mean or median.

Student *t*-Distribution

The statistic

$$t_c = \frac{\bar{d} - d_0}{S_d / \sqrt{n}} \tag{10}$$

is known to follow the Student's t-distribution with $n - 1$ degrees of freedom under the null hypothesis that the two characteristics have the same mean or median (or are drawn from a population with the same median), where

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i, \quad S_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}$$

and d_0 (usually assumed zero under H_0: $d = d_0$) is the value of the man or median of the deviations under the null hypothesis. The null hypothesis (H_0) is rejected at α level of significance if $|t_c| > t_{1-\alpha/2}$, where $t_{1-\alpha/2}$ is the $100(1 - \alpha)$ percentile of the Student's t-distribution with $n - 1$ degrees of freedom.

Sign Test

The test statistic for the sign test is k, the smaller of the number of positive signs n_+ and the number of negative signs n_-. That is

$$k = \min\left(n_+, n_-\right) \tag{11}$$

Under the null hypothesis that the medians of the two variates are equal, the random variable k follows the binomial distribution with parameters n and $p = 0.5$. That is, the number of positive signs (n_+) and negative (n_-) signs are expected to be equal.

For smaller sample sizes (i.e., $0 < n < 25$), the observed value of k is compared with the critical value (k_α) and the null hypothesis (H_0) is rejected at α level of significance if $k < k_\alpha$, where k_α is computed from the binomial probability function as

$$\mathrm{F}\left(k_\alpha; n, p\right) = \Pr\left(X \le k_\alpha\right) = \sum_{x=1}^{\lfloor k_\alpha \rfloor} \frac{n!}{x!\left(n-x\right)!} p^x \left(1-p\right)^{n-x} \le \alpha \tag{12}$$

where $p = 0.5$ and $\lfloor k_\alpha \rfloor$ is the "floor" under k_α, i.e., the greatest integer less than or equal to k_α (Corder & Foreman, 2014).

For larger sample sizes (i.e., $n \ge 25$), Corder and Foreman (2014) recommended the use of z_c, given as

$$z_c = \frac{\left(k' - 0.5\right) - 0.5n}{0.5\sqrt{n}} \tag{13}$$

where $k' = \max(n_+, n_-)$. This approximately follows the standard normal distribution under the null hypothesis. The null hypothesis (H_0) is rejected at α level of significance if $k' > z_\alpha$ and accepted otherwise.

Wilcoxon Signed-Ranks Test

For small sample sizes (i.e., $n \le 30$), the Wilcoxon Signed-Ranks test statistic is given by

$$T_c = \mathrm{Min}\left\{\sum_{i=1}^{n_+} R_{d_i^+}, \sum_{i=1}^{n_-} \left|R_{d_i^-}\right|\right\} \tag{14}$$

where $\sum_{i=1}^{n_+} R_{d_i^+}$ is the sum of the positive ranks of non-zero differences and $\left|\sum_{i=1}^{n_-} R_{d_i^-}\right|$ is the sum of the absolute values of the negative ranks of non-zero differences. If the null hypothesis (H_0) is true, these sums are expected to be equal.

For large sample sizes (i.e., $n > 30$), the Wilcoxon Signed-Ranks test statistic is given by Corder and Foreman (2014) as

$$z_c = \frac{T_c - \mu_T}{\sigma_T} \tag{15}$$

where n is the number of matched pairs of data for which their differences is not zero and

$$\mu_T = \frac{n(n+1)}{4}, \quad \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

i.e., the mean and standard deviation of T_c, respectively, under the null hypothesis. The null hypothesis (H_0) is rejected at α level of significance if $z_c < z_\alpha$ and accepted otherwise.

When the usual parametric assumptions (difference scores are normally and identically distributed in the population from which the sample was drawn and that they are measured on at least an interval scale) are met, the Student's *t*-distribution is used. The sign test and the Wilcoxon signed-ranks test are used when the usual assumptions of parametric tests are not met. It is important to note that the sign test and Wilcoxon signed-ranks test require only that the distribution of the study data be symmetric.

For detection of the presence of seasonal effect in a time series data, we let U_i denote the row variance in the presence of the seasonal effect and V_i denote row variance in the absence of the seasonal effect.

For example:

(a) For the linear trend-cycle component, in the presence of seasonal effect, the row variance is

$$U_i(L) = \hat{\sigma}_i^2(L) = b^2\left(\frac{s(s+1)}{12}\right) + \left(\frac{2b}{s-1}\right)\sum_{j=1}^{s} jS_j + \frac{1}{s-1}\sum_{j=1}^{s} S_j^2 \tag{16}$$

When there is no seasonal effect, $S_j = 0\ \forall j = 1, 2, \ldots, s$, and so

$$V_i(L) = b^2\left(\frac{s(s+1)}{12}\right) \tag{17}$$

and

$$d_i(L) = U_i(L) - V_i(L) = \left(\frac{2b}{s-1}\right)\sum_{j=1}^{s} jS_j + \frac{1}{s-1}\sum_{j=1}^{s} S_j^2 \tag{18}$$

which is zero under null hypothesis (H_0: $S_j = 0$).

(b) For the Quadratic trend-cycle component, in the presence of seasonal effect, the row variance is

$$U_i(Q) = \hat{\sigma}_i^2(Q)$$
$$= \left\{ \begin{array}{l} \dfrac{s(s+1)}{180}\left\{(2s-1)(8s-11)c^2 - 30(s-1)bc + 15b^2\right\} \\ + \dfrac{1}{s-1}\left\{\sum_{j=1}^{s} S_j^2 + 2[b-2cs]C_1 + 2cC_2\right\} \\ + \left\{\dfrac{s^2(s+1)}{3}\left[bc - c^2(s-1) + \dfrac{4csC_1}{s-1}\right]\right\} i + \left[\dfrac{s^3(s+1)c^2}{3}\right] i^2 \end{array} \right\} \tag{19}$$

When there is no seasonal effect, $S_j = 0$ $\forall j = 1, 2, \ldots, s$, $C_1 = C_2 = \sum_{j=1}^{s} S_j^2 = 0$. Hence

$$V_i(Q) = \left\{ \begin{array}{l} \dfrac{s(s+1)}{180}\left\{(2s-1)(8s-11)c^2 - 30(s-1)bc + 15b^2\right\} \\ + \dfrac{s^2(s+1)}{3}\left[bc - c^2(s-1)\right] i + \left(\dfrac{s^3(s+1)c^2}{3}\right) i^2 \end{array} \right\} \tag{20}$$

$$d_i(Q) = U_i(Q) - V_i(Q)$$
$$= \frac{1}{s-1}\left[\sum_{j=1}^{s} S_j^2 + 2[b-2cs]C_1 + 2cC_2\right] + \frac{s^2(s+1)}{3}\left(\frac{4csC_1}{s-1}\right) i \tag{21}$$

which is zero under null hypothesis (H_0: $S_j = 0$).

(c) For the Exponential trend-cycle component, in the presence of seasonal effect, the row variance is

$$U_i(E) = \sigma_i^2(E)$$
$$= b^2 e^{2c[(i-1)s+1]}\left[\left(\frac{1-e^{2cs}}{1-e^{2c}}\right) - \frac{1}{s}\left(\frac{1-e^{cs}}{1-e^{c}}\right)\right] + \sum_{j=1}^{s} S_j^2 + 2be^{c(i-1)s}\sum_{j=1}^{s} e^{cj} S_j \tag{22}$$

When there is no seasonal effect, $S_j = 0$ $\forall j = 1, 2, \ldots, s$, $\sum_{j=1}^{s} S_j^2 = \sum_{j=1}^{s} e^{cj} S_j = 0$. Hence

$$V_i(E) = \left\{ b^2 e^{2c[(i-1)s+1]} \left[\left(\frac{1-e^{2cs}}{1-e^{2c}} \right) - \frac{1}{s} \left(\frac{1-e^{cs}}{1-e^{c}} \right) \right] \right\} \tag{23}$$

$$d_i(E) = U_i(E) - V_i(E) = \sum_{j=1}^{s} S_j^2 + 2b e^{c(i-1)s} \sum_{j=1}^{s} e^{cj} S_j \tag{24}$$

Which again is zero under the null hypothesis (H_0: $S_j = 0$).

It is clear from $d_i = U_i - V_i$ (see (18), (21), and (24)) that when the trend dominates the series, the presence of the seasonal effect in a series will be difficult to detect. Therefore, it is advisable to isolate the trend before embarking on test for presence of seasonal effect in a series. It is important to note that the d_i represented by (18), (21), and (24) for linear, quadratic, and exponential curves respectively are functions of the seasonal components only when the trend is removed.

Empirical Examples

This section presents some empirical examples to illustrate the application of the tests for seasonality in time series data discussed previously, and to compare the powers of the tests in the detection of the presence of seasonal effects in a series. The data used consists of 106 data sets of 120 observations each, simulated using the MINITAB software from: (a) $X_t = (a + bt) + S_t + e_t$ with $a = 1$ and $b = 0.02$, 0.20, and 2.00, for the linear trend-cycle component; (b) $X_t = (a + bt + ct^2) + S_t + e_t$ with $a = 1$, $b = 2.0$, and $c = 3$ for the Quadratic trend-cycle component; and (c) $X_t = (be^{ct}) + S_t + e_t$ with $b = 10$ and $c = 0.02$ for exponential trend-cycle component. In each case it is assumed that $e_t \sim N(0, 1)$ and S_j, $j = 1, 2, \ldots, 12$ are as shown in Table 2. Meteorological data were collected from the meteorlogical station in Owerri, southeastern Nigeria, for the period of 1990-2010 with the assistance of the computer unit of the Federal Meteorological Centre Oshodi, Lagos. The weather parameters collected are mean monthly values of air temperature, relative humidity, and rainfall. Data on monthly U.S. male (16 to 19 years) unemployment figures (in thousands) for the period 1948 to 1981, monthly gasoline demand Ontario (gallon millions) for the period 1960 to 1975, monthly production of Portland cement (thousands of tons) for the period 1956 to 1970, and monthly milk production (pounds per cow) for the period 1962 to 1975, sourced from Hyndman (2014), were used to further illustrate the application of the proposed tests for seasonality in real life time series data.

Table 2. Seasonal indices used for simulation

j	1	2	3	4	5	6	7	8	9	10	11	12
S_j	-0.89	-1.22	0.10	-0.15	-0.09	1.16	2.34	1.95	0.64	-0.73	-2.14	-0.97

Table 3. Summary results of tests for seasonality when b = 0.02, 0.20, and 2.00 for linear trend curve

		1% (0.01)		5% (0.05)		10% (0.10)	
Slope	**Test Statistic**	**%Pass**	**%Fail**	**%Pass**	**%Fail**	**%Pass**	**%Fail**
b = 0.02	*t*-test	100.00	0.00	100.00	0.00	100.00	0.00
	S-R test	100.00	0.00	100.00	0.00	100.00	0.00
	Sign test	84.91	15.09	99.06	0.04	99.06	0.04
b = 0.20	*t*-test	100.00	0.00	100.00	0.00	100.00	0.00
	S-R test	74.53	25.47	100.00	0.00	100.00	0.00
	Sign test	76.41	23.59	99.06	0.04	99.06	0.04
b = 2.00	*t*-test	67.92	32.62	74.53	25.47	82.08	17.92
	S-R test	60.38	39.62	74.53	25.47	80.11	19.81
	Sign test	47.17	52.83	65.09	34.91	65.09	34.91

The summary of the results of the application of the three tests for the presence of seasonal effects in the simulated series are shown in Table 3 when the trend-cycle component is present for linear trend curve and Table 4 when trend-cycle component is absent for linear, quadratic, and exponential trend curves.

As Table 3 shows, when the slope b is 0.02, the t-test and Wilcoxon signed-ranks test performed equally well (100% of the time) in detecting the presence of seasonal effect at the three chosen levels of significance (i.e. 1%, 5%, and 10%). The sign test was able to detect the presence of seasonal effect from at least 84.91% of the time at 1% level of significance to about 99.06% of the time at both 5% and 10% levels of significance. When the slope b is increased to 0.20, the t-test was able to detect the presence of seasonality 100% of the times at the three chosen levels of significance. The Wilcoxon signed-ranks test was able to detect the presence of seasonality 100% of the time at 5% and 10% levels of significance and less than 75% of the time at 1% levels of significance. The sign test, on the other hand, was able to detect the presence of seasonal effect about 99.06% of the time at both 5% and 10% levels of significance but at about 76.41% of the time at 1% level of significance. For b = 2.00, all three tests did not perform as well in detection of the presence of seasonal effects in a series.

Table 4. Summary results of tests for seasonality for the de-trended series for linear, quadratic, and exponential trend curves

Trend Component	Test Statistic	1% %Pass	1% %Fail	5% %Pass	5% %Fail	10% %Pass	10% %Fail
Linear:	*t*-test	100.00	0.00	100.00	0.00	100.00	0.00
a = 1.0,	S-R test	85.85	14.15	100.00	0.00	100.00	0.00
b = 2.0	Sign test	85.85	14.15	99.06	0.04	99.06	0.04
Quadratic:	*t*-test	100.00	0.00	100.00	0.00	100.00	0.00
a = 1.0, b = 2.0,	S-R test	100.00	0.00	100.00	0.00	100.00	0.00
c = 3.0	Sign test	84.91	15.09	99.06	0.04	99.06	0.04
Exponential:	*t*-test	100.00	0.00	100.00	0.00	100.00	0.00
b = 10,	S-R test	100.00	0.00	100.00	0.00	100.00	0.00
c = 0.02	Sign test	96.23	4.77	100.00	0.00	100.00	0.00

The best, the *t*-test, was able to detect the presence of seasonal effects at most 82% of the time at 10% level of significance and less than 75% of the time at 1% and 5% levels of significance.

In summary, the performances of all the tests (*t*-test, Wilcoxon signed-ranks test, and sign test) appear to have decreased with increasing dominance of the trend-cycle component in the simulated series and increased with increasing levels of significance. The *t*-test was observed to have performed better than the other two statistical tests applied while the Sign test appears to be trailing behind others.

The results also appear to support the claim made by Iwueze and Nwogu (2014) that it is necessary to de-trend time series data before conducting test for seasonality. This claim was supported by results of (18), (21), and (24). In other to assess the authenticity of this claim, the three tests (*t*-test, Wilcoxon signed-ranks test, and sign test) were applied to the de-trended series from the simulated series with b = 2.0. The results of these are shown in Table 4.

The results in Table 4 show that the *t*-test and Wilcoxon signed-ranks test are equal and perfect in performance (100% all through) in detecting the presence of seasonal effects, although the sign test has performance percentages of about 85.85% at 1% level of significance and 99.06% at both 5% and 10% significance levels. These are in line with the results obtained when the slope b = 0.02, and supports the claim that dominance of a series by trend can obscure the presence of seasonal effect in a series.

Table 5. Results of tests for seasonality using real life time series data

	t-Test	Wilcoxon S-R Test	Sign Test
Weather Parameter		**P-Value Sig. (2-tailed)**	
Air Temperature	0.003	0.001	0.000
Relative Humidity	0.000	0.000	0.000
Rain Fall	0.000	0.000	0.000
US Male (16-19 years) Unemployment	0.000	0.003	0.006
Gasoline Demand	0.000	0.002	0.000
Production of Portland Cement	0.004	0.008	0.006
Milk Production	0.000	0.002	0.000

The summary of the results of the application of the three tests for presence of seasonal effect in the real life time series are shown in Table 5. The three proposed tests (*t*-test, Wilcoxon signed-ranks test, and sign test) performed well in the detection of the presence of seasonal effects in all the real life time series data used, even at 1% significance level.

Concluding Remark

In this study, three tests (*t*-test, Wilcoxon signed-ranks test, and sign test for paired sample data) for detection of seasonal effects in a time series data have been proposed. The tests were developed using the row variances of the Buys-Ballot table when the model structure is additive, and for selected trending curves. The performances of the tests were assessed using simulated series with different trending curves and at different levels of significance, and with real life time series data.

The results of the analysis from the simulated series show that the performances of all three tests to have decreased with increasing dominance of the trend-cycle component in the simulated series, and increased with increasing levels of significance. The *t*-test was observed to have performed better than the other two statistical tests applied, while the sign test appears to be trailing behind others.

When the tests were applied to the de-trended series from a trend dominated series (simulated series with $b = 2.00$), the results are in line with the results obtained when the slope b is 0.02. This supports the claim by Iwueze and Nwogu (2014) that dominance of a series by trend can obscure the presence of seasonal effect in a series and that it is necessary to de-trend a time series data before conducting test for seasonality.

In view of these results, it has been recommended that the Student's *t*-test and Wilcoxon signed-ranks test be used for the detection of the presence of seasonal effects in time series data when the model structure is additive until further studies prove otherwise. It has also been recommended that the tests be applied to the de-trended series when a series is dominated by trend. Preliminary assessments like the time plot of the study series can offer a guide to determining when a series is dominated by the trend.

Furthermore, when real life time series data were used, the three proposed tests (*t*-test, Wilcoxon signed-ranks test, and sign test) performed well in detection of the presence of seasonal effect even at 1% significance level.

References

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control*. Englewood Cliffs, NJ: Prentice Hall.

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). *Graphical methods for data analysis*. Bellmont, CA: Wadsworth.

Chatfield, C. (2004). *The analysis of time series: An introduction*. Boca Raton, FL: Chapman & Hall/CRC.

Cleveland, W. S. (1993). *Visualizing data*. Murray Hill, NJ: AT & T Bell Laboratories.

Corder, G. W., & Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Davey, A. M., & Flores, B. E. (1993). Identification of seasonality in time series: A note. *Mathematical and Computer Modelling, 18*(6), 73-81. doi: 10.1016/0895-7177(93)90126-J

Edwards, J. H. (1961). The recognition and estimation of cyclic trends. *Annals of Human Genetics, 25*(1), 83-86. doi: 10.1111/j.1469-1809.1961.tb01501.x

Flisher, A. J. C., Parry, D. H., Bradshaw, D., & Juritz. J. M. (1996). Seasonal variation of suicide in South Africa. *Biological Psychiatry, 39*(1), 522-523. doi: 10.1016/0006-3223(96)84029-7

Fomby, T. B. (2008). Buys ballot plots: Graphical methods for detecting seasonality in time series (Unpublished manuscript). Retrieved from http://faculty.smu.edu/tfomby/eco5375/data/season/buys%20ballot%20plots.pdf

Fomby, T. B. (2010). Stable seasonal pattern (SSP) model (Unpublished manuscript). Retrieved from http://faculty.smu.edu/tfomby/eco5375/data/notes/Stable%20Seasonal%20Pattern%20Model_v2.pdf

Freedman, L. S. (1979). The use of a Kolmogorov-Smirnov type statistic in testing hypotheses about seasonal variation. *Journal of Epidemiology and Community Health, 33*(3), 223-228. doi: 10.1136/jech.33.3.223

Hakko, H. (2000). *Seasonal variation of suicides and homicides in Finland, with special attention to statistical techniques used in seasonality studies* (Unpublished doctoral thesis). University of Oulu, Fnland.

Herring, D. A., & Hoppa. R. D. (1997). Changing patterns of mortality seasonality among the western James Bay Cree. *International Journal of Circumpolar Health, 56*(4), 121-133.

Hewitt, D., Milner, J., Csima, A., & Pakula, A. (1971). On Edwards' criterion of seasonality and a non-parametric alternative. *British Journal of Preventive Social Medicine, 25*, 174-176.

Horn, S. D. (1977). Goodness-of-fit tests for discrete data: A review and an application to a health impairment scale. *Biometrics, 33*(1), 237-247. doi: 10.2307/2529319

Hyndman, R. (2014). *Time series data library: Data on monthly U.S. male (16 to 19 years) unemployment figures (in thousands), monthly gasoline demand Ontario (gallon millions), monthly production of Portland cement (thousands of tonnes), and monthly milk production (pounds per cow)* [Data files]. Retrieved from http://datamarket.com/data/list/?q=provider:tsdl

Iwueze, I. S., & Nwogu, E. C. (2004). Buys-Ballot estimates for time series decomposition. *Global Journal of Mathematics, 3*(2), 83-98. doi: 10.4314/gjmas.v3i2.21356

Iwueze, I. S., & Nwogu, E. C. (2014). Framework for choice of models and detection of seasonal effect in time series. *Far East Journal of Theoretical Statistics, 48*(1), 45-66.

Iwueze, I. S., Nwogu, E. C., Ohakwe, J, & Ajaraogu, J. C. (2011). Uses of the Buys-Ballot table in time series analysis. *Applied Mathematics, 2*, 633-645. doi: 10.4236/am.2011.25084

Kendall, M., & Ord, J. K. (1990). *Time series* (3rd ed.). London: Griffin.

McLaren, C. E., Legler, J. M., & Brittenham, G. M. (1994). The generalized χ^2 goodness-of-fit test. *Journal of the Royal Statistical Society. Series D (The Statistician), 43*(2), 247-258. doi: 10.2307/2348342

Rau, R. (2005). *Seasonality in human mortality. A demographic approach*. (Unpublished doctoral thesis). Universität Rostock, Rostock, Germany.

Rihmer, Z., Rutz, W., Pihlgren, H., & Pestiality, P. (1998). Decreasing tendency of seasonality in suicide may indicate lowering rate of depressive suicides in the population. *Psychiatry Research, 81*(2), 233-240. doi: 10.1016/S0165-1781(98)00106-1

Rogerson, P. A. (1996). A Generalization of Hewitt's test for seasonality. *International Journal of Epidemiology, 25*(3), 644-648. 10.1093/ije/25.3.644

Sheth, T., Nair, C., Muller, J., & Yusuf, S. (1999). Increased winter mortality from acute myocardial infarction and stroke: The effect of age. *Journal of the American College of Cardiology, 33*(7), 1916-1919. 10.1016/S0735-1097(99)00137-0

Underwood, J. H. (1991). Seasonality of vital events in a Pacific island population. *Social Biology, 38*(1-2), 113-126. doi: 10.1080/19485565.1991.9988775

Verdoux, H., Takei, N., Cassou de Saint-Mathurin, R., & Bourgeois, M. (1997). Analysis of the seasonal variation of schizophrenic births using a Kolmogorov-Smirnov type statistic. *European Psychiatry, 12*(3), 111–116. doi: 10.1016/S0924-9338(97)80198-8

Villa, S., Guisecafré, H., Martinez, H., & Muñoz, O. (1999). Seasonal diarrhoeal mortality among Mexican children. *Bulletin of the World Health Organization, 77*(5), 375-380.