PAPER • OPEN ACCESS

Modeling Autoregressive Integrated Moving Average (ARIMA) and Forecasting of PT Unilever Indonesia Tbk Share Prices During the COVID-19 Pandemic Period

View the article online for updates and enhancements.

Modeling Autoregressive Integrated Moving Average (ARIMA) and Forecasting of PT Unilever Indonesia Tbk Share Prices During the COVID-19 Pandemic Period

R N Putri[1,2]**, M Usman**[2]**, Warsono**[2]**, Widiarti**[2]**, E Virginia**[3]

[1]Graduate School of Mathematics, University of Lampung, Jl. Sumantri Brojonegoro no 1, Bandar Lampung, Indonesia

[2]Departement of Matemathics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Lampung, Indonesia

[3]Graduate Student, Department of Accounting, Faculty of Economics and Bussiness, Universitas Indonesia, Depok, Indonesia

email: rimanugrahiya@gmail.com[1]

Abstract. ARIMA method is one method that can be used in predicting the movement of company shares. This study aims to obtain a time series model with the ARIMA method and predict stock price data of PT Unilever Indonesia Tbk from January 2020 to June 2020. The best model that fits the data based on the MSE value is the ARIMA(1,1,1) model. The ARIMA model (1,1,1) shows a match between real data and the predicted value. This model is then used for forecasting the next 14 days. Data on UNVR stock price from January 2020 to June 2020 are below 8000, this seems to correlate with the current conditions, namely the Covid-19 pandemic. Forecasting for the next 14 days (two weeks) from July 1, 2020 to July 14, 2020, the forecast values have a trend decrease, the trend of PT. Unilever Indonesia Tbk has been going down since January 2020. This seems to have occurred as an implication of the Covid-19 pandemic from January 2020 to the present.

Keywords: time series, ARIMA, COVID-19, forecasting

1. Introduction

Time series analysis has become a topic and study that has attracted the attention of many researchers in statistics, economics, finance, population and other fields of science. The application of time series analysis is mainly used for forecasting in various fields. Time series data is data that is recorded during a certain period. Usually in the form of daily, weekly, monthly, six-monthly and annual data. The pattern can be in the form of repetition of the past or does not have a pattern. Time series data that has a repetition pattern is called seasonal time series, for example is data on a company's stock movements. In the case of non-seasonal time series, the Box Jenkins method is modelled by determining several criteria which are then known as the ARMA and ARIMA models. These criteria include the Autocorrelation (ACF) and Partial Autocorrelation (PACF) functions. Similar to the case of seasonal time series, Box Jenkins models using the same criteria. Forecasting is a conjecture or

estimate about the occurrence of a situation in the future by using certain methods. Forecasting is done by using the best information so that the desired goals can be achieved.

Forecasting methods commonly used are Autoregressive Integrated Moving Average (ARIMA). The Integrated Moving Average Autoregressive Model (ARIMA) is a model that uses dependent variables or data in the past and completely ignores the independent variables. This method has several advantages, namely not requiring stationary data patterns and can be used on data that form seasonal patterns [1]. Therefore ARIMA is a statistic that is suitable to be used in predicting a number of variables quickly, simply and accurately because it only requires variable data to be predicted. Therefore in this study the ARIMA model for time series data analysis and forecasting using PT Unilever Indonesia Tbk (UNVR) stock price data from January 2, 2020 to June 30, 2020.

2. Statistical Model

2.1 Data Time series
The time series is a collection of observations X_t, each observation that is collected at time t. The time series model in observational data X_t is a specification of the joint distribution (or maybe only the mean and covariance) of the random variable array X_t The most important part of time series analysis is the selection of possible models that fit the data. Time series data itself is data collected from time to time for an individual [2].

2.2 Stationarity
Stationary means there is no drastic change in the data. Data fluctuations are around a constant average value, independent of the time and variance of these fluctuations. The assumption of stationarity in time series data analysis is fundamental and must be checked before analyzing data. Several methods are available to check stationarity of time series data, based on data plots or through the test Augmented Dickey-Fuller (ADF). The ADF test process is as follows, for example $y_1, y_2, \ldots, y_n$ becomes a time series, assuming that $\{y_t\}$ follows the AR(p) model with the given μ:

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \ldots + \phi_p(y_{tp} - \mu) + \varepsilon_t \tag{1}$$

where:
ε_t is the white noise average 0 and variance σ^2, and $\varepsilon_t \sim N(0, \sigma^2)$.
Testing non-stationary data in equation (1) using the ADF test or tau test (τ) is carried out as follows:
$H_0 : \phi_1^* = 0$ (there are root units or non-stationary time series)
$H_1 : \phi_1^* < 0$ (no root units or stationary time series)
Test statistics:
ADF test

$$(\tau) = \frac{\hat{\phi}_1^*}{\widehat{Se}_{\phi_1^*}} \tag{2}$$

for significance level ($\alpha = 0.05$), reject H_0 if $\tau < -2.57$ or if p-value <0.05 [2],[8].

2.3 Model Autoregressive (AR)
Model autoregressive (AR) is a model that illustrates that present value is influenced by past value. The AR model with order p is denoted by AR (p). The general form of the AR (p) model is:

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \tag{3}$$

Where:
C : Constant,
ϕ_p : Coefficient parameter AR with order p,
Y_t : Data at time t,
ε_t : Error at time t.

Order in the model AR is often used in time series analysis is p=1 or p=2 [3].

2.4 Model Moving Average (MA)
Model Moving Average has the form as follows:

$$Y_t = C + \varepsilon_t - \theta_1\varepsilon_1 - \theta_2\varepsilon_2 - \cdots - \theta_q\varepsilon_q \tag{4}$$

Where:

C : Constant,

θ_q : Coefficient parameter MA of order q,

Y_t : Data at time t,

ε_t : error at time t.

From equation (4), it can be seen that Y_t is the weighted average error of q periods backwards. The number of errors q used in this equation indicates the level of the moving average model.

2.5 Model ARIMA

The Integrated Moving Average Autoregressive Model (ARIMA) was developed by George E.P. Box and Gwilym M. Jenkins (1976), so ARIMA is also called the Box-Jenkins time series method. The ARIMA model is divided into 3 elements, namely the Autoregressive (AR) and Moving Average (MA) and Integrated (I) models. The ARIMA model is a model that completely ignores independent variables in making forecasting. ARIMA uses past and present values of the dependent variable to produce accurate short-term forecasting. ARIMA is very good accuracy for short-term forecasting, while for long-term forecasting is not good. Usually it will tend to be flat (horizontal / constant) for a fairly long period. The general form of the ARIMA model can be stated in the following equation [4]:

$$\phi_p(B)\nabla^d Y_t = C + \theta_q(B)\varepsilon_t \tag{5}$$

Where:

Y_t : Data at time t,

ϕ_p : Parameter Autoiregressive (AR)

B : Backward Operator,

d : Order of differencing,

C : Parameter constant,

θ_q : Parameter Moving Average (MA),

ε_t : error at time t.

The ARIMA model assumes that the input data must be stationary. Stationary means that there is no growth or decrease in the data. The data must be roughly horizontal along the time axis. In other words, data fluctuations are around an average value and a variance that is constant with time. If the input data is not stationary, adjustments are needed to produce stationary data. One of the commonly used methods is the differencing method. This method is done by reducing the value of data in a period with the data value of the previous period. One method that can be used to estimate model parameters is Least Squares (Conditional Least Squares) [5]. The least squares method is done by minimizing the number of error squares. The steps for applying the ARIMA method in a row are model identification, model parameter estimation, model evaluation.

2.6 Test for Stability of the model

The Eigen value of matrix F satisfies

$$\left|I_n\lambda - \phi_1\lambda^{p-1} - \phi_1\lambda^{p-2} - \cdots - \phi_p\right| = 0 \tag{6}$$

is covariance stationary as long as $|\lambda| < 1$ for all values of λ. Or equivalently, the VAR is covariance stationary if all values of z are satisfying

$$\left|I_n - \phi_1 z - \phi_1 z^2 - \cdots - \phi_p z^p\right| = 0 \tag{7}$$

It lies outside the unit circle ([9],[10],[11]).

2.7 The steps of ARIMA method

The steps in applying the ARIMA method in succession are:

2.7.1 Data stationarity test

The first thing to do at this stage is whether the time series data is stationary or non-stationary and that the AR and MA aspects of the ARIMA model only relate to stationary time series [6]. The stationarity of a time series can be seen from the ACF plot, where the autocorrelation coefficient decreases rapidly to zero, usually after the 2nd or 3rd lag. If the data is not stationary then a distinction can be made, the order of differentiation until the series becomes stationary can be used to determine the value of d on ARIMA (p, d, q).

2.7.2   Parameter estimation model
Parameter estimation by trying various ARIMA models at each level of each order which is likely to be an ARIMA parameter.

2.7.3   Model Evaluation
The second step is to estimate the autoregressive parameters and moving average parameters based on the order obtained at the identification stage. A good estimation model can be seen from the significance of the estimated parameters, and the smallest mean square error (MSE) value [7].

2.7.4   Forecasting
The last stage is forecasting, which is to make predictions or estimates of data based on the selected ARIMA model.

3.   Results and Discussion

In doing time series modeling, the first step that must be taken is the data stationarity test. Stationary testing can be done in three ways, namely viewing the time series plot, using the Autocorrelation Function (ACF) graph, and the unit root test (unit root test).
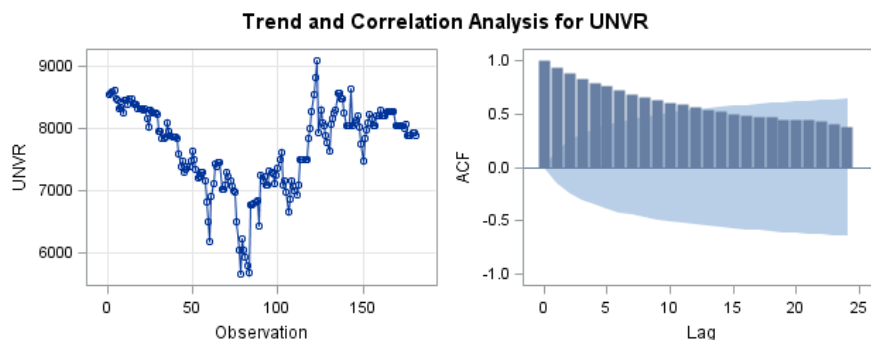

Figure 1. Plot *time series* and ACF Data UNVR

Based on Figure 1, it can be seen that the four data are not stationary in the average or variety because they have a tendency patte
rn. Furthermore, looking at the ACF graphs from the four images also shows the instability because it has a pattern of decreasing exponentially close to zero. Then finally the stationary testing with the unit root test.

Table 1. Augmented Dickey-Fuller Unit Root Tests

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	3	-0.1497	0.6479	-0.46	0.5122		
Single Mean	3	-9.0190	0.1653	-2.19	0.2120	2.43	0.4514
Trend	3	-9.3204	0.4742	-2.27	0.4494	2.90	0.5987

Hypothesis test:
H_0: has a root unit (not stationary)

H_1: has no unit root (stationary)
Conclusion: Based on Table 1, it can be seen in the statistical p-value that Tau (τ) all types of testing for each variable are greater than the significant level used, which is $\alpha = 0.05$, so that starting with H0 is not stationary.

Because the data is not stationary, differencing must be performed on the data, then re-testing of stationary testing using time series plots, ACF charts and unit root tests.

Figure 2. Plot *time series* and ACF Data share price UNVR after differencing (d=1)

Based on Figure 2, it can be seen that the data shows that they are stationary in mean and variance because they do not have a tendency pattern, then the stationary testing is done by unit root test.

Table 2. Augmented Dickey-Fuller Unit Root Tests after differencing 1

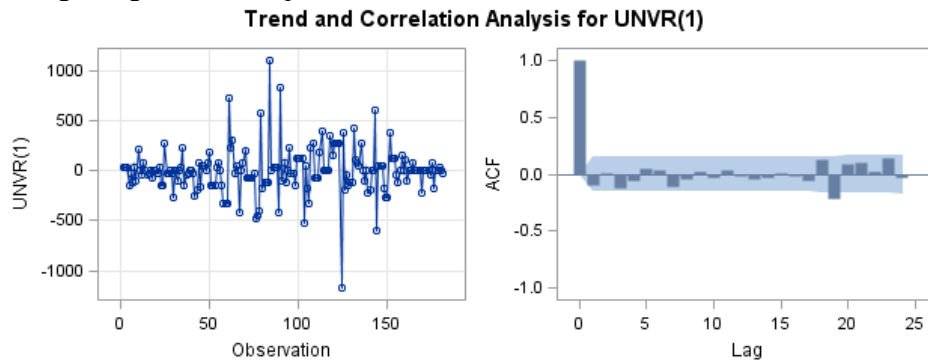| Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	3	-493.318	0.0001	-7.97	<.0001		
Single Mean	3	-496.399	0.0001	-7.95	<.0001	31.60	0.0010
Trend	3	-524.471	0.0001	-7.98	<.0001	31.84	0.0010

Hypothesis test:
H_0: has a root unit (not stationary)
H_1: has no unit root (stationary)
Conclusion: Based on Table 2, it can be seen in the statistical p-value Tau (τ) that all types of tests for each variable are smaller than the significant level used, which is $\alpha = 0.05$, so reject H0 which means the data has been stationary. After stationary data with the first differencing (d = 1), it can then determine the ARIMA model by looking at the partial autocorrelation (PACF) value and the autocorrelation (ACF) value using the first differencing data.

Table 3. Values of PACF and ACF

Lag	PACF	T	ACF	T	Lag	PACF	T	ACF	T
1	-0,097	-1,3	-0,097	-1,3	16	-0,03097	-0,42	-0,01828	-0,23
2	-0,00265	-0,04	0,006785	0,09	17	-0,09798	-1,31	-0,05759	-0,73
3	-0,12933	-1,74	-0,12851	-1,71	18	0,113378	1,52	0,127839	1,63
4	-0,07878	-1,06	-0,05175	-0,68	19	-0,21548	-2,89	-0,22114	-2,77
5	0,030104	0,4	0,04354	0,57	20	0,011074	0,15	0,08996	1,08
6	0,025691	0,34	0,035005	0,46	21	0,131691	1,77	0,09685	1,16
7	-0,12518	-1,68	-0,11086	-1,44	22	-0,01332	-0,18	0,016706	0,2
8	-0,07151	-0,96	-0,0501	-0,64	23	0,122557	1,64	0,138303	1,64
9	0,028733	0,39	0,0276	0,35	24	0,02544	0,34	-0,03014	-0,35
10	-0,06341	-0,85	-0,03682	-0,47	25	0,064905	0,87	0,012023	0,14

11	-0,00615	-0,08	0,040073	0,51	26	-0,04815	-0,65	-0,03297	-0,38
12	-0,00575	-0,08	-0,01645	-0,21	27	-0,09531	-1,28	-0,07085	-0,83
13	-0,04668	-0,63	-0,04235	-0,54	28	0,065394	0,88	0,009657	0,11
14	-0,05841	-0,78	-0,02922	-0,37	29	0,118954	1,6	0,132628	1,54
15	-0,01335	-0,18	0,008527	0,11	30	0,064168	0,86	0,019758	0,23

Based on Table 3, to find out the order on AR (p) and MA (q) can be seen from the T value in each table with a limit of $\pm$ 1.96. The T value on PACF indicates the order p. Because there are 3 values in the initial lags that exceed the $\pm$ 1.96 limit, the order p is 1. While the T value in ACF indicates the order q, where there is 1 value in the initial lags that exceeds the limit of $\pm$ 1.96, meaning that the order q is 1.

After getting each order on AR and MA, then you can try various ARIMA models on each order that are likely to be the parameters of the ARIMA model. Next do a significance test to choose which model is suitable to use. The model is said to be significant, if the p-value parameter is smaller than α = 0.05. The possibilities that can be used to become the parameters of the ARIMA model and the results of the significance tests on each of the possible ARIMA models are as follows.

Table 4. The result of the possible significance
and MSE value of ARIMA model

Model	Significant	MSE
ARIMA(1,1,0)	No	53340
ARIMA(0,1,1)	No	53329
ARIMA(1,1,1)	Yes	52694
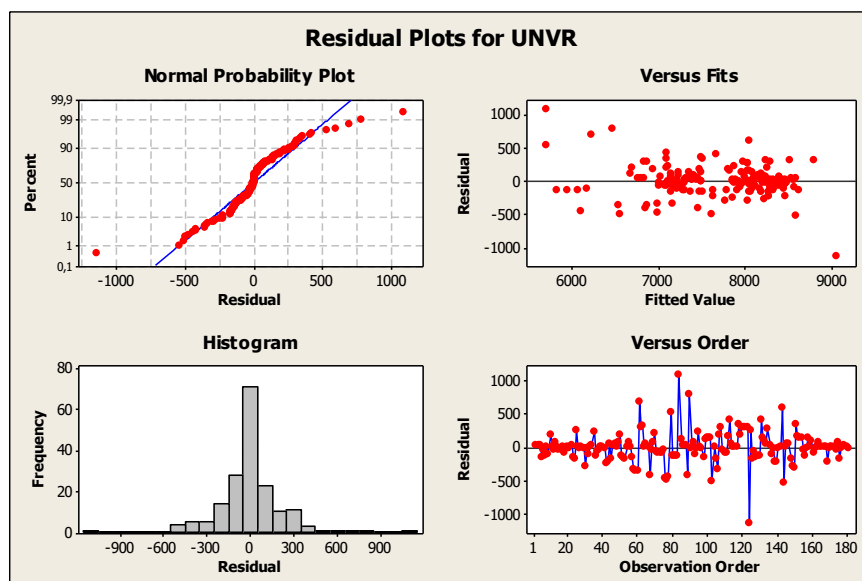
Figure 3. Plot Residual Data Share price UNVR

Based on Table 4, the significant ARIMA model (p-value <0.05) is ARIMA (1,1,1). And after obtaining some significant ARIMA models, the thing to do next is to evaluate the model to get the best model. The best model is by looking at the smallest MSE value. Based on Table 4 it can be seen that the one with the smallest MSE value among all possible models is ARIMA (1,1,1) with MSE value = 52694. Figure 3 shows the residual plot for UNVR share price data from January 2, 2020 to June 30, 2020. There is an outline of a straight line that can be observed from the normality plot. This indicates that the error is close to normal with some outliers. Therefore, the normal assumptions are met, besides that it can be seen from the histogram plot that the graph looks symmetrical and converges in the middle and shows a slight spread, thus supporting the normality assumption. Therefore ARIMA

(1,1,1) was chosen as the best model that can be used if you want to forecast. The results of parameter determination obtained by ARIMA (1,1,1) with coefficients C = -0,736, AR (1) = 0.7684, and MA (1) = 0.8747. Therefore the ARIMA model (1,1,1) can be stated as follows.

$$Y_t = -0.736 + 0.7684Y_{t-1} - 0.8747\varepsilon_{t-1} \tag{8}$$

And the following is the result of forecasting from the stock price data of PT Unilever Indonesia Tbk from July 1 to July 14, 2020:

Table 5. Forecasting data PT Unilever Indonesia Tbk from July 1 to July 14, 2020:

Period	Forecast	Lower	Upper	Real
182	7906,18	7456,17	8356,19	7800
183	7910,19	7306,66	8513,72	8050
184	7912,54	7207,00	8618,07	7900
185	7913,60	7131,64	8695,56	7900
186	7913,69	7070,49	8756,88	7925
187	7913,02	7018,45	8807,59	7900
188	7911,77	6972,60	8850,93	7925
189	7910,07	6931,15	8888,99	8100
190	7908,03	6892,94	8923,12	7975
191	7905,73	6857,19	8954,27	7925
192	7903,22	6823,34	8983,11	8025
193	7900,56	6791,01	9010,11	8075
194	7897,78	6759,93	9035,64	8025
195	7894,91	6729,88	9059,94	8075

From the analysis above model (8), ARIMA (1,1,1) can be relied upon to be used in further studies, especially for forecasting purposes. Based on the results of forecasting, it can be seen that the real and forecast data are almost close to each other.

4. Conclusion

The UNVR stock price data from January 2020 to June 2020 is modeled with the time series analysis approach and the best model that fits the data based on the MSE value is the ARIMA(1,1,1) model. The ARIMA model (1,1,1) shows a match between real data and the predicted value. This model is then used for forecasting the next 14 days. Data on UNVR stock price from January 2020 to June 2020 are below 8000, this seems to correlate with the current conditions, namely the Covid-19 pandemic. Forecasting for the next 14 days (two weeks) from July 1, 2020 to July 14, 2020, the forecast values have a trend decrease, the trend of PT. Unilever Indonesia Tbk has been going down since January 2020. This seems to have occurred as an implication of the Covid-19 pandemic from January 2020 to the present.

References
[1]   Hadiansyah F N 2017 *Prediksi Harga Cabai dengan Pemodelan Time Series ARIMA* vol 2 Issue 1
[2]   Brockwell, PJ and Davis RA 2002 *Introduction to Time Series and Forecasting* (New York: Springer-Verlag)
[3]   Pankratz A 1991 *Forecasting with Dynamic Regression* (Canada : Intersciences Publication)
[4]   Sartono B 2006 *Modul Kuliah Pelatihan Time Series Analysis* (Bogor : IPB Press)
[5]   Cyrer J D dan Chan K S 2008 *Time Series Analysis: With Aplication in R* Second Edition (USA : Springer Science dan Businiess Media)
[6]   Makridakis S Wheelwright S C dan McGee V E 1999 *Metode dan Aplikasi Peramalan* Ed. ke-2

Terjemahan Ir Untung Sus Ardiyanto M.Sc dan Ir. Abdul Basith M.Sc (Jakarta : Erlangga)

[7] Juanda B dan Junaidi 2012 *Ekonometrika Deret Waktu* (Bogor : IPB Press)

[8] Tsay R S 2005 *Analysis of Financial Time Series* (New York : A John Wiley & Sons, Inc. Publication)

[9] Hamilton J 1994 *Time Series Analysis* (New Jersey : Princeton University Press)

[10] Warsono, Russel, E., Wamiliana, Widiarti, Usman, M 2019 *Vector autoregressive with exogenous variable model and its application in modeling and forecasting energy data: Case study of PTBA and HRUM energy*. International Journal of Energy Economics and Policy, 9(2), 390-398

[11] Warsono, Russel, E., Wamiliana, Widiarti, And Usman, M 2019 *Modeling and Forecasting by the Vector Autoregressive Moving Average Model for Export of Coal and Oil Data*. International Journal of Energy Economics and Policy, 9(4), 240-247