

Bab 7. Model ARIMA

Data yang digunakan merupakan data saham bulanan PT Indofood Sukses Makmur dari tahun 2002 hingga 2022. Akan dilakukan peramalan data timeseries dengan menggunakan ARIMA (Autoregressive Integrated Moving Average) untuk 1 periode ke depan.

A. Stasioneritas

- Hipotesis
 H_0 : data tidak stasioner terhadap mean
 H_1 : data stasioner terhadap mean
- Tingkat Signifikansi
 $\alpha = 0.1$
- Statistik Uji

```
Augmented Dickey-Fuller Test  
data: data  
Dickey-Fuller = -1.5934, Lag order = 6, p-value = 0.747  
alternative hypothesis: stationary
```

- Daerah Kritik
 H_0 ditolak, jika $P\text{-Value} < \alpha = 0.1$
- Kesimpulan
Karena nilai $P\text{-Value} = 0.747 > \alpha = 0.1$, maka H_0 tidak ditolak. Oleh karena itu, dapat disimpulkan bahwa data tidak stasioner terhadap mean.

Interpretasi:

Akan dilakukan uji kestasioneran data dengan menggunakan Dickey-Fuller Test dengan hipotesis awal yaitu data tidak stasioner terhadap mean dan hipotesis alternatif yaitu data stasioner terhadap mean. Didapatkan nilai $P\text{-Value} = 0.747$ yang mana nilainya lebih dari $\alpha = 0.1$, sehingga H_0 tidak ditolak. Oleh karena itu, dapat disimpulkan bahwa data tidak stasioner terhadap mean. Untuk membuat data menjadi stasioner, akan dilakukan transformasi terhadap data.

B. Transformasi

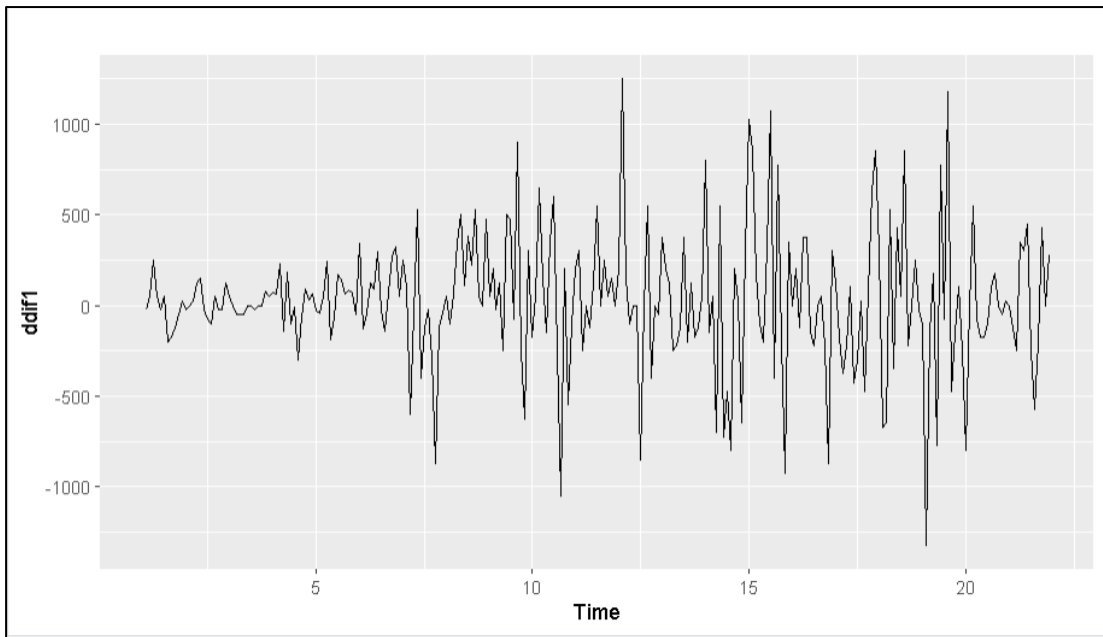
Transformasi akan dilakukan dengan menggunakan differencing terhadap data yang digunakan. Untuk mengetahui berapa kali differencing dilakukan, akan dilakukan perhitungan dengan RStudio sebagai berikut.

```
> ndiffs(log(data), "adf")  
[1] 1
```

Didapatkan bahwa differencing yang perlu dilakukan yaitu 1 kali. Selanjutnya, dilakukan differencing sebanyak 1 kali pada data.

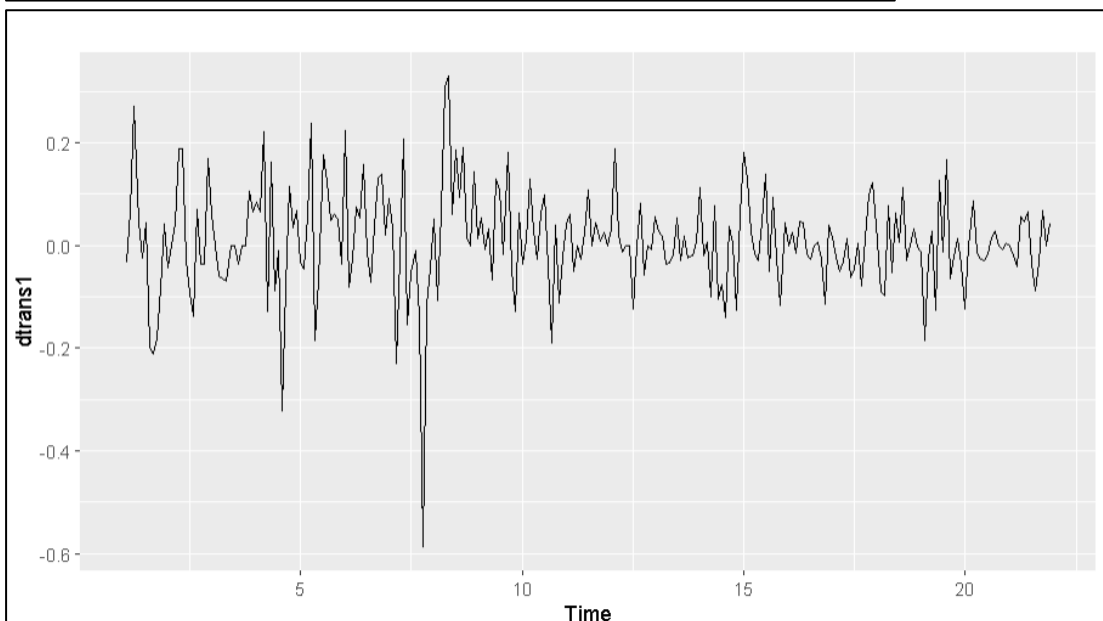
- Differencing

```
Augmented Dickey-Fuller Test  
data: ddif1  
Dickey-Fuller = -6.3629, Lag order = 6, p-value = 0.01  
alternative hypothesis: stationary
```



- Differencing dengan Log-Trans

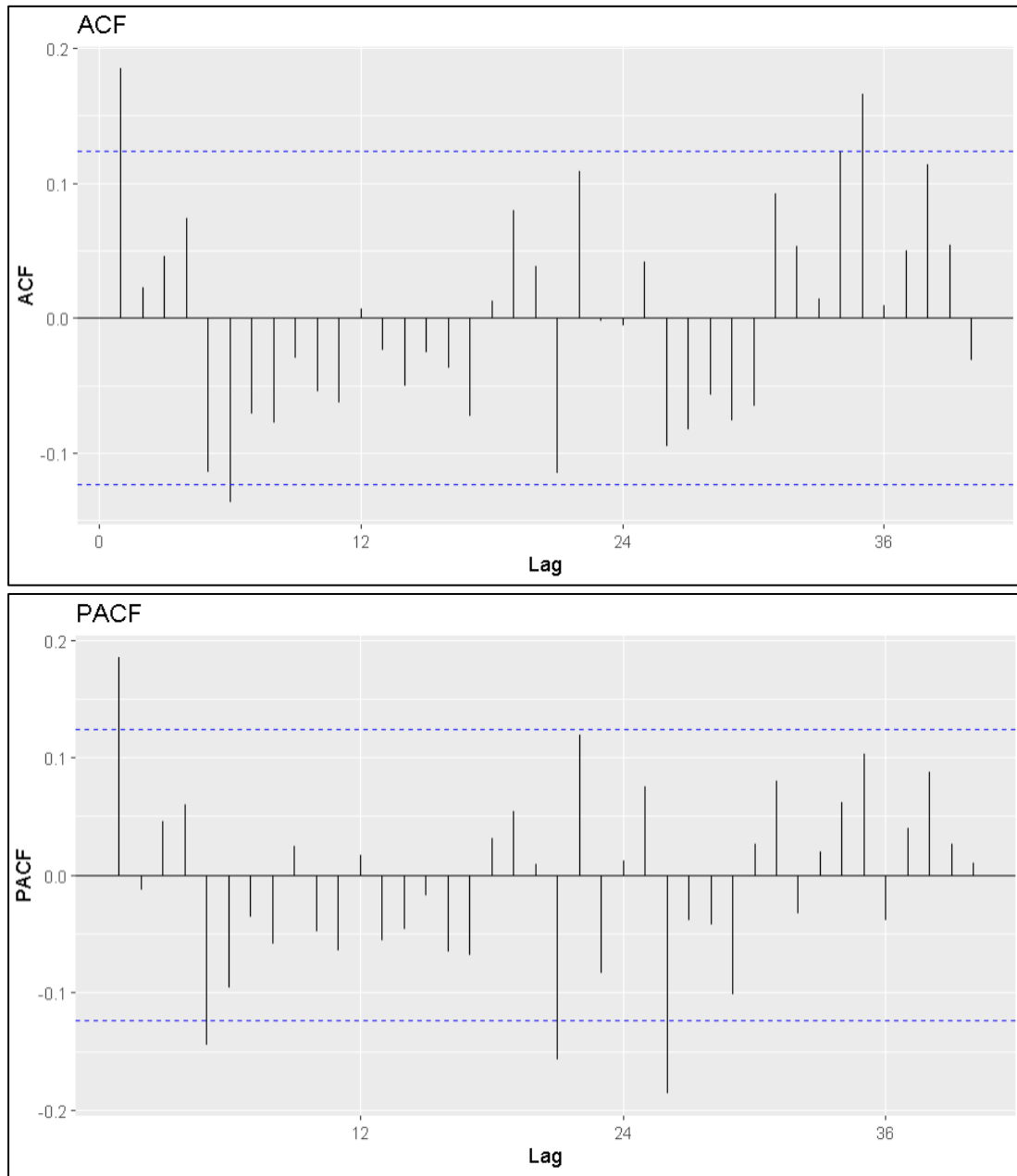
```
Augmented Dickey-Fuller Test  
data: dtrans1  
Dickey-Fuller = -6.681, Lag order = 6, p-value = 0.01  
alternative hypothesis: stationary
```



Interpretasi:

Didapatkan nilai P-Value yang sama, yaitu 0.01 yang mana nilainya lebih kecil dari $\alpha = 0.1$, sehingga H_0 tidak ditolak. Oleh karena itu, dapat disimpulkan bahwa data tidak stasioner terhadap mean. Untuk menentukan data differencing mana yang akan digunakan, akan dipilih nilai |Dickey-Fuller| yang paling besar di antara dua metode. Didapatkan bahwa nilai |Dickey-Fuller| terbesar yaitu pada data differencing dengan Log-Trans, sehingga data tersebut yang akan digunakan pada identifikasi model selanjutnya.

C. Identifikasi Model



Interpretasi:

Untuk mengetahui model ARIMA apakah yang akan digunakan, dilakukan pencarian ACF, PACF, dan differencing yang dilakukan pada data. Sebelumnya, sudah didapatkan nilai differencing yang perlu dilakukan untuk menjadikan data stasioner, yaitu 1. Nilai differencing ini akan digunakan sebagai nilai d pada model ARIMA (p,d,q).

Selanjutnya, untuk mengetahui nilai p pada model, akan dilihat dari plot PACF data differencing tadi, dilihat lag terakhir yang melewati batas pada 4 lag pertama. Pada plot yang sudah terbentuk di atas, dapat diketahui bahwa lag terakhir yang melewati batas yaitu lag 1, sehingga $p = 1$.

Kemudian, untuk mengetahui nilai q pada model, akan dilihat dari plot ACF data differencing tadi, dilihat lag terakhir yang melewati batas pada 4 lag pertama. Pada plot yang sudah terbentuk di atas, dapat diketahui bahwa lag terakhir yang melewati batas yaitu lag 1, sehingga $q = 1$.

Didapatkan bahwa model ARIMA yang akan digunakan yaitu ARIMA(1,1,1).

D. Underfitting dan Pemodelan

- Underfitting

p	d	q	c/tc
1	1	1	c
1	1	0	c
0	1	1	c
1	1	1	tc
1	1	0	tc
0	1	1	tc

- Pemodelan

- o Hipotesis

H_0 : parameter tidak signifikan terhadap model

H_1 : parameter signifikan terhadap model

- o Tingkat Signifikansi

$\alpha = 0.1$

- o Statistik Uji

```
> coeftest(c111)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1    0.0840999  0.3676358  0.2288  0.8191
ma1    0.1036889  0.3680212  0.2817  0.7781
drift  0.0086000  0.0073925  1.1634  0.2447
```

```
> coeftest(c110)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1    0.1845106  0.0619204  2.9798 0.002884 **
drift  0.0086004  0.0075224  1.1433 0.252911
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> coeftest(c011)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ma1    0.1849614  0.0608748  3.0384 0.002378 **
drift  0.0086021  0.0072711  1.1831 0.236786
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> coeftest(tc111)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1    0.121698   0.391900  0.3105  0.7562
ma1    0.070587   0.396201  0.1782  0.8586
```

```
> coefstest(tc110)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.190612    0.061845  3.0821 0.002056 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coefstest(tc011)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ma1  0.189243    0.060521  3.1269 0.001767 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Daerah Kritik
 H_0 ditolak, jika $P\text{-Value} < \alpha = 0.1$
- Kesimpulan

Model	Variabel	P-Value	Kesimpulan
c111	ar1	0.8191	P-Value $> \alpha = 0.1$, H_0 tidak ditolak, sehingga parameter tidak signifikan terhadap model.
	ma1	0.7781	P-Value $> \alpha = 0.1$, H_0 tidak ditolak, sehingga parameter tidak signifikan terhadap model.
	drift	0.2447	P-Value $> \alpha = 0.1$, H_0 tidak ditolak, sehingga parameter tidak signifikan terhadap model.
c110	ar1	0.002884	P-Value $< \alpha = 0.1$, H_0 ditolak, sehingga parameter signifikan terhadap model.
	drift	0.252911	P-Value $> \alpha = 0.1$, H_0 tidak ditolak, sehingga parameter tidak signifikan terhadap model.
c011	ma1	0.002327	P-Value $< \alpha = 0.1$, H_0 ditolak, sehingga parameter signifikan terhadap model.
	drift	0.236786	P-Value $> \alpha = 0.1$, H_0 tidak ditolak, sehingga parameter tidak signifikan terhadap model.
tc111	ar1	0.7562	P-Value $> \alpha = 0.1$, H_0 tidak ditolak, sehingga parameter tidak signifikan terhadap model.
	ma1	0.8586	P-Value $> \alpha = 0.1$, H_0 tidak ditolak, sehingga parameter tidak signifikan terhadap model.
tc110	ar1	0.002056	P-Value $< \alpha = 0.1$, H_0 ditolak, sehingga parameter signifikan terhadap model.
tc011	ma1	0.001767	P-Value $< \alpha = 0.1$, H_0 ditolak, sehingga parameter signifikan terhadap model.

Interpretasi:

Dilakukan underfitting dari model ARIMA(1,1,1), didapatkan 6 model. Model tersebut akan diuji satu persatu untuk mengetahui apakah parameter signifikan terhadap model. Dengan hipotesis awal yaitu parameter tidak signifikan terhadap model dan hipotesis alternatif yaitu parameter signifikan terhadap model, didapatkan nilai P-Value masing-masing variabel dari setiap model yang terbentuk. Setelah diuji, didapatkan bahwa hanya 2 model yang memiliki $P\text{-Value} < \alpha = 0.1$, maka H_0 ditolak, sehingga parameter signifikan terhadap model. Model tersebut adalah tc110 dan tc011. Selanjutnya, kedua model ini akan masuk ke diagnostic checking.

E. Diagnostic Checking

- Model tc110

- Uji autokorelasi

- Hipotesis

- H_0 : tidak terjadi autokorelasi pada residual

- H_1 : terjadi autokorelasi pada residual

- Tingkat Signifikansi

- $\alpha = 0.1$

- Statistik Uji

```
Box-Ljung test
```

```
data: tc110$residuals
```

```
X-squared = 0.0023859, df = 1, p-value = 0.961
```

- Daerah Kritik

- H_0 ditolak, jika $P\text{-Value} < \alpha = 0.1$

- Kesimpulan

- Karena $P\text{-Value} = 0.961 > \alpha = 0.1$, maka H_0 tidak ditolak. Oleh karena itu, dapat disimpulkan bahwa tidak terjadi autokorelasi pada residual.

- Uji homoskedastisitas

- Hipotesis

- H_0 : residual bersifat homoskedastik

- H_1 : residual tidak bersifat homoskedastik

- Tingkat Signifikansi

- $\alpha = 0.1$

- Statistik Uji

```
Box-Ljung test
```

```
data: (tc110$residuals)^2
```

```
X-squared = 0.66094, df = 1, p-value = 0.4162
```

- Daerah Kritik

- H_0 ditolak, jika $P\text{-Value} < \alpha = 0.1$

- Kesimpulan

- Karena $P\text{-Value} = 0.4162 > \alpha = 0.1$, maka H_0 tidak ditolak. Oleh karena itu, dapat disimpulkan bahwa residual bersifat homoskedastik.

- Uji normalitas

- Hipotesis

- H_0 : residual berdistribusi normal

- H_1 : residual tidak berdistribusi normal

- Tingkat Signifikansi

- $\alpha = 0.1$

- Statistik Uji

```
Jarque Bera Test
```

```
data: tc110$residuals
```

```
X-squared = 253.34, df = 2, p-value < 2.2e-16
```

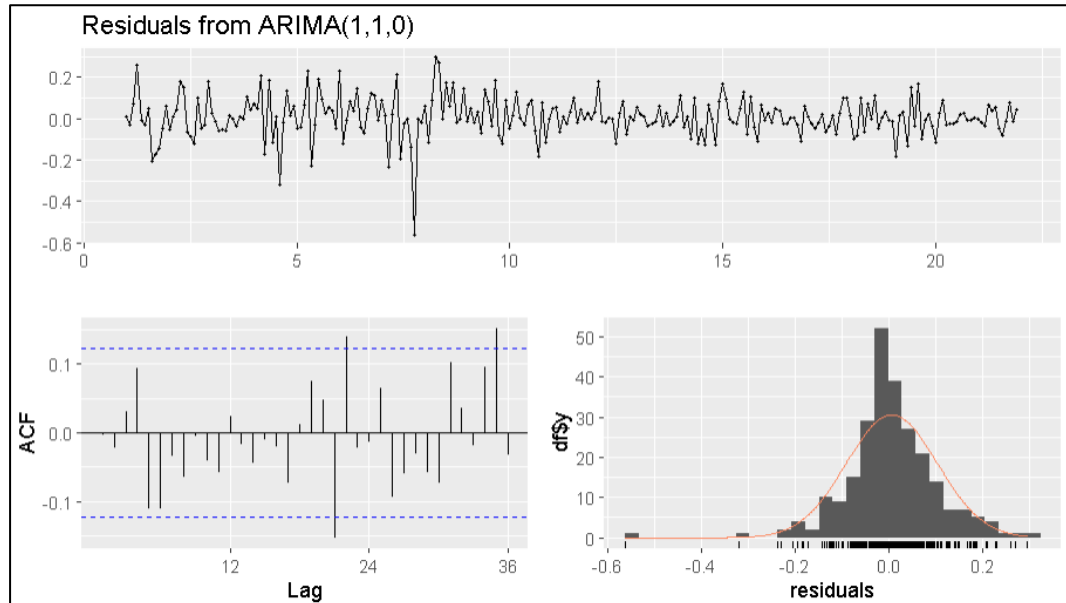
- Daerah Kritik

H_0 ditolak, jika $P\text{-Value} < \alpha = 0.1$

- Kesimpulan

Karena $P\text{-Value} < 2.2e-16 < \alpha = 0.1$, maka H_0 ditolak. Oleh karena itu, dapat disimpulkan bahwa residual tidak berdistribusi normal.

- Residual



Interpretasi:

Dilakukan diagnostic checking terhadap model tc110 dengan menguji autokorelasi, homoskedastik, dan normalitas. Didapatkan bahwa nilai $P\text{-Value}$ uji autokorelasi dan homoskedastik $> \alpha = 0.1$, maka H_0 tidak ditolak. Oleh karena itu, dapat disimpulkan bahwa tidak terjadi autokorelasi pada residual dan residual bersifat homoskedastik. Selain itu, didapatkan nilai $P\text{-Value}$ uji normalitas $< \alpha = 0.1$, maka H_0 ditolak. Oleh karena itu, dapat disimpulkan bahwa residual tidak berdistribusi normal.

- Model tc011

- Uji autokorelasi

- Hipotesis

H_0 : tidak terjadi autokorelasi pada residual

H_1 : terjadi autokorelasi pada residual

- Tingkat Signifikansi

$\alpha = 0.1$

- Statistik Uji

```
Box-Ljung test
data:  tc011$residuals
X-squared = 0.0001776, df = 1, p-value = 0.9894
```

- Daerah Kritik

H_0 ditolak, jika $P\text{-Value} < \alpha = 0.1$

- Kesimpulan

Karena $P\text{-Value} = 0.9894 > \alpha = 0.1$, maka H_0 tidak ditolak. Oleh karena itu, dapat disimpulkan bahwa tidak terjadi autokorelasi pada residual.

- Uji homoskedastisitas

- Hipotesis

H_0 : residual bersifat homoskedastik

H_1 : residual tidak bersifat homoskedastik

- Tingkat Signifikansi

$\alpha = 0.1$

- Statistik Uji

```
Box-Ljung test

data: (tc011$residuals)^2
X-squared = 0.64568, df = 1, p-value = 0.4217
```

- Daerah Kritik

H_0 ditolak, jika $P\text{-Value} < \alpha = 0.1$

- Kesimpulan

Karena $P\text{-Value} = 0.4217 > \alpha = 0.1$, maka H_0 tidak ditolak. Oleh karena itu, dapat disimpulkan bahwa residual bersifat homoskedastik.

- Uji normalitas

- Hipotesis

H_0 : residual berdistribusi normal

H_1 : residual tidak berdistribusi normal

- Tingkat Signifikansi

$\alpha = 0.1$

- Statistik Uji

```
Jarque Bera Test

data: tc011$residuals
X-squared = 257.78, df = 2, p-value < 2.2e-16
```

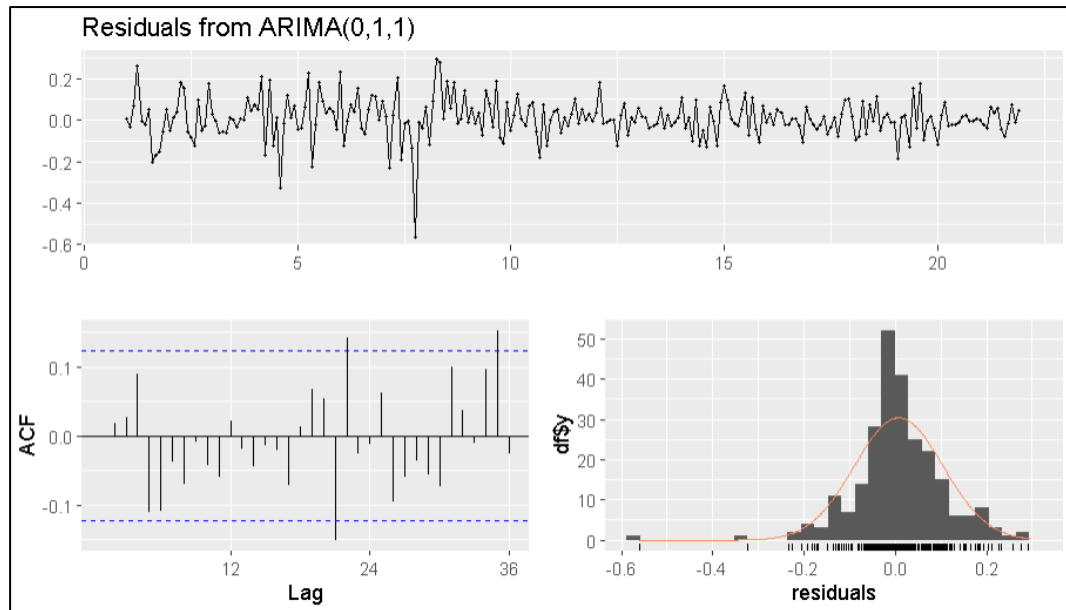
- Daerah Kritik

H_0 ditolak, jika $P\text{-Value} < \alpha = 0.1$

- Kesimpulan

Karena $P\text{-Value} < 2.2e-16 < \alpha = 0.1$, maka H_0 ditolak. Oleh karena itu, dapat disimpulkan bahwa residual tidak berdistribusi normal.

- Residual



Interpretasi:

Dilakukan diagnostic checking terhadap model tc011 dengan menguji autokorelasi, homoskedastik, dan normalitas. Sama seperti model tc110, didapatkan bahwa nilai P-Value uji autokorelasi dan homoskedastik $> \alpha = 0.1$, maka H_0 tidak ditolak. Oleh karena itu, dapat disimpulkan bahwa tidak terjadi autokorelasi pada residual dan residual bersifat homoskedastik. Selain itu, didapatkan nilai P-Value uji normalitas $< \alpha = 0.1$, maka H_0 ditolak. Oleh karena itu, dapat disimpulkan bahwa residual tidak berdistribusi normal.

F. Pemilihan Model Terbaik dan Penulisan Model

- Pemilihan Model Terbaik

```
> model_selection
```

	Model	LogLik	AIC	BIC
1	tc110	228.0727	-452.1453	-445.0944
2	tc011	228.0263	-452.0527	-445.0018

Interpretasi:

Untuk melakukan pemilihan model terbaik, akan dilihat nilai loglikelihood, AIC, dan BIC dari setiap model. Model terbaik berlaku apabila nilai loglikelihood merupakan nilai terbesar serta nilai AIC dan BIC merupakan nilai terkecil. Pada output di atas, dapat dilihat nilai model tc110 memenuhi syarat model terbaik (nilai loglikelihood terbesar serta nilai AIC dan BIC terkecil), sehingga model tc110 merupakan model terbaik.

- Penulisan Model

```
> tc110
Series: data
ARIMA(1,1,0)
Box Cox transformation: lambda= 0

Coefficients:
      ar1
      0.1906
s.e.    0.0618

sigma^2 = 0.009549: log likelihood = 228.07
AIC=-452.15   AICC=-452.1   BIC=-445.09
```

$$(1 - a_1B - a_2B^2 - \dots - a_pB^p)(1 - B)^dX_t = \mu + (1 + b_1B + b_2B^2 + \dots + b_qB^q)\varepsilon_t$$

$$(1 - 0.1906B)(1 - B)^1X_t = 0 + (1 - 0)\varepsilon_t$$

$$(1 - 1.1906B + 0.1906B^2)X_t = \varepsilon_t$$

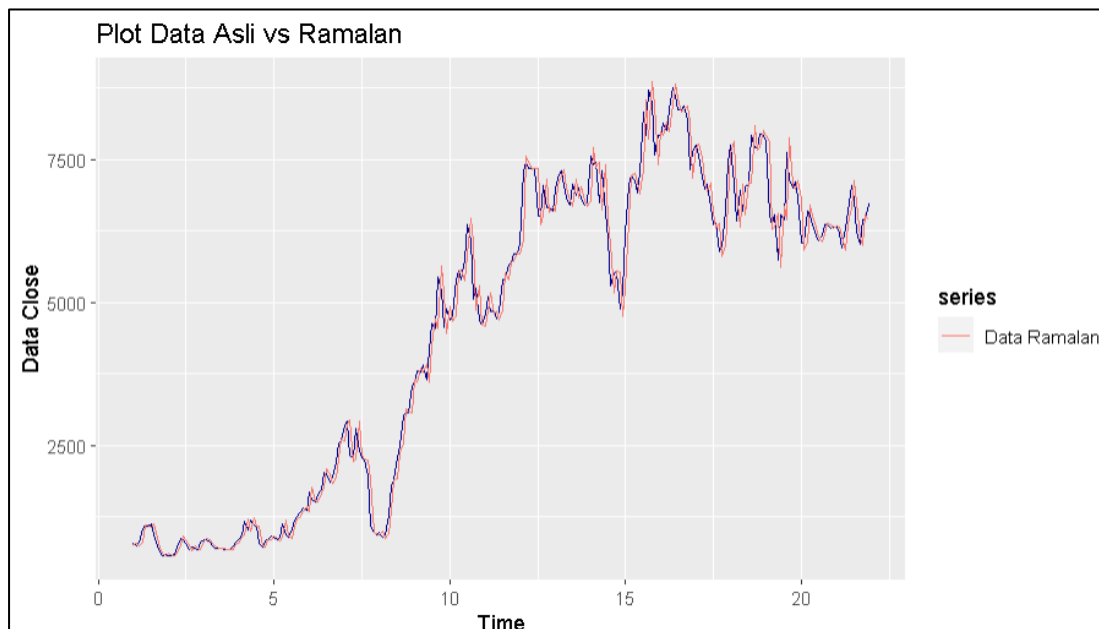
$$X_t - 1.1906X_{t-1} + 0.1906X_{t-2} = \varepsilon_t$$

$$X_t = 1.1906X_{t-1} - 0.1906X_{t-2} + \varepsilon_t$$

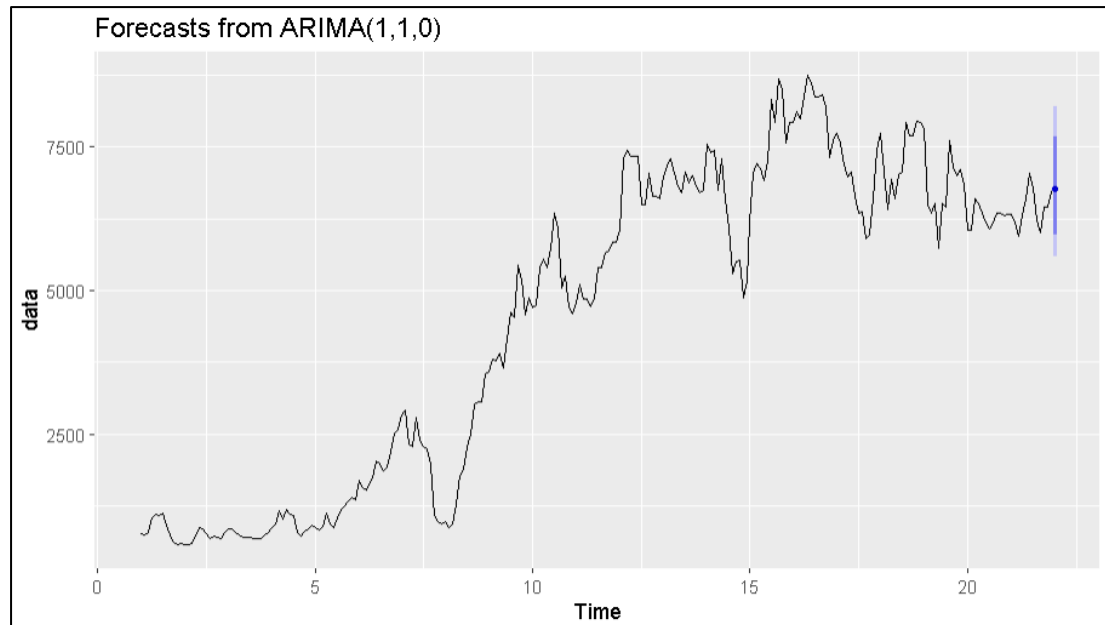
Interpretasi:

- Setiap kenaikan satu satuan data pada 1 periode yang lalu, akan berakibat pada kenaikan data ramalan sebesar 1.1906 satuan dengan menganggap variabel lain konstan.
- Setiap kenaikan satu satuan data pada 2 periode yang lalu, akan berakibat pada penurunan data ramalan sebesar 0.1906 satuan dengan menganggap variabel lain konstan.
- Data ramalan akan bernilai galat atau error apabila setiap satuan data, baik 1 periode lalu maupun 2 periode lalu, bernilai 0.

G. Forecast dan Akurasi



```
> forecast(tc110,1)
      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Jan 22      6778.734 5980.827 7683.09 5597.193 8209.691
```



```
> accuracy(tcl10)
```

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	16.21838	368.2645	255.5229	0.2168174	6.839652	0.3024983	-0.1594408

Interpretasi:

Dilakukan peramalan selama 1 periode menggunakan model ARIMA(1,1,0) terhadap data saham PT Indofood Sukses Makmur dari tahun 2002 hingga 2022, didapatkan plot seperti di atas. Data ramalan yang dihitung merupakan data ramalan untuk bulan Januari 2023, didapatkan nilai 6778.734 yang digambarkan pada plot Forecasts from ARIMA(1,1,0). Selanjutnya, akan dihitung akurasi dari peramalan tersebut. Didapatkan nilai RMSE (Root Mean Square Error), yaitu ukuran rata-rata kesalahan prediksi dalam skala yang sama dengan variabel target. Ini menghitung akar kuadrat rata-rata dari selisih antara nilai prediksi dan nilai sebenarnya. Semakin kecil nilainya RMSE suatu ramalan, maka tingkat kesalahan prediksi semakin rendah. RMSE yang didapat dari ramalan di atas yaitu 368.2645.

Sintaks:

```
#Loading the required library
library(ggplot2)
library(forecast)
library(tseries)
library(lmtest)

#Importing the data
INDF<-read.delim('clipboard')
data<-ts(INDF$Close,freq=12)
View(data)

#Creating the plot
autoplot(data)+xlab("Waktu (Bulan)") +
  ylab("Data close") +
  ggtitle("Plot Data Asli") + geom_point()

#Stationarity Test
adf.test(data)

#Transforming data
##Checking the number of differences required
ndiffs(log(data),"adf")

##Diff 1
ddif1 = diff(data, differences=1)
adf.test(ddif1)
autoplot(ddif1)

#Diff 2 with Log-Trans
dtrans1 = diff(log(data), differences=1)
adf.test(dtrans1)
```

```
autoplot(dtrans1)

#Identifying the model
#Membuat plot ACF dan PACF
ggAcf(dtrans1,lag.max = 40) + ggtitle("ACF")
ggPacf(dtrans1,lag.max = 40) + ggtitle("PACF")
##ARIMA(1,1,1)##

#Underfitting
c111<-Arima(data,order=c(1,1,1),include.constant = T, lambda = 0)
coeftest(c111)
c110<-Arima(data,order=c(1,1,0),include.constant = T, lambda = 0)
coeftest(c110)
c011<-Arima(data,order=c(0,1,1),include.constant = T, lambda = 0)
coeftest(c011)

tc111<-Arima(data,order=c(1,1,1),include.constant = F, lambda = 0)
coeftest(tc111)
tc110<-Arima(data,order=c(1,1,0),include.constant = F, lambda = 0)
coeftest(tc110)
tc011<-Arima(data,order=c(0,1,1),include.constant = F, lambda = 0)
coeftest(tc011)

#Diagnostic Checking
Box.test(tc110$residuals,type="Ljung") #uji autokorelasi
Box.test((tc110$residuals)^2,type="Ljung") #uji homoskedastik
jarque.bera.test(tc110$residuals) #uji normalitas
checkresiduals(tc110)

Box.test(tc011$residuals,type="Ljung") #uji autokorelasi
Box.test((tc011$residuals)^2,type="Ljung") #uji homoskedastik
jarque.bera.test(tc011$residuals) #uji normalitas
```

```
checkresiduals(tc011)

#Best model selection
mod_tc110 = data.frame(Model = "tc110", LogLik = logLik(tc110), AIC =
AIC(tc110), BIC = BIC(tc110))
mod_tc011 = data.frame(Model = "tc011", LogLik = logLik(tc011), AIC =
AIC(tc011), BIC = BIC(tc011))

model_selection = rbind(mod_tc110, mod_tc011)
model_selection

#Penulisan Model
tc110

#Best model
##Plot##
autoplot(tc110$x, col="darkblue") +
  autolayer(fitted(tc110), series = "Data Ramalan") +
  ylab("Data Close") +
  ggtitle("Plot Data Asli vs Ramalan")
##Forecast and Accuracy##
accuracy(tc110)
forecast(tc110,1)
autoplot(forecast(tc110,1))
```