Model Identification and Derivation for Double Seasonal Integrated Moving Average (DSARIMA) Model

Puteri Aiman Syahirah Rosman(✉) and Nur Haizum Abd Rahman

Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia
GS58462@student.upm.edu.my, nurhaizum_ar@upm.edu.my

Abstract. Double Seasonal Autoregressive Integrated Moving Average (DSARIMA) model is an extension of the single SARIMA that is incorporated in modelling data with two seasonality. Model identification, parameter estimation and diagnostic checking are the steps in the modelling. However, the model identification is the most crucial stage as it provides the information used in the next step. Thus, this study extended the derivation of the model identification for DSARIMA in all three models which are additive, multiplicative and subset. The daily and weekly seasonality which can be indicated by 24 and 128 were used in this study with the derivation involving correlation and covariance from the general form of both seasonal and non-seasonal parts. The derivation results were shown for ARIMA (0, 0, 1) $(0, 0, 1)^{24}(0, 0, 1)^{168}$, ARIMA (0, 0, [1, 24, 25, 168, 169, 192, 193]) and ARIMA (0, 0, [1, 24, 168]) for multiplicative, subset and additive models, respectively. In conclusion, this study gives a valuable insight into the model identification step in DSARIMA models.

Keywords: DSARIMA · identification · additive · multiplicative · subset

1 Introduction

A time series is a collection of observations made over a period of time. A monthly sequence of the quantity of items delivered from a factory, a weekly series of the number of road accidents, daily rainfall amounts and hourly observations of the yield of a chemical process are all examples of time series dataset [1]. In 1970, George Box and Gwilym Jenkins proposed the Box-Jenkins method. The method is based on the assumption that the process that created the time series may be approximated using either an ARMA or an ARIMA model, depending on whether it is stationary or non-stationary. ARIMA model can only be applied to stationary time series data. If the data is not stationary, differencing need to be done first to make the data stationary [2]. Box-Jenkins model is an iterative process that consist of three steps which are identification, estimation and diagnostic checking. Figure 1 shows the step of Box-Jenkins model [3].

ARIMA method are widely used method to time series analysis. Humaira, Nursuprianah and Darwan (2020) used the Box-Jenkins method of the time series analysis to

N. Wahi et al. (Eds.): ICMSS 2022, ACSR 98, pp. 141–149, 2023.
https://doi.org/10.2991/978-94-6463-014-5_14

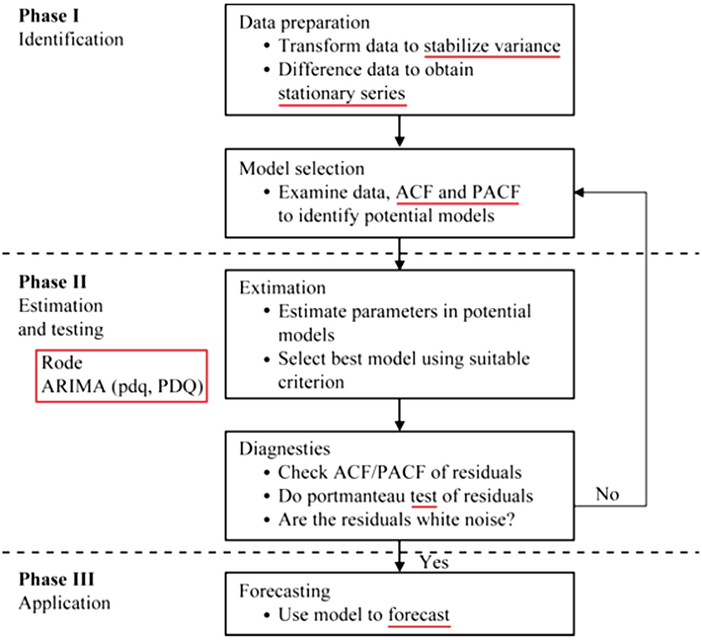

Fig. 1. Box-Jenkins Methodology

forecast the number of schizophrenia disorder disease. Meanwhile, this method are also being used to forecast the exchange rate of the Jordanian Dinar versus the US Dollar [4]. Multiplicative SARIMA is being applied to forecasting the solar radiation [5].

Although many previous papers have concentrated on model estimation, model identification is actually the most crucial stage in building ARIMA models, because false model identification will cause the wrong stage of model estimation and increase the cost of reidentification. In particular of DSARIMA models, most of previous papers usually used directly the multiplicative model without testing whether the multiplicative parameter was significant. It means that the multiplicative DSARIMA models assume that there is a significant parameter as a result of multiplicative between non-seasonal and seasonal parameters [6]. However, the subset and additive relationship may exist in DSARIMA model. Thus, the objective of this paper is to obtain the significance lag that will produce in subset, additive and multiplicative DSARIMA model.

The paper is organized as follows: a brief theoretical review about the time series, Box-Jenkins methodology, ARIMA, SARIMA, DSARIMA, autocorrelation (ACF) and partial autocorrelation (PACF) functions of subset, multiplicative, and additive DSARIMA models and conclusion.

2 Methodology

2.1 DSARIMA Model

Autoregressive Integrated Moving Average, or ARIMA, is a forecasting univariate time series method introduced in the 1970s. ARIMA models are a form of Box-Jenkins model where the terms ARIMA and Box-Jenkins are used interchangeably. One of the attractive features of the Box-Jenkins approach for forecasting is that ARIMA processes are a very rich class of possible models and it is usually possible to find a process which provides an adequate description to the data. This model has originated from the autoregressive model (AR), the moving average model (MA) and the combination of the AR and MA, the ARMA models [6]. The time series is assumed to be stationary in the Box-Jenkins model. To attain stationarity, Box and Jenkins advocate differencing non-stationary series one or more times. As a result, an ARIMA model is created, with the "I" standing for "Integrated." [7]. The full model can be written as follows [8]:

$$\emptyset_p(B)(1-B)^d Z_t = \theta_q(B)\alpha_t \tag{1}$$

where $\emptyset_p(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p$, $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$. Z_t is appropriately transformed load demand in period t; $(1-B)^d$ is the non-seasonal differencing operator, B is the backward shift operator; and α_t is the purely random process. We call this an $ARIMA(p, d, q)$ model, where p is the order of the autoregressive part, d is the degree of first differencing involved and q is the order of the removing average. In addition to the general ARIMA model, namely non-seasonal $ARIMA(p, d, q)$ model, we should also consider some periodical time series. The periodicity of periodic time series is usually due to seasonal changes [9]. This is an extension of ARIMA which is also known as Seasonal ARIMA (SARIMA) that explicitly supports univariate time series data with a seasonal component. It adds three new hyper parameters to specify the autoregressive (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality. This seasonality includes year, month, days etc. To deal with seasonality, the ARIMA model is extended to a general multiplicative seasonal ARIMA (SARIMA) model which is defined as follows [1]:

$$\emptyset_p(B)\Phi_p(1-B)^d\left(1-B^s\right)^D Z_t = \theta_q(B)\Theta_Q\left(B^s\right)\alpha_t \tag{2}$$

where

$\emptyset_p(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p$
$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$
$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$
$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$.

Z_t is appropriately transformed load demand in period t; $(1-B)^d$ and $(1-B^s)^D$ are the nonseasonal and seasonal differencing operators respectively; B is the backward shift operator; and α_t is the purely random process. If the integer D is not zero, then seasonal differencing is involved. The above model is called a SARIMA model of order $(p, d, q)\times(P, D, Q)_s$. If d is non-zero, then there is a simple differencing to remove trend, while seasonal differencing, $(1-B^s)^D$ may be used to remove seasonality. In certain type of data, there exist a double seasonal pattern and therefore a Double SARIMA (DSARIMA) Model is developed to produce a more accurate forecast. The general multiplicative double seasonal ARIMA model is as follows [1]:

$$\begin{aligned}&\emptyset_p(B)\Phi_{P1}\left(B^{S1}\right)\prod\nolimits_{P2}\left(B^{S2}\right)(1-B)^d\left(1-B^{S1}\right)^{D1}\left(1-B^{S2}\right)^{D2}Z_t\\&=\theta_q(B)\Theta_{Q1}\left(B^{s1}\right)\Psi_{Q2}\left(B^{s1}\right)\alpha_t\end{aligned} \tag{3}$$

where

$\phi_p(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \cdots - \phi_p B^p$
$\Phi_{P1}(B^{S^1}) = 1 - \Phi_1 B^{S^1} - \Phi_2 B^{2S^1} - \cdots - \Phi_{P1} B^{P1S^1}$
$\Pi_{P2}(B^{S^2}) = 1 - \Pi_1 B^{S2} - \Pi_2 B^{2S^2} - \cdots - \Pi_{P2} B^{P2S^2}$
$\theta_q(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - \cdots - \theta_q B^q$
$\Theta_{Q1}(B^{S^1}) = 1 - \Theta_1 B^{S^1} - \Theta_2 B^{2S^1} - \cdots - \Theta_{Q1} B^{Q1S^1}$
$\Psi_{Q2}(B^{S^2}) = 1 - \Psi_1 B^{S^2} - \Psi_2 B^{2S^1} - \cdots - \Psi_{Q2} B^{Q^2S^1}$

Z_t is appropriately transformed load demand in period t; B is the backward shift operator; $\emptyset_p(B)$ and θ_q (B are regular autoregressive and moving average polynomials of orders p and q; $\Phi_{P1}(B^{S1})$, $\prod_{P2}(B^{S2})$, $\Theta_{Q1}(B^{s1})$ and $\Psi_{Q2}\left(B^{s1}\right)$ are autoregressive and moving average polynomials of orders P_1, P_2, Q_1 and Q_2; S_1 and S_2 are the seasonal periods; d, D_1 and D_2 are the orders of integration; at is a white noise process with zero mean and constant variance. In this study, we choose the simplest form of DSARIMA model and we let $p = 0, q = 1, P_1 = 0, P_2 = 0, Q_1 = 1, Q_2 = 1, S_1 = 24$ and $S_2 = 168$.

2.2 ACF and PACF

The stationarity of the time series is tested as the first step in the modelling procedure. To acquire a fair estimate of stationarity, utilise the partial auto correlation function (PACF) and auto correlation function (ACF) plots of the time series. The ACF measures the correlation of a time series value with other values from the same time series at various delays. PACF evaluates the connection between a time series value and a value with a different lag. PACF, on the other hand, ignores other values at other delays when determining the correlation for a particular lag value [10]. We have a stationary time series on our hands if the ACF does not reflect any meaningful value after a few delays or the PACF has a sharp cutoff after the initial value [11].

3 Results and Discussion

Three forms of DSARIMA models which are additive, multiplicative and subset were selected. The theoretical explanation about ACF and PACF for these three models was focusing on non-seasonal and the double seasonal moving average orders. i.e.

ARIMA (0, 0, [1, 24, 25, 168, 169, 192, 193]), ARIMA (0, 0, 1)(0, 0, 1)^{24}(0, 0, 1)^{168}, and ARIMA (0, 0, [1, 24, 168]),for subset, multiplicative and additive model, respectively.

3.1 Subset DSARIMA Model

The generalized form of ARIMA (0, 0, [1, 24, 25, 168, 169, 192, 193]) model, also known as subset DSARIMA, can be written as

$$\begin{aligned} z_t - \mu = {} & \alpha_t - \theta_1\alpha_{t-1} - \theta_{24}\alpha_{t-24} + \theta_{25}\alpha_{t-25} - \theta_{168}\alpha_{t-168} + \theta_{169}\alpha_{t-169} \\ & + \theta_{192}\alpha_{t-192} - \theta_{193}\alpha_{t-193} \end{aligned} \tag{4}$$

where $\theta_1, \theta_{24}, \theta_{25}, \theta_{168}, \theta_{169}, \theta_{192}$ and θ_{193} denotes the parameters of MA orders. From Eq. (4), this subset model needs to estimate seven number of parameters. By using mathematical statistics, the following ACF of this model is obtained:

$$\rho_k = \begin{cases} \frac{-\theta_1+\theta_{24}\theta_{25}+\theta_{168}\theta_{169}+\theta_{192}\theta_{193}}{1+\theta_1^2+\theta_{24}^2+\theta_{25}^2+\theta_{168}^2+\theta_{169}^2+\theta_{192}^2+\theta_{193}^2}, & k = 1 \\ \frac{\theta_1\theta_{24}+\theta_{169}\theta_{192}}{1+\theta_1^2+\theta_{24}^2+\theta_{25}^2+\theta_{168}^2+\theta_{169}^2+\theta_{192}^2+\theta_{193}^2}, & k = 23 \\ \frac{-\theta_{24}+\theta_1\theta_{25}+\theta_{168}\theta_{192}+\theta_{169}\theta_{193}}{1+\theta_1^2+\theta_{24}^2+\theta_{25}^2+\theta_{168}^2+\theta_{169}^2+\theta_{192}^2+\theta_{193}^2}, & k = 24 \\ \frac{-\theta_{25}+\theta_{168}\theta_{193}}{1+\theta_1^2+\theta_{24}^2+\theta_{25}^2+\theta_{168}^2+\theta_{169}^2+\theta_{192}^2+\theta_{193}^2}, & k = 25 \\ \frac{\theta_1\theta_{168}+\theta_{24}\theta_{192}+\theta_{25}\theta_{193}}{1+\theta_1^2+\theta_{24}^2+\theta_{25}^2+\theta_{168}^2+\theta_{169}^2+\theta_{192}^2+\theta_{193}^2}, & k = 167 \\ \frac{-\theta_{168}+\theta_1\theta_{169}+\theta_{24}\theta_{193}}{1+\theta_1^2+\theta_{24}^2+\theta_{25}^2+\theta_{168}^2+\theta_{169}^2+\theta_{192}^2+\theta_{193}^2}, & k = 168 \\ \frac{-\theta_{169}}{1+\theta_1^2+\theta_{24}^2+\theta_{25}^2+\theta_{168}^2+\theta_{169}^2+\theta_{192}^2+\theta_{193}^2}, & k = 169 \\ \frac{\theta_1\theta_{191}}{1+\theta_1^2+\theta_{24}^2+\theta_{25}^2+\theta_{168}^2+\theta_{169}^2+\theta_{192}^2+\theta_{193}^2}, & k = 191 \\ \frac{-\theta_{192}+\theta_1\theta_{193}}{1+\theta_1^2+\theta_{24}^2+\theta_{25}^2+\theta_{168}^2+\theta_{169}^2+\theta_{192}^2+\theta_{193}^2}, & k = 192 \\ \frac{-\theta_{193}}{1+\theta_1^2+\theta_{24}^2+\theta_{25}^2+\theta_{168}^2+\theta_{169}^2+\theta_{192}^2+\theta_{193}^2}, & k = 193 \\ 0, & others. \end{cases} \tag{5}$$

The theoretical ACF and PACF of Eq. (5) are presented in Fig. 2.

From the Fig. 2, we can clearly see that the lag is significant at lag number 1, 23, 24, 25, 167, 168, 169, 191, 192 and 193. Thus this additive subset DSARIMA model needs to estimate seven different number of parameters at ten number of lags.

3.2 Multiplicative DSARIMA Model

The generalized form of ARIMA $(0, 0, 1)(0, 0, 1)^{24}(0, 0, 1)^{168}$ model, also known as multiplicative DSARIMA, can be written as

$$\begin{aligned} z_t - \mu = {} & \alpha_t - \theta_1\alpha_{t-1} - \theta_{24}\alpha_{t-24} + \theta_1\theta_{24}\alpha_{t-25} - \theta_{168}\alpha_{t-168} + \theta_1\theta_{168}\alpha_{t-169} \\ & + \theta_{24}\theta_{168}\alpha_{t-192} - \theta_1\theta_{24}\theta_{168}\alpha_{t-193} \end{aligned} \tag{6}$$

where θ_1, θ_{24} and θ_{168} represents the parameters of non-seasonal, first seasonal and second seasonal MA order, respectively. This model is the same as the subset DSARIMA model in Eq. (4) when $\theta_{25} = -\theta_1\theta_{24}, \theta_{169} = -\theta_1\theta_{168}, \theta_{192} = -\theta_{24}\theta_{168}$ and $\theta_{193} = \theta_1\theta_{24}\theta_{168}$. Therefore, it can be concluded that multiplicative model is a part of subset

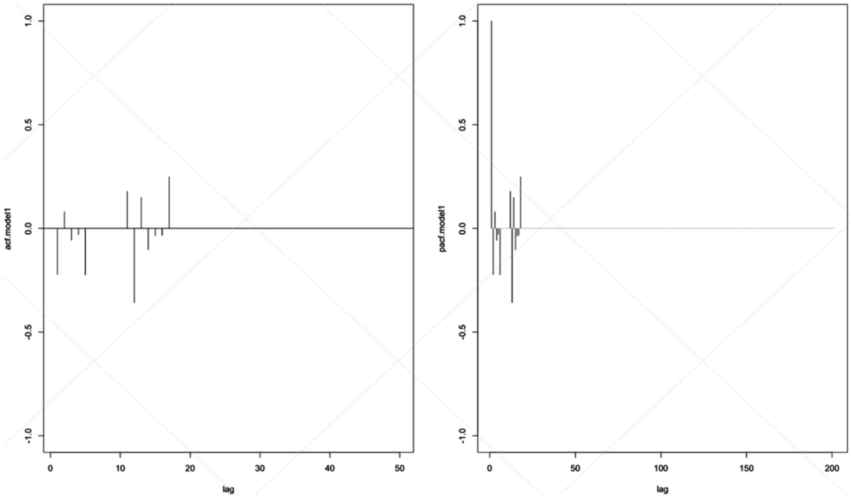

Fig. 2. Theoretical ACF and PACF of subset DSARIMA

model. From Eq. (6), this model needs to estimate seven different parameters. By using mathematical statistics, the following ACF of this model is obtained:

$$\rho_k = \begin{cases} \frac{-\theta_1}{1+\theta_1^2}, & \text{k} = 1 \\ \frac{\theta_1\theta_{24}}{(1+\theta_1^2)(1+\theta_{24}^2)}, & \text{k} = 23, 25 \\ \frac{-\theta_{24}}{(1+\theta_{24}^2)}, & \text{k} = 24 \\ \frac{\theta_1\theta_{168}}{(1+\theta_1^2)(1+\theta_{168}^2)}, & \text{k} = 167, 169 \\ \frac{-\theta_{168}}{(1+\theta_{168}^2)}, & \text{k} = 168 \\ \frac{-\theta_1\theta_{24}\theta_{168}}{(1+\theta_1^2)(1+\theta_{24}^2)(1+\theta_{168}^2)}, & \text{k} = 191, 193 \\ \frac{\theta_{24}\theta_{168}}{(1+\theta_{24}^2)(1+\theta_{168}^2)}, & \text{k} = 192 \\ 0, & \text{others} \end{cases} \tag{7}$$

Equation (7) shows that the significant lag is at lag number 1, 23, 24, 25, 167, 168, 169, 191, 192 and 193 with lag 167 is equal with lag 169 and ACF values at lag 191 is equal with lag 193. The theoretical ACF and PACF of Eq. (7) are presented in Fig. 3.

Based on the above plot, multiplicative DSARIMA model need to estimate seven number of parameters at ten number of different lags.

3.3 Additive DSARIMA Model

The generalized form of ARIMA (0, 0, [1, 24, 168]) model, also known as additive DSARIMA, can be written as

$$z_t - \mu = \alpha_t - \theta_1\alpha_{t-1} - \theta_{24}\alpha_{t-24} - \theta_{168}\alpha_{t-168} \tag{8}$$

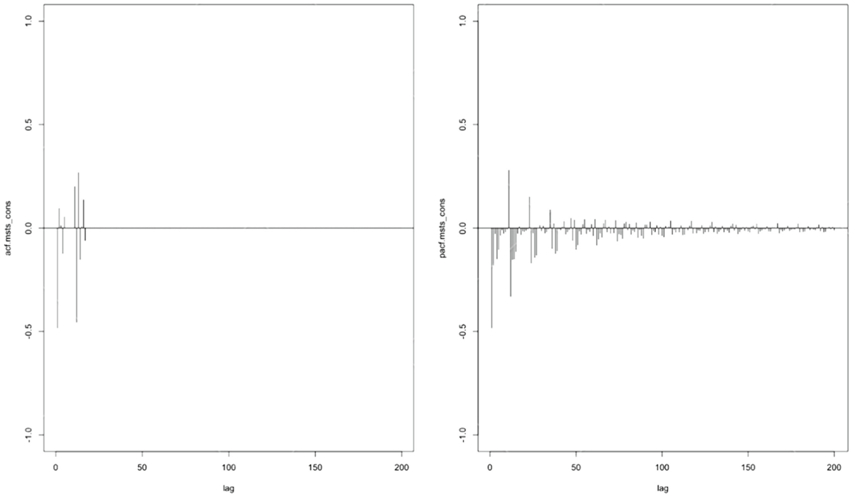

Fig. 3. Theoretical ACF and PACF of multiplicative DSARIMA

where θ_1, θ_{24} and θ_{168} represents the parameters of non-seasonal, first seasonal and second seasonal MA order, respectively. This model is the same with subset DSARIMA model in Eq. (4) when $\theta_{25} = \theta_{169} = \theta_{192} = \theta_{193} = 0$. Therefore, it can be concluded that additive model is also a part of subset model. In addition, this additive model in Eq. (8) could also be seen as subset ARIMA model with lower order than model in Eq. (4). Thus, by using mathematical statistics, the following ACF of this model is obtained:

$$\rho_k = \begin{cases} \frac{-\theta_1}{1+\theta_1^2+\theta_{24}^2+\theta_{168}^2}, & \text{k} = 1 \\ \frac{\theta_1\theta_{24}}{1+\theta_1^2+\theta_{24}^2+\theta_{168}^2}, & \text{k} = 23 \\ \frac{-\theta_{24}}{1+\theta_1^2+\theta_{24}^2+\theta_{168}^2}, & \text{k} = 24 \\ \frac{\theta_1\theta_{168}}{1+\theta_1^2+\theta_{24}^2+\theta_{168}^2}, & \text{k} = 167 \\ \frac{-\theta_{168}}{1+\theta_1^2+\theta_{24}^2+\theta_{168}^2}, & \text{k} = 168 \\ 0, & \text{others} \end{cases} \tag{9}$$

Equation (8) shows that the main difference between additive and the other two models (subset and multiplicative) is this model needs to estimate only three parameters where the significant lag is at lag number 1, 23, 24, 167 and 168 only. The theoretical ACF and PACF of Eq. (9) are presented in Fig. 4.

Based on the above plot, additive DSARIMA model need to estimate three number of parameters at five number of different lags.

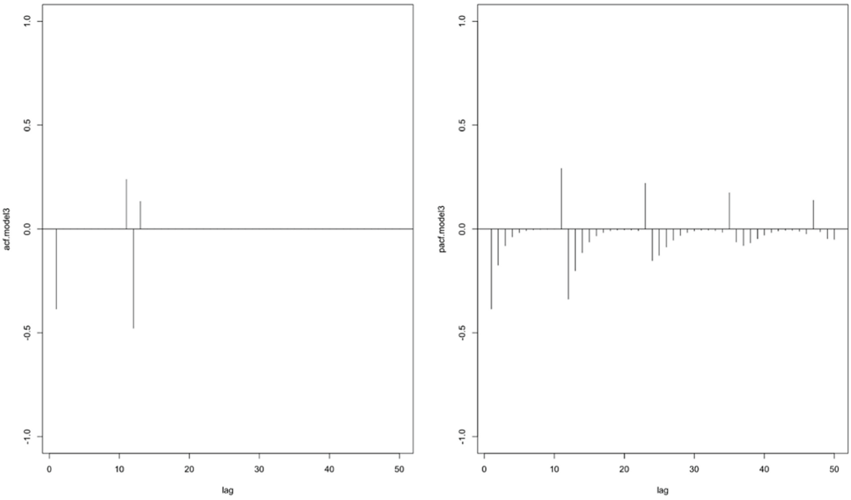

Fig. 4. Theoretical ACF and PACF of additive DSARIMA

4 Conclusion

Accurate forecasting is very crucial in order to produce a correct result and this need to start from model identification since it is the first step of Box-Jenkins modeling. In this paper we have discussed the model identification of Double seasonal ARIMA (DSARIMA) for all three models which are multiplicative, additive and subset. Often researchers directly used multiplicative DSARIMA model without checking if the multiplicative model is the best model fit to the data. This will definitely produce a less accurate forecasting and will result in poor decision making. It proved in our result and discussion where we can see from the obtained ACF, each model needs to estimate different number of parameters at different number of lags. Subset DSARIMA model need to estimate seven parameters at lag number 1, 23, 24, 25, 167, 168, 169, 191, 192 and 193. Meanwhile for multiplicative DSARIMA model, it needs to estimate also seven parameters with different value as the subset DSARIMA model at the same number of lags also. As for the subset DSARIMA model, it needs to estimate three parameters at significant lag number 1, 23, 24, 167, 168. This clearly shows that the different model produces different lag with different parameter used. Hence, it is very important to use the correct model that fit best to the data in order to increase forecasting accuracy.

Acknowledgments. The author acknowledges the support given by Universiti Putra Malaysia under project number GP-IPM/2017/9587700.

References

1. Box, G.E., Jenkins, G., Reinsel, G., Ljung, G.: Time Series Analysis, 5th edn. (2015)
2. Mado, I., Soeprijanto, A., Suhartono, S.: Applying of double seasonal ARIMA model for electrical power demand forecasting at PT. PLN Gresik Indonesia. Int. J. Electr. Comput. Eng. (IJECE) **8**(6), 4892 (2018). https://doi.org/10.11591/ijece.v8i6.pp4892-4901
3. Makridakis, S., Wheelwrigt, S.C., Mcgee, V.: Forecasting: methods and applications, 2nd edn. Wiley, New York (1983)
4. Humaira, S.P., Nursuprianah, I., Darwan, D.: Forecasting of the number of schizophrenia disorder by using the Box-Jenkins of time series analysis. J. Robot. Control (JRC) **1**(6), 213–219 (2020). https://doi.org/10.18196/jrc.1640
5. Farhan, D., Fakhir, M.: Forecasting the exchange rates of the Iraqi Dinar against the US Dollar using the time series model (ARIMA). Int. J. Eng. Manag. Res. **09**(06), 50–55 (2019). https://doi.org/10.31033/ijemr.9.6.8
6. Shadab, A., Said, S., Ahmad, S.: Box–Jenkins multiplicative ARIMA modeling for prediction of solar radiation: a case study. Int. J. Energy Water Resour. **3**(4), 305–318 (2019). https://doi.org/10.1007/s42108-019-00037-5
7. Suhartono: Time series forecasting by using seasonal autoregressive integrated moving average: subset, multiplicative or additive model. J. Math. Stat. **7**(1), 20–27 (2011)
8. Stephanie: ARIMA (Box-Jenkins Models): Autoregressive Integrated Moving Average. Statistics How To (2016). https://www.statisticshowto.com/arima/
9. Wei, W.W.: Time Series Analysis, Univariate and Multivariate Methods, 2nd edn. Pearson, Addison Wesley, New York (2006)
10. Hyndman, R.J., Athanasopoulos, G.: Forecasting: Principles and Practice, 2nd edn. OTexts, Melbourne. https://otexts.com/fpp2/. Accessed 31 Jan 2022
11. Flores, J.H.F., Engel, P.M., Pinto, R.C.: Autocorrelation and partial autocorrelation functions to improve neural networks models on univariate time series forecasting. In: The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2012)