Seasonal test for non-stationary time series data by periodogram analysis

Gumgum Darmawan, Budhi Handoko and Defi Yusti Faidah

Department of Statistics, Faculty of Mathematics and Natural Science,Universitas Padjadjaran,Jatinangor-Sumedang 45363,Indonesia

E-mail : gumgum@unpad.ac.id

Abstract. Seasonal phenomenon is a common occurence in our daily activities. Many business and economic time series contain a seasonal phenomenon that repeats itself after a regular period of time. The smallest time period for this repetitive phenomenon is called the seasonal period. Seasonal test for time series data is well identified by Fisher's exact test in Periodogram Analysis. However, this seasonal test is only accurate for stationary Seasonal time series data. So, in this research we apply seasonal test for non-stationary time series data from generated data. Performance of this test is determined by the percentage of fit identification. We apply this Periodogram Analysis to real data.

1. Introduction

In our daily life, seasonal phenomenon is commonly happened. Seasonal behaviour occurs in many periods, such as monthly, weekly and daily. Seasonal or periodic changes of days, weeks or months within the annual cycle. Periodicity means that the statistical characteristic changes periodically within the year. For example, in hydrologic data concerning river flows, we expect high runoff periods in the spring and flow periods in the summer. Thus, the river flow correlations between spring months may be different from the correlations between summer months.

In Economic area, most economic time series are likely to exhibit some degree of seasonal variation. An obvious example, known to everyone, is the existence of a 'high' and 'low' season for air transportation and other recreational activities. Perhaps less obvious, but equally important, is the presence of a seasonal pattern in most economic aggregates such as the index of production, price indicates, the unemployment rate and so on.

The concept of periodically correlated processes was introduced by [1]. He gave a formal definition of periodic stationarity for a general periodic process. The first application of periodic time series models seems to have been by hydrologists [2]. They used the lag-one autoregressive (AR) modelling monthly stream flow. Since then there have been many discussion and summaries about periodic time series models For example, [3] demonstrate the superiority of periodic autoregressive models among several other competitors in forecasting thirty monthly river flow time series. [4], studies some properties of the periodic autoregressive models using a related multivariate autoregressive representation. [5] show how periodic autoregressive moving average models may be as heterogeneous models.

[6] dealt with moment estimation of parameters in periodic autoregressive (PAR) models. He showed that estimates of the seasonal parameters obtained by using the seasonal Yule-Walker equations possess many desirable properties, including maximum asymptotic efficiency under normality. [7] suggest estimating the parameters of PARMA models using the seasonal Yule-Walker equations. [8] proposed an algorithm for the maximum likelihood estimation for periodic ARMA models. [9] developed an algorithm for exact likelihood of periodic moving average models. [10] developed the innovations algorithm for estimation of PARMA model parameters.

[11] examine the recursive prediction and likelihood evaluation techniques for PARMA models, which are extensions of [12] for ARMA series. [13] study correlation and partial auto-correlation properties PARMA time series models.

ISBN: 978-979-3812-53-3

However, to analyze the Data, it is important to check seasonality of time series data. [14] credits these arrangements of the table to Buys-Ballot; hence, the table has also been called the Buys-Ballot table in the literature. Interested readers are referred to an excellent collection of articles edited by [15] and to some research papers on the topic such as [16], [17], [18], [19], and [20].

Based on these papers, we proposed exact identification of seasonal time series data by periodogram analysis. [21] applied seasonal test for data that available in R software such as Lynx Pelt sales, UKDriverDeaths, Lynx, Nottem co2 and AirPassegers. So,The purposes of this study are to determine accuracy of periodogram analysis in detecting non-stationary seasonal time series data and to applicate in Indonesia economic data.

2. Method

In this study we use two methods, Periodogram Analysis and Fractionally difference. Periodogram analysis is used to detect hidden periodicity in time series and fractionally difference is used to detect rate of differencing (d).

1.1. Periodogram

The. periodogram is the Fourier transform of the autocovariance function**.** A periodogram is used to identify the dominant periods (or frequencies) of a time series. This can be a helpful tool for identifying the dominant cyclical behaviour in a series, particularly when the cycles are not related to the commonly encountered monthly or quarterly seasonality. The equation of standard periodogram is as follows;

$$I(\omega_k) = \begin{cases} n a_0^2 & k = 0, \\ \frac{n}{2}\left(a_k^2 + b_k^2\right) & k = 1,..,[(n-1)/2], \\ n a_{n/2}^2 & k = \frac{n}{2} \text{ when } n \text{ is even.} \end{cases} \tag{1}$$

It was introduced by [22] to search a periodic component in a series.

1.2. Fractionally Difference

To identify rate of trend in time series data, we need tool for identifying this. Here, we use Geweke and Porter-Hudak method for identification rate of trend (*d).* Calculation of coefficient d is determined by a regression method.

$$\hat{d} = \frac{\sum_{j=1}^{m}\left(X_j - \bar{X}\right)\left(Y_j - \bar{Y}\right)}{\sum_{j=1}^{m}\left(X_j - \bar{X}\right)^2} \tag{2}$$

Where;

$$X_j = \ln\left[\frac{1}{4\left[\sin\left(\omega_j/2\right)\right]^2}\right]$$

$$Y_j = \ln I_z(\omega_j)$$
$$\omega_j = 2\pi j / n, \quad j = 1,2,\ldots \lfloor \sqrt{n} \rfloor$$

2. Algorithm for non-stationary seasonal data detection.

Step of seasonal test by Periodogram analysis as follow; :

1. Given a time series of N observation, real or generated data. The specification of data is trend and seasonal pattern.
2. Identify the non-stationary of data by GPH-Method, formula (2) . If the value of the *d* (differencing coefficient) is higher 0.5, then the data must be differenced.
3. Fit the data to the second equation:

$$Z_t = \sum_{k=0}^{[n/2]} (a_k \cos \omega_k t + b_k \sin \omega_k t) \ , \qquad (3)$$

   where k = 0,1,…,[n/2] and ω_k is fourier frequency determined by ω_k = 2π.k / n.

4. Compute a_k and b_k by these formulas :

$$a_k = \begin{cases} \frac{1}{n}\sum_{t=1}^{n} z_t \cos\omega_k t, & k=0 \text{ and } k=\frac{n}{2} \text{ if } n \text{ even} \\ \frac{2}{n}\sum_{t=1}^{n} z_t \cos\omega_k t, & k=1,2,\ldots,\frac{(n-1)}{2} \end{cases} \qquad (4)$$

$$b_k = \frac{2}{n}\sum_{t=1}^{n} z_t \sin \omega_k t, \qquad \text{k = 1, 2, ..., } \frac{(n-1)}{2} \qquad (5)$$

5. Compute value of ordinate $I(\omega_k)$ by formula (1)
6. Test for significance of every fourier frequency as follow;

   H_0: $\alpha = \beta = 0$ (data don’t have seasonal pattern)

   H_1: $\alpha \neq 0$ or $\beta \neq 0$ (data have seasonal pattern)

   Statistic test:

$$T = \frac{I^{(1)}(\omega_{(1)})}{\sum_{k=1}^{[n/2]} I(\omega_k)} \qquad (6)$$

   where,

   $I^{(1)}(\omega_{(1)})$: maximum ordinate of periodogram of fourier frequency

   $I(\omega_k)$: Value of periodogram ordinate at k-th’ fourier frequency .

   Test criteria:

   Reject H_0 if T >g_α with α = significant level. Value of g_α can be seen at table Fisher (Wei,2006).

   Statistic test as follow:

$$F = \frac{(n-3)\left(a_k^2 + b_k^2\right)}{2\sum_{\substack{j=1 \\ j\neq k}}^{[n/2]} \left(a_j^2 + b_j^2\right)} \qquad (7)$$

   Where j = 1,2, …, (n-1)/2 and k=n/2.

   Test criteria:

   reject H_0 if F >F-table (2, n-3; α) where α = significant level.

7. Build periodogram according to [22] to calculate the value of seasonal periodic:

Table 1 Periodogram Table

K	Frequency (ω_k)	Period (P)	$I(\omega_k)$	F
1	ω_1	P_1	$I(\omega_1)$	f_1
2	ω_2	P_2	$I(\omega_2)$	f_2
:	:	:	:	:
:	:	:	:	:
n/2	$\omega_{n/2}$	$P_{n/2}$	$I(\omega_{n/2})$	$f_{n/2}$

Where $P = \frac{2\pi}{\omega_k}$.

8. Base on Periodogram table, we know the result of computation and compare F with F-table, υ1=2, υ2=n-3 and α = significant level. If H_0 is significant then seasonal pattern is available in this time series data.
9. Test for the value of period as follow ;

$H_0: \alpha = \beta = 0$

H1: $\alpha \neq 0$ or $\beta \neq 0$

Based on Equation (1) where $I^{(1)}(\omega_{(1)})$ has got by the formula:

$$I^{(1)}(\omega_{(1)}) = \max \{ I(\omega_k) \}. \quad (8)$$

Test criteria T from equation (6) where g_α can be seen from table Fisher.

Value of g_α from table above divided by two classes :

a. g_α *by exact formula* by formula as follows:

$$P(T > g) = \sum_{j=1}^{m} (-1)^{(j-1)} \binom{N}{j} (1 - jg)^{N-1} , \quad (9)$$

Where N=[n/2], g>0 , and m is the largest than 1/g. Thus, for any given significance level α, we can use equation (9) to find the critical value g_α such that :

$P(T > g_\alpha) = \alpha$.

b. If the T value calculated from series is larger than g_α , then we rejected the null hypothesis and conclude that the series Z_t contains a periodic component. This test procedure is known as Fisher's test. The critical values of T for the significance level α =0.05 as shown in table Fisher..

The third column in table Fisher is an approximation obtained by using only the first term in (9),

$$P(T > g) \cong N(1 - g)^{N-1} . \quad (10)$$

The approximation is very close to exact result, hence, for most practical purposes, we can use equation (11) to derive the critical value g_α for the test [23]. After comparing T with T-table, the result is H_0 is rejected or no in α =0.05, then the conclusion is that time series data have seasonal pattern or not in period p.

3. SIMULATION STUDY AND APLICATION

A detailed simulation study was conducted to evaluate accuracy of algorithm seasonal detection. Data from several different SARIMA (Seasonal Autoregressive Integrated Moving Average) models were generated. For each model, N = 500 ,100 and 60 with 100 repetitions.

Intercept 0 and 10, differencing coefficient *d* =1 and D = {0;1} and P (Period) {6;12}. ARIMA coefficient we used first $\phi = \{0;0.4\}$ $\theta = \{0;0.4\}$ $\Theta = \{0;0.3\}$ $\Phi = \{0;0.3\}$, with

ISBN: 978-979-3812-53-3

$e \sim IIDN(0,1)$. Open source software R 3.4.1 (OSSR) program was used to generate the SARIMA data. Some general conclusions can be drawn from this study. We used two packages for running this R macro, package fracdiff and package gsarima.

Package fracdiff is used to detect trend pattern in data and package gsarima is used to generate non-stationary seasonal time series data. Both packages are external package so we should download from R website.

In application, we used GDP (Gross Domestic Product) data from 2000 to 2016. Gross Domestic Product (GDP) is a monetary measure of the market value of all final goods and services produced in a period (quarterly or yearly) of time. Nominal GDP estimates are commonly used to determine the economic performance of a whole country or region, and to make international comparisons.

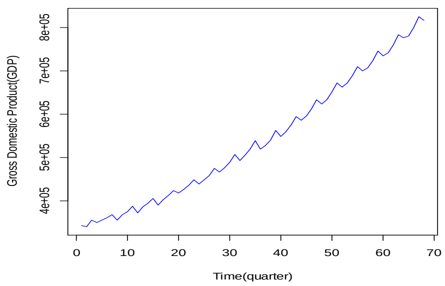

Figure 1 Gross Domestic Product 2000-2016

4. Result

In this session, we show two results, the first is the result for simulation study and the second is about analysis for real data (GDP). According to session 4, Simulation study produced four output tables, the first D=1,d =1,intercept =10, the second D=0,d =1,intercept =10, the third D=1,d =1,intercept =0 and the last D=0,d =1,intercept =0.

A detailed simulation study was conducted to demonstrate the effectiveness of the results of the previous section (algorithm for non-stationary seasonal data detection) using simulated data from SARIMA Models. For each model, individual realization of N = 500,100 and 60 months of data were simulated and the innovation algorithm was used to obtain parameter estimates for each realization. In each case, generating data were obtained for k=100 iterations. A software R program was used to simulated the SARIMA samples as to make all the necessary calculations.

Table 2. D=1, *d* =1,intercept =10

Data Model	Period	Sample		
		N=500	N=100	N=60
phi<-c(0) theta<-c(0.4) Phi<-c(0) Theta<-c(0.3)	P=12	51	47	40
	P=6	82	52	55
phi<-c(0,4) theta<-c(0) Phi<-c(0.3) Theta<-c(0)	P=12	42	30	29
	P=6	52	44	40

ISBN: 978-979-3812-53-3

It is observed that in model for *D*=1, *d*=1 and intercept =10, accuracy of detection seasonal models under 50% . However, for N=500 and P=6, The algorithm was accurate enough to detect non-stationary-seasonal models,82% of seasonal model can be detected by the algorithm. It must be noted that N = 500 is the best result than any other sample. Moreover, Accuracy of algorithm for non-stationary seasonal detection for the model with P=6 is better than that the model with P=12.

Table 3. D=0, *d* =1,intercept =10

Data Model		Sample		
	Period	N=500	N=100	N=60
phi<-c(0) theta<-c(0.4) Phi<-c(0) Theta<-c(0.3)	P=12	64	54	52
	P=6	78	50	43
phi<-c(0,4) theta<-c(0) Phi<-c(0.3) Theta<-c(0)	P=12	67	42	31
	P=6	19	58	47

It is be seen that in Model *D*=0, *d*=1, and intercept =10, the best result was in Moving average model in P=6 and N =500 with value =79 and the worst result was in Autoregressive model P=6 and N =500 .However, In N=100 the accuracy relatively stable the value between 42 and 58.

Table 4. D=1, *d* =1,intercept =0

Data Model		Sample		
	Period	N=500	N=100	N=60
phi<-c(0) theta<-c(0.4) Phi<-c(0) Theta<-c(0.3)	P=12	51	38	20
	P=6	74	54	62
phi<-c(0,4) theta<-c(0) Phi<-c(0.3) Theta<-c(0)	P=12	39	21	28
	P=6	85	41	36

It is observed that in table 4, The Algorithm has good performance for identifying Autoregressive and Moving average models with N=500 and P =6 . Both models have values 74 and 85. It is mean that for 100 seasonal data that had been generated from R macro, the macro of algorithm had successfully detected 74 and 85 seasonal data respectively. However, the macro of algorithm has bad performance for autoregressive models in both N=100 and N=60. Moreover, Moving average models was fail to detect by this macro, all values have under 50 in P = 12.

Table 5. D=0, *d* =1,intercept =0

Data Model	Period	Sample N=500	N=100	N=60
phi<-c(0) theta<-c(0.4) Phi<-c(0) Theta<-c(0.3)	P=12	49	36	33
	P=6	84	42	42
phi<-c(0,4) theta<-c(0) Phi<-c(0.3) Theta<-c(0)	P=12	63	37	48
	P=6	21	46	33

It is be seen that for Models D=0,d =1 and intercept =0, the performance of macro algorithm has bad performance . Almost all of the values have under 50%. However, the best performance of this macro shown for Autoregressive model with N=500 and P=6. This data model has value 84.

Analysis for real data, we use GDP data from 2000 to 2016. The first step is to identify trend from data by equation 2, the result is $d = 1.022$. This value show us, that GDP data is non stationer, so we should difference this data with d=1.

After differencing data, we should calculate the Periodgram as shown in figure 2. The Value of Periodogram is used for hypothesis testing of availability seasonal pattern. The macro algorithm detected that this data were seasonal pattern (figure 3).

```
> periodogram
 [1]   6539985.16   1343866.08    521169.15   1187589.04    139315.33
 [6]    905984.29    157074.98     53894.23    105288.30    511236.36
[11]     40608.74    665595.99    623330.91   4267859.63   5865783.07
[16]  40750300.95 111092681.04   5098006.99   6330039.02   1168810.26
[21]    662200.57    411403.70    766525.46    793486.59    132708.62
[26]    110921.89    339575.86    362195.09    104824.52    240400.57
[31]   5220720.92   3632438.49  67579660.36
```

Figure 2 Value of Periodogram from Data

```
> Thitung
[1] 0.41495
> Ttabel
galpha
0.13135
[1] "Data mengikuti pola musiman"
> round(Periode)
[1] 4
```

Figure 3. Output Of R Macro Algorithm

It can be seen from figure 3, the last output is 4. It is mean that the seasonal period of the data is 4. This result was done by hypothesis testing with T distribution as explained in session 3 above.

5. Conclusion

According simulation study, Seasonal Testing with Periodogram Analysis approach has fairly good accuracy for seasonal time series data with period 6 and large sample (N=500). For non-stationary Seasonal time series data with Period 12, the algorithm should be modified with another periodogram. The last analysis, in real data, the algorithm could detect seasonal patterns in the data. compared to the seasonal view of plot data this result was accurate.

ISBN: 978-979-3812-53-3

Acknowledgments
The author would like to thank the financial support to the University of Padjadjaran, so this paper can be presented in the seminar.

6. References

[1] Gladyshev 1961 Periodically correlated random sequences *Sov.Math **2** 3*85-388.
[2] Fiering 1962 Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation *Myron* 459-493
[3] Noakes *et al* 1985 Forecasting monthly riverflow time series *International Journal of Forecasting* **1** 179-190
[4] Troutman 1979 Some Results in periodic autoregression *Biometrika* **6** 219-228
[5] Tiao and Grupe 1980 Hidden periodic autoregressive moving average models in time series data *Biometrika* **67** 365-373
[6] Pagano 1978 On periodic and multiple autoregressions”,*Annals of Statistics* **6** 1310-1317
[7] Salas et al. 1982 Estimation of ARMA models with seasonal parameters *Water Resource Research* **18** 1006-1010
[8] Vecchia 1985 Periodic autoregressive-moving average (PARMA) modelling with application to water resources *Water Resources Bulletin* **21** 721-730
[9] W K Li and Y V Hui 1988 An algorithm for the exact likelihood of periodic autoregressive moving average models *Communication in Statistics* **17** 1483-1494
[10] P LAnderson, M M Meerschaert and A V Vecchia 1999 Innovations algorithm for periodically stationary time series *Stochastic Processes and their Applications* **83** 149-169
[11] Lund and Basawa 2000 Recursive prediction and likelihood evaluation for periodic ARMA models Journal *of Time Series Analysis* **20** 75-93
[12] Ansley 1979 An algorithm for the exact likelihood of a mixed autoregressive moving average process *Biometrika* **66** 59-65
[13] Shao and Lund 2004 Computation and characteristization of autocorrelations and partial autocorrelations ini periodic ARMA models *Journal of Time Series Analysis* **25** 359-372
[14] Wold 1938 A study in the analysis of stationary Time series (Upsala)
[15] Zellner 1978 *Seasonal Analysis of Economic Time Series* (U.S Department of commerce Washongton DC)
[16] Dagum 1900 *The X-11-ARIMA Seasonal Adjusment method,Catalogue,12-564* (Ontario : Statistics Canada)
[17] Pierce 1980 A Survey of recent developments in seasonal adjustment *The American Statistician* **34** 125-134
[18] Hilmer and Tiao 1982An ARIMA based approach to seasonal adjustment *Am. statist. Assoc.* **77** 63-70
[19] Bell and Hillmer 1984 Issues Involved with the Seasonal Adjustment of Economic Time Series *Journal of Business & Economic Statistics* **2**
[20] Cupingood and Wei 1986 Seasonal Adjustment of Time Series Using One-Sided Filters *Journal of Business & Economic Statistics October* **4**
[21] G Darmawan, , B Handoko and Suparman 2016 *Proceeding ICAS II* (Bandung : Statistics Department Padjadjaran University)
[22] Schuster 1898 On the investigation Hidden Periodicities with application to a supposed 26-days period of meteorogical phenomena *Terr.mag.Atmos,elect* **3** 13-41
[23] Wei W W S 2006 Time series Analysis univariate and multivariate method (USA : Department of Statistics Temple University)
[24] Meerschaert and vecchia 1999 *Innovations algorithm for periodically stationary time series*