Latihan Pertemuan 6 : Pendugaan Parameter - Adinda Shabrina Putri Salsabila

2023-09-30

# Packages

library(ggplot2)  
library(tsibble)  
library(tseries)  
library(MASS)  
library(forecast)  
library(TSA)  
library(TTR)  
library(aTSA)  
library(graphics)

# Data

Digunakan data harga cabai rawit di wilayah Jakarta yang dalam hal ini hanya digunakan data 140 periode akhir

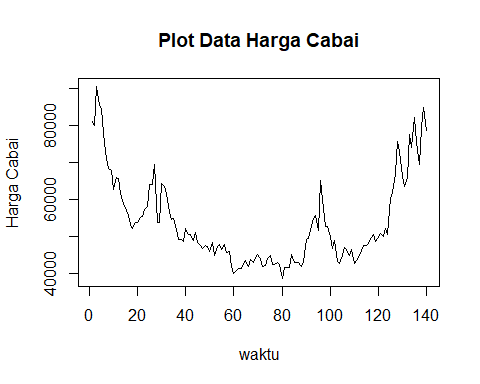
data <- read.csv("C:/Users/hp/Documents/datampdw6.csv")  
datacabai.ts <- ts(data)  
head(datacabai.ts)

## Time Series:  
## Start = 1   
## End = 6   
## Frequency = 1   
## harga  
## [1,] 81090  
## [2,] 80000  
## [3,] 90670  
## [4,] 85730  
## [5,] 84340  
## [6,] 75040

# Eksplorasi Data

## Plot Data Penuh

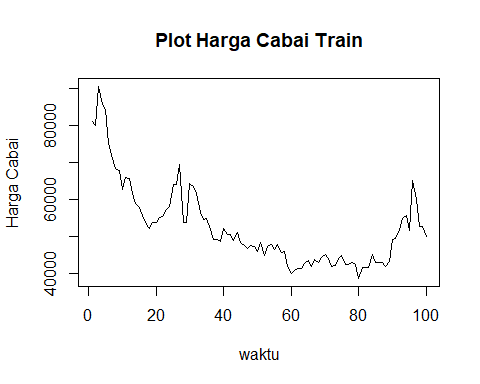
plot.ts(datacabai.ts, lty=1, xlab="waktu", ylab="Harga Cabai", main="Plot Data Harga Cabai")



Berdasarkan plot data deret waktu, terlihat adanya tren jangka panjang yang fluktuatif. Data dimulai dengan harga tinggi, kemudian menurun secara umum hingga sekitar titik data ke-80, lalu mulai menunjukkan tren kenaikan kembali hingga akhir periode.Juga adanya indikasi pola siklis, ditunjukkan oleh fluktuasi yang berulang dalam jangka waktu yang lebih panjang. Terlihat beberapa siklus naik-turun sepanjang periode waktu yang ditampilkan. Berdasarkan pola data, pembagian data latih dan data uji ditetapkan dengan proporsi 71%:29%.

## Plot Data Latih

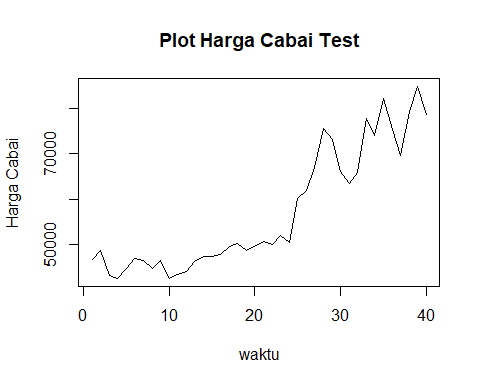
cabaitrain<-datacabai.ts[1:100]  
train.ts<-ts(cabaitrain)  
plot.ts(train.ts, lty=1, xlab="waktu", ylab="Harga Cabai", main="Plot Harga Cabai Train")



Berdasarkan plot data deret waktu pada data latih, terlihat bahwa data cenderung memiliki trend yang naik-turun dan cenderung tidak bergerak pada nilai tengah tertentu. Hal ini mengindikasikan bahwa data tidak stasioner dalam rataan.

## Plot Data Uji

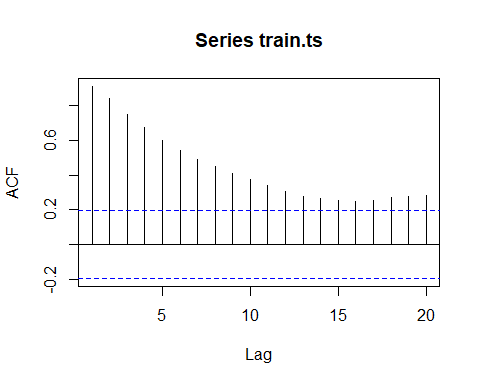
cabaitest<-datacabai.ts[101:140]  
test.ts<-ts(cabaitest)  
plot.ts(test.ts, lty=1, xlab="waktu", ylab="Harga Cabai", main="Plot Harga Cabai Test")



# Uji Stasioneritas Data

## Plot ACF

acf(train.ts)



Berdasarkan plot ACF, terlihat bahwa plot ACF data menurun secara perlahan (*tails of slowly*). Hal ini juga menjadi indikasi bahwa data tidak stasioner dalam rataan

## Uji ADF

tseries::adf.test(train.ts)

##   
## Augmented Dickey-Fuller Test  
##   
## data: train.ts  
## Dickey-Fuller = -3.1755, Lag order = 4, p-value = 0.09586  
## alternative hypothesis: stationary

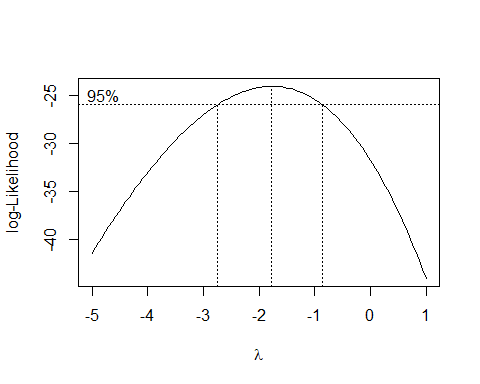
: Data tidak stasioner dalam rataan

: Data stasioner dalam rataan

Berdasarkan uji ADF tersebut, didapat *p-value* sebesar 0.09586 yang lebih besar dari taraf nyata 5% sehingga tak tolak dan menandakan bahwa data tidak stasioner dalam rataan. Hal ini sesuai dengan hasil eksplorasi menggunakan plot time series dan plot ACF, sehingga ketidakstasioneran model kedepannya harus ditangani

## Plot Box-Cox

index <- seq(1:100)  
bc = boxcox(train.ts~index, lambda = seq(-5,by=1))



#Nilai Rounded Lambda  
lambda <- bc$x[which.max(bc$y)]  
lambda

## [1] -1.787879

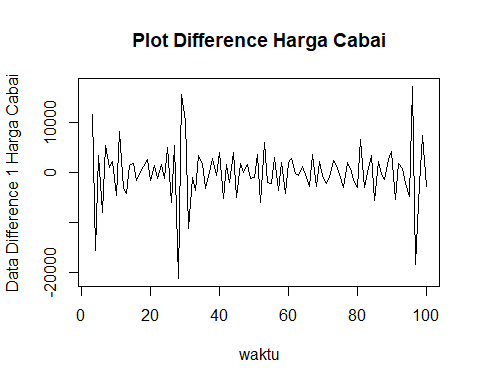
#SK  
bc$x[bc$y > max(bc$y) - 1/2 \* qchisq(.95,1)]

## [1] -2.6969697 -2.6363636 -2.5757576 -2.5151515 -2.4545455 -2.3939394  
## [7] -2.3333333 -2.2727273 -2.2121212 -2.1515152 -2.0909091 -2.0303030  
## [13] -1.9696970 -1.9090909 -1.8484848 -1.7878788 -1.7272727 -1.6666667  
## [19] -1.6060606 -1.5454545 -1.4848485 -1.4242424 -1.3636364 -1.3030303  
## [25] -1.2424242 -1.1818182 -1.1212121 -1.0606061 -1.0000000 -0.9393939  
## [31] -0.8787879

Plot Boxcox menunjukkan nilai *rounded value* () optimum sebesar **-1.787879** dan pada selang kepercayaan 95% nilai memiliki batas bawah **-2.6969697** dan batas atas **-0.8787879**. Selang tersebut tidak memuat nilai satu sehingga dapat dikatakan bahwa data tidak stasioner dalam ragam.

# Penanganan Ketidakstasioneran Data

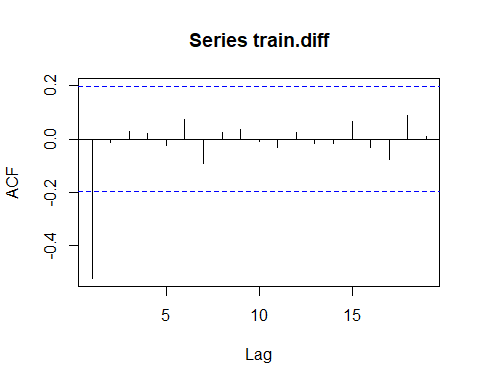
train.diff<-diff(train.ts,differences = 2)   
plot.ts(train.diff, lty=1, xlab="waktu", ylab="Data Difference 1 Harga Cabai", main="Plot Difference Harga Cabai")



Berdasarkan plot data deret waktu, terlihat bahwa data sudah stasioner dalam rataan ditandai dengan data bergerak pada nilai tengah tertentu (tidak terdapat trend ataupun musiman pada data)

## Plot ACF

acf(train.diff)



Berdasarkan plot tersebut, terlihat bahwa plot ACF cuts off pada lag ke 1. Hal ini menandakan data sudah stasioner dalam rataan dan ketidakstasioneran data telah berhasil tertangani.

## Uji ADF

tseries::adf.test(train.diff)

## Warning in tseries::adf.test(train.diff): p-value smaller than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: train.diff  
## Dickey-Fuller = -8.4919, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary

: Data tidak stasioner dalam rataan

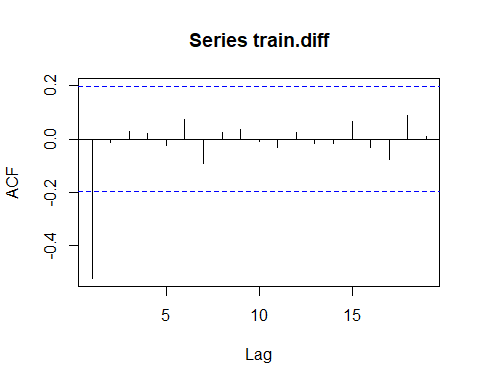
: Data stasioner dalam rataan

Berdasarkan uji ADF tersebut, didapat *p-value* sebesar 0.01 yang lebih kecil dari taraf nyata 5% sehingga tolak atau data stasioner dalam rataan. Hal ini sesuai dengan hasil eksplorasi menggunakan plot time series dan plot ACF, sehingga dalam hal ini ketidakstasioneran data sudah berhasil ditangani dan dapat dilanjutkan ke pemodelan

# Identifikasi Model

## Plot ACF

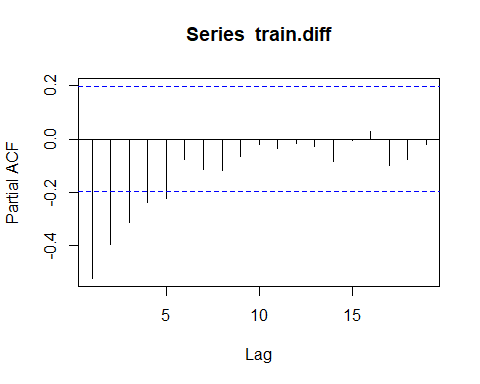
acf(train.diff)



Berdasarkan plot tersebut, terlihat bahwa plot ACF cenderung *cuts off* pada lag ke 1, sehingga jika plot PACF dianggap *tails of*, maka model tentatifnya adalah ARIMA(0,2,1)

## Plot PACF

pacf(train.diff)



Berdasarkan plot tersebut, terlihat bahwa plot PACF cenderung *cuts off* pada lag ke 5, sehingga jika plot ACF dianggap *tails of*, maka model tentatifnya adalah ARIMA(5,2,0).

Jika baik plot ACF maupun plot PACF keduanya dianggap tails of, maka model yang terbentuk adalah ARIMA(5,2,1)

## Plot EACF

eacf(train.diff)

## AR/MA  
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13  
## 0 x o o o o o o o o o o o o o   
## 1 x x o o o o o o o o o o o o   
## 2 x o x o o o o o o o o o o o   
## 3 x x o x o o o o o o o o o o   
## 4 x o o o o o o o o o o o o o   
## 5 x x x o o o o o o o o o o o   
## 6 x o o o o o o o o o o o o o   
## 7 x x o o o o o o o o o o o o

Identifikasi model menggunakan plot EACF dilakukan dengan melihat ujung segitiga pada pola segitiga nol. Dalam hal ini model tentatif yang terbentuk adalah ARIMA(0,2,1), ARIMA(0,2,2), ARIMA(1,2,2), ARIMA(2,2,3), ARIMA(4,2,1)

# Pendugaan Parameter Model Tentatif

## ARIMA(0,2,1)

model1.da=Arima(train.diff, order=c(0,0,1),method="ML")  
summary(model1.da) #AIC=1893.61

## Series: train.diff   
## ARIMA(0,0,1) with non-zero mean   
##   
## Coefficients:  
## ma1 mean  
## -1.0000 16.1206  
## s.e. 0.0265 12.6682  
##   
## sigma^2 = 13219384: log likelihood = -943.81  
## AIC=1893.61 AICc=1893.87 BIC=1901.37  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -30.86232 3598.555 2373.288 -Inf Inf 0.364166 -0.1434246

lmtest::coeftest(model1.da) #seluruh parameter signifikan

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ma1 -1.00000 0.02646 -37.7932 <2e-16 \*\*\*  
## intercept 16.12064 12.66817 1.2725 0.2032   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## ARIMA(5,2,0)

model2.da=Arima(train.diff, order=c(5,0,0),method="ML")  
summary(model2.da) #AIC=1908.64

## Series: train.diff   
## ARIMA(5,0,0) with non-zero mean   
##   
## Coefficients:  
## ar1 ar2 ar3 ar4 ar5 mean  
## -1.0057 -0.9562 -0.7802 -0.5505 -0.2824 -11.6436  
## s.e. 0.0989 0.1430 0.1618 0.1582 0.1192 85.6004  
##   
## sigma^2 = 15261584: log likelihood = -947.32  
## AIC=1908.64 AICc=1909.89 BIC=1926.74  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -58.56193 3785.129 2622.455 Inf Inf 0.402399 -0.007891049

lmtest::coeftest(model2.da) #seluruh parameter signifikan

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 -1.005687 0.098853 -10.1736 < 2.2e-16 \*\*\*  
## ar2 -0.956161 0.143007 -6.6861 2.291e-11 \*\*\*  
## ar3 -0.780199 0.161814 -4.8216 1.424e-06 \*\*\*  
## ar4 -0.550459 0.158162 -3.4803 0.0005008 \*\*\*  
## ar5 -0.282426 0.119166 -2.3700 0.0177874 \*   
## intercept -11.643643 85.600405 -0.1360 0.8918029   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## ARIMA(5,2,1)

model3.da=Arima(train.diff, order=c(5,0,1),method="ML")  
summary(model3.da) #AIC=1899.01

## Series: train.diff   
## ARIMA(5,0,1) with non-zero mean   
##   
## Coefficients:  
## ar1 ar2 ar3 ar4 ar5 ma1 mean  
## -0.1933 -0.1651 -0.0943 -0.0639 0.0102 -1.000 18.8908  
## s.e. 0.1026 0.1107 0.1160 0.1169 0.1240 0.029 8.5315  
##   
## sigma^2 = 13183762: log likelihood = -941.51  
## AIC=1899.01 AICc=1900.63 BIC=1919.69  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 4.992139 3498.866 2416.909 -Inf Inf 0.3708592 0.005667689

lmtest::coeftest(model3.da) #terdapat parameter tidak signifikan

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 -0.193255 0.102628 -1.8831 0.05969 .   
## ar2 -0.165103 0.110728 -1.4911 0.13594   
## ar3 -0.094303 0.115999 -0.8130 0.41624   
## ar4 -0.063925 0.116907 -0.5468 0.58452   
## ar5 0.010216 0.123967 0.0824 0.93432   
## ma1 -0.999986 0.028964 -34.5250 < 2e-16 \*\*\*  
## intercept 18.890834 8.531463 2.2143 0.02681 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## ARIMA(0,2,2)

model4.da=Arima(train.diff, order=c(0,0,2),method="ML")  
summary(model4.da) #AIC=1892.73

## Series: train.diff   
## ARIMA(0,0,2) with non-zero mean   
##   
## Coefficients:  
## ma1 ma2 mean  
## -1.1994 0.1994 16.8496  
## s.e. 0.1201 0.1170 10.0711  
##   
## sigma^2 = 12907690: log likelihood = -942.36  
## AIC=1892.73 AICc=1893.16 BIC=1903.07  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -43.16505 3537.309 2437.761 -Inf Inf 0.3740588 0.03292367

lmtest::coeftest(model4.da) #seluruh parameter signifikan

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ma1 -1.19939 0.12007 -9.9891 < 2e-16 \*\*\*  
## ma2 0.19940 0.11700 1.7043 0.08833 .   
## intercept 16.84957 10.07107 1.6731 0.09431 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## ARIMA(1,2,2)

model5.da=Arima(train.diff, order=c(1,0,2),method="ML")  
summary(model5.da) #AIC=1890.3

## Series: train.diff   
## ARIMA(1,0,2) with non-zero mean   
##   
## Coefficients:  
## ar1 ma1 ma2 mean  
## 0.7783 -1.9989 1.0000 16.3161  
## s.e. 0.0765 0.0595 0.0595 3.0042  
##   
## sigma^2 = 11892664: log likelihood = -940.15  
## AIC=1890.3 AICc=1890.96 BIC=1903.23  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 81.33425 3377.462 2283.513 -Inf Inf 0.3503905 -0.03784076

lmtest::coeftest(model5.da) #seluruh parameter signifikan

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 0.778303 0.076545 10.1679 < 2.2e-16 \*\*\*  
## ma1 -1.998881 0.059523 -33.5815 < 2.2e-16 \*\*\*  
## ma2 0.999957 0.059537 16.7957 < 2.2e-16 \*\*\*  
## intercept 16.316128 3.004177 5.4311 5.599e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## ARIMA(4,2,1)

model6.da=Arima(train.diff, order=c(4,0,1),method="ML")  
summary(model6.da) #AIC=1897.02

## Series: train.diff   
## ARIMA(4,0,1) with non-zero mean   
##   
## Coefficients:  
## ar1 ar2 ar3 ar4 ma1 mean  
## -0.1933 -0.1642 -0.0949 -0.0660 -1.000 18.9301  
## s.e. 0.1026 0.1107 0.1148 0.1147 0.029 8.4628  
##   
## sigma^2 = 13039218: log likelihood = -941.51  
## AIC=1897.02 AICc=1898.26 BIC=1915.11  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 5.188913 3498.699 2414.611 -Inf Inf 0.3705067 0.006381861

lmtest::coeftest(model6.da) #terdapat parameter tak signifikan

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 -0.193303 0.102623 -1.8836 0.05961 .   
## ar2 -0.164153 0.110714 -1.4827 0.13816   
## ar3 -0.094906 0.114764 -0.8270 0.40826   
## ar4 -0.066019 0.114668 -0.5757 0.56479   
## ma1 -0.999997 0.029023 -34.4555 < 2e-16 \*\*\*  
## intercept 18.930097 8.462832 2.2369 0.02530 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Penentuan Model Terbaik

perbandingan <- matrix(c("1893.61","semua\_signifikan",  
 "1908.64","semua\_signifikan",  
 "1899.01","4\_parameter\_tidak\_signifikan",  
 "1892.73","semua\_signifikan",  
 "1890.3","Semua\_signifikan",  
 "1897.02","3\_parameter\_tidak\_signifikan"),nrow=6,ncol=2,byrow = T)  
colnames(perbandingan) <- c("AIC", "Parameter")  
row.names(perbandingan) <- c("ARIMA(0,2,1)","ARIMA(5,2,0)","ARIMA(5,2,1)","ARIMA(0,2,2)","ARIMA(1,2,2)","ARIMA(4,2,1)")  
perbandingan

## AIC Parameter   
## ARIMA(0,2,1) "1893.61" "semua\_signifikan"   
## ARIMA(5,2,0) "1908.64" "semua\_signifikan"   
## ARIMA(5,2,1) "1899.01" "4\_parameter\_tidak\_signifikan"  
## ARIMA(0,2,2) "1892.73" "semua\_signifikan"   
## ARIMA(1,2,2) "1890.3" "Semua\_signifikan"   
## ARIMA(4,2,1) "1897.02" "3\_parameter\_tidak\_signifikan"

Berdasarkan pendugaan parameter di atas dengan differencing sebanyak dua kali, nilai AIC terkecil dimiliki oleh model ARIMA(1,2,2) dan parameter model ARIMA(1,2,2) juga seluruhnya signifikan sehingga model yang dipilih adalah model ARIMA(1,2,2).