Statistika Inferensial Lanjut Teori & Praktik

Pertemuan 13

Model $\boldsymbol{AR(p)}$

AR(p) adalah model linear yang paling dasar untuk proses stasioner. Model ini dapat diartikan sebagai proses hasil regresi dengan dirinya sendiri. Secara matematis dapat dituliskan:

$$X_t = \emptyset_0 + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_p X_{t-p} + a_t$$

keterangan:

X_t = data pada periode t, t =1, 2, 3, …, n

X_{t-i}= data pada periode t-i, i =1, 2, 3, …, p

a_t = error pada periode t

$\emptyset_0$ =konstanta

$\emptyset_i$= koefisien AR, i =1, 2, 3, …, p

Model $AR(1)$

Model autoregresi tingkat 1 atau proses AR(1), secara matematis didefinisikan:

$$X_t = \emptyset_0 + \emptyset_1 X_{t-1} + a_t$$

keterangan:

X_t = data pada periode t, t =1, 2, 3, …, n

X_{t-1}= data pada periode t-1

a_t = error pada periode t

$\emptyset_0$ =konstanta

$\emptyset_1$= koefisien AR ke-1

Model $AR(2)$

Model autoregresi tingkat 2 atau proses AR(2), secara matematis didefinisikan sebagai:

$$X_t = \emptyset_0 + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + a_t$$

keterangan:

X_t = data pada periode t, t =1, 2, 3, …, n

X_{t-1}= data pada periode t-1

X_{t-2}= data pada periode t-2

a_t = error pada periode t

$\emptyset_0$ =konstanta

$\emptyset_1$= koefisien AR ke-1

$\emptyset_2$= koefisien AR ke-2

Model $MA(q)$

Bentuk umum dari model moving average tingkat q atau MA(q) didefinisikan sebagai:

$$X_t = \theta_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

keterangan:

X_t = data pada periode t, t =1, 2, 3, …, n

a_{t-i}= error pada periode t-i, i= 1, 2, 3, …, q

θ_0 =konstanta

θ_i= koefisien MA, i =1, 2, 3, ..., q

Model $MA(1)$

Model moving average juga diawali dengan tingkat 1 atau proses MA(1), didefinisikan sebagai

$$X_t = \theta_0 + a_t - \theta_1 a_{t-1}$$

keterangan:

X_t = data pada periode t, t =1, 2, 3, …, n

a_{t-1}= error pada periode t-1

a_t= error pada periode t

θ_0 =konstanta

θ_1= koefisien MA, ke-1

Model $MA(2)$

Model moving average tingkat 2 atau proses MA(2), didefinisikan sebagai:

$$X_t = \theta_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

keterangan:

X_t = data pada periode t, t =1, 2, 3, …, n

a_{t-1}= error pada periode t-1

a_t= error pada periode t

a_{t-2}= error pada periode t-2

θ_0 =konstanta

θ_1= koefisien MA, ke-1

θ_2= koefisien MA, ke-2

Model $\boldsymbol{ARMA(p,q)}$

Model ini merupakan gabungan antara AR(p) dengan MA(q), sehingga dinyatakan sebagai ARMA(p,q), dengan bentuk umumnya:

$$X_t = \emptyset_0 + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_p X_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

Keterangan:

X_t = data pada periode t, t =1, 2, 3, …, n

X_{t-i}= data pada periode t-i, i =1, 2, 3, …, p

a_{t-i}= error pada periode t-i, i= 1, 2, 3, …, q

θ_0 =konstanta

θ_i= koefisien MA, i =1, 2, 3, ..., q

$\emptyset_i$= koefisien AR, i =1, 2, 3, …, p

Model $\boldsymbol{ARMA(1,1)}$

Model ini merupakan kombinasi antara AR(1) dan MA(1), matematisnya dapat didefinisikan sebagai:

$$X_t = \emptyset_0 + \emptyset_1 X_{t-1} + a_t - \theta_1 a_{t-1}$$

Keterangan:

X_t = data pada periode t, t =1, 2, 3, …, n

X_{t-1}= data pada periode t-1

a_t= error pada periode t

a_{t-1}= error pada periode t-1

θ_0 =konstanta

θ_1= koefisien MA, Ke-1

$\emptyset_1$= koefisien AR ke-1

Model $\boldsymbol{ARMA(2,1)}$

Secara matematis, model ARMA(2,1) dapat didefinisikan sebagai berikut:

$$X_t = \emptyset_0 + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + a_t - \theta_1 a_{t-1}$$

Keterangan:

X_t = data pada periode t, t =1, 2, 3, …, n

X_{t-1}= data pada periode t-1

a_t= error pada periode t

a_{t-1}= error pada periode t-1

θ_0 =konstanta

θ_i= koefisien MA, i =1, 2, 3, ..., q

$\emptyset_i$= koefisien AR, i =1, 2, 3, …, p

Model $\boldsymbol{ARIMA(p, d, q)}$

ARIMA sering juga disebut metode runtun waktu Box-Jenkins. ARIMA sangat baik ketepatannya untuk peramalan jangka pendek, sedangkan untuk peramalan jangka panjang ketepatan peramalannya kurang baik. Biasanya akan cenderung flat (mendatar/konstan) untuk periode yang cukup panjang.

Model Autoregresif Integrated Moving Average (ARIMA) adalah model yang secara penuh mengabaikan independen variabel dalam membuat peramalan. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. ARIMA cocok jika observasi dari deret waktu (time series) secara statistik berhubungan satu sama lain (dependent).

Lanjutan..

Tujuan model ini adalah untuk menentukan hubungan statistik yang baik antar variabel yang diramal dengan nilai historis variabel tersebut sehingga peramalan dapat dilakukan dengan model tersebut.

Model ARIMA terdiri dari tiga langkah dasar, yaitu tahap identifikasi, tahap penaksiran dan pengujian, dan pemeriksaan diagnostik. Selanjutnya model ARIMA dapat digunakan untuk melakukan peramalan jika model yang diperoleh memadai.

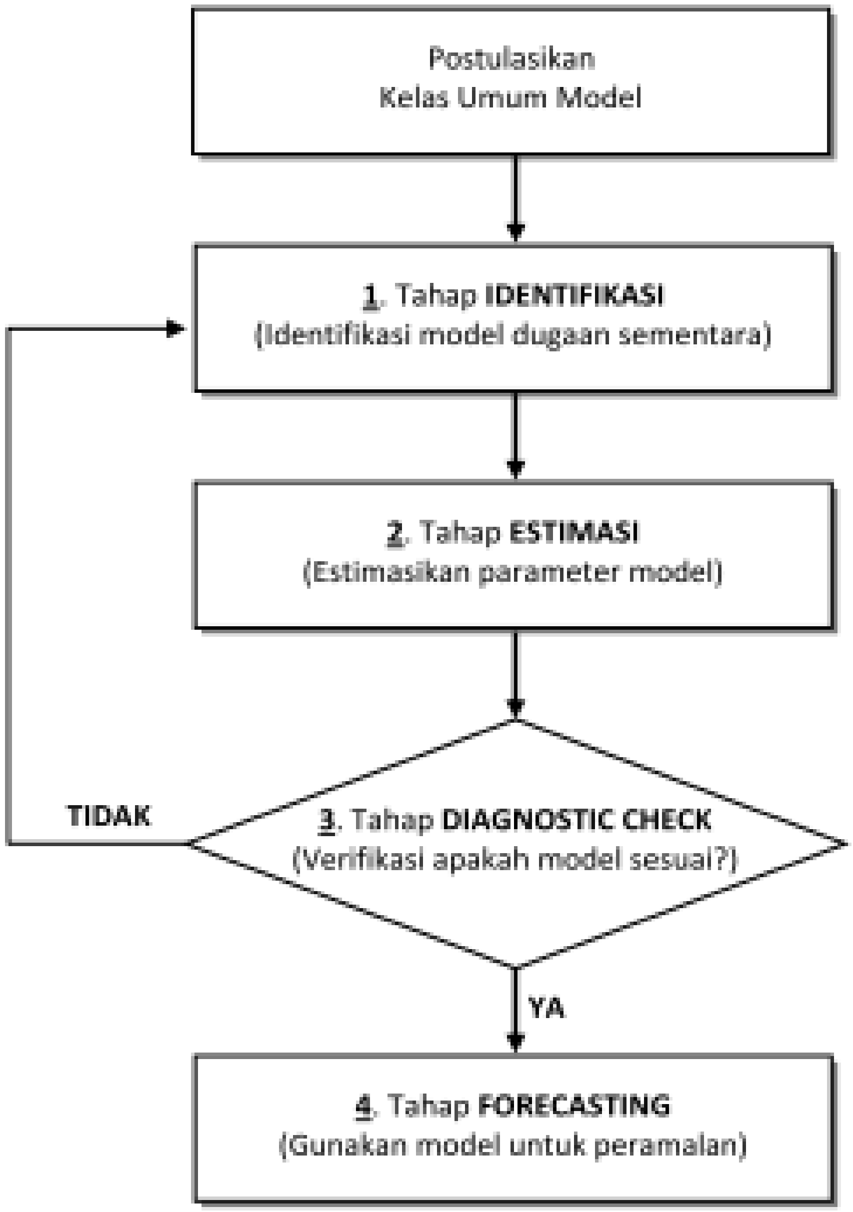
Stationeritas

Hal yang perlu diperhatikan adalah bahwa kebanyakan deret berkala bersifat nonstasioner dan bahwa aspek-aspek AR dan MA dari model ARIMA hanya berkenaan dengan deret berkala yang stasioner.

Stasioneritas berarti tidak terdapat pertumbuhan atau penurunan pada data. Data secara kasarnya harus horizontal sepanjang sumbu waktu. Dengan kata lain, fluktuasi data berada di sekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan varians dari fluktuasi tersebut pada pokoknya tetap konstan setiap waktu.

Suatu deret waktu yang tidak stasioner harus diubah menjadi data stasioner dengan melakukan differencing. Yang dimaksud dengan differencing adalah menghitung perubahan atau selisih nilai observasi. Nilai selisih yang diperoleh dicek lagi apakah stasioner atau tidak. Jika belum stasioner maka dilakukan differencing lagi. Jika varians tidak stasioner, maka dilakukan transformasi logaritma.

Skema Box-Jenkins



Proses $\boldsymbol{ARIMA(p,d,q)}$

Secara umum, bentuk matematis dari model ARIMA(p,d,q) dapat ditulis sebagai berikut (Cryer, 1986; Wei, 2006)

$$(1-\phi_1 B - \ldots - \phi_p B^p)(1-B)^d Z_t = \theta_0 + (1-\theta_1 B - \ldots - \theta_q B^q)a_t \,,$$

dengan *B* adalah operator mundur, yaitu $B^k Z_t = Z_{t-k}$. Penentuan orde p dan q dari model ARIMA pada suatu data runtun waktu dilakukan dengan mengidentifikasi plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) dari data yang sudah stasioner. Berikut ini adalah petunjuk umum untuk penentuan orde p dan q pada suatu data runtun waktu yang sudah stasioner.

Tabel Pola

ACF-PACF

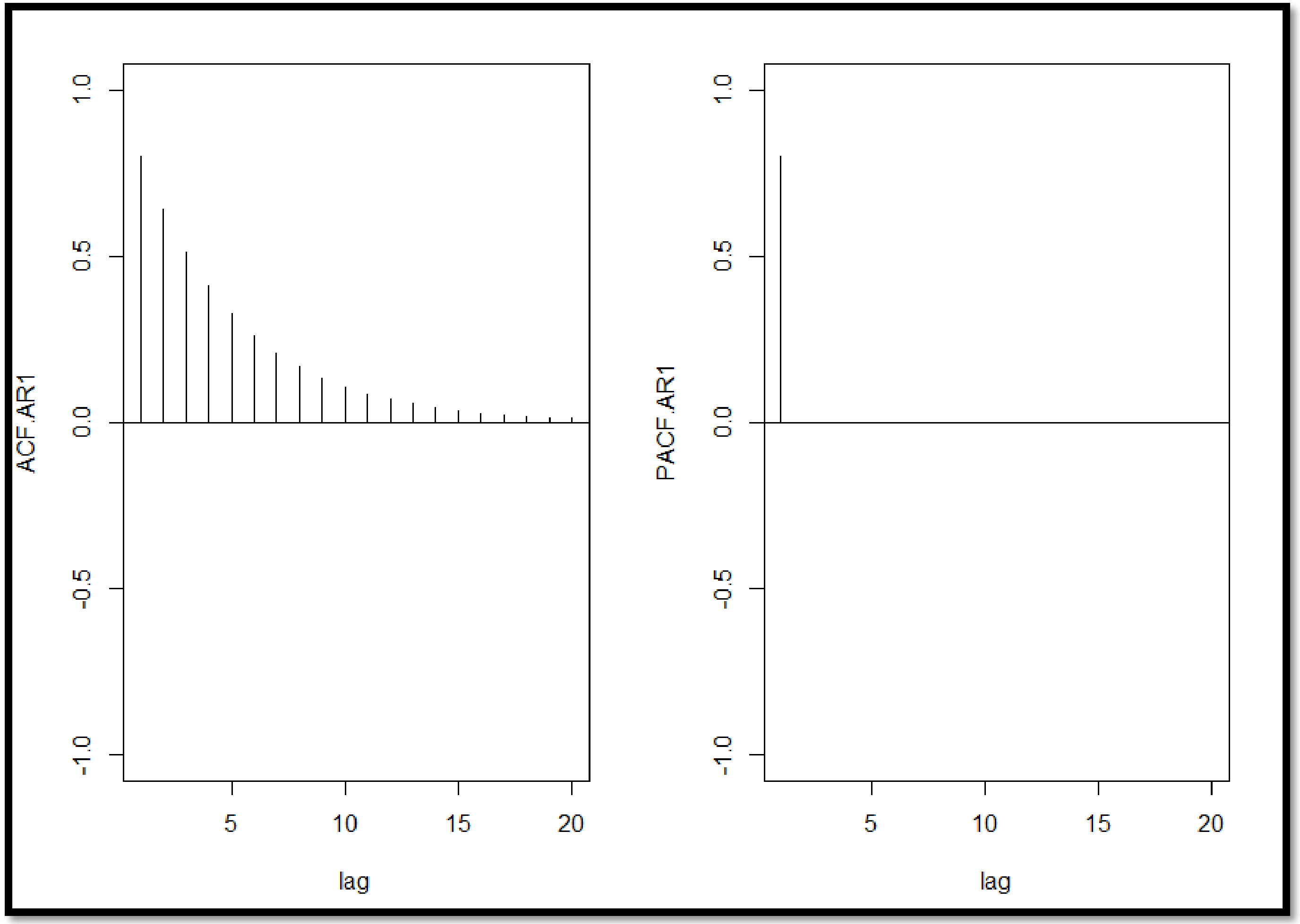
Tabel 11.1. Pola teoritis ACF dan PACF dari proses yang stasioner

Proses	ACF	PACF
AR(p)	*Dies down* (turun cepat secara eksponensial / sinusoidal)	*Cuts off after lag p* (terputus setelah lag p)
MA(q)	*Cuts off after lag q* (terputus setelah lag q)	*Dies down* (turun cepat secara eksponensial / sinusoidal)
ARMA(p,q)	*Dies down* (turun cepat secara eksponensial / sinusoidal))	*Dies down* (turun cepat secara eksponensial / sinusoidal))
AR(p) atau MA(q)	*Cuts off after lag q* (terputus setelah lag q)	*Cuts off after lag p* (terputus setelah lag p)
White noise (Random)	Tidak ada yang signifikan (tidak ada yang keluar batas)	Tidak ada yang signifikan (tidak ada yang keluar batas)

Contoh

```
> #ACF & PACF Teoritis Untuk AR(1)
> ACF.AR1 = ARMAacf(ar=0.8, ma=0, 20)
> PACF.AR1 = ARMAacf(ar=0.8, ma=0, 20, pacf=TRUE)
> ACF.AR1 = ACF.AR1[2:21]
> c1 = ACF.AR1
> c2 = PACF.AR1
> AR1 = cbind(c1,c2)
> AR1 # Nilai Teoritis untuk ACF & PACF
> par(mfrow=c(1,2))
> plot(ACF.AR1, type="h", xlab="lag", ylim=c(-1,1))
> abline(h=0)
> plot(PACF.AR1, type="h", xlab="lag", ylim=c(-1,1))
> abline(h=0)
```

	c1	c2
1	0.80000000	8.000000e-01
2	0.64000000	-3.083953e-16
3	0.51200000	6.853229e-17
4	0.40960000	3.083953e-17
5	0.32768000	6.167906e-17
6	0.26214400	-6.167906e-17
7	0.20971520	1.541976e-17
8	0.16777216	-4.625929e-17
9	0.13421773	3.083953e-17
10	0.10737418	7.709882e-18
11	0.08589935	-2.312965e-17
12	0.06871948	1.541976e-17
13	0.05497558	3.854941e-18
14	0.04398047	7.709882e-18
15	0.03518437	-7.709882e-18
16	0.02814750	1.927471e-18
17	0.02251800	-5.782412e-18
18	0.01801440	3.854941e-18
19	0.01441152	-3.854941e-18
20	0.01152922	9.637353e-19

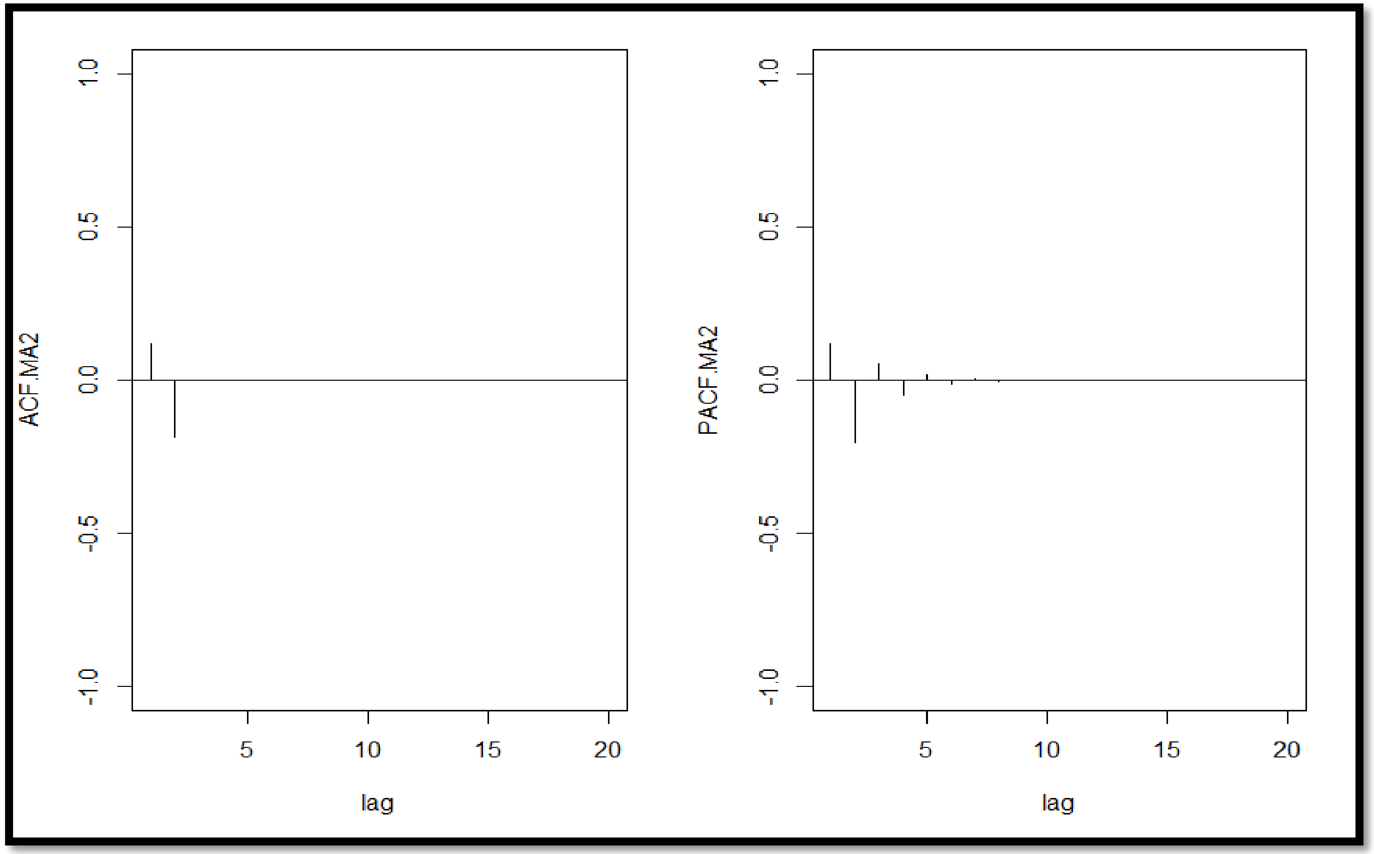


Dari gambar diatas dapat dijelaskan bahwa plot ACF pada model AR(1) dengan koefisien parameter positif adalah *dies down* (turun cepat secara eksponensial) dengan nilai ACF yang selalu positif. Sedangkan PACF menunjukkan pola yang terputus setelah lag 1 seperti petunjuk pada Tabel 11.1.

Contoh

```
> #ACF & PACF Teoritis Untuk MA(2)
> ACF.MA2 = ARMAacf(ar=0, ma=c(1.5, -0.7), 20)
> PACF.MA2 = ARMAacf(ar=0, ma=c(1.5, -0.7), 20,
pacf=TRUE)
> ACF.MA2 = ACF.MA2[2:21]
> c1 = ACF.MA2
> c2 = PACF.MA2
> MA2 = cbind(c1,c2)
> MA2
par(mfrow=c(1,2))
> plot(ACF.MA2, type="h", xlab="lag", ylim=c(-1,1))
> abline(h=0)
> plot(PACF.MA2, type="h", xlab="lag", ylim=c(-1,1))
> abline(h=0)
```

	c1	c2
1	0.1203209	1.203209e-01
2	-0.1871658	-2.046050e-01
3	0.0000000	5.480013e-02
4	0.0000000	-4.926042e-02
5	0.0000000	1.876542e-02
6	0.0000000	-1.282006e-02
7	0.0000000	5.793512e-03
8	0.0000000	-3.483878e-03
9	0.0000000	1.713071e-03
10	0.0000000	-9.689988e-04
11	0.0000000	4.967519e-04
12	0.0000000	-2.727639e-04
13	0.0000000	1.427302e-04
14	0.0000000	-7.724470e-05
15	0.0000000	4.082966e-05
16	0.0000000	-2.194069e-05
17	0.0000000	1.165482e-05
18	0.0000000	-6.241274e-06
19	0.0000000	3.323392e-06
20	0.0000000	-1.776688e-06

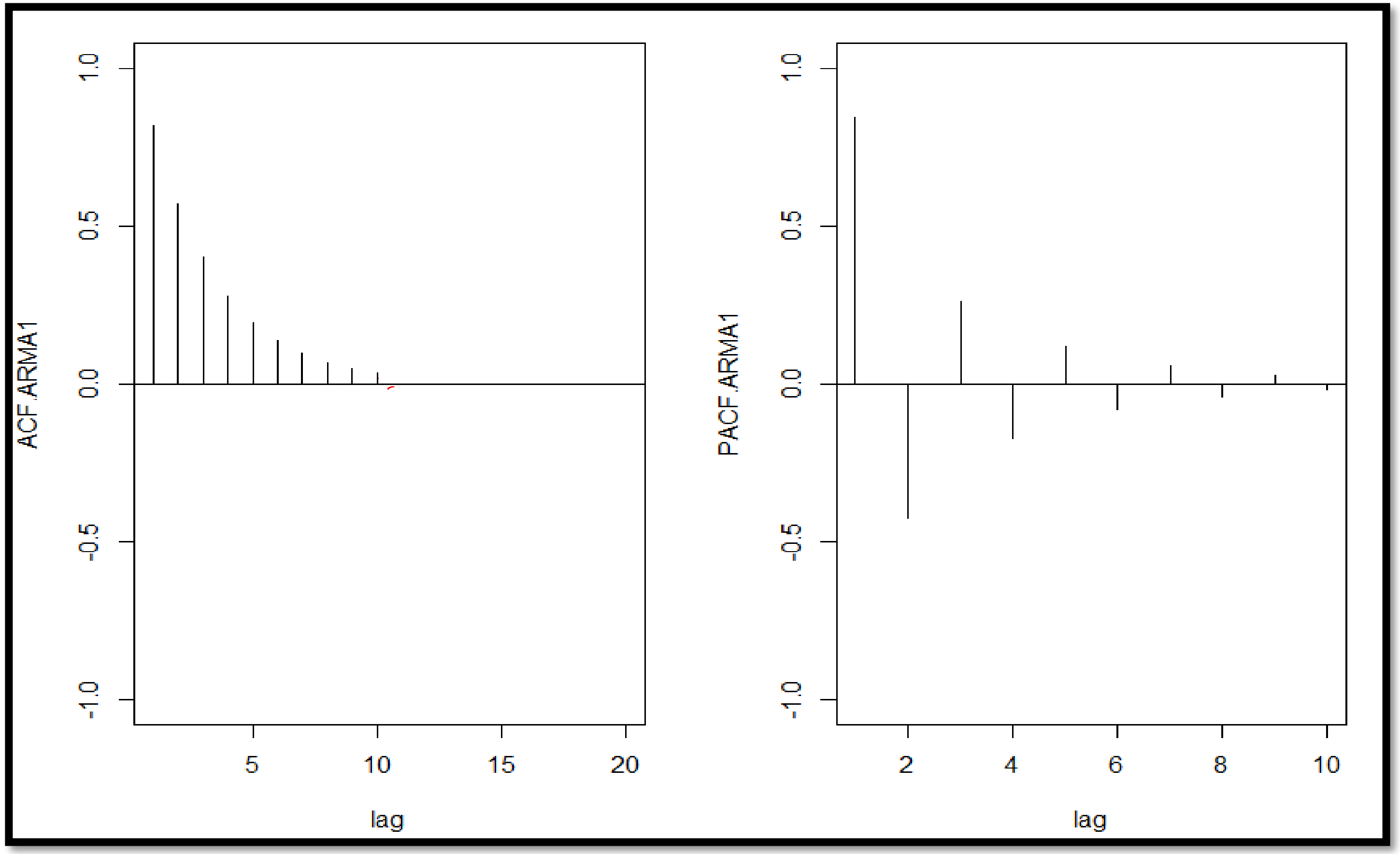


Berdasarkan pola pada gambar diatas dapat dijelaskan bahwa plot ACF pada model MA(2) dengan koefisien parameter positif 1,5 (*tetha1*) dan -0,7 (*tetha2*) adalah terputus setelah lag 2. Sedangkan PACF menunjukkan pola yang *dies down* (turun cepat secara sinusoidal) dengan nilai PACF yang berubah dari positif ke negatif seperti petunjuk pada Tabel 11.1 diatas.

Contoh

```
> #ACF & PACF Teoritis Untuk ARMA(1,1)
> ACF.ARMA1 = ARMAacf(ar=0.7, ma=0.4, 10)
> PACF.ARMA1 = ARMAacf(ar=0.7, ma=0.7, 10, pacf=TRUE)
> ACF.ARMA1 = ACF.ARMA1[2:21]
> c1 = ACF.ARMA1
> c2 = PACF.ARMA1
> ARMA1 = cbind(c1,c2)
> ARMA1
> par(mfrow=c(1,2))
> plot(ACF.ARMA1, type="h", xlab="lag", ylim=c(-1,1))
> abline(h=0)
> plot(PACF.ARMA1, type="h", xlab="lag", ylim=c(-1,1))
> abline(h=0)
```

```
              c1          c2
1     0.81860465  0.84453441
2     0.57302326 -0.42566465
3     0.40111628  0.26202026
4     0.28078140 -0.17317748
5     0.19654698  0.11799726
6     0.13758288 -0.08153457
7     0.09630802  0.05671637
8     0.06741561 -0.03957986
9     0.04719093  0.02766439
10    0.03303365 -0.01935086
<NA>          NA  0.84453441
<NA>          NA -0.42566465
<NA>          NA  0.26202026
<NA>          NA -0.17317748
<NA>          NA  0.11799726
<NA>          NA -0.08153457
<NA>          NA  0.05671637
<NA>          NA -0.03957986
<NA>          NA  0.02766439
<NA>          NA -0.01935086
```

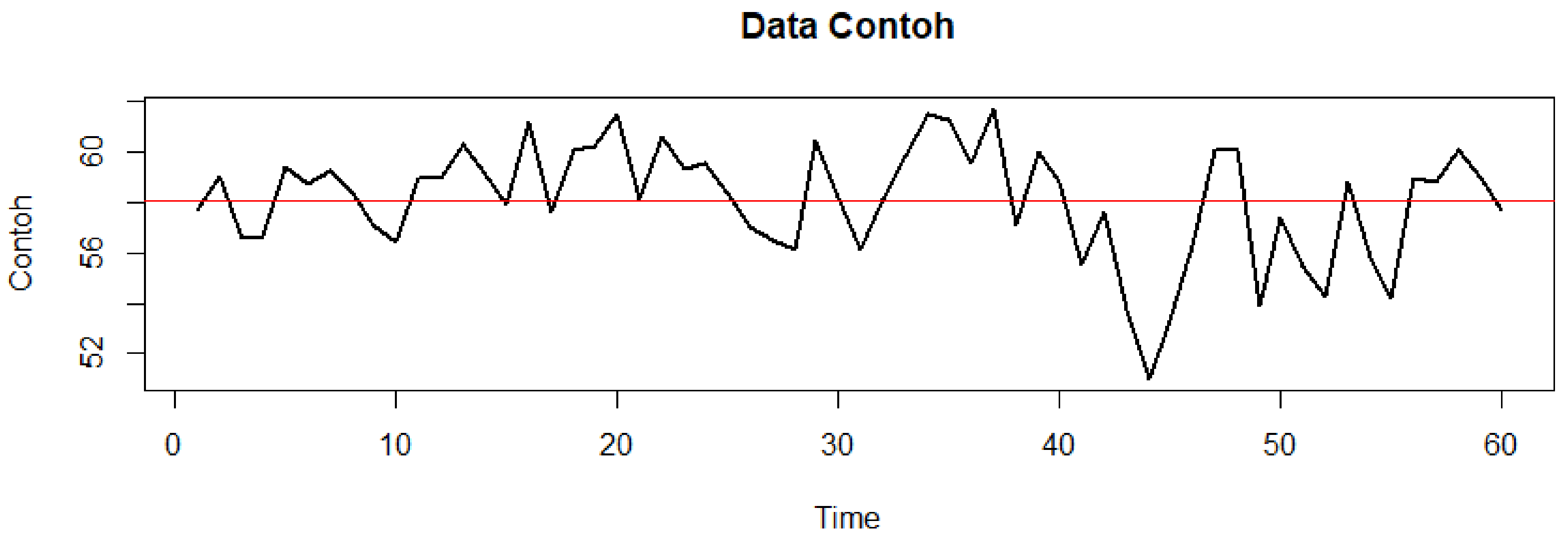


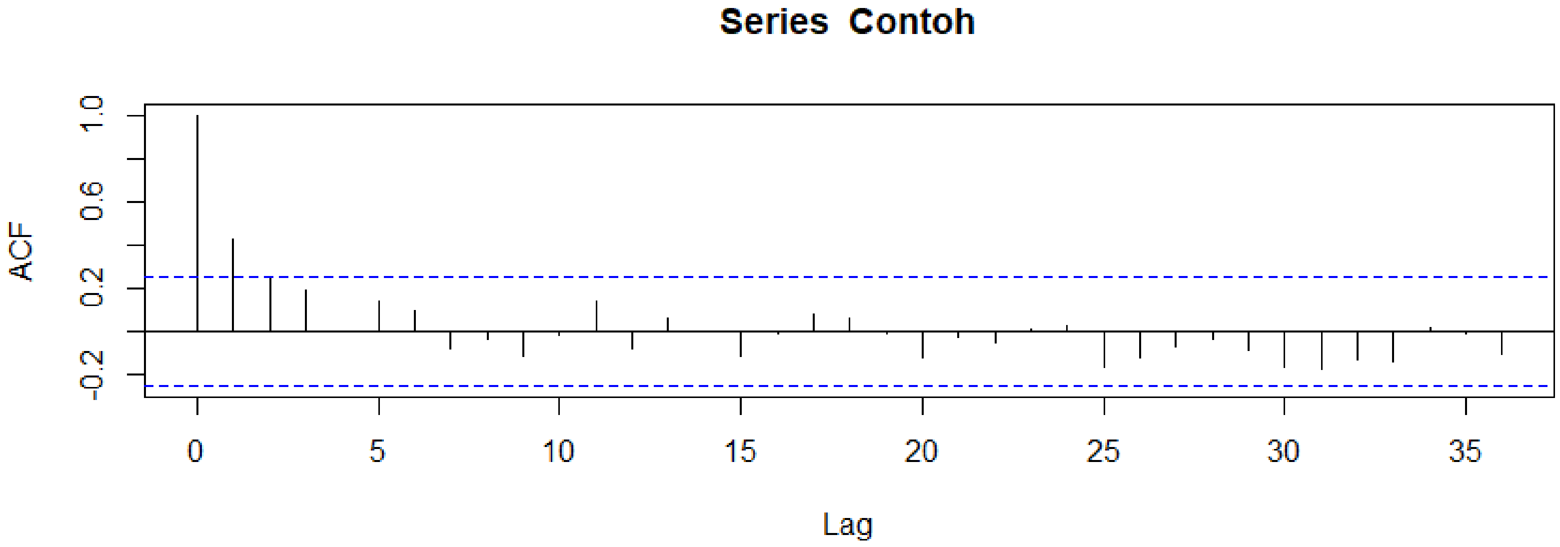
Gambar diatas menunjukkan bahwa plot ACF pada model ARMA(1,1) dengan koefisien parameter p *phi* 0,7 dan *tetha* 0,4 adalah *dies down* (turun cepat secara eksponensial). Pola yang sama juga ditunjukkan oleh plot PACF yaitu *dies down* (turun cepat secara sinusoidal) dengan nilai PACF yang berubah dari positif ke negatif seperti petunjuk pada Tabel 11.1 diatas.

Contoh Model ARIMA Stasioner

Diberikan data Contoh ARIMA (dapat di download di link berikut https://docs.google.com/spreadsheets/d/1Kj8RN4VMAAyAIYPK-YZeIqqN4a336e4a/edit#gid=1791075545)

```
> library(readxl)
> Contoh <- read_excel("FILE DARI DOCUMENT/MATERI AJAR UTY 2019/SEMESTER 1
20212022/Statistika Inferensia Lanjut Praktikum/file latihan/Data Contoh ARIMA.xlsx")
> View(Contoh)
>
> #Mengubah Data Menjadi Data Time Series
>
> Contoh <- ts(Contoh$Zt)
>
> #Menggambarkan Grafik Data Contoh
>
> par(mfrow=c(2,1))
> plot(Contoh, lwd=2, main="Data Contoh")
> abline(h=mean(Contoh), col='red')
> acf(Contoh, lag.max=36)
```


Data Contoh
Contoh
52
56
60
0
10
20
30
40
50
60
Time


Series Contoh
ACF
-0.2
0.2
0.6
1.0
0
5
10
15
20
25
30
35
Lag

Uji Stasioneritas Dengan Augmented Dickey-Fuller (ADF Test)

```
> #Menguji Stasioneritas Data Contoh
> library(tseries)
Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo

    'tseries' version: 0.10-49

    'tseries' is a package for time series analysis and computational finance.

    See 'library(help="tseries")' for details.

Warning message:
package 'tseries' was built under R version 4.0.5
> adf.test(Contoh)

        Augmented Dickey-Fuller Test

data:  Contoh
Dickey-Fuller = -3.7023, Lag order = 3, p-value = 0.03218
alternative hypothesis: stationary
```

Install package : tseries, lmtest, forecast

1. **Hipotesis** :

H_0 : Data tidak stasioner vs H_1 : Data Stasioner

2. **Tingkat Signifikansi :** $\boldsymbol{\alpha = 0,05}$

3. **Kriteria penolakan** $\boldsymbol{H_0}$

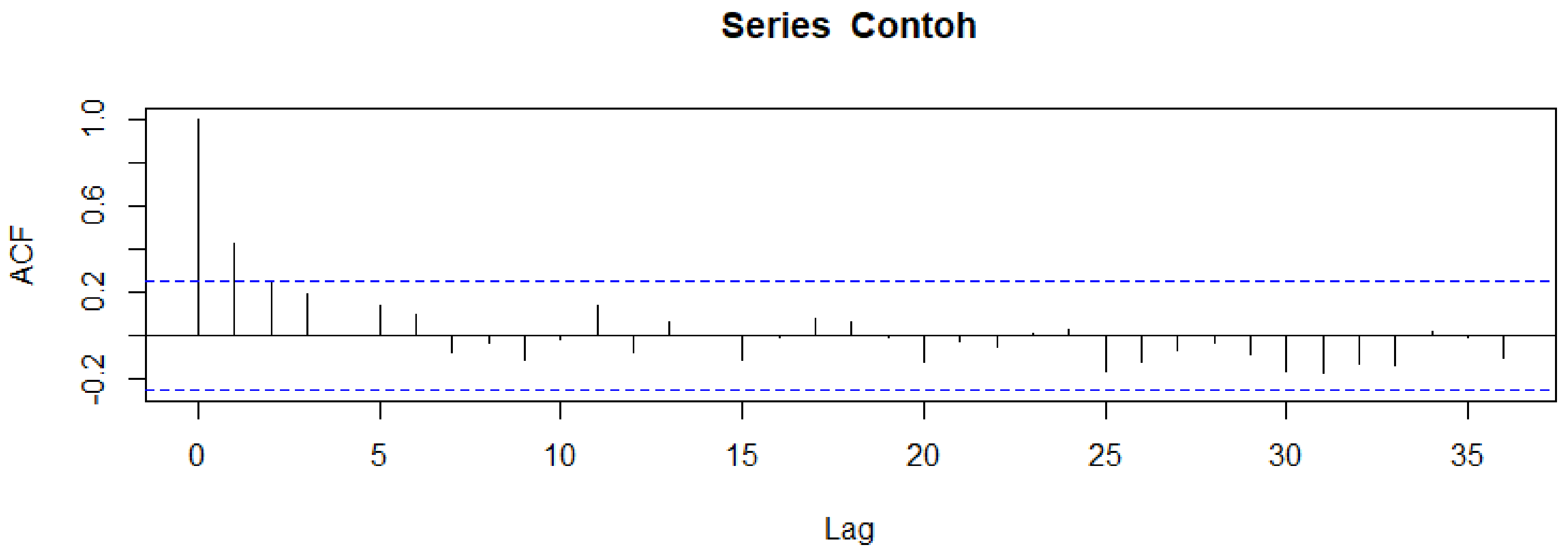
H_0 ditolak jika $p-value < \alpha$. Karena $p-value = 0{,}03218 < \alpha = 0{,}05$ maka H_0 ditolak.

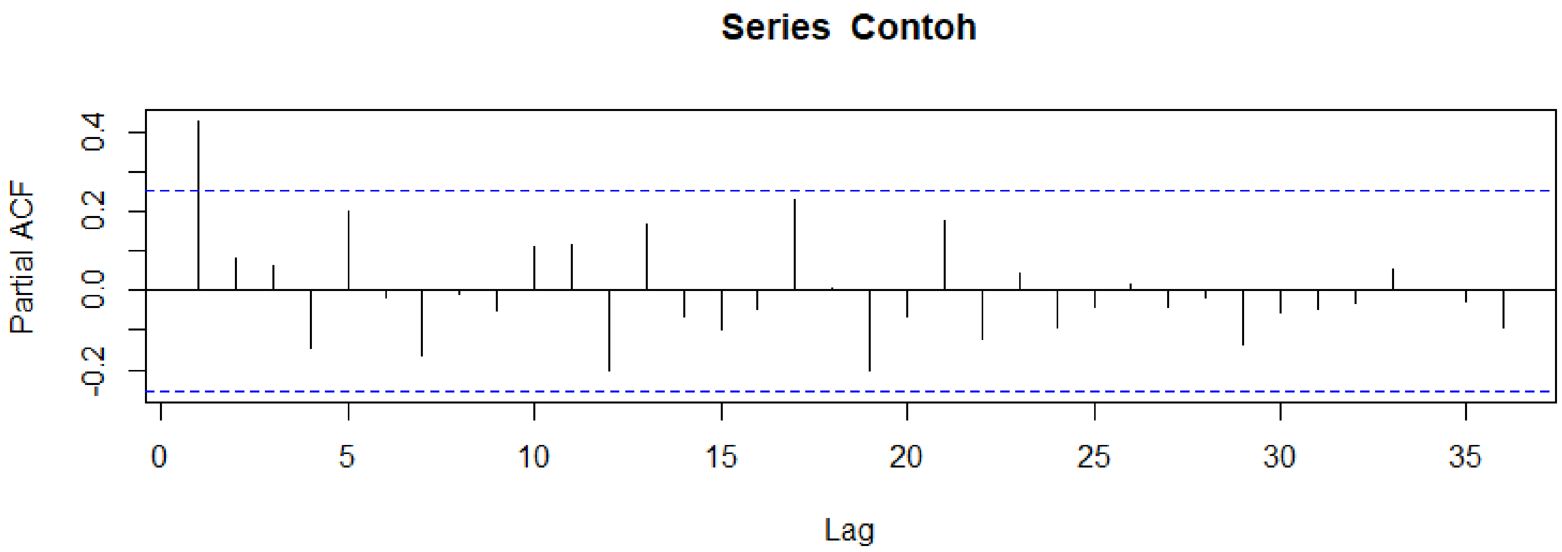
4. **Kesimpulan :**

Jadi, Data Stasioner

Spesifikasi Model

```
> #Spesifikasi Model
>
> par(mfrow=c(2,1))
> acf(Contoh, lag.max=36)
> abline(h=0)
> pacf(Contoh, lag.max=36, pacf=TRUE)
```





Pada Grafik ACF terlihat bahwa data tersebut meluruh pada saat lag 1 dan grafik PACF menunjukkan bahwa ada kecenderungan data meluruh atau naik turun secara sinusoidal pada saat lag 1.

Dengan demikian model yang dapat digunakan adalah

1. AR(1)
2. MA(1)
3. ARMA(1,1)

library(lmtest)

```
> AR1 <- arima(Contoh, order=c(1,0,0))
> MA1 <- arima(Contoh, order=c(0,0,1))
> ARMA1 <- arima(Contoh, order=c(1,0,1))
> #menguji signifikansi model
>
> coeftest(AR1)

z test of coefficients:

           Estimate Std. Error z value  Pr(>|z|)
ar1         0.41965    0.11545   3.635  0.000278 ***
intercept  58.08327    0.44446 130.682 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> coeftest(MA1)

z test of coefficients:

           Estimate Std. Error  z value  Pr(>|z|)
ma1         0.37971    0.11900   3.1909  0.001418 **
intercept  58.08456    0.36503 159.1222 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> coeftest(ARMA1)

z test of coefficients:

          Estimate Std. Error  z value Pr(>|z|)
ar1        0.58753    0.23428   2.5078  0.01215 *
ma1       -0.20737    0.28623  -0.7245  0.46877
intercept 58.08663    0.49025 118.4842  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Dari hasil uji koefisien tersebut, terlihat bahwa nilai $P - value$ untuk koefisien model AR(1) dan MA(1) menunjukkan hasil yang kurang dari α sehingga model AR(1) dan MA(1) signifikan. Sedangkan model ARMA(1,1) koefisien yang signifikan hanyalah AR(1). Sehingga ada 2 model yang mungkin dapat digunakan yaitu AR(1) dan MA(1)

```
> library(forecast)
This is forecast 8.15
  Want to meet other forecasters? Join the International Institute of Forecasters:
  http://forecasters.org/
Warning message:
package ‘forecast’ was built under R version 4.0.5
> AR1 <- arima(Contoh, order=c(1,0,0))
> AR1

Call:
arima(x = Contoh, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.4197    58.0833
s.e.  0.1154     0.4445

sigma^2 estimated as 4.088:  log likelihood = -127.47,  aic = 260.95
> MA1<- arima(Contoh, order=c(0,0,1))
> MA1

Call:
arima(x = Contoh, order = c(0, 0, 1))

Coefficients:
         ma1  intercept
      0.3797    58.0846
s.e.  0.1190     0.3650

sigma^2 estimated as 4.238:  log likelihood = -128.54,  aic = 263.08
> 
```

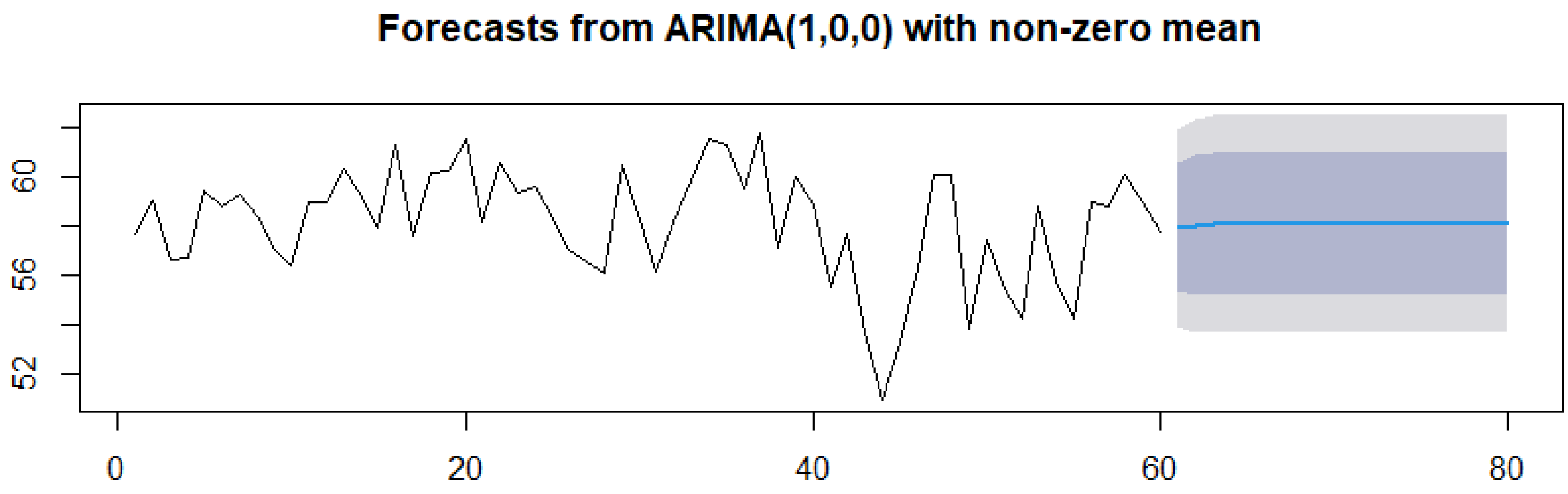
Untuk menentukan model terbaik, perhatikan nilai AIC (Akaike Information’s Criterion) yang terkecil. Karena AIC AR(1) adalah yang terkecil maka model yang digunakan adalah model AR(1).

```
> checkresiduals(AR1)

	Ljung-Box test

data:  Residuals from ARIMA(1,0,0) with non-zero mean
Q* = 8.6857, df = 8, p-value = 0.3695

Model df: 2.   Total lags used: 10

> accuracy(AR1)
                     ME     RMSE      MAE        MPE     MAPE     MASE        ACF1
Training set 0.00361766 2.021856 1.646999 -0.1192122 2.866773 0.828836 -0.02782611
```

Dari cek residu, terlihat bahwa Residu dari model AR(1) tidak nol dan memiliki akurasi dengan MAPE sebesar 2,87% artinya model yang kita bentuk cukup baik.

```
> Prediksi = forecast(ts(Contoh), model= AR1, h=20)
> Prediksi
   Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
61       57.92744 55.33633 60.51856 53.96468 61.89021
62       58.01788 55.20786 60.82790 53.72032 62.31543
63       58.05583 55.20900 60.90266 53.70198 62.40968
64       58.07175 55.21849 60.92502 53.70806 62.43544
65       58.07844 55.22404 60.93283 53.71302 62.44386
66       58.08124 55.22665 60.93583 53.71552 62.44697
67       58.08242 55.22779 60.93705 53.71664 62.44820
68       58.08291 55.22828 60.93755 53.71712 62.44870
69       58.08312 55.22848 60.93776 53.71733 62.44891
70       58.08321 55.22857 60.93784 53.71742 62.44900
71       58.08324 55.22861 60.93788 53.71745 62.44903
72       58.08326 55.22862 60.93789 53.71747 62.44905
73       58.08326 55.22863 60.93790 53.71747 62.44905
74       58.08327 55.22863 60.93790 53.71748 62.44906
75       58.08327 55.22863 60.93790 53.71748 62.44906
76       58.08327 55.22863 60.93790 53.71748 62.44906
77       58.08327 55.22863 60.93791 53.71748 62.44906
78       58.08327 55.22863 60.93791 53.71748 62.44906
79       58.08327 55.22863 60.93791 53.71748 62.44906
80       58.08327 55.22863 60.93791 53.71748 62.44906
> 
```

Hasil Peramalan (forecasting) dengan model AR(1)

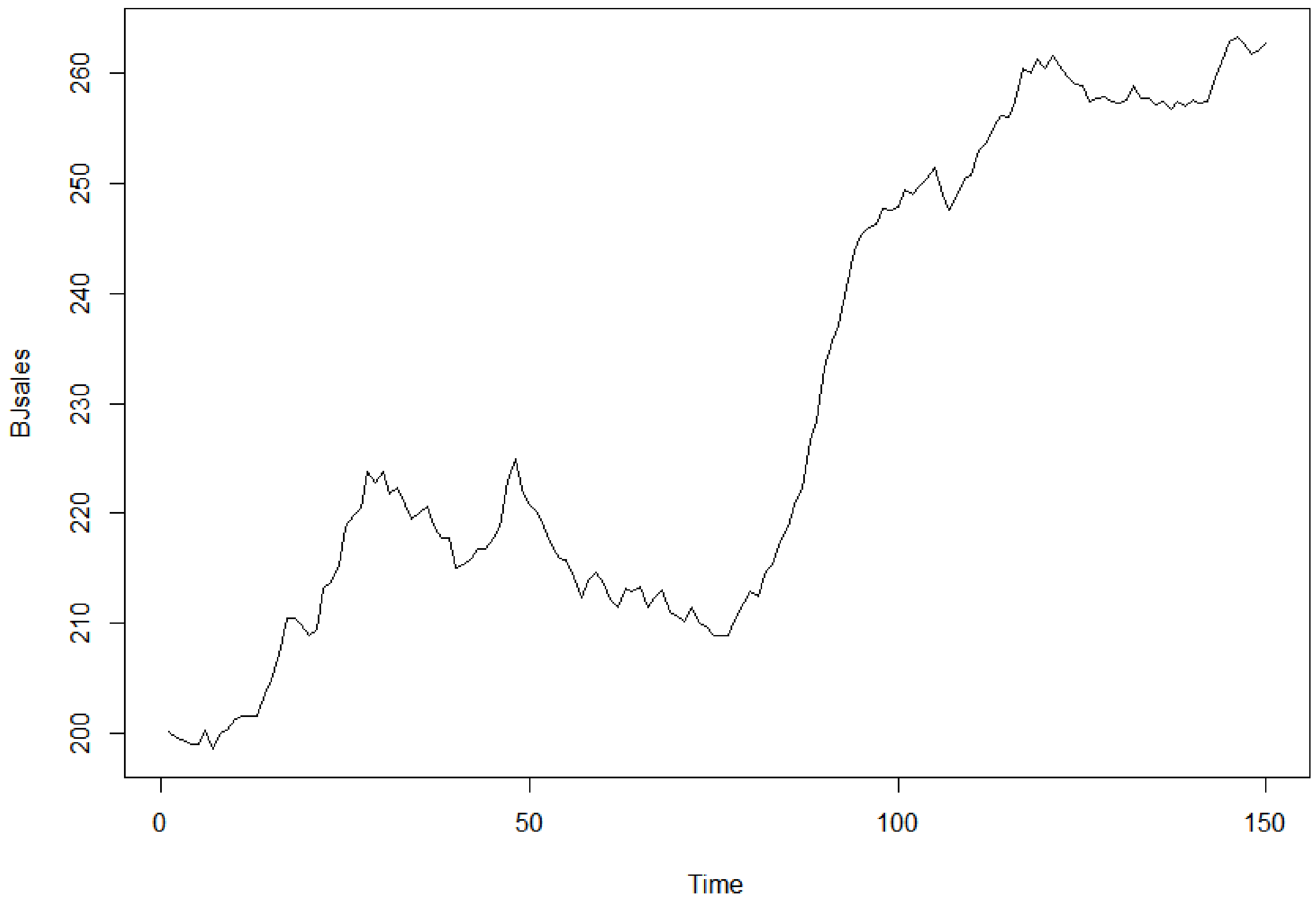

Forecasts from ARIMA(1,0,0) with non-zero mean
52
56
60
0
20
40
60
80

Contoh Model ARIMA Non Stasioner

Menggunakan data Bjsales yang sudah tersedia di Rstudio.

```
> BJsales
Time Series:
Start = 1
End = 150
Frequency = 1
  [1] 200.1 199.5 199.4 198.9 199.0 200.2 198.6 200.0 200.3 201.2 201.6 201.5 201.5 203.5 204.9
 [16] 207.1 210.5 210.5 209.8 208.8 209.5 213.2 213.7 215.1 218.7 219.8 220.5 223.8 222.8 223.8
 [31] 221.7 222.3 220.8 219.4 220.1 220.6>
>
> #plotting data BJsales
>
> plot(BJsales)
>
```

Hasil Plotting Data



Uji Stasioneritas Dengan Adf Test

```
> #Uji Stasioneritas Data
> library(tseries)
Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo

    ‘tseries’ version: 0.10-49

    ‘tseries’ is a package for time series analysis and computational finance.

    See ‘library(help="tseries")’ for details.

Warning message:
package ‘tseries’ was built under R version 4.0.5
> adf.test(BJsales)

        Augmented Dickey-Fuller Test

data:  BJsales
Dickey-Fuller = -2.1109, Lag order = 5, p-value = 0.5302
alternative hypothesis: stationary
```

1. **Hipotesis** :

H_0 : Data tidak stasioner vs H_1 : Data Stasioner

2. **Tingkat Signifikansi :** $\boldsymbol{\alpha = 0,05}$

3. **Kriteria penolakan** $\boldsymbol{H_0}$

H_0 ditolak jika $p - value < \alpha$. Karena $p - value = 0{,}5302 > \alpha = 0{,}05$ maka H_0 tidak ditolak.

4. **Kesimpulan :**

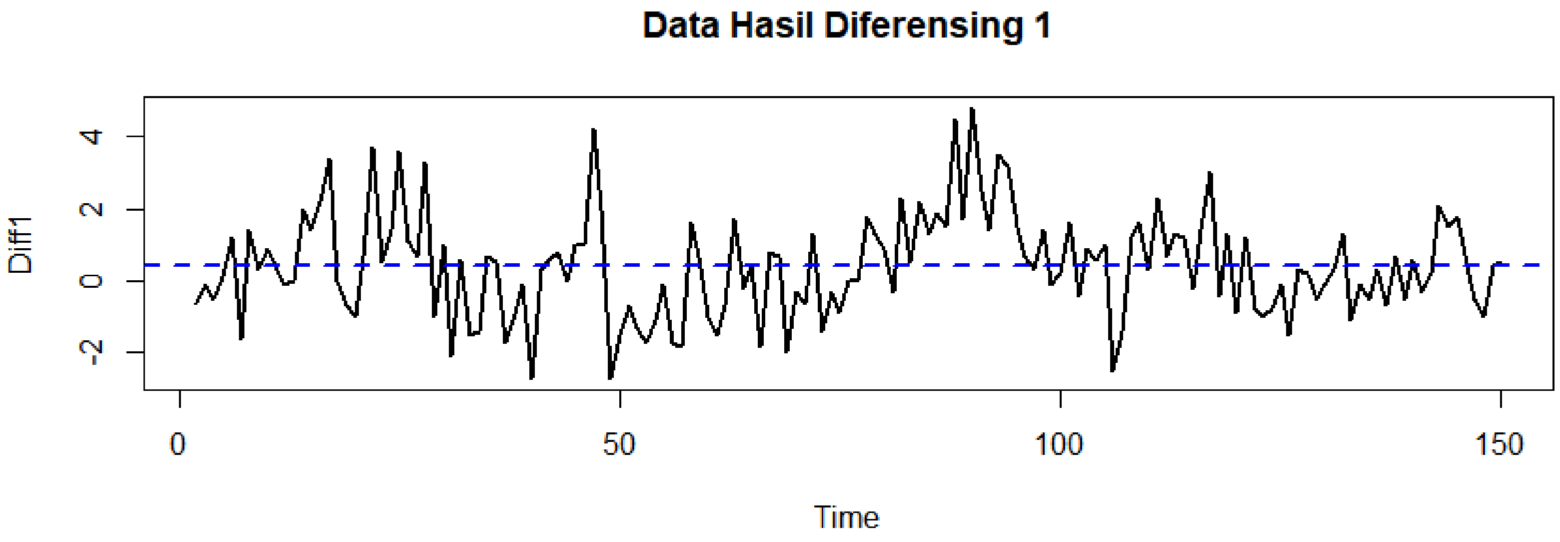
Jadi, Data Tidak Stasioner

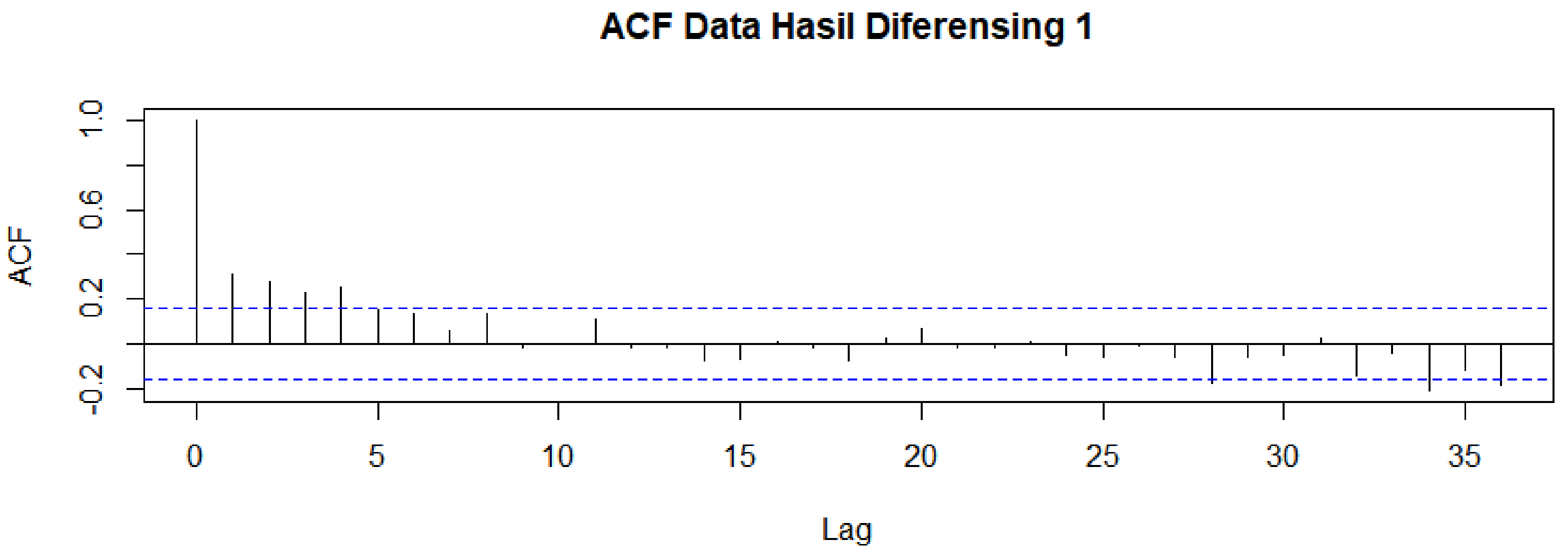
Proses Differencing Karena Data Tidak Stasioner

```
> #Proses Differencing diff(BJsales)
>
> Diff1 = diff(BJsales)
> Diff1
Time Series:
Start = 2
End = 150
Frequency = 1
  [1] -0.6 -0.1 -0.5  0.1  1.2 -1.6  1.4  0.3  0.9  0.4 -0.1  0.0  2.0  1.4  2.2  3.4  0.0 -0.7 -1.0
 [20]  0.7  3.7  0.5  1.4  3.6  1.1  0.7  3.3 -1.0  1.0 -2.1  0.6 -1.5 -1.4  0.7  0.5 -1.7 -1.1 -0.1
 [39] -2.7  0.3  0.6  0.8  0.0  1.0  1.0  4.2  2.0 -2.7 -1.5 -0.7 -1.3 -1.7 -1.1 -0.1 -1.7 -1.8  1.6
 [58]  0.7 -1.0 -1.5 -0.7  1.7 -0.2  0.4 -1.8  0.8  0.7 -2.0 -0.3 -0.6  1.3 -1.4 -0.3 -0.9  0.0  0.0
 [77]  1.8  1.3  0.9 -0.3  2.3  0.5  2.2  1.3  1.9  1.5  4.5  1.7  4.8  2.5  1.4  3.5  3.2  1.5  0.7
 [96]  0.3  1.4 -0.1  0.2  1.6 -0.4  0.9  0.6  1.0 -2.5 -1.4  1.2  1.6  0.3  2.3  0.7  1.3  1.2 -0.2
[115]  1.4  3.0 -0.4  1.3 -0.9  1.2 -0.8 -1.0 -0.8 -0.1 -1.5  0.3  0.2 -0.5 -0.1  0.3  1.3 -1.1 -0.1
[134] -0.5  0.3 -0.7  0.7 -0.5  0.6 -0.3  0.2  2.1  1.5  1.8  0.4 -0.5 -1.0  0.4  0.5
>
> #Plotting Data Hasil Diferencing
>
> par(mfrow=c(2,1))
> plot(Diff1, lwd=2, main="Data Hasil Diferensing 1")
> abline(h=mean(Diff1), lwd=2, lty=2, col="blue")
```

```
> acf(Diff1, main="ACF Data Hasil Diferensing 1", lag.max=36)
>
```

Hasil Plotting Data Hasil Differencing 1





Uji Stasioneritas Data Diff1

```
 #Uji Stasioneritas data Diff1
>
> adf.test(Diff1)

        Augmented Dickey-Fuller Test

data:  Diff1
Dickey-Fuller = -3.3485, Lag order = 5, p-value = 0.06585
alternative hypothesis: stationary
```

1. **Hipotesis** :

H_0 : Data tidak stasioner vs H_1 : Data Stasioner

2. **Tingkat Signifikansi :** $\boldsymbol{\alpha = 0,05}$

3. **Kriteria penolakan** $\boldsymbol{H_0}$

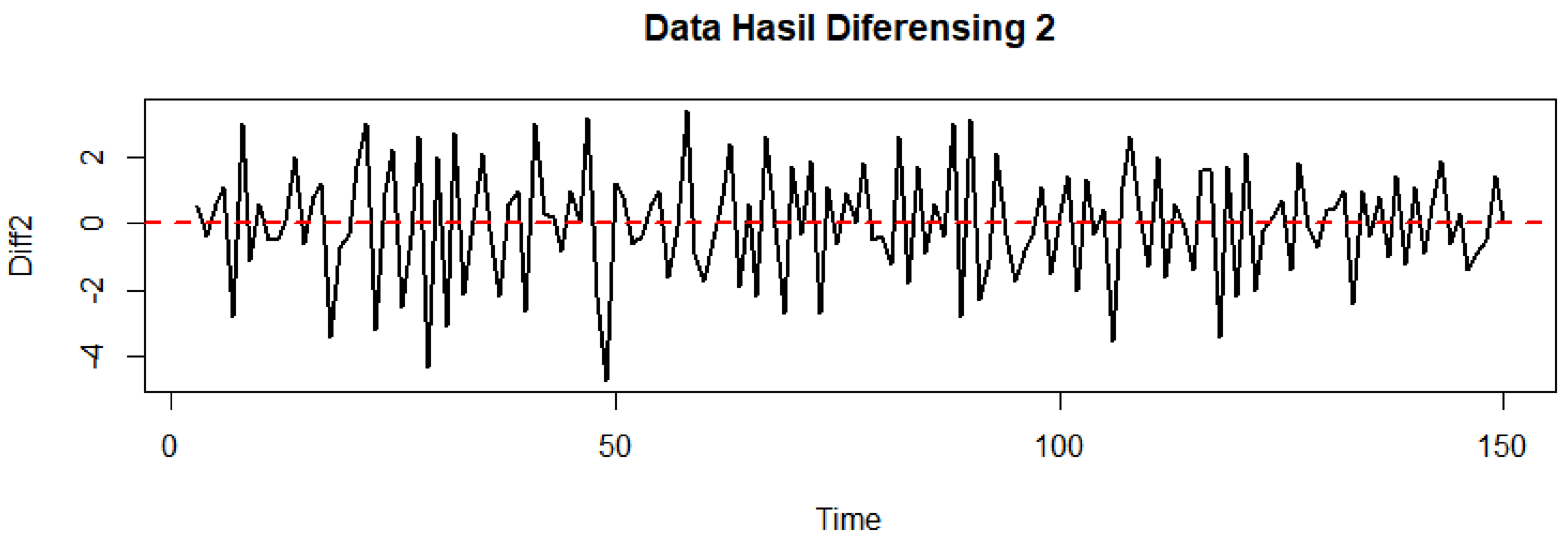
H_0 ditolak jika $p - value < \alpha$. Karena $p - value = 0{,}06585 > \alpha = 0{,}05$ maka H_0 tidak ditolak.

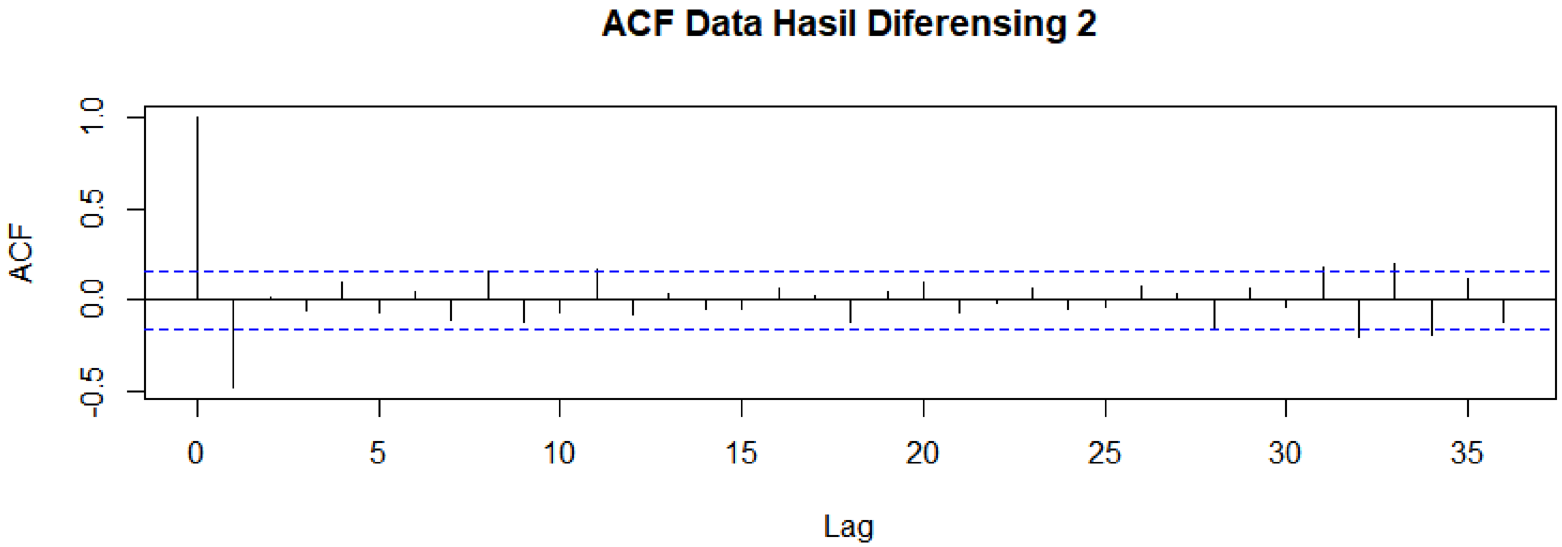
4. **Kesimpulan :**

Jadi, Data Tidak Stasioner

Proses Differencing 2 Karena Data Diff 1 Tidak Stasioner

```
#Proses Differencing 2 Karena Data Diff1 masih belum stasioner
>
> Diff2=diff(Diff1)
> Diff2
Time Series:
Start = 3
End = 150
Frequency = 1
  [1]  0.5 -0.4  0.6  1.1 -2.8  3.0 -1.1  0.6 -0.5 -0.5  0.1  2.0 -0.6  0.8  1.2 -3.4 -
0.7 -0.3  1.7
 [20]  3.0 -3.2  0.9  2.2 -2.5 -0.4  2.6 -4.3  0.7 -1.4  1.8 -0.1 -0.7  0.4  0.4  1.0 -
2.4  1.0 -0.4
[134]  0.8 -1.0  1.4 -1.2  1.1 -0.9  0.5  1.9 -0.6  0.3 -1.4 -0.9 -0.5  1.4  0.1
>
> #Plotting Data Diff2
>
> par(mfrow=c(2,1))
> plot(Diff2, lwd=2, main="Data Hasil Diferensing 2")
> abline(h=mean(Diff2), lwd=2, lty=2, col="red")
> acf(Diff2, main="ACF Data Hasil Diferensing 2", lag.max=36)
```


Data Hasil Diferensing 2
Diff2
-4
-2
0
2
0
50
100
150
Time


ACF Data Hasil Diferensing 2
ACF
-0.5
0.0
0.5
1.0
0
5
10
15
20
25
30
35
Lag

Uji Stasioneritas Data Diff 2

```
 #Uji Stasioneritas data Diff2
>
> adf.test(Diff2)

        Augmented Dickey-Fuller Test

data:  Diff2
Dickey-Fuller = -6.562, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(Diff2) : p-value smaller than printed p-value
```

1. **Hipotesis** :

H_0 : Data tidak stasioner vs H_1 : Data Stasioner

2. **Tingkat Signifikansi :** $\boldsymbol{\alpha = 0,05}$

3. **Kriteria penolakan** $\boldsymbol{H_0}$

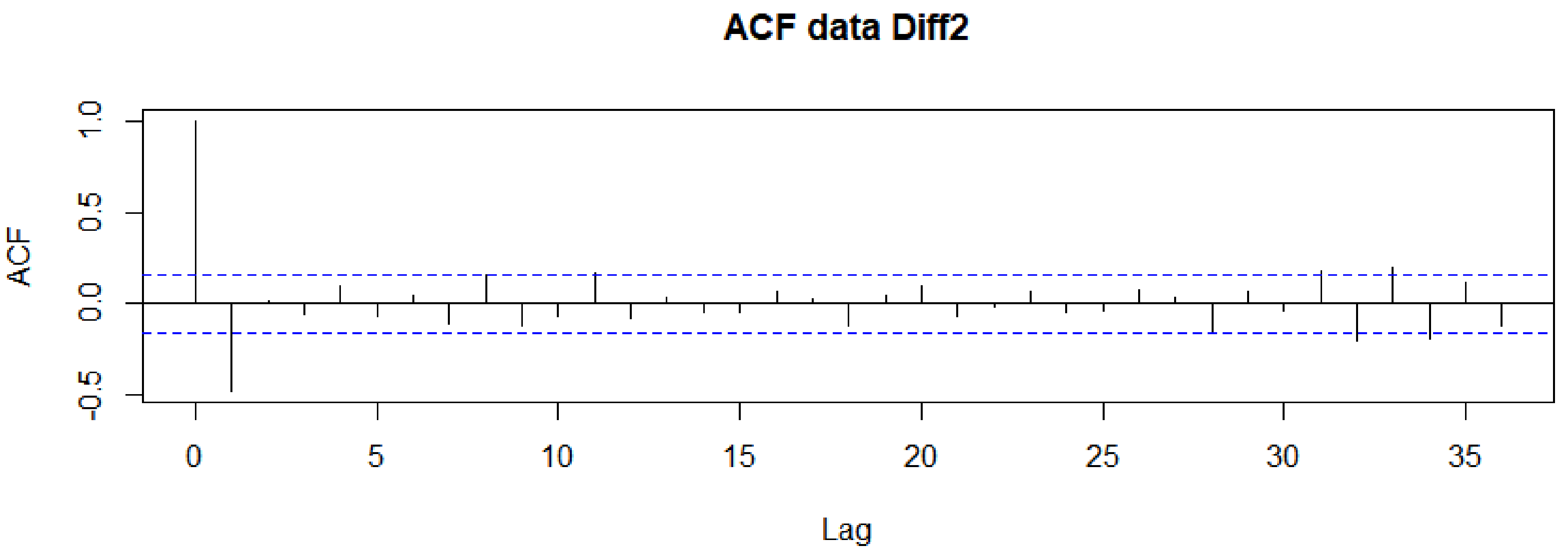
H_0 ditolak jika $p - value < \alpha$. Karena $p - value = 0{,}01 < \alpha = 0{,}05$ maka H_0 ditolak.

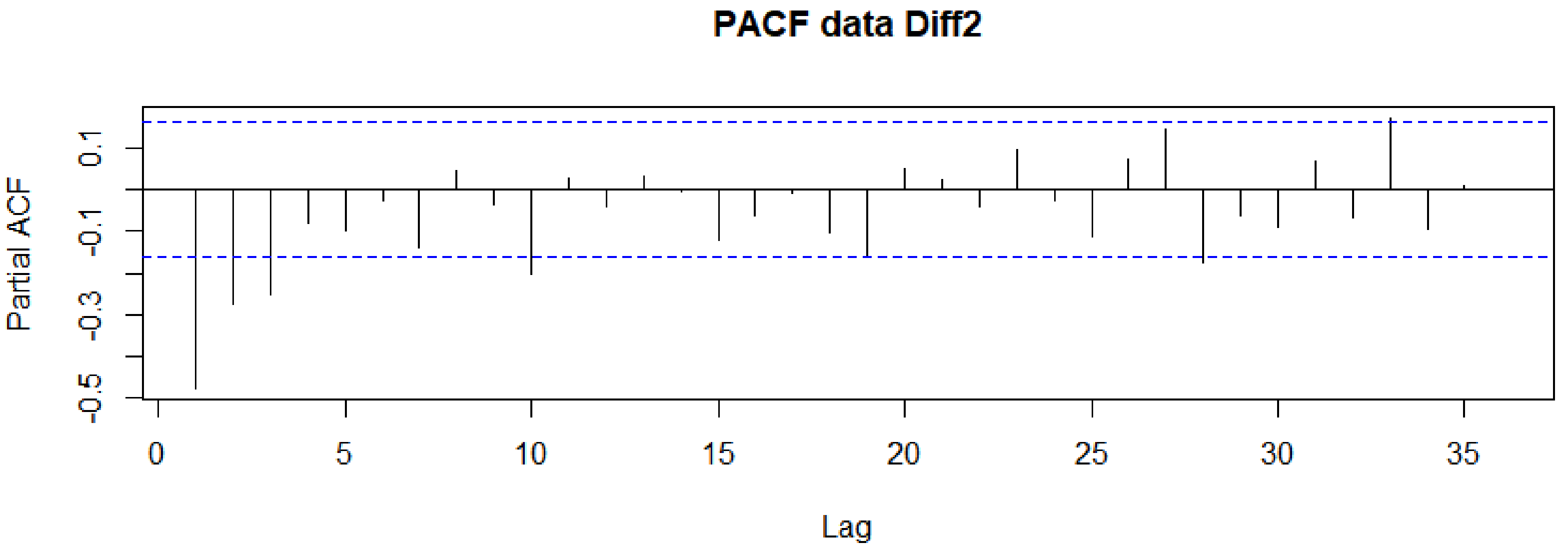
4. **Kesimpulan :**

Jadi, Data Stasioner

Penentuan Model

```
 #Penentuan Model
>
> par(mfrow=c(2,1))
> acf(Diff2, main="ACF data Diff2", lag.max=36)
> pacf(Diff2, main="PACF data Diff2", lag.max=36, pacf=TRUE)
```





Pada Grafik ACF terlihat bahwa data tersebut meluruh secara sinusoidal pada saat lag 1 dan grafik PACF menunjukkan bahwa ada kecenderungan data meluruh atau naik turun secara sinusoidal pada saat lag 1.

Dengan demikian model yang dapat digunakan adalah

1. ARIMA(1,2,1)
2. ARIMA(0,2,1)
3. ARIMA(1,2,0)

Menguji Signifikansi Koefisien Dari Model

```
> #Menguji SIgnifikansi Koefisien
>
> coeftest(ARIMA121)
    Estimate Std. Error z value  Pr(>|z|)
ar1  0.05277    0.13126  0.4020    0.6877
ma1 -0.78009    0.10037 -7.7721 7.718e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> coeftest(ARIMA120)

z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ar1 -0.472647   0.071981 -6.5663 5.159e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> coeftest(ARIMA021)

z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ma1 -0.747976   0.066166 -11.305 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Dari hasil uji koefisien tersebut, terlihat bahwa nilai $P - value$ untuk koefisien model ARIMA(1,2,0) dan ARIMA(0,2,1) menunjukkan hasil yang kurang dari α sehingga model ARIMA(1,2,0) dan ARIMA(0,2,1) signifikan. Sedangkan model ARIMA(1,2,1) koefisien yang signifikan hanyalah MA(1). Sehingga ada 2 model yang mungkin dapat digunakan yaitu ARIMA(1,2,0) dan ARIMA(0,2,1)

Menentukan Model Terbaik

```
 #Menentukan Model Terbaik
>
> library(lmtest)
>
> ARIMA121 = arima(BJsales, order=c(1,2,1))
> ARIMA121
>
> ARIMA021 = arima(BJsales, order=c(0,2,1))
> ARIMA021
>
> ARIMA120 = arima(BJsales, order=c(1,2,0))
> ARIMA120
```

```
> ARIMA121 = arima(BJsales, order=c(1,2,1))
> ARIMA121

Call:
arima(x = BJsales, order = c(1, 2, 1))

Coefficients:
         ar1      ma1
      0.0528  -0.7801
s.e.  0.1313   0.1004

sigma^2 estimated as 1.863:  log likelihood = -256.49,  aic = 518.97
>
> ARIMA021 = arima(BJsales, order=c(0,2,1))
> ARIMA021
  -0.7480
s.e.   0.0662

sigma^2 estimated as 1.866:  log likelihood = -256.57,  aic = 517.14
>
> ARIMA120 = arima(BJsales, order=c(1,2,0))
> ARIMA120

Call:
arima(x = BJsales, order = c(1, 2, 0))

Coefficients:
          ar1
      -0.4726
s.e.   0.0720

sigma^2 estimated as 2.215:  log likelihood = -268.98,  aic = 541.96
>
```

Untuk menentukan model terbaik, perhatikan nilai AIC (Akaike Information’s Criterion) yang terkecil. Karena AIC ARIMA(0,2,1) adalah yang terkecil maka model yang digunakan adalah model ARIMA(0,2,1).

Model Yang Digunakan

Model ARIMA(0,2,1)

$$\Phi_0(B)(1-B)^2 X_t = \theta_0 + (1-\theta_1 B)\varepsilon_t$$

$$1 \cdot (1-2B+B^2)X_t = (1-\theta_1 B)\varepsilon_t$$

$$X_t - 2BX_t + B^2 X_t = \varepsilon_t - \theta_1 B\varepsilon_t$$

$$X_t - 2X_{t-1} + X_{t-2} = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$$X_t = 2X_{t-1} - X_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

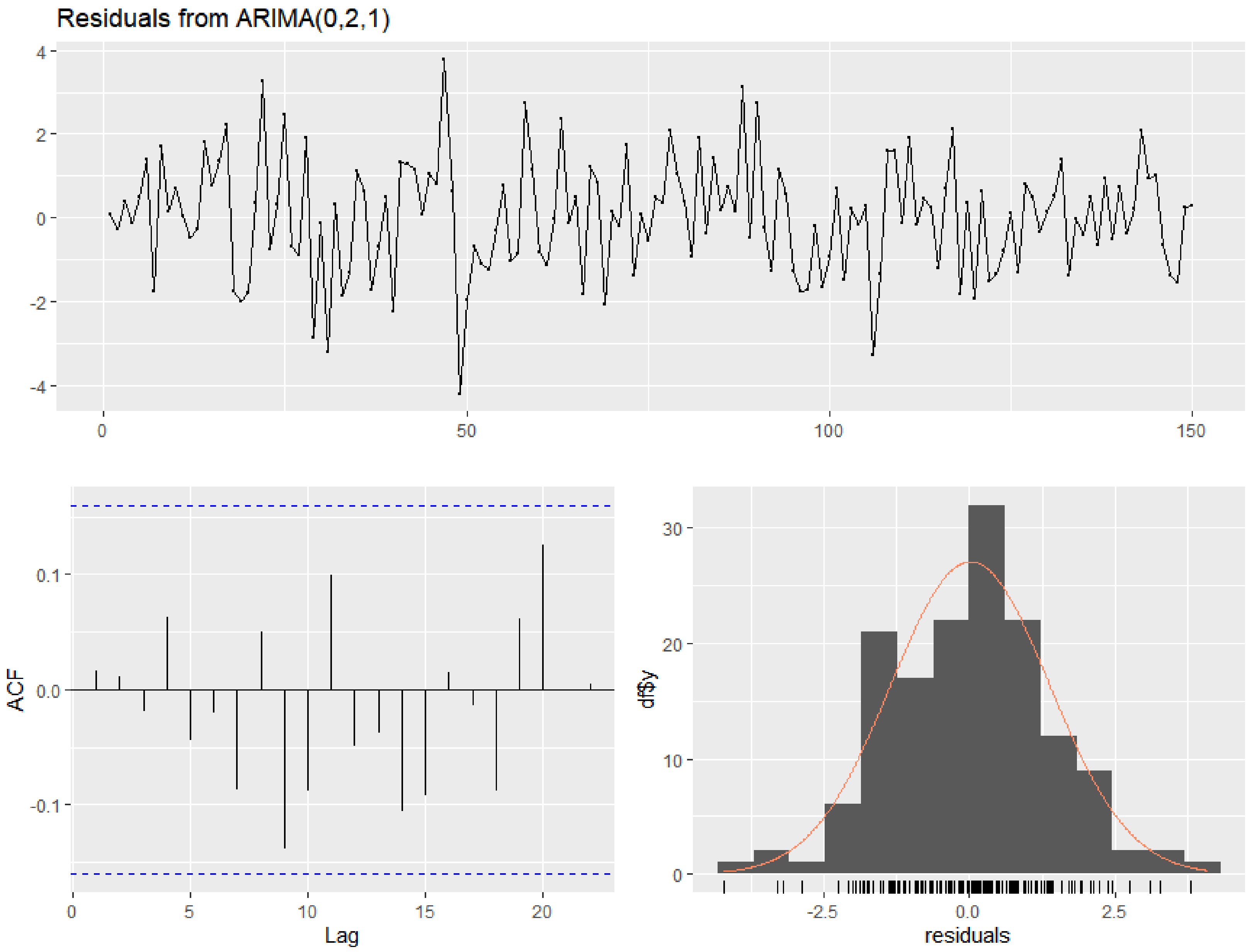
Karena $\theta_1 = -0{,}7480$ maka model ARIMA(0,2,1) menjadi

$$X_t = 2X_{t-1} - X_{t-2} + \varepsilon_t + 0{,}7480\varepsilon_{t-1}$$

Diagnostik & Akurasi Model

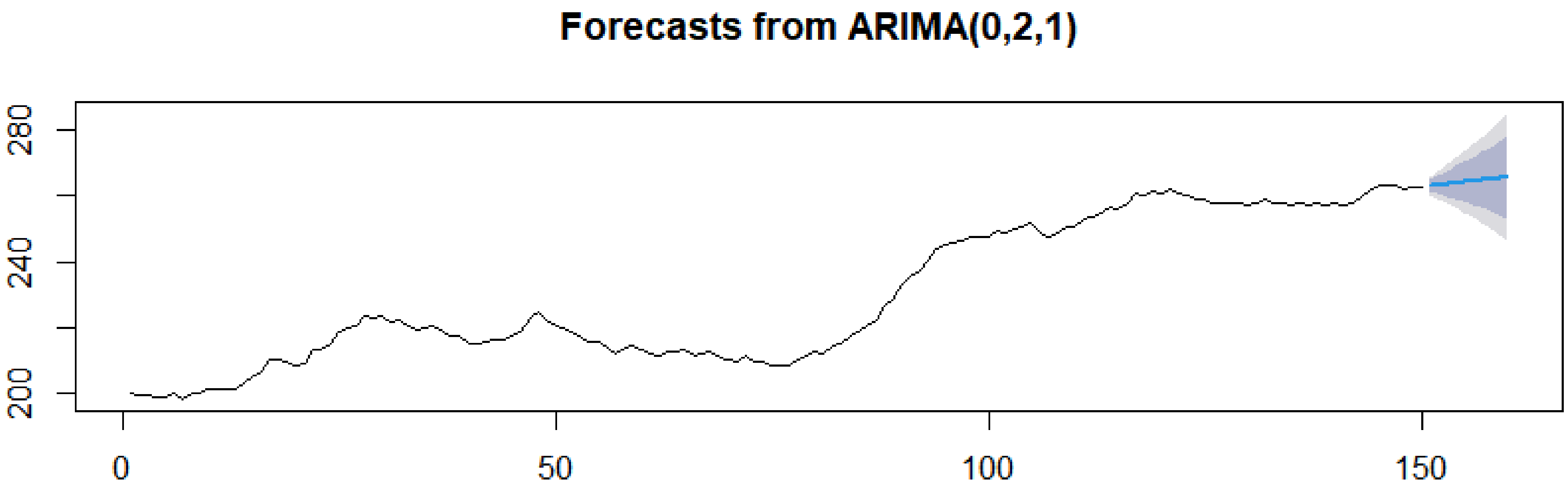
```
> checkresiduals(ARIMA021)

	Ljung-Box test

data:  Residuals from ARIMA(0,2,1)
Q* = 7.0422, df = 9, p-value = 0.6327

Model df: 1.   Total lags used: 10

> accuracy(ARIMA021)

                    ME     RMSE      MAE        MPE      MAPE      MASE       ACF1
Training set 0.0164288 1.357029 1.066835 0.01198676 0.4695181 0.9188348 0.01634351
```

Artinya residu dari model ARIMA(0,2,1) memenuhi syarat independent satu sama lain sehingga memenuhi syarat white noise. Dengan akurasi MAPE 46,95%


Residuals from ARIMA(0,2,1)
ACF
Lag
df$y
residuals

Peramalan

```
> forecast(BJsales, model=ARIMA021, h=10)
    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
151       262.9837 261.2331 264.7343 260.3065 265.6610
152       263.2674 260.4624 266.0724 258.9775 267.5573
153       263.5511 259.7040 267.3982 257.6674 269.4348
154       263.8348 258.9103 268.7593 256.3035 271.3662
155       264.1185 258.0681 270.1690 254.8652 273.3719
156       264.4022 257.1730 271.6314 253.3461 275.4583
157       264.6859 256.2242 273.1476 251.7448 277.6270
158       264.9696 255.2221 274.7172 250.0620 279.8772
159       265.2533 254.1676 276.3391 248.2991 282.2075
160       265.5370 253.0620 278.0120 246.4582 284.6159

> plot(forecast(BJsales, model=ARIMA021, h=10))
```


Forecasts from ARIMA(0,2,1)
280
240
200
0
50
100
150

Latihan

Diberikan data `sunspots` (yang ada di Rstudio) mengenai fenomena Sunspot pada tahun 1755 – 1831.

a. Plot data tersebut dan ujilah apakah data tersebut stasioner atau tidak

b. Tentukan model ARIMA (p,d,q) yang terbaik dan akurasinya

c. Lakukan peramalan untuk 10 tahun berturut-turut (h=10)