Accuracy of Periodogram Analysis for Identification of Multiplicative SARIMA Models

Gumgum Darmawan*, Nuning Kurniasih†, & Ansari Saleh Ahmar‡

1Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Padjadjaran, Jatinangor-Sumedang 45363, Indonesia
2Faculty of Communication Sciences, Library and Information Science program, Universitas Padjadjaran, Jatinangor-Bandung 45363, Indonesia
3Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar, Makassar, Indonesia.
gumgum@unpad.ac.id

Abstract. In everyday life, seasonal events are common, seasonal behavior is common in various periods, such as daily, weekly or monthly. Business, economics and financial cases for example, we often encounter a seasonal phenomenon, namely data that repeated over the same period. However, this seasonal test is only accurate for stationary Seasonal time series data. So, in this research we apply exact identification for multiplicative 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, AirPassengers and UKgas data. All simulation and real data analysis were done by open source Software R (OSSR). It is observed that in all model for Period 12, accuracy of detection seasonal models under 50%. However, for P=6, The algorithm was accurate enough to detect multiplicative seasonal models, 70% of seasonal model can be detected by the algorithm.

1. Introduction
Seasonal phenomenon is commonly happened, in our daily life. Seasonal behavior occurs in many periods, such as daily, weekly and monthly. 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. In this way, the stream relationship between spring months might be not the same as the relationships between mid year months.

In Economic area, seasonal variation usually has an effect on economic time series, such as on the production index and the price of goods and services, the highest and lowest time for the number of tourist visits, unemployment, etc.

The concept of periodically correlated processes was introduced by Gladyshev [1]. This study provides a formal definition for a seasonal process in time series data that is often used by researchers in the field of hydrology [2]. They used the lag-one autoregressive (AR) modelling monthly stream flow. Due to rapid development, the researchers developed periodic data models that have seasonal patterns. For example, Noakes [3] demonstrate the superiority of periodic autoregressive models among several other competitors in forecasting thirty monthly river flow time series. Troutman [4], studies some properties of the periodic autoregressive models using a related multivariate
autoregressive representation. Tiao and Grupe [5] show how periodic autoregressive moving average models may be as heterogeneous models.

In Pagano [6], this research explains how to estimate periodic autoregressive mosel parameters (PAR). According to this study the Yule-Walker method of estimation gives the desired result, including maximum asymptotic efficiency under the assumption of normality. Whereas at Salas et al [7], Salas et. al suggested estimation of PARMA model parameters using the seasonal Yule-Walker equation. Then, Vecchia [8] proposes an algorithm for maximum probability estimation for seasonal ARMA models. Li and Hui [9] developed an algorithm for the accurate possibilities of the seasonal moving average model. Then on Anderson, Meerschaert and Vecchia [10] discussed how to estimate PARMA parameters through the algorithm they created.

Lund and Basawa [11] proposes a recursive forecasting method and a maximum probability technique for the PARMA model, which is a continuation of Ansley [12] for ARMA data. Shao and Lund [13] study correlation and partial auto-correlation properties PARMA time series models. However, to analyze the Data, it is important to check seasonality of time series data [14][15]. Wold [16] applied Buys-Ballot methods to detect seasonal patterns; therefore, in the table literature it is also called the Buys-Ballot table. Interested readers are referred to an excellent collection of articles edited by Burman, et. al. [17] and to some research papers on the topic such as Dagum [18], Pierce [19], Hilmer and Tiao [20], Bell and Hillmer [21], and Cupingood and Wei [22].

Based on these papers, we proposed exact identification of multiplicative seasonal time series data by periodogram analysis. Darmawan, Handoko , and Suparman [23] have 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 multiplicative seasonal time series data and to applicate in real data, AirPassengers and UKgas.

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 ($d$).

2.1. Periodogram

Periodograms are commonly used to identify the hidden periodicity of time series data. This tool can detect repetitive behaviours that occur in time series data, including repetitive cycle patterns over long periods of time. The equation of standard periodogram is as follows;

$$I(\omega_i) = \left\{ \begin{array}{ll}
na_0^2 & i = 0 \\
\frac{n}{2} (\hat{\alpha}_i + \hat{\beta}_i^2) & i = 1, \ldots, \left[ (n-1) / 2 \right] \\
n\alpha_2 / 2 & i = \frac{n}{2} \text{ when } n \text{ is even}
\end{array} \right. \quad (1)$$

It was introduced by Schuster [24] to search a periodic component in a series.

2.2. 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.
\[
\hat{d} = \frac{\sum_{j=1}^{m}(X_j - \bar{X})(Y_j - \bar{Y})}{\sum_{j=1}^{m}(X_j - \bar{X})^2}
\]  
(2)

Where;

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

\[
Y_j = \ln I_i(\omega_j)
\]

\[
\omega_j = 2\pi j / n, \quad j = 1, 2, ..., \left[\sqrt{n}\right]
\]

3. Algorithm for Non-stationary Seasonal Data Detection.

Procedure of pattern identification by periodogram analysis;

a) Input Real or generated data \((n)\). The specification of data is trend and seasonal pattern.

b) Difference the time series data by Geweke and Porter-Hudak (GPH) method (formula 2), if the data is non-stationary, then the value of GPH is greater then 0.5.

c) Use fourier series in equation (3) to decompose the time series data;

\[
X_i = \sum_{j=0}^{[n/2]} \left( \hat{\alpha}_i \cos \omega_i t + \hat{\beta}_i \sin \omega_i t \right).
\]  
(3)

where \(i = 0, 1, ..., [n/2] \) and \(\omega_i\) is fourier frequency determined by \(\omega_i = 2\pi i / n\).

d) Calculate two parameter above ( \(\hat{\alpha}_i\) and \(\hat{\beta}_i\) ) by (4):

\[
\hat{\alpha}_i = \begin{cases} 
\frac{1}{n} \sum_{i=1}^{n} x_i \cos \omega_i t & \text{if } n \text{ even } \\
\frac{2}{n} \sum_{i=1}^{n} x_i \cos \omega_i t & \text{if } n \text{ even }
\end{cases}
\]  
(4)

\[
\hat{\beta}_i = \frac{2}{n} \sum_{i=1}^{n} x_i \sin \omega_i t, \quad i = 1, 2, ..., \left(\frac{n-1}{2}\right)
\]  
(5)

e) Calculate Periodogram \(I(\omega_i)\) by equation (1)

f) Proceed the inferential procedure for these two parameters;

\(H_0: \alpha = \beta = 0\) ; the time series data don’t have sinusoidal pattern

\(H_1: \alpha \neq 0 \text{ or } \beta \neq 0\) ; the time series data don’t have sinusoidal pattern

Statistic test:
\[ T = \frac{I^{(1)}(\omega_{1})}{\sum_{i=1}^{n/2} I(\omega_{i})} \]  

(6)

Where:

\( I^{(1)}(\omega_{1}) \) : the ordinate peak value of the periodogram of the fourier frequency

\( I(\omega_{i}) \) : Value of periodogram ordinate at \( i \)-th fourier frequency.

Test criteria:

Reject \( H_{0} \) if \( T > g_{\alpha} \) with \( \alpha \) = significant level. Value of \( g_{\alpha} \) can be seen in Wei [25].

Statistic test as follow:

\[ F = \frac{(n-3)(\hat{\alpha}^{2} + \hat{\beta}^{2})}{2\sum_{k=1}^{n/2} (\hat{\alpha}_{k}^{2} + \hat{\beta}_{k}^{2})} \]  

(7)

Where \( k = 1, 2, ..., (n-1)/2 \) and \( i=n/2 \).

Test criteria:

reject \( H_{0} \) if \( F > F \)-table \( (2, n-3; \alpha) \) where \( \alpha \) = significant level.

\[ \begin{array}{c|c|c|c|c}
\hline
I & Frequency (\( \omega_{i} \)) & Period (S) & I(\( \omega_{i} \)) & F \\
\hline
1 & \( \omega_{1} \) & s_{1} & I(\( \omega_{1} \)) & f_{1} \\
2 & \( \omega_{2} \) & s_{2} & I(\( \omega_{2} \)) & f_{2} \\
\hline
n/2 & \( \omega_{n/2} \) & s_{n/2} & I(\( \omega_{n/2} \)) & f_{n/2} \\
\hline
\end{array} \]

Table 1. Periodogram Table

Where \( s = \frac{2\pi}{\omega_{j}} \).

h) According to table 1 above, the \( F \) value of the calculation results is compared with \( F \)-table with \( F \)-table, \( v1 = 2 \), \( v2 = n-3 \) and \( \alpha \) = significant level. If the result of the calculation gives a significant result then the time series data indicates a seasonal pattern. Test for the value of period as follow;

\( H_{0} \): \( \alpha = \beta = 0 \)

\( H_{1} \): \( \alpha \neq 0 \) or \( \beta \neq 0 \)

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

\[ I^{(1)}(\omega_{1}) = \max \{ I(\omega_{i}) \} \]  

(8)

i) Test criteria \( T \) from equation (6) where \( g_{\alpha} \) can be seen from table Fisher [25].

4. Simulation Study and Application

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, \( N = 100 \) and 60 with 1000 repetitions.
Differencing coefficient \( d = \{0,1\} \) and \( D = \{0;1\} \) and \( P \) (Period) \{6;12\}. SARIMA coefficient we used first \( \phi = 0.8 \; \Theta = 0.4 \; \Phi = 0.3 \), with \( e \sim IIDN(0,1) \). Open source software R 3.4.1 (OSSR) program was used to generate the SARIMA data. In this research can be drawn some general conclusions. We used three packages for running this R macro, package fracdiff, package sarima and TSA.

Mathematical formula of three multiplicative SARIMA models, as follows;

The First Model, SARIMA\((1,1,1)(1,0,1)\),
\[
z_t = (1 + \phi)z_{t-1} - \phi z_{t-2} - \Phi z_{t-s} + (1 + \phi)\Phi z_{t-(s+1)} + \phi \Phi z_{t-(s+2)} + a_t - \Theta a_{t-1} - \Theta a_{t-s} + \Theta a_{t-(s+1)}
\]

Second Model, SARIMA\((1,0,1)(1,0,1)\),
\[
z_t = \phi z_{t-1} + \phi z_{t-2} - \phi \Phi z_{t-s} + a_t - \Theta a_{t-1} - \Theta a_{t-s} + \Theta a_{t-(s+1)}
\]

The third model, SARIMA\((1,1,1)(0,0,1)\),
\[
z_t = (1 + \phi)z_{t-1} - \phi z_{t-2} - \Phi z_{t-s} + (1 + \phi)\Phi z_{t-(s+1)} + \phi \Phi z_{t-(s+2)} + a_t - \Theta a_{t-s}
\]

Where;

\( Z_t = \) time series data
\( \theta = \) Moving Average coefficient
\( \phi = \) Autoregressive coefficient
\( \Theta = \) Coefficient of seasonal moving average
\( \Phi = \) Coefficient of seasonal Autoregressive
\( a_t = \) residual
\( s = \) Period

Package fracdiff was used to detect trend pattern in data and package sarima was used to generate multiplicative seasonal time series data and package TSA was used to activate periodogram syntax. Three packages were external package so we should download from R website.

In application, we used AirPassengers and UKgas data. Both data have multiplicative seasonal pattern and available in R software (see figure 1 and figure 2).

Figure 1. AirPassengers
5. 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 (AirPassengers and UKgas). Both models have multiplicative seasonal pattern and available in R Software. In simulation study, generated data from three models were identified availability of period. We used three models, SARIMA(1,1,1)(1,0,1)s, SARIMA(1,0,1)(1,0,1)s, and SARIMA(1,1,1)(0,0,1)s, with parameters were ar=0.8, ma=0.4, sar=0.3 and sma=-0.7.

To show the effectiveness of the algorithm above the simulation details can be seen below as implement of section before (algorithm for multiplicative seasonal data detection) using simulated data from SARIMA Models. In the above algorithm, to get parameter estimation, each model is determined N = 100 and 60 (month). In each case, generating data were obtained for k=1000 iterations with periods s=6 and s=12. In this simulation, we used R software to simulate data that follows the seasonal ARIMA model.

Packages for running simulation we used sarima, fracdiff and TSA. Package sarima was needed for generating sarima data, package fracdiff for identification trend of time series data and package TSA was used to activated periodogram in R macro.

| Model | N  | Period | 6  | 12  |
|-------|----|--------|----|-----|
| 1     | 100|        | 0.704 | 0.147 |
|       | 60 |        | 0.662 | 0.123 |
| 2     | 100|        | 0.746 | 0.081 |
|       | 60 |        | 0.768 | 0.227 |
| 3     | 100|        | 0.671 | 0.093 |
|       | 60 |        | 0.78 | 0.119 |

It is observed that in all model for Period 12, accuracy of detection seasonal models under 50%. However, for P=6, The algorithm was accurate enough to detect multiplicative-seasonal models, 70%
of seasonal model can be detected by the algorithm. It must be noted that $N = 100$ is the best result than any other sample. Moreover, Accuracy of algorithm for multiplicative seasonal detection for the model with $P=6$ is better than that the model with $P=12$.

Analysis for real data, we use AirPassengers and UKgas. The first step is to identify trend from data by equation 2, the result is $d = 0.651$ (AirPassengers) and $d=1.000$ (UKgas). This value show us, that Both data are non stationer, so we should difference this data with its $d$, respectively.

After differencing data, we should calculate the Periodogram as shown in figure 2. The Value of Periodogram is used for hypothesis testing of availability seasonal pattern.

It can be seen from figure 3 and figure 4, Periodogram of both data after differencing by $d$. The differencing process should be done to the data because periodogram cannot detect hidden periodicity in non-stationary time series data.

![Figure 3. Periodogram of AirPassengers](image1.png)  
![Figure 4. Periodogram of UKgas](image2.png)

6. Conclusion
According simulation study, Seasonal Testing with Periodogram Analysis approach has fairly good accuracy for seasonal time series data with period 6. 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 identify sinusoidal effect from real data series. If we look at the figure above, the result is fit.

References
[1] Gladyshev E ~G., 1961 Periodically Correlated Random Sequences Sov. Math. 2 p. 385–388.
[2] Thomas H A and Fiering M B, 1962, Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins by Simulation, in Design of Water Resource Systems, p. 460–493.
[3] Noakes D J McLeod A I and Hipel K W, 1985 Forecasting monthly riverflow time series Int. J. Forecast. 1, 2 p. 179–190.
[4] Troutman B M, 1979 Some results in periodic autoregression Biometrika 66, 2 p. 219–228.
[5] Tiao G C and Grupe M R, 1980 Hidden periodic autoregressive-moving average models in time series data Biometrika 67, 2 p. 365–373.
[6] Pagano M, 1978 On periodic and multiple autoregressions Ann. Stat. p. 1310–1317.
[7] Salas J D Boes D C and Smith R A, 1982 Estimation of ARMA Models with seasonal parameters Water Resour. Res. 18, 4 p. 1006–1010.
[8] Vecchia A V., 1985 Periodic autoregressive-moving average (PARMA) modelling with application to water resources JAWRA J. Am. Water Resour. Assoc. 21, 5 p. 721–730.
[9] Li W K and Hui Y V., 1988 An algorithm for the exact likelihood of periodic autoregressive
moving average models Commun. Stat. - Simul. Comput. 17, 4 p. 1483–1494.
[10] Anderson P L Meerschaert M M and Vecchia A V., 1999 Innovations algorithm for periodically stationary time series Stoch. Process. their Appl. 83, 1 p. 149–169.
[11] Lund R and Basawa I V., 2000 Recursive prediction and likelihood evaluation for periodic ARMA models J. Time Ser. Anal. 21, 1 p. 75–93.
[12] Ansley C F, 1979 An algorithm for the exact likelihood of a mixed autoregressive-moving average process Biometrika 66, 1 p. 59–65.
[13] Shao Q and Lund R, 2004 Computation and characterization of autocorrelations and partial autocorrelations in periodic arma models J. Time Ser. Anal. 25, 3 p. 359–372.
[14] Ahmar A S et al., 2018 Modeling Data Containing Outliers using ARIMA Additive Outlier (ARIMA-AO) J. Phys. Conf. Ser. 954, 1.
[15] Ahmar A S, 2018 A Comparison of α-Sutte Indicator and ARIMA Methods in Renewable Energy Forecasting in Indonesia Int. J. Eng. Technol. 7, 1 p. 9–11.
[16] Wold H, 1938 A Study in the Analisys of Stationery: Time Series Almqvist & Wiksells Boktryckeri.
[17] P Burman J et al., 1978, Contributed Comments to" Seasonal Analysis of Economic Time Series", in Seasonal Analysis of Economic Time Series, (NBER), p. 461–479.
[18] Dagum E B, 1980 The X-II-ARIMA seasonal adjustment method Statistics Canada, Seasonal Adjustment and Time Series Staff.
[19] Pierce D A, 1980 A survey of recent developments in seasonal adjustment Am. Stat. 34, 3 p. 125–134.
[20] Hillmer S C and Tiao G C, 1982 An ARIMA-model-based approach to seasonal adjustment J. Am. Stat. Assoc. 77, 377 p. 63–70.
[21] Bell W R and Hillmer S C, 1984 Issues involved with the seasonal adjustment of economic time series J. Bus. Econ. Stat. 2, 4 p. 291–320.
[22] Cupingood L A and Wei W W S, 1986 Seasonal Adjustment of Time Series Using One-Sided Filters J. Bus. Econ. Stat. 4, 4 p. 473–484.
[23] Darmawan G Handoko B and Suparman Y, 2017 Seasonal Test for Non-Stationary Time Series Data by Means of Periodogram Analysis in The 2nd International Conference On Applied Statistics (ICAS 2016) p. 74–81.
[24] Schuster A, 1898 On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena Terr. Magn. 3, 1 p. 13–41.
[25] Wei W W S, 2006 Time Series Analysis : Univariate and Multivariate Methods 2nd ed. New York, NY: Pearson Addison Wesley.