Journal of Economics, Finance and Accounting Studies (JEFAS)
www.jefas.one

Original Research Article

Modeling and Forecasting the Algerian Stock Exchange Using the Box-Jenkins Methodology

Mohamed Samir Boudrioua[1]* & Abderrahmane Boudrioua[2]
[1]*Ronin institute, Montclair, NJ, 07043, USA*
[2]*Universiti Utara Malaysia, Sintok, Kedah, 06010, MY*
Corresponding Author: Mohamed Samir Boudrioua, E-mail: mohamedsamirboudrioua@gmail.com

ARTICLE INFO

Article History
Received: May 03, 2020
Accepted: June 19, 2020
Volume: 1
Issue: 1

KEYWORDS

The Algiers Stock Exchange; Forecasting; Box-Jenkins methodology; SARIMA model; Akaike Information Criterion (AIC); Autocorrelation, Heteroskedasticity

ABSTRACT

The Algiers Stock Exchange (ASE) is the only stock exchange in Algeria. It's one of the newest and smallest emerging stock exchanges in the world. The focus of this paper is to model and forecast monthly returns of the ASE index (DZAIRINDEX) using The Box- Jenkins methodology. The period of this study is from June 2010 to May 2020. We split the data into training and testing returns datasets. According to Akaike's Information Criterion (AIC) estimator, the Seasonal Autoregressive Integrated Moving Average SARIMA (2,0,0)(0,0,1)12 with zero mean is chosen as the best model for forecasting the monthly DZAIRINDEX returns. Diagnostic tests show that the fitted model is adequate, where the residuals of this model are normally distributed with no autocorrelation and no heteroskedasticity. We evaluate our model forecasts on returns testing datasets for 11 steps ahead. We get white noise residuals without heteroskedasticity, which confirm the adequacy of our model. Based on different measures of forecast accuracy such as ME, MAE, RMSE, MASE, we show that the forecast accuracy of SARIMA (2,0,0)(0,0,1)12 is acceptable and this model performs much better than a naïve model. The forecast of the whole returns dataset for one year ahead using this model shows a slight increasing fluctuations trend. These results could be used by the financial communities in Algeria to deal with stock exchange risks and to improve their decisions.

Introduction

The Algiers Stock Exchange (ASE) is one of the newest and smallest stock exchanges in the world. Due to this fact, there aren't a lot of studies that focused on this emerging stock exchange. According to the ASE official website (http://www.sgbv.dz), the official listing of Algiers Stock Exchange for transferable securities includes an equity securities market and a debt securities market. Where the market for equity securities comprises a Main market for large enterprises, and Small and medium-sized enterprises Market reserved for the Small and Medium Enterprises.

The main market includes five companies: SAIDAL Group activating in the pharmaceutical sector, EGH EL AURASSI activating in the tourism sector, ALLIANCE INSURANCE activating in the insurance sector, NCA-Rouiba activating in the agri-food sector and BIOPHARM activating in the pharmaceutical sector. The Algiers Stock Exchange uses a reference index called DZAIRINDEX includes the total listing companies in the main market.

The concept of forecasting stock market returns has become very important in finance, where it help investors to understand the stock market trend and make their decisions, minimize investment risks associated with the market and improve their returns (Lin and Yu, 2009). There are several methods of forecasting time series such as exponential smoothing methods, simultaneous-equation regression models and autoregressive integrated moving average models (ARIMA) models, Neural networks models and others (Gujarati, 2004). The most used and popular model in forecasting financial time series over Short time periods is the Box-Jenkins model or ARIMA model (Box and Jenkins ,1970). The emphasis of these methods is on analyzing the stochastic properties of economic time series on their own (Gujarati, 2004). Box and Jenkins (1976) developed an extension of ARIMA model which is seasonal ARIMA (SARIMA) model to forecast seasonal time series. This model requires that the data be seasonally differenced to achieve stationarity condition.

Several studies have used the Box-Jenkins methodology for forecasting different stock markets in the world. Al-Shiab (2006) used the Box-Jenkins method to forecast the Amman stock exchange. Katircioglu1 and Al-khaza'leh (2016) applied the same method for forecasting services sector volatility in the Amman Stock Exchange. Wahyudi (2017) used also ARIMA model to forecast the Indonesia Stock Price. Ashik and Kannan (2017) forecasted National Stock Exchange in India using the same models. There are also many studies have applied this methodology in various fields of forecasting. For example, Wagner (2010) compared seasonal ARIMA model and vector time series model for forecasting daily demand in cash supply chains. Chang et al (2012) forecasted monthly precipitation in Yantai, China with seasonal ARIMA model. The results of these studies show that this model performed well and gives less error compared with other models over short time periods.

The objective of this study is to fit the best forecasting model for the Algiers Stock Exchange returns using the Box -Jenkins methodology. Monthly returns for the index DZAIRINDEX were used in this task over the period (June 2010- May 2020). We evaluate the forecast of the chosen model seasonal ARIMA $(2,0,0)(0,0,1)_{12}$ on a test returns datasets from July 2019 to May 2020, then we forecast the whole returns datasets for one year ahead. This forecast could be helpful for policymakers and investors in Algeria to make their decisions.

The rest of the paper is outlined as follows: Some literature review in section two. The third section describes the Box-Jenkins methodology. The fourth section presents the empirical findings with discussions. Finally, we present the conclusion with some perspectives in the fifth section.

Literature Review

There is a large literature on forecasting stock exchanges using the Box-Jenkins methodology. Al-Shiab (2006) used a general daily index of the Amman Stock Exchange, over the period (4/1/2004 - 10/8/2004) to examine the univariate ARIMA forecasting model. He found that the forecast was not consistent with actual performance since the Amman Stock Exchange follow most closely the Efficient Market Hypothesis in its weak form, during the period of the prediction (11/8/2004–19/8/2004). Similarly, to Al–Shiab (2006), Katircioglu and Al-khaza'leh (2016) forecasted services sector volatility in Amman Stock Exchange with ARIMA models, using historical indices data accumulated daily over the period (3/1/2010-10/5/2015). They found that the best model for forecasting this stock exchange is ARIMA (0, 0, 1).

Wahyudi (2017) conducted the prediction of Indonesia stock price using Autoregressive Integrated Moving Average model. He used the daily Indonesia Composite Stock Price Index (CSPI) over the period (04/01/2010 -05/12/2014). He found that ARIMA model has a strong potential for short-term prediction and can compete favorably with the existing techniques for stock price prediction.

Ashik and Kannan (2017) evaluated and predicted the trend of upcoming trading days of the Nifty 50 stock market in India using the Box-Jenkins methodology. They found that the prediction accuracy is more suitable for the Nifty 50 closing stock price. They concluded that the closing stock price of Nifty 50 taken in this study shows a slight decreasing fluctuations trend for upcoming trading days.

Concerning examples of the other fields of forecasting with the Box-Jenkins method, Wagner (2010) forecasted Daily Demand in Cash Supply Chains using a SARIMA model and vector time series model. He found that the seasonal ARIMA model resulted in a higher forecasting accuracy compared to the vector time series model. According to the cited author, this result confirms the benefit of advanced forecasting techniques for daily forecasts.

Chang et al (2012) predicted the monthly precipitation in Yantai, China, over the period from 1961 to 2011with seasonal ARIMA model. Their results showed that the model fitted the data well and the stochastic seasonal fluctuation was successfully modeled. Chang et al (2012) concluded that Seasonal ARIMA model is a proper method for modeling and predicting time series of monthly precipitation.

Methodology

This study adopts the Box Jenkins methodology (Box and Jenkins, 1970). This method consists of the following four steps:

Model identification

Through this step the degree of ARIMA (p,d,q)(P,D,Q)$_S$ model is determined, where:
S is the periodicity, p is the autoregressive (AR) order, d is the non-seasonal differencing order, q is the moving average (MA) order, P is the seasonal AR order, D is the seasonal differencing and Q is the seasonal MA order.
Recall that ARIMA (p,d,q) model is an integrated ARMA(p,q) including differencing in order to get a stationary time series. The general form of the model is written as follow (Box et al, 2015):

$$\varphi(B)x_t = \phi(B)\nabla^d x_t = \theta_0 + \theta(B)\omega_t . \qquad (1)$$

Where:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q.$$
$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p.$$

With: x_t is the studied time series, $\varphi(B)$ is called the generalized autoregressive operator, $\omega_t \sim wn(0, \sigma^2)$ is the Gaussian white noise, ∇^d is the Differencing operator of order d given by $\nabla^d = (1 - B)^d$. B is the backshift operator defined by $B^k x_t = x_{t-k}$, B^k means backshift k times.

When $\theta_0 = 0$ the general model can be written as (Box et al, 2015):

$$x_t = \varphi_1 x_{t-1} + \ldots + \varphi_{p+d} x_{t-p-d} - \theta_1 \omega_{t-1} - \ldots - \theta_q \omega_{t-q} + \omega_t . \qquad (2)$$

The multiplicative seasonal autoregressive integrated moving average model, or SARIMA (p, d, q) $(P, D, Q)_S$ model is given by (Box et al, 2015):

$$\Phi_p(B^S)\phi_P(B)\nabla_S^D \nabla^d x_t = \Theta_Q(B^S)\theta_q(B)\omega_t . \qquad (3)$$

Where:

The seasonal difference component is represented by: $\nabla_S^D = (1 - B^S)^D$.

$\Phi_p(B^S)$ and $\Theta_Q(B^S)$ are polynomials in B represents respectively the seasonal autoregressive and moving average components, given as follow:

$$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \ldots - \Phi_P B^{PS}.$$
$$\Theta(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \ldots - \Theta_Q B^{QS}.$$

Stationarity

The Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979; Said and Dickey, 1984) was employed to determine the order of difference d and to check the stationarity of our time series by testing the presence of unit roots in it.

The objective of this test is to examine the null hypothesis that a unit root is present in a time series x_t , which means that x_t is not stationary. The ADF test is performed in the follows form:

$$\nabla x_t = \alpha + \beta t + \gamma x_{t-1} + \delta_1 \nabla x_{t-1} + \ldots + \delta_p \nabla x_{t-p} + \omega_t . \qquad (4)$$

Where: α is a constant, β the coefficient on a time trend and *p* the lag order of the autoregressive process Cheung and Lai, 1995). If x_t has no trend and no drift then the test can be performed as follows:

$$\nabla x_t = \gamma x_{t-1} + \delta_1 \nabla x_{t-1} + \ \dots + \ \delta_p \nabla x_{t-p} + \ \omega_t . \qquad (5)$$

According to (Brooks, 2008, page 329) the frequency of the data can be used to determine the optimal numbers of lags p for the dependent variable. So, if the data are monthly, we use 12 lags, if the data are quarterly, we use 4 lags, and so on.

The test of HEGY (Hylleberg et al, 1990) can be used to show if there are some seasonal patterns in our time series. The null hypothesis of this test is that a seasonal unit root exists (López-de-Lacalle and Boshnakov, 2019).

Order selection

The ACF and PACF plot of a simulated time series can be used to determine the orders p and q of our model. According to Tsay (2005), if the ACF cuts off at lag q, this lag is the order of MA. If the PACF cuts off at lag p, this lag is the order of AR.

The seasonal MA and AR orders can be determined by showing a significant spike in the seasonal lags of the ACF and PACF functions respectively (Hyndman & Athanasopoulos, 2018).

The Akaike information criteria (AIC)) (Akaike, 1974) was used in order to fit the best model and get the appropriate orders p,d,q and P,D,Q. The model that gives the lowest AIC value would be selected as the best one (Box et al, 2015)[7]. The (AIC) is defined as follows (Box et al, 2015):

$$\text{AIC= (-2) log (maximum likelihood) + 2K} \approx n \log(\hat{\sigma}^2) + 2K .$$

Where:
K is the number of independently adjusted parameters within the model
$\hat{\sigma}^2$ is the maximum likelihood estimate of σ^2.

Model estimation

Maximum likelihood estimation was used to estimate the parameters of the fitted model. The Gaussian Likelihood for a Gaussian ARMA Process with zero mean is given by (Brockwell & Davis, 2002):

$$L(\phi, \theta, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n} r_0 \dots r_{n-1}} exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{n} \frac{(x_j - \hat{x}_j)^2}{r_{j-1}} \right\}.$$

Where: $\hat{x}_j$ and r_j are independent of σ^2.

r_n is the mean squared error given by $r_n = E(y_{n+1} - \hat{y}_{n+1})^2$, with y_t is the transformed process of x_t that verify:

$$y_t = \frac{x_t}{\sigma}, \mathrm{t} = 1, \dots, \mathrm{m}; \mathrm{m} = \max(\mathrm{p}, \mathrm{q}),$$

$$y_t = \frac{\phi(B) x_t}{\sigma}, \mathrm{t} > \mathrm{m} .$$

The maximum likelihood estimators $\hat{\phi}$, $\hat{\theta}$ and $\hat{\sigma}^2$ satisfy the following equations (Brockwell and Davis, 2002, page160):

$$\hat{\sigma}^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n} .$$

Where: $S(\hat{\phi},\hat{\theta})=\sum_{j=1}^{n}\frac{(x_j-\hat{x}_j)^2}{r_{j-1}}$.

$\hat{\phi}$ and $\hat{\theta}$ are the values that minimize $\ell(\phi,\theta)=\ln(\frac{S(\phi,\theta)}{n})+\frac{\sum_{j=1}^{n}r_{j-1}}{n}$.

For further details see (Brockwell and Davis, 2002)[8].

Model diagnostic

The diagnostic checking is necessary to test the adequacy of the selected model. Three diagnostic tests were used in this study:

Autocorrelation test

We used the Ljung-Box test to check autocorrelation between residuals. The statistic of this test is given as follow (Ljung & Box, 1978):

$$\tilde{Q}(r)=n(n+2)\sum_{i=1}^{m}\frac{r_k^2}{(n-k)} .$$

Where*:* $\tilde{Q}(r)$ is distributed as χ_m^2 for a large n,

n is the number of observations and *m* is the maximum number of lags (the degree of freedom),

r_k is the autocorrelation function (for lag k), given by: $r_k=\sum_{i=k+1}^{n}\frac{a_i a_{i-k}}{\sum_{i=1}^{n}a_i^2}$,

$\{a_i\}$ is a sequence of independent and identically distributed $N\left(0,\ \sigma^2\right)$ random deviates.

The hypotheses of Ljung - Box test are:

$H_0: r_1=r_2=...=r_n=0$ (No residuals autocorrelation).

$H_a: r_i\neq 0$ (Residuals autocorrelation).

The decision rule is to reject H_0 if $\tilde{Q}(r)>\chi_m^2$ (Tsay, 2005).

.

Heteroskedasticity test

The Engle's ARCH test (Engle, 1982) was used to show weather the residuals are heteroskedastic or not. The alternative hypothesis of this Lagrange multiplier (LM) test is that ARCH (p) effects are present in a series of residuals (Drachal, 2017)[15].
The procedure of this test consists of the following steps (Engle, 1982, page1000):
Running the OLS regression and saving the residuals.
Regress the squared residuals on a constant and p lags and test TR^2 as χ_p^2 .

Where T is the number of observations and R^2 is the coefficient of multiple correlations.

Normality test

Different tests of normality were used in this study such as kolmogorov-Smirnov (Frank and Massey, 1951), Shapiro-Wilk, Anderson-darling (Stephens, 1986), to check the distribution of our time series and residuals. The null hypothesis of these tests is that the sample is normally distributed.
The Q-Q plot and density plot can be used also to check the distribution of time series and residuals.

Forecasting

Forecasting is predicting the future values of a time series, based on its available current and past values. According to Box et al (2015) an observation x_{t+l} generated by the ARIMA process can be expressed directly in terms of the difference equation as follow:

$$x_{t+l} = \varphi_1 x_{t+l-1} + \ldots + \varphi_{p+d} x_{t+l-p-d} - \theta_1 \omega_{t+l-1} - \ldots - \theta_q \omega_{t+l-q} + \omega_{t+l}, \qquad (6)$$

by taking conditional expectations at time t in the previous equation (6) we get the following forecasts from difference equation (Box et al, 2015):

$$[x_{t+l}] = \hat{x}_t(l) = \varphi_1[x_{t+l-1}] + \ldots + \varphi_{p+d}[x_{t+l-p-d}] - \theta_1[\omega_{t+l-1}] - \ldots - \theta_q[\omega_{t+l-q}] + [\omega_{t+l}],$$

Where: l is the lead time,

$$[x_{t+l}] = E_t[x_{t+l}] = E[x_{t+l} / x_t, x_{t-1}, \ldots].$$

$$[\omega_{t+l}] = E_t[\omega_{t+l}].$$

Different measures of forecast accuracy were used to evaluate our forecast are (Hyndman and Athanasopoulos , 2018)

- Mean error: ME=mean (e_t) .

Where: $e_t = x_t - \hat{x}_t$, x_t is an observed value at time t and $\hat{x}_t$ is the predicted value at the same time t.

- Mean absolute error: MAE=mean (| e_t |).
- Root mean squared error: RMSE= $\sqrt{\text{mean}(e_t^2)}$.

The mean absolute scaled error (Hyndman and Koehler, 2006)*:* MASE=mean (| q_j |), where:

$$q_j = \frac{e_j}{\frac{1}{T-1}\sum_{t=2}^{T} x_t - x_{t-1}}$$

, is the scaled error for non-seasonal time series,

$$q_j = \frac{e_j}{\frac{1}{T-m}\sum_{t=m+1}^{T} x_t - x_{t-m}}$$

, is the scaled error for a seasonal time series.

m is the seasonal period.

Note that the Mean absolute percentage error (MAPE) can't be used in this case study because our data contains zero values.

Results and discussion

4.1. Exploratory data analysis

The data were collected from the official website of Algiers Stock Exchange. Due to a lot of missing values of daily closing prices for the ASE index (DZAIRINDEX), monthly observations were used by taking the closing price values of the ASE index on each month-end. We split the data into training datasets from June 2010 to June 2019, and testing datasets from July 2019 to May 2020. Following Chou(1988) and Emenike (2010), Returns (r_t) were calculated as follow:

$$r_t = \log\left(P_t / P_{t-1}\right).$$

P_t and P_{t-1} represent the closing price of the ASE index for the current month and the previous respectively.

We can see from the plot of training returns time series (r_t) in Figure 1, that our time series has no trend and no drift. We deduce also easily from the Q-Q plot (Figure 2) that r_t time series is not normally distributed.

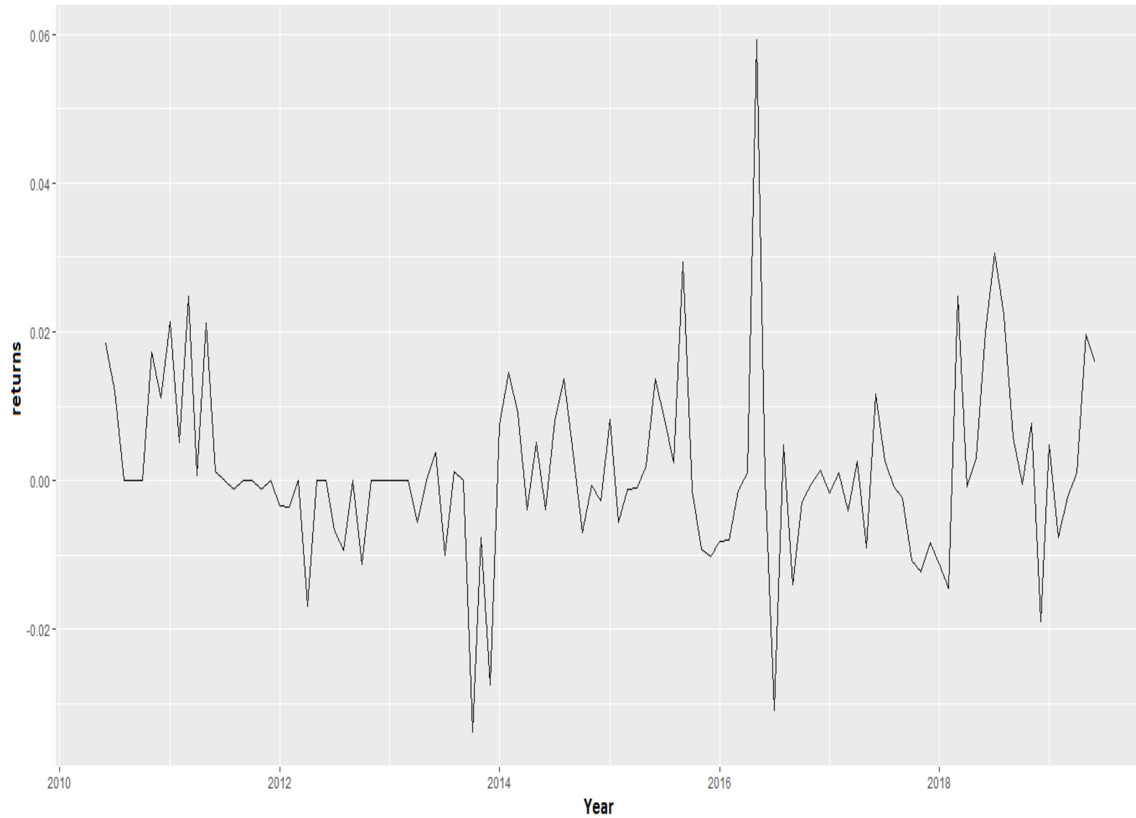

Figure 1: Time series plot of the monthly DZAIRINDEX training returns

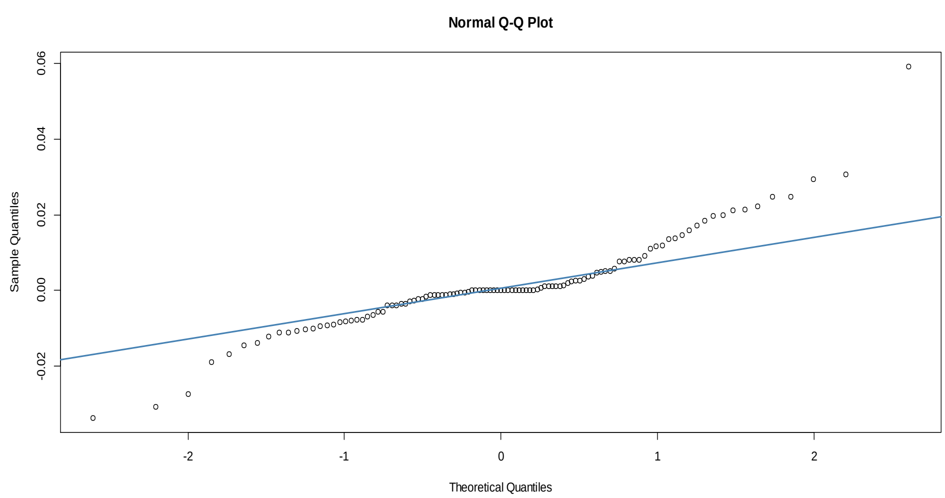

Figure 2: Q-Q Plot of the monthly DZAIRINDEX training returns time series

Table1below shows a summary statistics of r_t time series.

Table 1: Summary statistics of the monthly DZAIRINDEX training returns

	Mean	Sd	Median	Min	Max	Range	Skew	kurtosis
returns	0	0.01	0	-0.03	0.06	0.09	0.89	4.18

The HEGY test (Table 2) shows that r_t time series has some seasonal patterns, where all the p-values of this test statistics are not significant at 0.05 level.

Table 2: HEGY test

HEGY test for unit roots		
Data: r_t		
	Statistic	p-value
t_1	-1.9654	0.2302
t_2	-1.0452	0.192
F_3:4	1.2583	0.2411

Before start applying the Box-Jenkins procedure, outliers were replaced using linear interpolation, then training returns time series was seasonally adjusted by removing the seasonal component using STL decomposition (Hyndman & Athanasopoulos , 2018). See also (Hyndman & Khandakar, 2008) and (Hyndman et al, 2019). Figure 4 shows the STL decomposition for r_t time series.

These two tasks transform r_t into normally distributed time series, which is important in ARIMA model estimation and order selection (Burnham & Anderson, 2002) and in the calculation of prediction intervals Hyndman and Athanasopoulos , 2018). The result of Kolmogorov -Smirnov test in Table 3 shows that r_t can be assumed normally distributed after replacing outliers and seasonality adjustment. The assumption of normality is essential to get efficient and consistent maximum likelihood estimation. Note that in this case study we can't use the Box-Cox transformation (Box & Cox, 1964) because our data contains zero values. Figure 5 shows the plot of training returns time series r_t before and after replacing outliers and seasonality adjustment. In the next of this study, the seasonally adjusted time series r_t with replaced outliers is noted by y_t.

Table 3: Kolmogorov-Smirnov test of normality for y_t time series

Test	Statistic	p-value
Kolmogorov-Smirnov	0.0753	0.5663

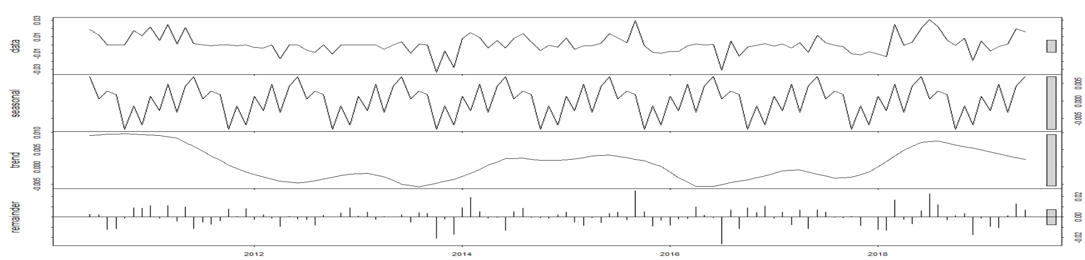

Figure 3: STL decomposition of the training returns r_t time series.

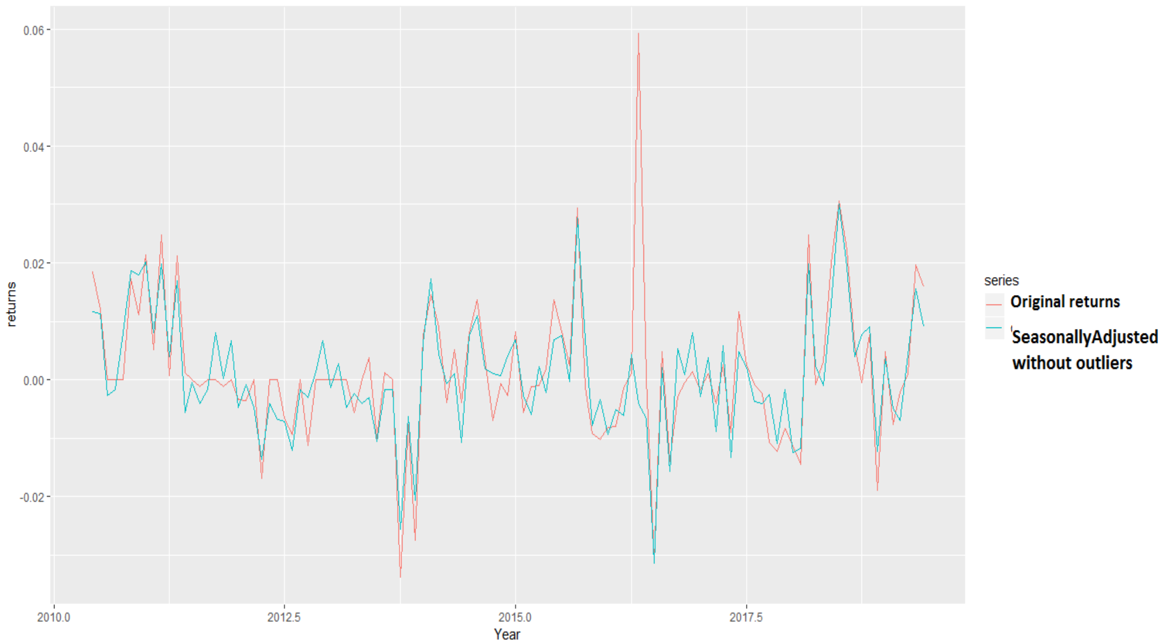

Figure 4: The training returns r_t time series after outliers replace and seasonality adjustment

The Mann-Kendall test McLeod, 2015) can be used to check if the time series y_t has a trend. The results of this test are given in Table 4. At the level of significance 0.05, the p-value of this test statistic is not significant. This leads us to accept the null hypothesis that y_t hasn't a trend.

Table 4: Mann- Kendall trend test

Mann-Kendall trend test		
data: y_t time series		
n = 109	z = -0.47136	p-value =0.6374

We can also confirm that y_t hasn't a drift by running an autoregressive model with drift then we check the significance of the intercept, as it is shown in Table 5. The result of this procedure shows that the intercept is not significant at the 0.05 level. This means that the time series y_t hasn't a drift.

Table 5: Test regression drift

Test regression drift				
Coefficients:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0002396	0.0009827	-0.244	0.80796
z.lag.1	-0.7335279	0.2617452	-2.802	0.00633
z.diff.lag1	-0.0300567	0.2456352	-0.122	0.90291

Model identification

Since y_t is a monthly time series with no drift and no trend we have chosen the ADF test with 12 lags and without a trend and without a drift. The result of this test is given in Table 6. From this table, we see that all the p-values of ADF test statistics are significant at 0.05 level. This means a rejection of the null hypothesis that y_t is not stationary. Hence y_t is stationary.

Figure 6 and Figure 7 show the correlograms of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) respectively. The summary of the candidate ARIMA models according to the AIC estimator is given in Table 7. Based on the results of this estimation, SARIMA $(0,0,2)(0,0,1)_{12}$ model with zero mean is selected as the best model because it shows the lowest AIC value. This means that the time series y_t still has some seasonal patterns.

Table 6: The augmented Dickey Fuller test of stationarity

Augmented Dickey-Fuller Test		
Alternative hypothesis: Stationary		
Type: No drift, no trend		
Lag	ADF	p-value
0	-7.52	0,01
1	-4.84	0,01
2	-4.28	0,01
3	-3.88	0,01
4	-3.81	0,01
5	-3.47	0,01
6	-3.64	0,01
7	-3.40	0,01
8	-3.28	0,01
9	-3.47	0,01
10	-2.90	0,01
11	-3.42	0,01

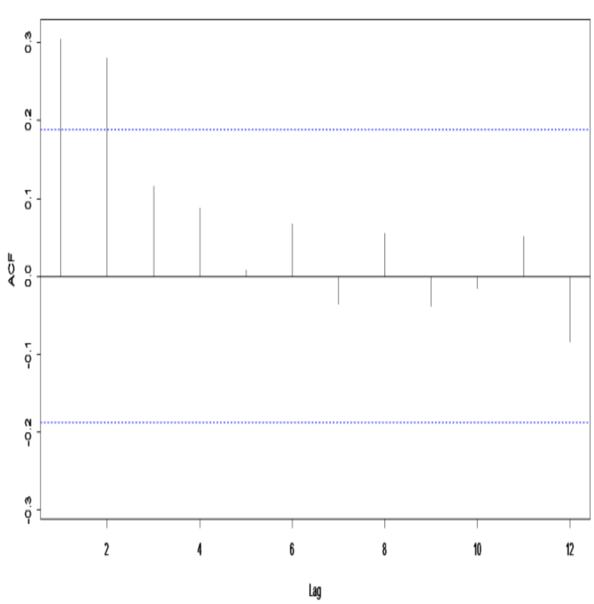

Figure 5: The autocorrelation function of y_t time series

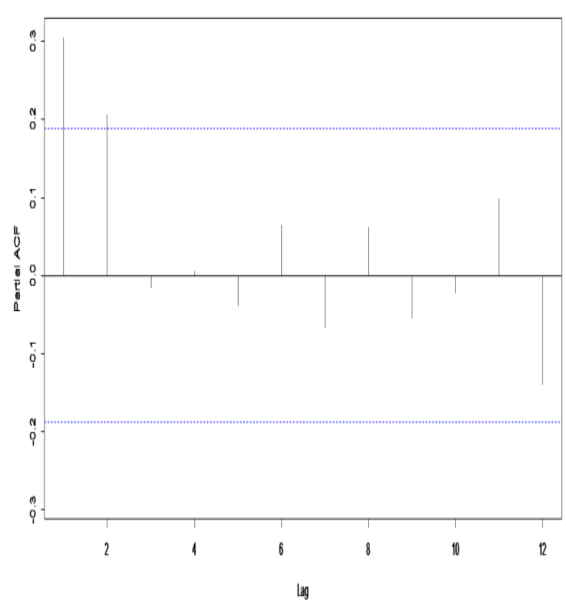

Figure 6:The partial autocorrelation function of y_t time series

Table 7: Summary of the fitted ARIMA models based on AIC estimation.

Models	AIC
ARIMA(2,0,2)(1,0,1)[12] with non-zero mean	Inf
ARIMA(0,0,0) with non-zero mean	-691.2342
ARIMA(1,0,0)(1,0,0)[12] with non-zero mean	-699.5073
ARIMA(0,0,1)(0,0,1)[12]with non-zero mean	-697.1523
ARIMA(0,0,0) with zero mean	-692.5961
ARIMA(1,0,0) with non-zero mean	-700.0314
ARIMA(1,0,0)(0,0,1)[12] with non-zero mean	-700.7933
ARIMA(1,0,0)(1,0,1)[12] with non-zero mean	Inf
ARIMA(1,0,0)(0,0,2)[12] with non-zero mean	Inf
ARIMA(1,0,0)(1,0,2)[12] with non-zero mean	Inf
ARIMA(0,0,0)(0,0,1)[12] with non-zero mean	-692.0527
ARIMA(2,0,0)(0,0,1)[12] with non-zero mean	-703.8522
ARIMA(2,0,0) with non-zero mean	-703.233
ARIMA(3,0,0)(0,0,1)[12] with non-zero mean	-702.2476
ARIMA(2,0,1)(0,0,1)[12] with non-zero mean	-702.8139
ARIMA(1,0,1)(0,0,1)[12] with non-zero mean	-702.0572
ARIMA(3,0,1)(0,0,1)[12] with non-zero mean	-701.4519
ARIMA(2,0,0)(0,0,1)[12] with zero mean	-705.6575
ARIMA(2,0,0) with zero mean	-704.8566
ARIMA(2,0,1)(0,0,1)[12] with zero mean	-704.6898
Best model: ARIMA(2,0,0)(0,0,1)[12] with zero mean	

Maximum likelihood estimation

The Maximum likelihood was used to estimate the parameters of our model SARIMA $(2,0,0)(0,0,1)_{12}$. The estimates of the parameters are presented in Table 8 below. We see that all the coefficients are significant at different levels. This implies the appropriateness of this model.

Table 8: Estimation of model parameters

z test of coefficients				
	Estimate	Std. Error	z value	Pr(>|z|)
ar1	0.248510	0.093472	2.6587	0.007845 **
ar2	0.215048	0.094735	2.2700	0.023208 *
sma1	-0.244925	0.150485	-1.6276	0.103616

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1

Model diagnostic

Box-Ljung and Engle'LM tests were used to check respectively the autocorrelation and the heteroskedasticity in residuals of the fitted model. The results of these two tests are respectively represented in Tables 9 and 10. We show that both p-values of these two tests statistics are not significant at 0.05 level. This implies an acceptance of the null hypotheses that residuals have no autocorrelation and no heteroskedasticity respectively.

Table 9: Box-Ljung test of autocorrelation.

Box-Ljung test		
Data: Residuals of SARIMA $(2,0,0)(0,0,1)_{12}$ with zero mean		
df = 12	χ -squared =4.4512	p-value = 0.9739

Table 10: Engle's LM Test of heteroskedasticity.

Engle's LM ARCH Test		
Data: Residuals of SARIMA $(2,0,0)(0,0,1)_{12}$ with zero mean		
Alternative hypothesis: ARCH effects of order 12 are present		
Lag = 12	Statistic= 15.614	p-value = 0.2096

Different tests of normality show also that the residuals of SARIMA $(2,0,0)(0,0,1)_{12}$ are normally distributed. See Table 11 below.

Table 11: Normality tests of residuals

	Test Statistic	p-value
Shapiro-Wilk	0.9804	0.1097
Kolmogorov-Smirnov	0.0801	0.4865
Anderson-Darling	0.6685	0.0788

All these diagnostic tests confirm the adequacy of our fitted model for forecasting the time series y_t.

Forecast

We evaluate the performance of our model on the real testing returns datasets from July 2019 to May 2020. The forecasts are shown in Figure 8.

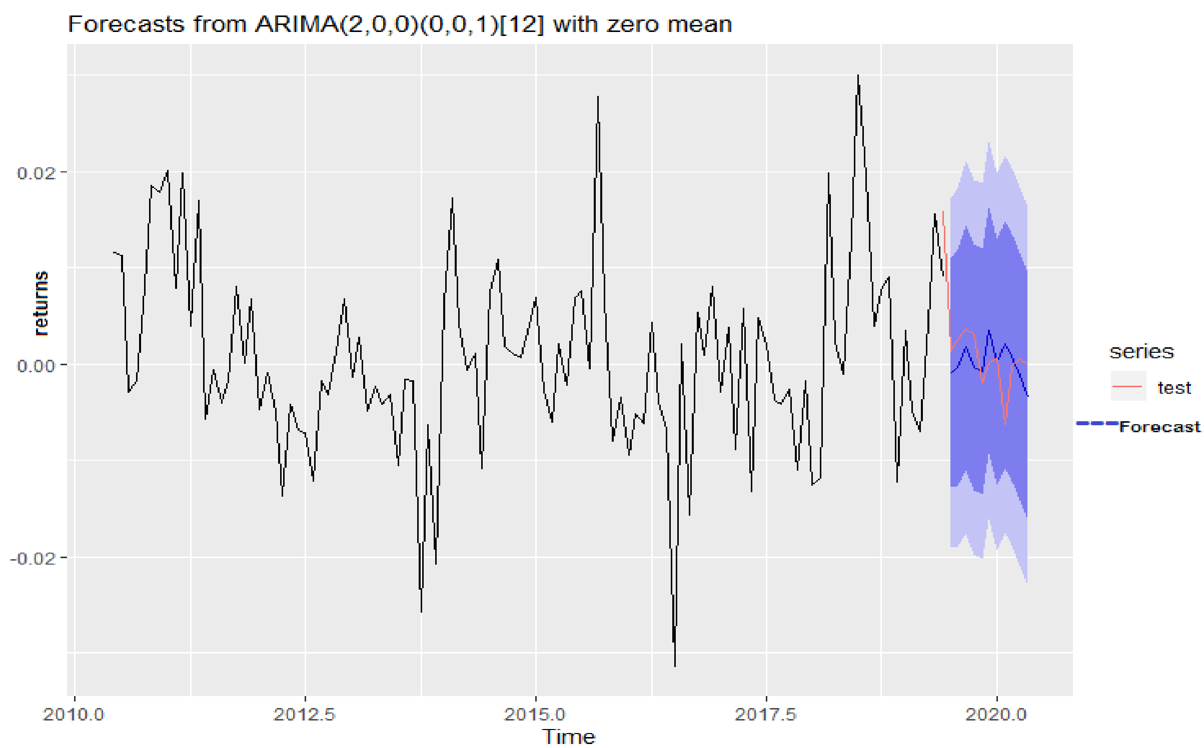

Figure 7: Forecasts from SARIMA $(2,0,0)(0,0,1)_{12}$ with zero mean model fitted to the training returns y_t time series.

Table 12 below shows the actual testing returns and their forecasted values.

Table 12: Actual test returns vs forecast values

Actual test returns	Forecasts
1.466847e-03	2.569530e-03
2.579572e-03	1.597536e-03
3.729525e-03	2.688951e-03
3.126945e-03	3.294262e-03
-1.894320e-03	3.222428e-04
8.516082e-05	6.069131e-03
7.076220e-04	7.129408e-05
-6.215200e-03	3.085934e-03
-1.329369e-04	-6.928065e-04
5.730178e-04	-1.196937e-04
0.000000e+00	-5.008601e-03

We can see by the Box-Ljung test and the Engle's LM ARCH Test in Tables 13 and 14 respectively that the residuals from forecasting training returns using this model still white noise and without heteroskedasticity. This implies that there is no information left in the residuals which should be used in computing forecasts (Hyndman & Athanasopoulos, 2018).

Table 13: Box-Ljung test of autocorrelation on residuals from forecasting monthly DZAIRINDEX returns using SARIMA $(2,0,0)(0,0,1)_{12}$

Box-Ljung test		
Data: Residuals		
df = 4	Chi-squared = 4.1088	p-value = 0.3915

Table 14: Engle's LM ARCH Test on residuals from forecasting monthly DZAIRINDEX returns using SARIMA $(2,0,0)(0,0,1)_{12}$

Engle's LM ARCH Test		
Data: Residuals		
Alternative hypothesis: ARCH effects of order 4 are present		
Lag = 4	Chi-squared = 5.8303	p-value = 0.2121

The forecast accuracy on the test returns set is represented in Table 15. Different error measurements show that the forecast accuracy of our model is acceptable. The MASE value is less than 1, which implies that SARIMA $(2,0,0)(0,0,1)_{12}$ performs much better than a naïve model in this case study (Hyndman & Athanasopoulos, 2018). The normal distribution of residuals is not necessary for forecasting (Hyndman and Athanasopoulos, 2018).

Table 15: Accuracy of forecasting the monthly DZAIRINDEX returns with SARIMA$(2,0,0)(0,0,1)_{12}$

Model	ME	RMSE	MAE	MASE
SARIMA $(2,0,0)(0,0,1)_{12}$	0.0002	0.0025	0.0019	0.79

The forecast of the whole real returns dataset for one year ahead using SARIMA $(2,0,0)(0,0,1)_{12}$ shows a slight increasing fluctuations trend in the upcoming months, see Figure 9 below.

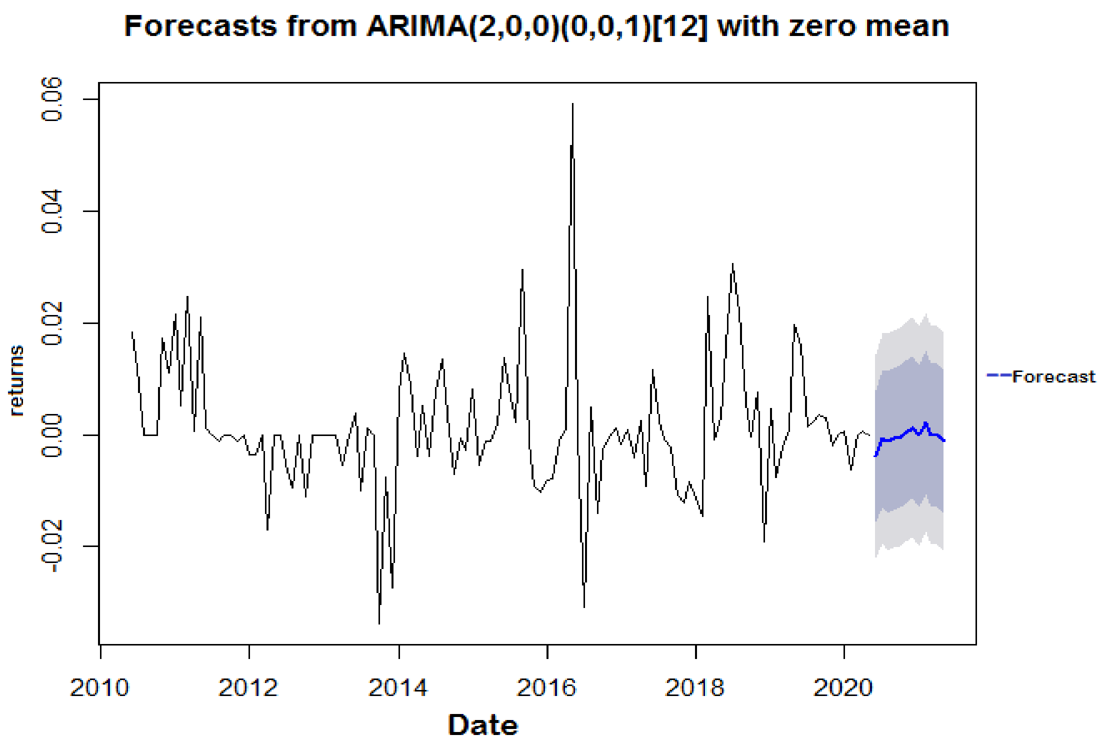

Figure 9: Forecast of the whole returns datasets for one year ahead using SARIMA $(2,0,0)(0,0,1)_{12}$

Conclusion

This study uses the Box-Jenkins methodology for forecasting the monthly DZAIRINDEX returns time series. Based on this approach, the seasonal ARIMA $(2,0,0)(0,0,1)_{12}$ is chosen as the best model to forecast the studied time series. We evaluate our model on a testing returns set. Different measures of forecast accuracy confirm that the forecast performance of our model is acceptable and it is better than a naïve model. The forecast of monthly returns for one year ahead shows a slight increasing fluctuations trend. We conclude from the results of this study that the seasonal ARIMA $(2,0,0)(0,0,1)_{12}$ model could be used by policymakers and investors in Algeria for forecasting the Algiers Stock Exchange, in order to get better investment decisions and returns.

The main limitation of the ARIMA model is the pre-assumed linear form of the model (Kumar & Thenmozhi, 2012). Most stock market time series are basically non-linear. So, the use of nonlinear models for forecasting the Algiers Stock Exchange might give a better forecast accuracy from the ARIMA model.

References

[1] Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control, **AC-19***, 716--723.
[2] Al-Shiab, M. (2006). The Predictability of the Amman Stock Exchange using the Univariate Autoregressive Integrated Moving Average (ARIMA) Model, *Journal of Economic and Administrative Sciences, Vol. 22* Iss 2 pp. 17 - 35
[3] Ashik, M. A and Kannan, S. K. (2017). Forecasting national stock price using ARIMA model. *Glob Stoch Anal 4*(1): 77-81.
[4] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society, Series B, 26*, 211-252.
[5] Box, G. E. P and Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
[6] Box, G. E. P and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*. San Fransisco: Holden Day.
[7] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: Forecasting and control (5th ed).* Hoboken, New Jersey: John Wiley & Sons.
[8] Brockwell, P.J and Davis, A.R. (2002). *An Introduction to Time Series and Forecasting*. 10.1007/978-1-4757-2526-1.

[9] Brooks, C. (2002). *Introductory econometrics for finance, first edition*, Cambridge University press, Cambridge.

[10] Burnham, K. P and Anderson, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach*. New York: Springer.

[11] Chang, X., Gao, M., Wang, Y., and Hou, X. (2012). Seasonal autoregressive integrated moving average model for precipitation time series. *Journal of Mathematics and Statistics. 8*. 500-505. 10.3844/jmssp.2012.500.505.

[12] Cheung, Y., Lai, K. (1995). Lag order and critical values of the augmented Dickey-Fuller test. *Journal of Business & Economic Statistics 13*(3), 277–280.

[13] Chou, R.Y. (1988). Volatility Persistence and Stock Valuations-Some Empirical Evidence Using GARCH. *Journal of Applied Econometrics, 3*, pp. 279-294.

[14] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimates for autoregressive time series with a unit root*, J. Am. Stat. Assoc.,* ***74***, 427--431.

[15] Drachal, K.(2017). fDMA: Dynamic Model Averaging and Dynamic Model Selection for continuous outcomes. https://CRAN.R-project.org/package=fDMA

[16] Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation. *Econometrica 50*, PP. 987–1008.

[17] Emenike, K. (2010). Modelling Stock Returns Volatility In Nigeria Using GARCH Models. *University Library of Munich, Germany, MPRA Paper*.

[18] Frank, J and Massey, Jr. (1951) The Kolmogorov-Smirnov Test for Goodness of Fit, *Journal of the American Statistical Association, 46*:253, 68-78.

[19] Gujarati, D.N. (2004). Basic Econometrics. 4th Edition, McGraw-Hill Companies. New York.

[20] Hylleberg, S., Engle, R., Granger, C. and Yoo, B. (1990). Seasonal integration and cointegration. *Journal of Econometrics* ***44***(1), pp. 215-238.

[21] Hyndman, R.J and Athanasopoulos, G. (2018). *Forecasting: principles and practice, 2nd edition*, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on August 2019.

[22] Hyndman, R. J and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *27(1), 1–22*.

[23] Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2019). forecast: Forecasting functions for time series and linear models. R package version 8.9, http://pkg.robjhyndman.com/forecast.

[24] Hyndman, R. J and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*, 679–688.

[25] Katircioglu, S.T and Al-khaza'leh, M.M. (2016). Modeling daily Amman Stock Exchange volatility for services sector, *Journal of Business, Economics and Finance (JBEF)*, ISSN: 2146 – 7943

[26] Kumar, M and Thenmozhi, M. (2012). A Comparison of Different Hybrid ARIMA - Neural Network Models for Stock Index Return Forecasting and Trading Strategy. *International Journal of Financial Management, 1*(1).

[27] Lin, T-W and Yu, C-C. (2009). Forecasting Stock Market with Neural Networks. Available at SSRN: https://ssrn.com/abstract=1327544

[28] Ljung, G. M, Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika 65*:297–303.

[29] López-de-Lacalle, J and Boshnakov, G. N. (2019). Package "uroot". Available at: https://cran.r-project.org/web/packages/uroot/uroot.pdf

[30] McLeod, A.I. (2015). Package "Kendall". Available at: https://cran.r- project.org /web/ packages/ Kendall/Kendall.pdf (accessed on 15 May 2019).

[31] Stephens, M.A. (1986). Tests based on EDF statistics. In: D'Agostino, R.B. and Stephens, M.A., eds.: Goodness-of-Fit Techniques. Marcel Dekker, New York.

[32] Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika,* ***71***, 599--607.

[33] Tsay, R. S. (2005). *Analysis of financial time series*. Hoboken, NJ: Wiley-Interscience. ISBN: 978-0-471-69074-0.

[34] Wahyudi, S. T. (2017). The ARIMA model for the Indonesia stock price. *International Journal of Economics and Management. 11*. 223-236.

[35] Wagner, M. (2010). Forecasting Daily Demand in Cash Supply Chains. *American Journal of Economics and Business Administration. 2*. 10.3844/ajebasp.2010.377.383