

“ANALISIS DATA DERET WAKTU MENGGUNAKAN R “



DISUSUN OLEH :

A ROFIQI MAULANA

125090500111025

DOSEN : SAMINGUN HANDOYO, S.SI, M.CS

PROGRAM STUDI STATISTIKA

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS BRAWIJAYA

2014

PENDAHULUAN

Pemodelan univariat dilakukan dengan mengidentifikasi terlebih dahulu proses yang ada pada data kemudian melakukan estimasi atas parameter yang bersesuaian $(\varphi, \theta, p, d, q)$. Pemodelan univariat bersifat seni daripada ilmiah karena memerlukan judgment dan suatu aktivitas trial error. Model yang terbaik adalah hasil suatu eksplorasi yang ekstensif.

Namun demikian terdapat suatu panduan dalam melakukan estimasi model univariat, yang disebut dengan prosedur Boix Jenkins. Prosedur ini terdiri atas tahapan berikut

1. Identifikasi model

Melihat plot data, mengevaluasi korelogram bagi data terkait. Proses manakah yang sesuai dengan data yang ada. Membandingkan korelogram untuk memilih orde yang terpilih. Model yang terpilih adalah model tentatif dan akan direvisi pada tahap selanjutnya

2. Estimasi model

Setelah identifikasi model, maka dapat melakukan estimasi model. Dilihat dari kriteria OLS dan MLE

3. Evaluasi model

Tahap ini memeriksa apakah model yang diestimasi telah memadai. Terdapat dua metode evaluasi, yaitu overfitting dan diagnose residual. Overfitting dilakukan dengan mengestimasi model yang lebih besar daripada model yang disarankan pada tahap 1. Serta melihat apakah term tambahan tersebut adalah signifikan. Sedangkan diagnose residual dilakukan dengan melihat apakah residual model telah bersifat white noise, dengan demikian tidak ada lagi informasi yang dapat digali dari data.

A. Membuka data di R

Software R menyediakan beberapa data yang langsung dapat dipakai dalam suatu analisis. Hal ini sangat berguna jika kita tidak mempunyai data yang cocok sesuai analisis kita.

1. Menampilkan daftar data yang tersedia di R

```
> #menampilkan daftar data di R
> data()
> |
```



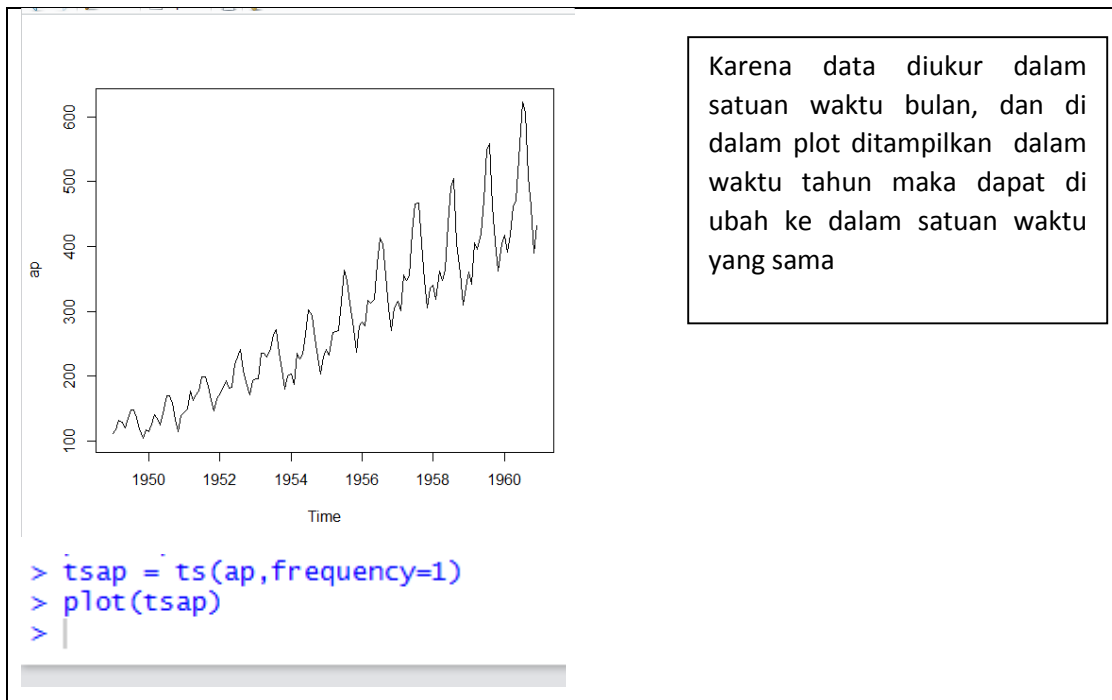
Data sets in package 'datasets':	
AirPassengers	Monthly Airline Passenger Numbers 1949-1960
BJsales	Sales Data with Leading Indicator
BJsales.lead (BJsales)	Sales Data with Leading Indicator
BOD	Biochemical Oxygen Demand
CO2	Carbon Dioxide Uptake in Grass Plants
ChickWeight	Weight versus age of chicks on different diets
DNase	Elisa assay of DNase
EuStockMarkets	Daily Closing Prices of Major European Stock Indices, 1991-1998
Formaldehyde	Determination of Formaldehyde
HairEyeColor	Hair and Eye Color of Statistics Students
Harman23.cor	Harman Example 2.3
Harman74.cor	Harman Example 7.4
Indometh	Pharmacokinetics of Indomethacin
InsectSprays	Effectiveness of Insect Sprays
JohnsonJohnson	Quarterly Earnings per Johnson & Johnson Share
LakeHuron	Level of Lake Huron 1875-1972
LifeCycleSavings	Intercountry Life-Cycle Savings Data
Loblolly	Growth of Loblolly pine trees
Nile	Flow of the River Nile
Orange	Growth of Orange Trees
OrchardSprays	Potency of Orchard Sprays
PlantGrowth	Results from an Experiment on Plant Growth
Puromycin	Reaction Velocity of an Enzymatic Reaction
Seatbelts	Road Casualties in Great Britain 1969-84
Theoph	Pharmacokinetics of Theophylline
Titanic	Survival of passengers on the Titanic
ToothGrowth	The Effect of Vitamin C on Tooth Growth in Guinea Pigs
UCBAdmissions	Student Admissions at UC Berkeley
UKDriverDeaths	Road Casualties in Great Britain 1969-84
UKGas	UK Quarterly Gas Consumption
USAccDeaths	Accidental Deaths in the US 1973-1978

Pilih data yang akan digunakan dengan cara *copy* dan *paste* nama data tersebut ke R Console.

```
> #Menyimpan data ke suatu variabel
> ap = AirPassengers
> |
```

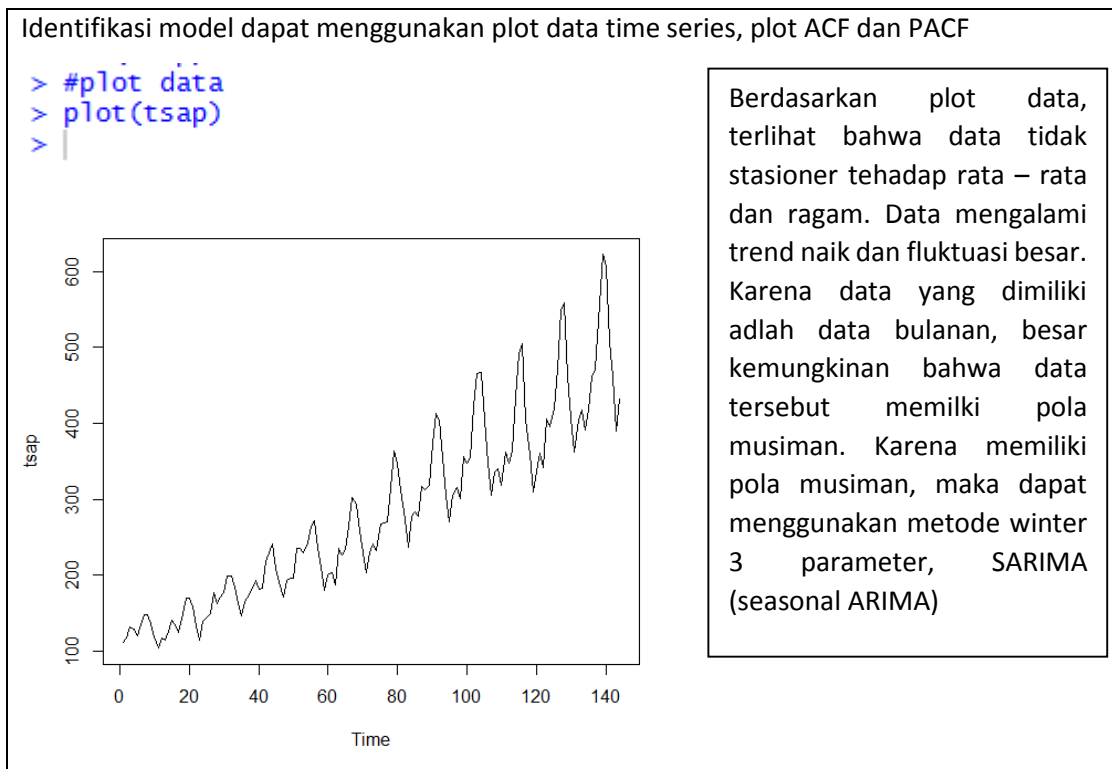
2. Melihat plot data

```
> ap
      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148 136 119 104 118
1950 115 126 141 135 125 149 170 170 158 133 114 140
1951 145 150 178 163 172 178 199 199 184 162 146 166
1952 171 180 193 181 183 218 230 242 209 191 172 194
1953 196 196 236 235 229 243 264 272 237 211 180 201
1954 204 188 235 227 234 264 302 293 259 229 203 229
1955 242 233 267 269 270 315 364 347 312 274 237 278
1956 284 277 317 313 318 374 413 405 355 306 271 306
1957 315 301 356 348 355 422 465 467 404 347 305 336
1958 340 318 362 348 363 435 491 505 404 359 310 337
1959 360 342 406 396 420 472 548 559 463 407 362 405
1960 417 391 419 461 472 535 622 606 508 461 390 432
> plot(ap)
> |
```



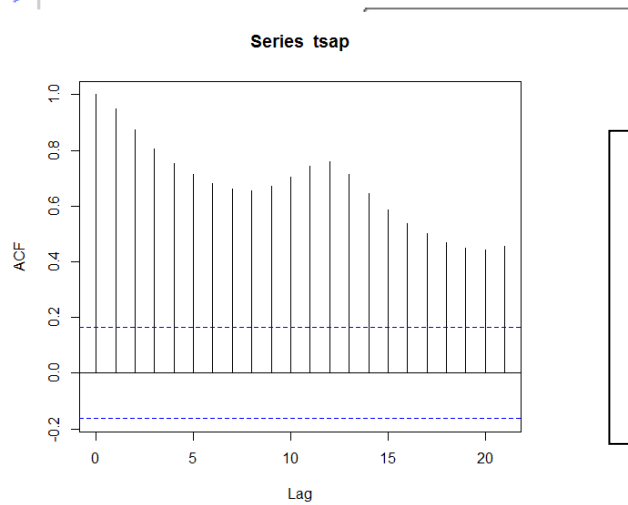
B. Analnsisi deret waktu dengan menggunakan R

1. Identifikasi model

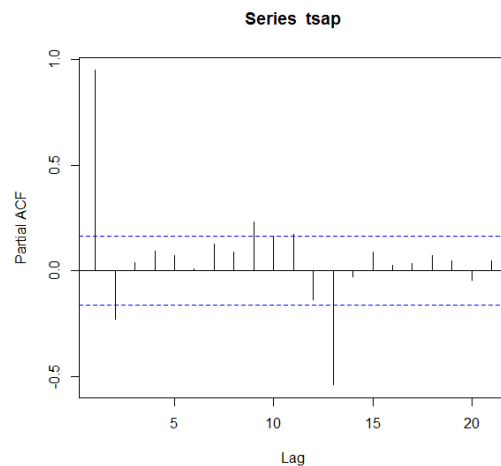


Untuk menentukan orde ARIMA(p,d,q) dapat menggunakan plot ACF dan PACF
 Plot ACF menentukan orde MA sedangkan plot PACF menentukan orde AR

```
> #Melihat plot ACF dan PACF
> acf(tsap)
> pacf(tsap)
> |
```



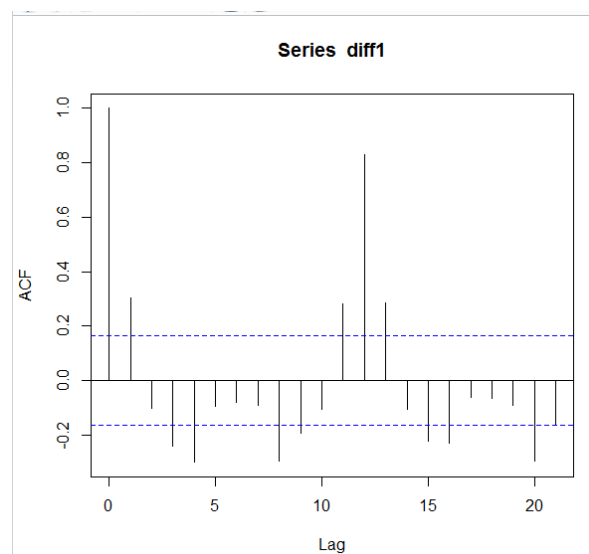
Berdasarkan plot ACF, dapat disimpulkan bahwa data belum stationer karena banyaknya lag yang keluar lebih dari 3 sehingga perlu differensiasi. ACF menurun secara lambat seiring waktu.



Differensi data

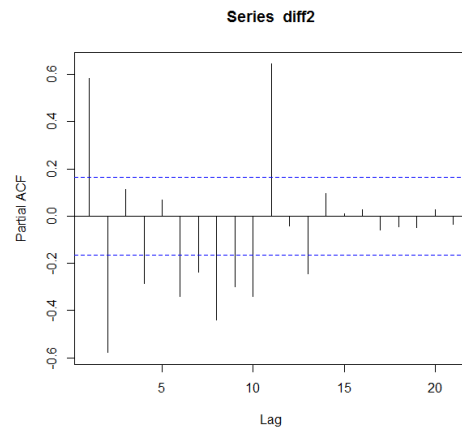
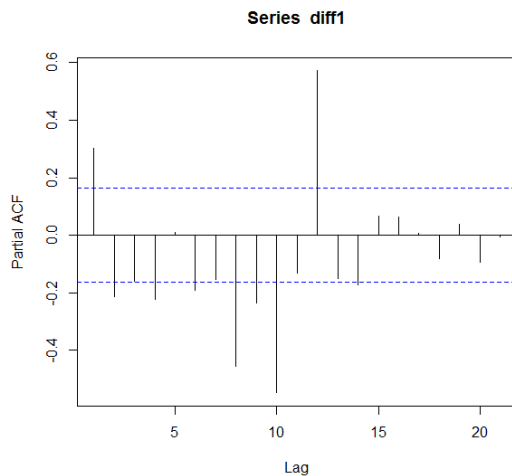
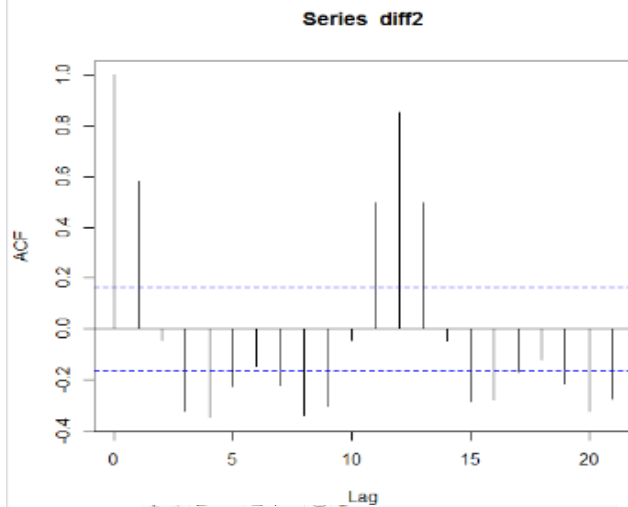
```
> diff1=diff(tsap,1)
> acf(diff1)
> |
```

Berdasarkan plot acf pada d=1
 Diperoleh bahwa data belum stationer



```
> diff2=diff(tsap,2)
> acf(diff2)
> |
```

Berdasarkan plot acf pada $d=2$ Diperoleh bahwa data belum stationer. Karena data masih belum stasioner, maka kemungkinan lainnya adalah data tersebut merupakan data musiman atau nonlinier.



Model ARIMA yang mungkin pada data tersebut adalah ARIMA(2,1,3). ARIMA(2,2,3). Model yang terpilih merupakan model awal yang akan di evaluasi pada tahap selanjutnya.

2. Pendugaan parameter

Pendugaan parameter dari model ARIMA yang terpilih.

```
> 'arima(x = tsap, order = c(2, 2, 3))
```

```
Call:
arima(x = tsap, order = c(2, 2, 3))
```

Coefficients:

	ar1	ar2	ma1	ma2	ma3
	0.3526	0.1293	-1.2031	-0.5841	0.7922
s.e.	0.1801	0.1688	0.1546	0.3033	0.1534

```
sigma^2 estimated as 787.2: log likelihood = -680.42, aic = 1372.84
```

```
> arima(x = tsap, order = c(2, 1, 3))

Call:
arima(x = tsap, order = c(2, 1, 3))

Coefficients:
      ar1      ar2      ma1      ma2      ma3
  1.6105 -0.8604 -1.6482  0.5017  0.2610
s.e.  0.0477  0.0476  0.1115  0.2066  0.1074

sigma^2 estimated as 639.2:  log likelihood = -668.8,  aic = 1349.61
> |
```

3. Evaluasi model

Evaluasi model terdapat dua metode yaitu overfitting dan diagnose residual. Dalam metode overfitting, mengestimasi model yang lebih besar daripada model yang disarankan. Kriteria pemilihan model dipilih berdasarkan nilai MSE, AIC, BIC, dan lainnya. Model yang baik untuk peramalan adalah model yang memiliki nilai AIC terkecil diantara model lainnya.

Pemilihan model dapat dilihat dari nilai AIC terkecil

```
> arima(x = tsap, order = c(2, 2, 0))
```

```
Call:
arima(x = tsap, order = c(2, 2, 0))
```

```
Coefficients:
      ar1      ar2
 -0.2494 -0.2527
s.e.  0.0834  0.0831
```

```
sigma^2 estimated as 1422:  log likelihood = -717.04,  aic = 1440.08
```

```
> arima(x = tsap, order = c(0, 2, 1))
```

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

```
Coefficients:
      ma1
 -1.000
s.e.  0.018
```

```
sigma^2 estimated as 1139:  log likelihood = -703.68,  aic = 1411.37
```

```
> arima(x = tsap, order = c(2, 2, 1))
```

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

```
Coefficients:
      ar1      ar2      ma1
  0.3845 -0.2259 -1.0000
s.e.  0.0828  0.0838  0.0177
```

```
sigma^2 estimated as 980.3:  log likelihood = -692.94,  aic = 1393.88
```

Dilakukan estimasi lainnya untuk model yang lebih besar untuk mendapatkan nilai AIC
Berkut merupakan overfitting model

Parameter	AIC	Parameter	AIC	Parameter	AIC
(1,0,0)	1409.75	(1,2,0)	1447.03	(1,3,0)	1524.05
(2,1,0)	1396.59	(2,2,0)	1440.08	(2,3,0)	1499.48
(3,1,0)	1395.02	(3,2,0)	1437.88	(3,3,0)	1496.8
(0,1,1)	1397.26	(0,2,1)	1411.37	(0,3,1)	1448.34
(1,1,1)	1394.68	(1,2,1)	1398.93	(1,3,1)	1445.18
(2,1,1)	1378.34	(2,2,1)	1393.88	(2,3,1)	1438.73
(3,1,1)	1379.61	(3,2,1)	1392.38	(3,3,1)	1436.88
(0,1,2)	1397.09	(0,2,2)	1394.37	(0,3,2)	1415.31
(1,1,2)	1385.5	(1,2,2)	1391.91	(1,3,2)	1401.83
(2,1,2)	1387.41	(2,2,2)	1392.66	(2,3,2)	1397.26
(3,1,2)	1377.09	(3,2,2)	1364.72	(3,3,2)	1396.08
(0,1,3)	1394.91	(0,2,3)	1394.41	(0,3,3)	1397.35
(1,1,3)	1386.89	(1,2,3)	1392.47	(1,3,3)	1395.26
(2,1,3)	1349.61	(2,2,3)	1372.84	(2,3,3)	1368.91
(3,1,3)	1345.91	(3,2,3)	1364.27	(3,3,3)	1396.44

Berdasarkan tabel di atas, diperoleh nilai AIC terkecil yaitu 1345,91 pada saat $p = 2, d = 1, \text{ dan } q = 3$. Sehingga model yang terpilih adalah ARIMA(3,1,3)

```
> arima(x = tsap, order = c(3, 1, 3))
```

call:

```
arima(x = tsap, order = c(3, 1, 3))
```

Coefficients:

```
      ar1      ar2      ar3      ma1      ma2      ma3
      1.0090  0.1678 -0.6237 -0.9497 -0.6697  0.9017
s.e.   0.0989  0.1667  0.1000  0.1052  0.1082  0.1023
```

```
sigma^2 estimated as 612.6: log likelihood = -665.95, aic = 1345.91
```

```
> |
```

$$Y_t - Y_{t-1} = \varphi_1(Y_{t-1} - Y_{t-2}) + \varphi_2(Y_{t-2} - Y_{t-3}) + \varphi_3(Y_{t-3} - Y_{t-4}) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3}$$

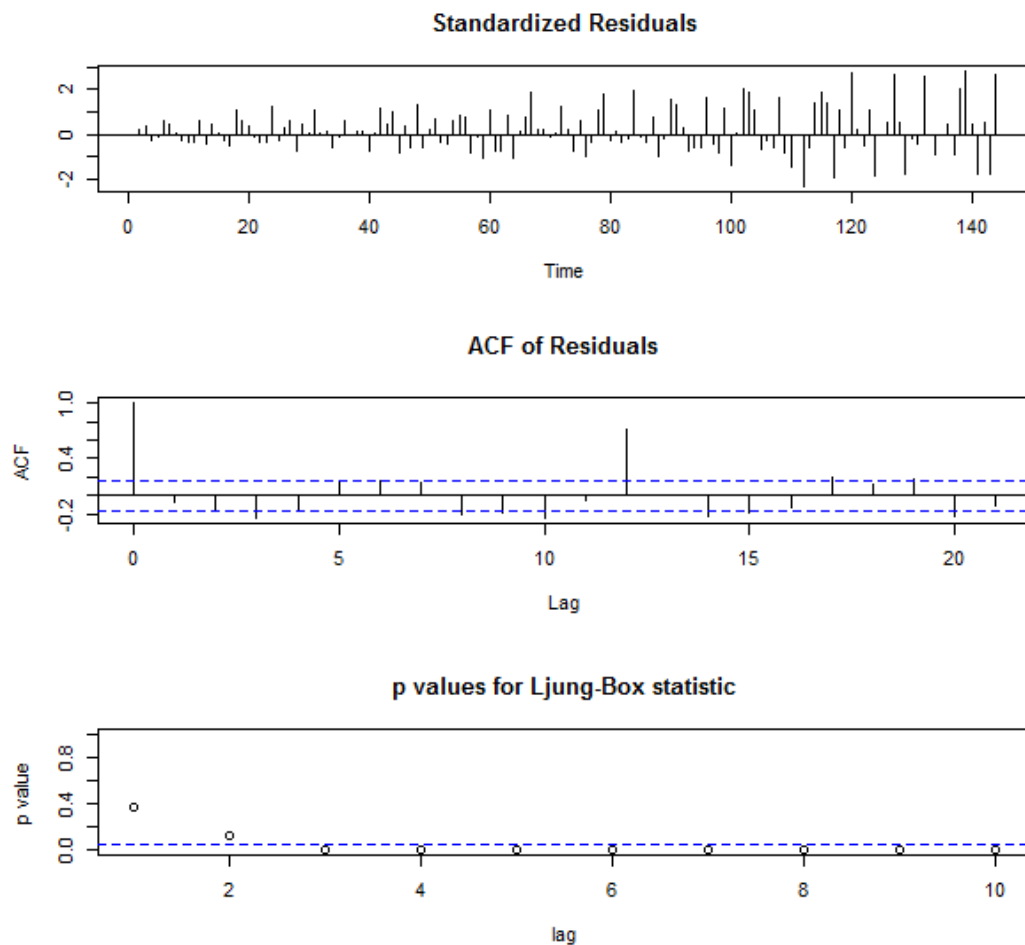
$$Y_t = (1 + \varphi_1)Y_{t-1} + (\varphi_2 - \varphi_1)Y_{t-2} + (\varphi_3 - \varphi_2)Y_{t-3} - \varphi_3 Y_{t-4} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3}$$

$$Y_t = (1 + 1,0090)Y_{t-1} + (0,1678 - 1,0090)Y_{t-2} + (-0,6237 - 0,1678)Y_{t-3} + 0,6237Y_{t-4} + e_t + 0,9497e_{t-1} + 0,6697e_{t-2} - 0,9017e_{t-3}$$

$$Y_t = 2,009Y_{t-1} - 0,841Y_{t-2} - 0,7915Y_{t-3} + 0,6237Y_{t-4} + e_t \\ + 0,9497e_{t-1} + 0,6697e_{t-2} - 0,9017e_{t-3}$$

Uji diagnosa residual

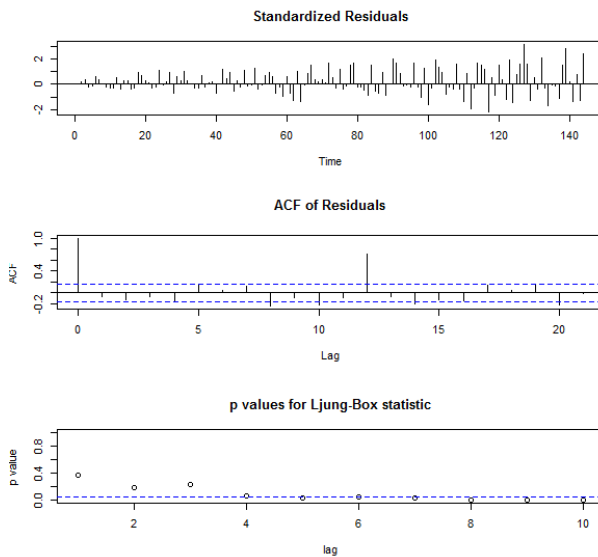
```
> model.ap=arima(x = tsap, order = c(3, 1, 3))
> x=model.ap$residual
> tsdiag(model.ap)
> |
```



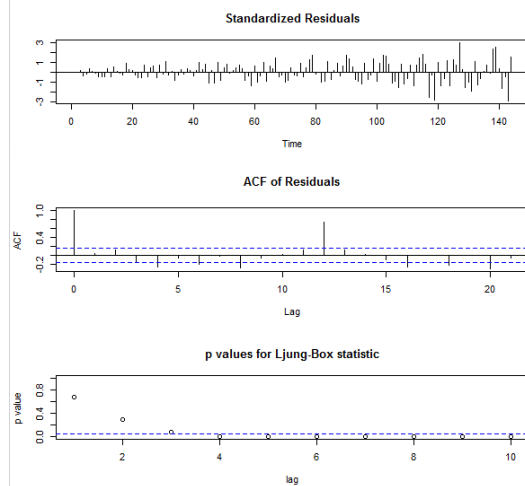
Dari plot ACF Residual, terlihat bahwa terdapat lag yang keluar batas. Residual bersifat white noise jika korelasi untuk setiap lag adalah nol sepanjang waktu. Sehingga dapat disimpulkan residual belum memenuhi sifat white noise sehingga perlu dilakukan metode lainnya misalnya SARIMA atau metode Winter 3 Parameter Holt

Uji diagnose residual ARIMA(2,2,3) dan ARIMA(2,1,3)

```
> model.ap1=arima(x = tsap, order = c(2, 1, 3))
> tsdiag(model.ap1)
> |
```



```
> model.ap2=arima(x = tsap, order = c(2, 2, 3))
Warning messages:
1: In log(s2) : NaNs produced
2: In log(s2) : NaNs produced
3: In log(s2) : NaNs produced
> tsdiag(model.ap2)
```



Dari plot ACF Residual, terlihat bahwa masih terdapat korelasi (plot ACF keluar batas), sehingga model belum memenuhi sifat white noise. Sehingga disarankan menggunakan alternative lain, misalnya menggunakan SARIMA, atau Winter 3 parameter.