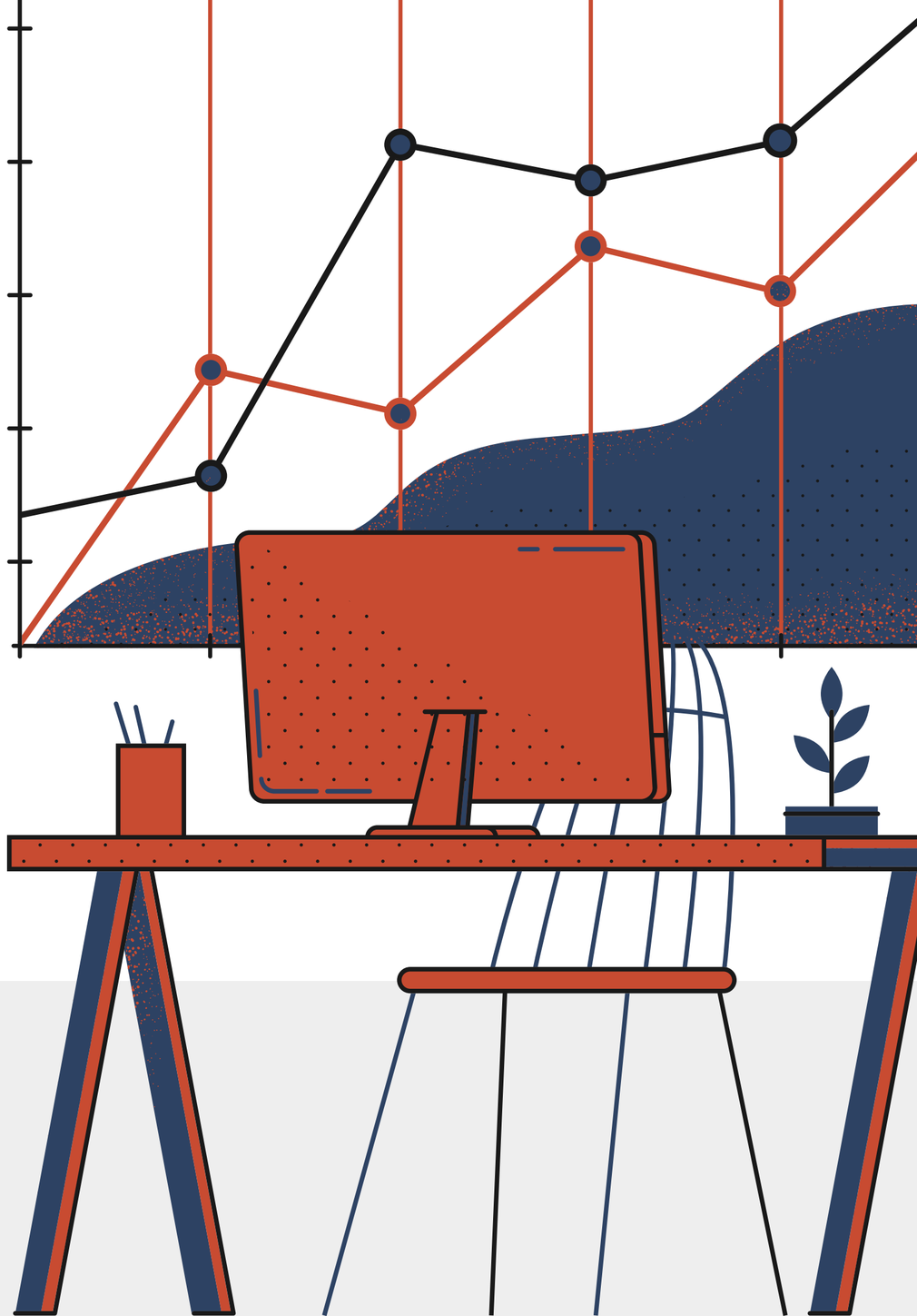
June, 2022

FMIPA, Universitas Indonesia

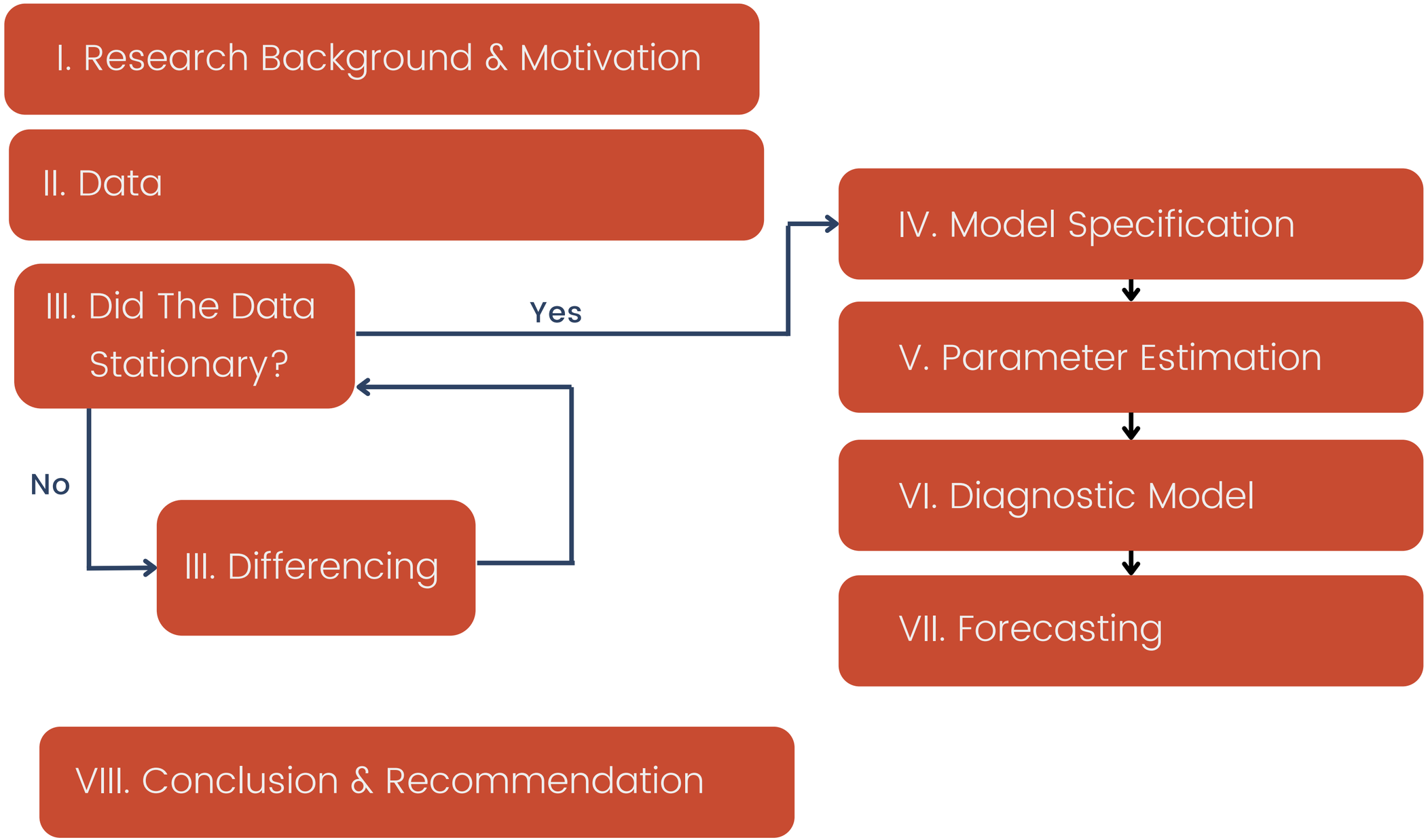
Alphabet Inc. (GOOG) Stock Forecasting



GROUP 9

Faza Muthmainnah
Raistra Aurel
Ria Rahma Septiani

Outline



Reported by CNBC Indonesia, this GOOG-coded stock price actually shows encouraging performance in the midst of the covid-19 pandemic, which has increased by 17.68% from year to year.

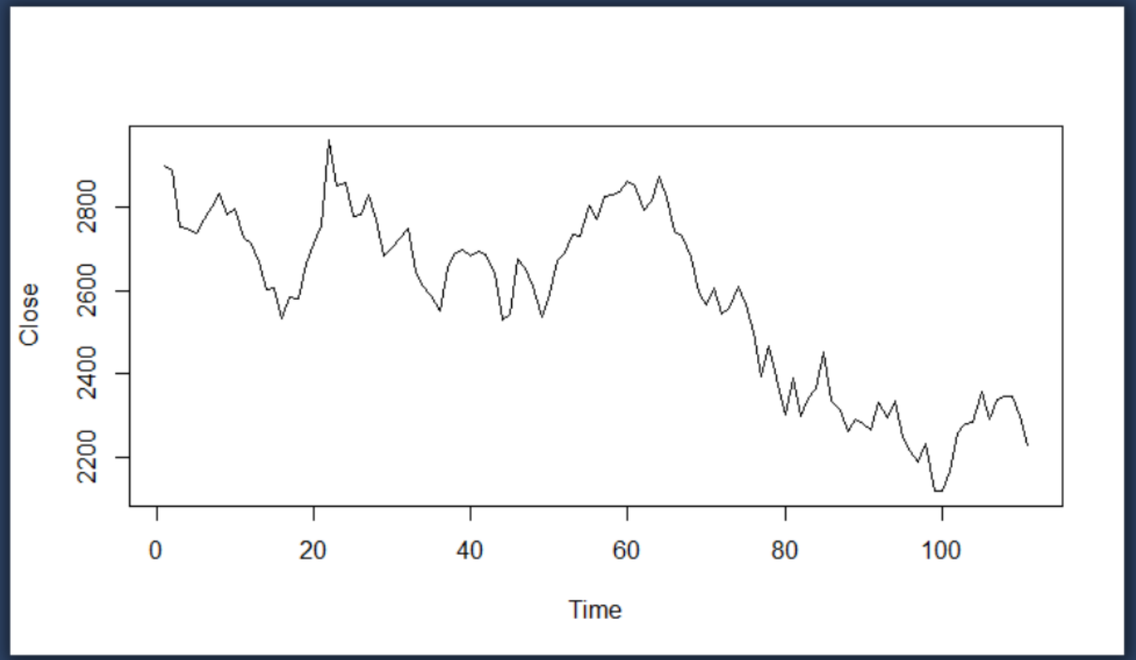
- **Further analyzing the GOOG stock after covid-19 pandemic with time series analysis.**
- **Predict the closing price for the next five days.**

Data

We use Alphabet Inc. (GOOG) stock historical data from yahoo! finance

Currency in USD — Download

Date	Open	High	Low	Close*	Adj Close**	Volume
Jun 10, 2022	2,255.63	2,269.94	2 217.22	2,228.55	2,228.55	1,566,200
Jun 09, 2022	2,326.83	2,367.00	2 297.34	2,298.36	2,298.36	1,157,100
Jun 08, 2022	2,337.53	2,372.92	2 333.94	2,344.76	2,344.76	1,127,200
Jun 07, 2022	2,312.96	2,354.97	2 302.51	2,344.59	2,344.59	1,320,700
Jun 06, 2022	2,334.85	2,387.97	2 330.57	2,340.21	2,340.21	1,189,300
Jun 03, 2022	2,319.85	2,327.29	2 273.36	2,291.28	2,291.28	1,252,600
Jun 02, 2022	2,283.76	2,357.96	2 266.16	2,354.92	2,354.92	1,373,600
Jun 01, 2022	2,298.63	2,347.98	2 271.01	2,282.74	2,282.74	1,431,500
May 31, 2022	2,261.58	2,328.67	2 251.45	2,280.78	2,280.78	2,565,100
May 27, 2022	2,195.77	2,257.36	2 191.00	2,255.98	2,255.98	1,496,200
May 26, 2022	2,121.01	2,179.10	2 109.76	2,165.92	2,165.92	1,514,400
May 25, 2022	2,102.84	2,130.89	2 084.23	2,116.79	2,116.79	1,895,000
May 24, 2022	2,127.55	2,127.90	2 044.16	2,118.52	2,118.52	3,019,300
May 23, 2022	2,202.08	2,240.11	2 183.08	2,233.33	2,233.33	1,577,900
May 20, 2022	2,241.71	2,251.00	2 127.46	2,186.26	2,186.26	1,879,300
May 19, 2022	2,236.82	2,271.75	2 209.36	2,214.91	2,214.91	1,459,600
May 18, 2022	2,304.75	2,313.91	2 242.84	2,248.02	2,248.02	1,399,100
May 17, 2022	2,344.55	2,344.55	2 306.75	2,334.03	2,334.03	1,078,800
May 16, 2022	2,307.68	2,332.15	2 286.70	2,295.85	2,295.85	1,164,100



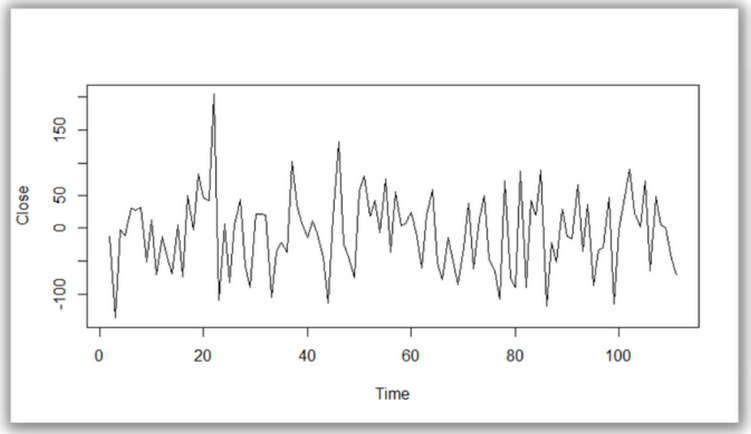
1 Dicky-Fuller Test 1

```
> adf.test(x)

        Augmented Dickey-Fuller Test

data:  x
Dickey-Fuller = -2.3145, Lag order = 4, p-value = 0.4461
alternative hypothesis: stationary
```

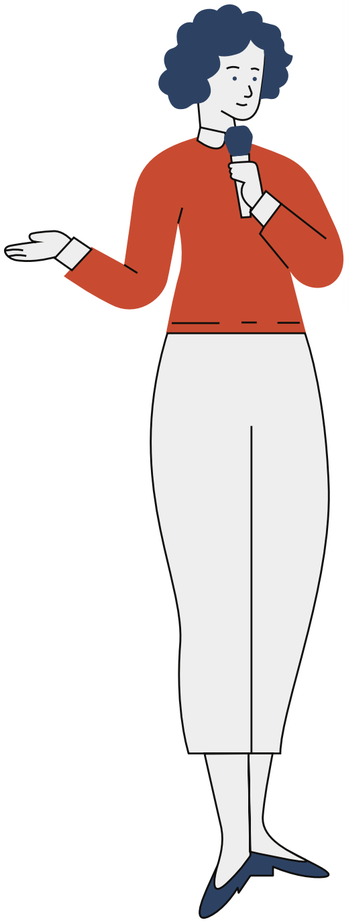
Because p-value > 0.05, then the data is not stationary, so we have to do differencing process of the data.

2 Differencing & Checking

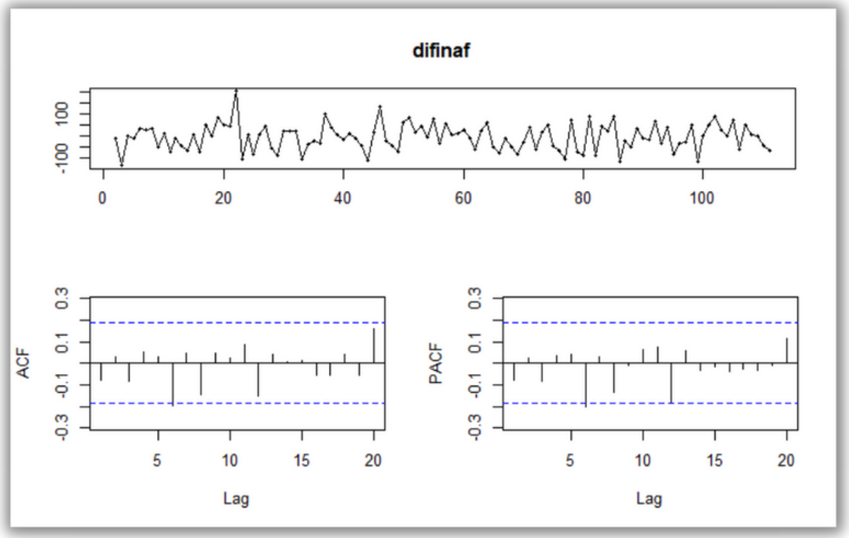
```
> difinaf<-diff(x,differences = 1)
> adf.test(difinaf)

        Augmented Dickey-Fuller Test

data:  difinaf
Dickey-Fuller = -4.3098, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(difinaf) : p-value smaller than printed p-value
```



We have to use the data after differencing (1) because its stationary indicated by the p-value < 0.05



Plot Model



From the ACF plot, it can be obtained the order of q=6 based on the correlogram line at lag 6. Meanwhile, from the PACF plot, the order p=6 is obtained. So the candidate model:

- ARIMA(0,1,6) based on ACF
- ARIMA (6,1,0) based on PACF

EACF Plot

```
> eacf(difinaf)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o o x o o o o o  o  o  o
1 x o o o o o o o o o o  o  o  o
2 x x o o o o o o o o o  o  o  o
3 x x o o o o o o o o o  o  o  o
4 x o x o o o o o o o o  o  o  o
5 x x o o x o o o o o o  o  o  o
6 o x o x o x o o o o o  o  o  o
7 x x o x o x o o o o o  o  o  o
```

For more candidate models, we observed from EACF. So the additional candidate model we obtain:

- ARIMA(0,1,1)
- ARIMA (1,1,1)

```
> #ARIMA Candidate Model Test
> model1 <- Arima(x,order=c(0,1,6),include.constant = TRUE)
> model2 <- Arima(x,order=c(6,1,0),include.constant = TRUE)
> model3 <- Arima(x,order=c(0,1,1),include.constant = TRUE)
> model4 <- Arima(x,order=c(1,1,1),include.constant = TRUE)
>
> cbind(model1,model2,model3,model4)
          model1       model2       model3       model4
coef      numeric,7    numeric,7    numeric,2    numeric,3
sigma2    3497.181     3571.609     3608.19      3478.988
var.coef  numeric,49   numeric,49   numeric,4    numeric,9
mask      logical,7    logical,7    logical,2    logical,3
loglik    -601.5318    -602.5638    -605.5784    -603.9828
aic       1219.064     1221.128     1217.157     1215.966
arma      integer,7    integer,7    integer,7    integer,7
residuals ts,111       ts,111       ts,111       ts,111
call      expression   expression   expression   expression
series    "x"          "x"          "x"          "x"
code      0            0            0            0
n.cond    0            0            0            0
nobs      110          110          110          110
model     list,10      list,10      list,10      list,10
aicc      1220.489     1222.553     1217.383     1216.347
bic       1240.667     1242.731     1225.258     1226.768
xreg      integer,111  integer,111  integer,111  integer,111
x         ts,111       ts,111       ts,111       ts,111
fitted    ts,111       ts,111       ts,111       ts,111
```

Fit Model

Because the aic, aicc, bic value of the model4 is the smallest then we will use model4 which is ARIMA $(1,1,1)$ to become a fit model.

```
> #Parameter Estimation
> #The Best Model
> fit <- Arima(x,order=c(1,1,1),include.constant = TRUE)
> fit
Series: x
ARIMA(1,1,1) with drift

Coefficients:
          ar1      ma1    drift
      0.9072  -1.0000  -5.6086
s.e.  0.0436   0.0258   1.4631

sigma^2 = 3479:  log likelihood = -603.98
AIC=1215.97   AICc=1216.35   BIC=1226.77
```

From ARIMA(1,1,1) output, we obtain Φ1=0.9072, θ1=-1, and θ0=-5.6086 Φ. So, we got the parameter estimation equation:

$$Y_t = -5.6086 + (1 + 0.9072)Y_{t-1} + a_t + a_{t-1}$$

$$Y_t = -5.6086 + 1.9072Y_{t-1} + a_t + a_{t-1}$$

RESIDUAL ANALYSIS

Stationary Test

```
> #DIAGNOSTIC MODEL CHECKING TEST
> #Residual Stationary Test
> adf.test(fit$residuals)

        Augmented Dickey-Fuller Test

data:  fit$residuals
Dickey-Fuller = -4.0493, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(fit$residuals) : p-value smaller than printed p-value
```

Because the p-value < 0.05, then the H0 is rejected. So we use alternative hypothesis which is **stationary**

Normality Test

```
> #DIAGNOSTIC MODEL CHECKING TEST
> #Normality Test
> #H0: Normal Distribution Data
> jb.norm.test(fit$residuals)

        Jarque-Bera test for normality

data:  fit$residuals
JB = 1.9025, p-value = 0.3035
```

Because the p-value > 0.05, then the H0 is not rejected. So we use the hypothesis that the data is **normal distributed**

Independency Test

Because the p-value > 0.05, then the H0 is not rejected. So we use the hypothesis which is **independent**

```
> #DIAGNOSTIC MODEL CHECKING TEST
> #Independency Test
> #ljung box H0: independent residual
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ARIMA(1,1,1) with drift
Q* = 8.3314, df = 7, p-value = 0.3043

Model df: 3.   Total lags used: 10
```

Because the fit model of the data meet all the Model Diagnostic Assumption which are **stationary**, **independen**t, and **normal**, then we can do further test which is Overfitting

In the overfitting test we only increase the order gradually, then compare the aic, aicc, and bic

and if we do the hypothetical test, if the critical area value > test statistics, then the overfitting is useless

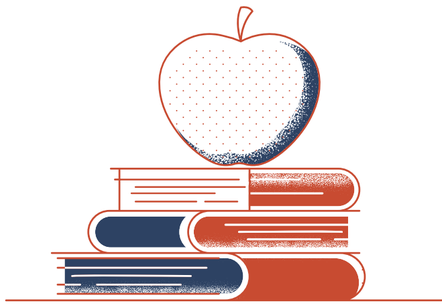
```
> #Overfitting
> overfit1 <-Arima(x,order=c(2,1,1),include.constant = TRUE)
> overfit2 <-Arima(x,order=c(1,1,2),include.constant = TRUE)
> cbind(fit,overfit1,overfit2)
           fit         overfit1    overfit2
coef       numeric,3   numeric,4   numeric,4
sigma2     3478.988    3665.991    3510.006
var.coef   numeric,9   numeric,16  numeric,16
mask       logical,3   logical,4   logical,4
loglik     -603.9828   -605.4274   -603.927
aic        1215.966    1220.855    1217.854
arma       integer,7   integer,7   integer,7
residuals  ts,111      ts,111      ts,111
call       expression  expression  expression
series     "x"         "x"         "x"
code       0           0           0
n.cond     0           0           0
nobs       110         110         110
model      list,10     list,10     list,10
aicc       1216.347    1221.432    1218.431
bic        1226.768    1234.357    1231.356
xreg       integer,111 integer,111 integer,111
x          ts,111      ts,111      ts,111
fitted     ts,111      ts,111      ts,111
```

ARIMA (2,1,1)

```
> critical_value; test_stat_ar2
[1] -1.981765
[1] 0.1523121
```

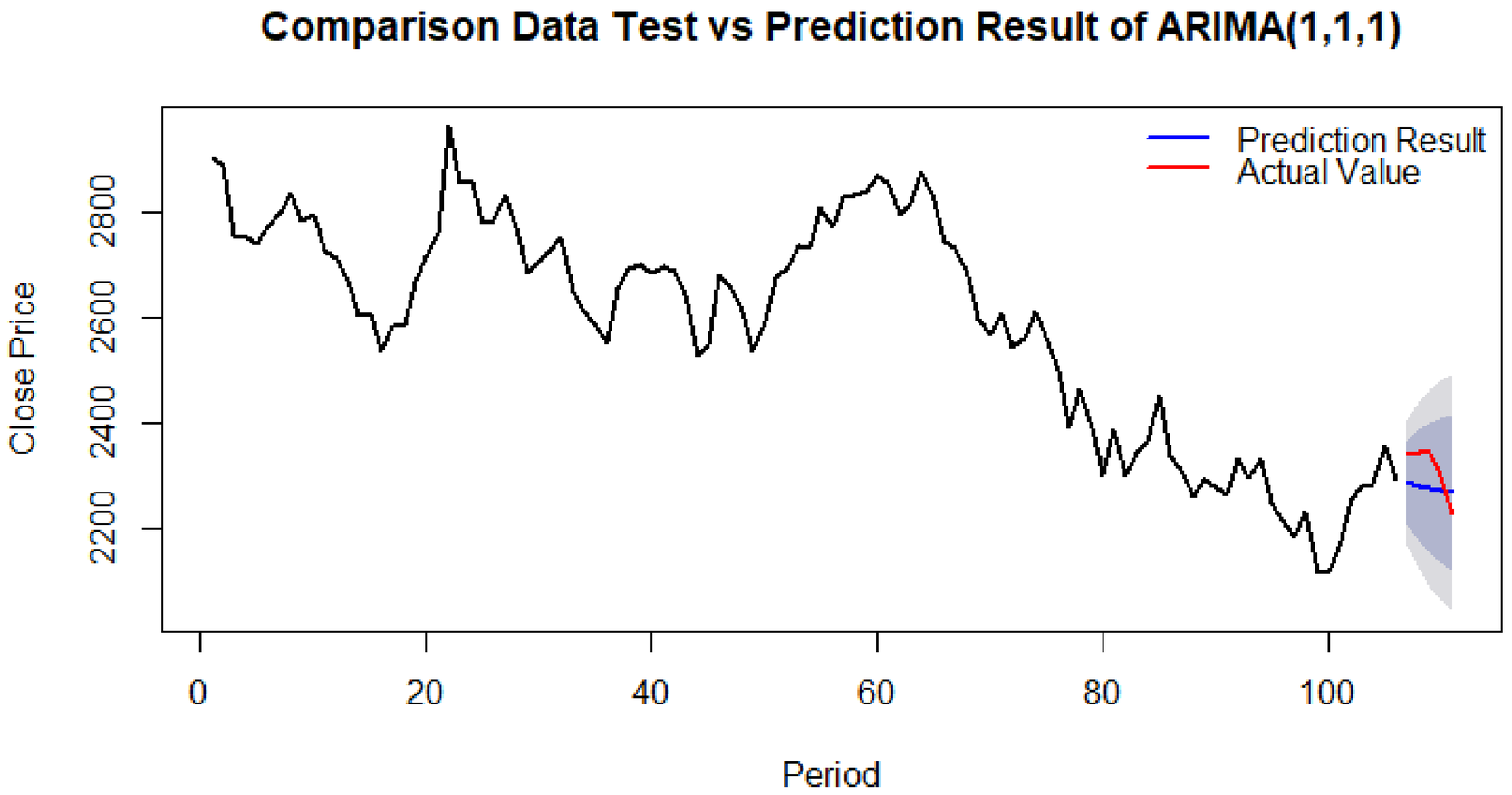
ARIMA (1,1,2)

```
> critical_value; test_stat_ma2
[1] -1.981765
[1] 0.3339732
```

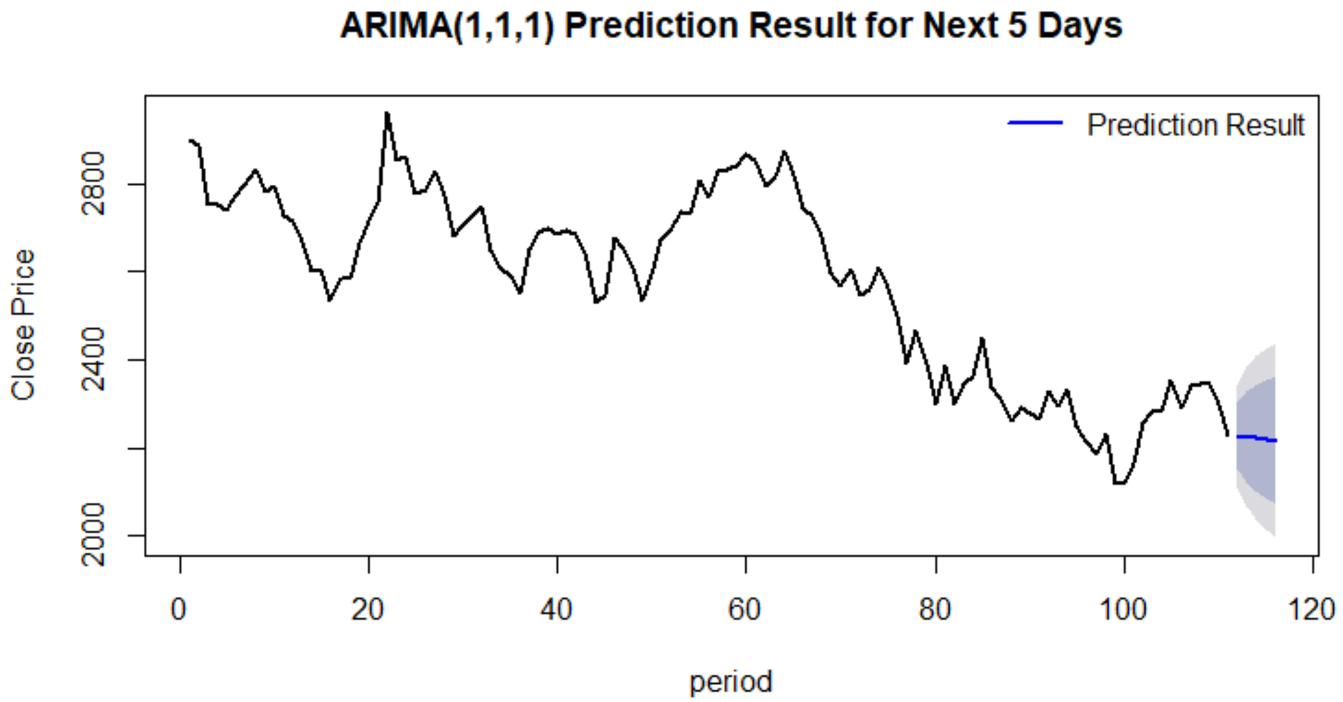


Because the fit model still has the lower aic, aicc, and bic, also the critical area value > test statistics in the hypothetical test of ARIMA(2,1,1) and ARIMA (1,1,2), then the overfitting is useless. So **we use the fit model**.

Data Training Forecasting



Actual Data Forecasting



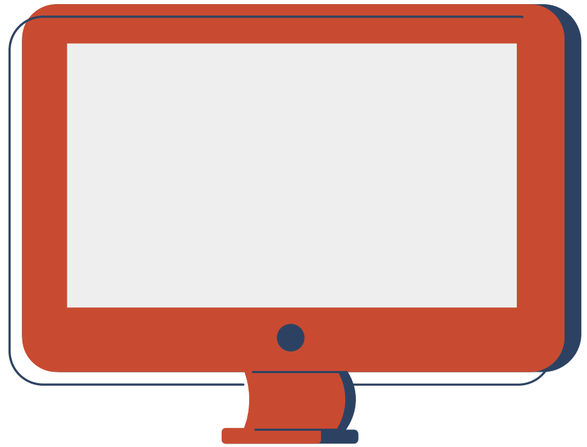
Date	Forecasting Result (USD)
June 11, 2022	2226.589
June 12, 2022	2224.288
June 13, 2022	2221.681
June 14, 2022	2218.796
June 15, 2022	2215.657

1

Based on the prediction result table for the next 5 days, we clearly see that there is a downward trend but not significant.

2

The model obtained is intended to be a reference for investors to invest in Alphabet Inc's Google stock.



3

Forecasting is not a definite value that can occur in the future, because it is necessary to pay attention to factors in the environment that can affect the company's stock price.

Thank you