Latihan Pertemuan 6 : Pendugaan Parameter

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

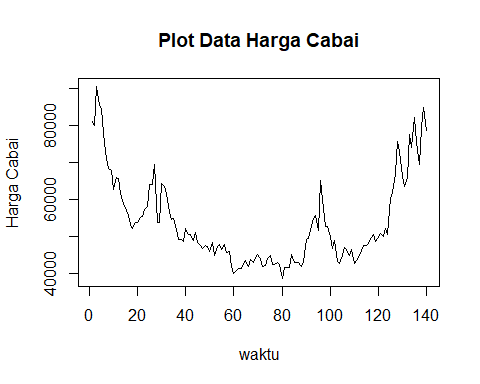
data <- read.csv("C:/Users/hp/Documents/datampdw6.csv")  
datacabai.ts <- ts(data)  
head(datacabai.ts)

## 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

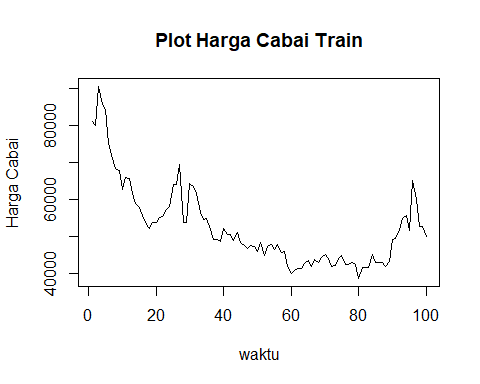
plot.ts(datacabai.ts, lty=1, xlab="waktu", ylab="Harga Cabai", main="Plot Data Harga Cabai")



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

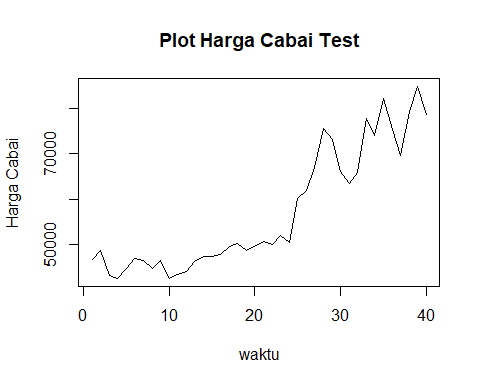
cabaitrain<-datacabai.ts[1:100]  
train.ts<-ts(cabaitrain)  
plot.ts(train.ts, lty=1, xlab="waktu", ylab="Harga Cabai", main="Plot Harga Cabai Train")



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

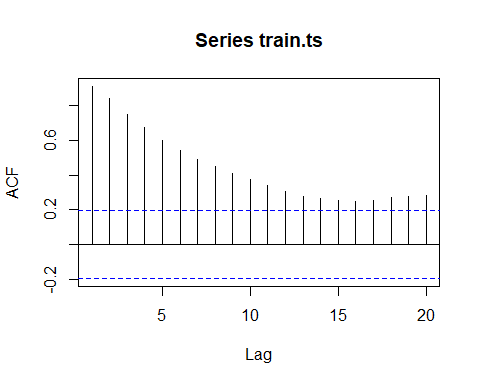
cabaitest<-datacabai.ts[101:140]  
test.ts<-ts(cabaitest)  
plot.ts(test.ts, lty=1, xlab="waktu", ylab="Harga Cabai", main="Plot Harga Cabai Test")



# 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

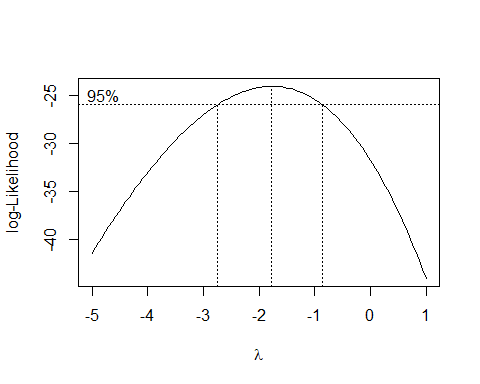
: Data tidak stasioner dalam rataan

: 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

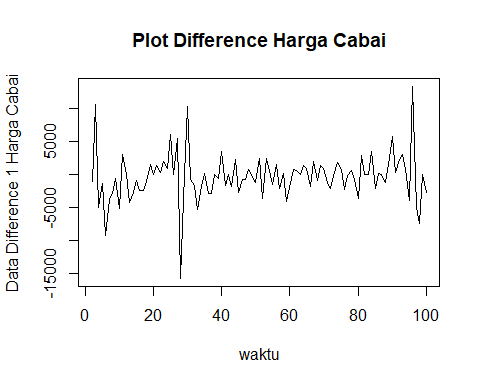
#SK  
bc$x[bc$y > max(bc$y) - 1/2 \* qchisq(.95,1)]

## [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

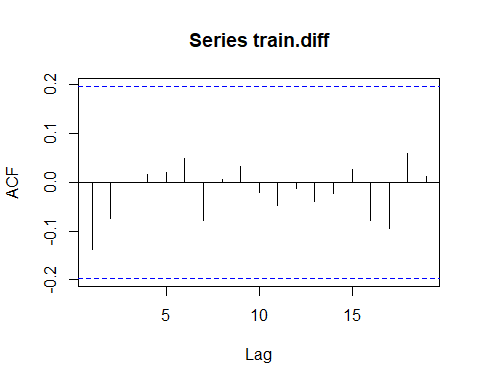
train.diff<-diff(train.ts,differences = 1)   
plot.ts(train.diff, lty=1, xlab="waktu", ylab="Data Difference 1 Harga Cabai", main="Plot Difference Harga Cabai")



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 tails off tetapi tidak signifikan pada 20 lag pertama. 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 = -4.4962, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary

: Data tidak stasioner dalam rataan

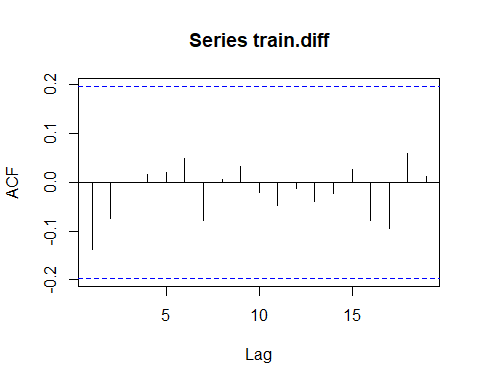
: Data stasioner dalam rataan

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

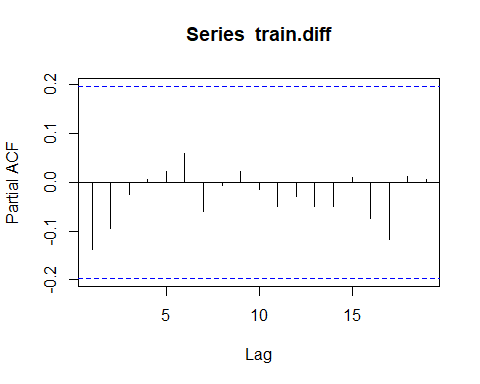
acf(train.diff)



Berdasarkan plot tersebut, terlihat bahwa plot ACF tidak *cuts off* pada 20 lag pertama, sehingga belum bisa ditentukan model tentatifnya.

## Plot PACF

pacf(train.diff)



Berdasarkan plot tersebut, terlihat bahwa plot PACF tidak *cuts off* pada lag ke 20 lag pertama, sehingga belum bisa ditentukan model tentatifnya.

## Plot EACF

eacf(train.diff)

## AR/MA  
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13  
## 0 o o o o o o 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 o o o o o o o o o o o o o   
## 3 o o o o o o o o o o o o o o   
## 4 o o o x o o o o o o o o o o   
## 5 x o o x o o o o o o o o o o   
## 6 x x x o o o o o o o o o o o   
## 7 o 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(0,1,1), ARIMA(1,1,1), ARIMA(2,1,1), ARIMA(2,1,2), ARIMA(3,1,0), dan ARIMA(3,1,1).

# Pendugaan Parameter Model Tentatif

## ARIMA(0,1,1)

model1.da=Arima(train.diff, order=c(0,1,1),method="ML")  
summary(model1.da) #AIC=1893.21

## Series: train.diff   
## ARIMA(0,1,1)   
##   
## Coefficients:  
## ma1  
## -0.9905  
## s.e. 0.0596  
##   
## sigma^2 = 13406660: log likelihood = -944.61  
## AIC=1893.21 AICc=1893.34 BIC=1898.38  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 353.5852 3624.337 2419.476 Inf Inf 0.6491219 -0.1362592

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

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ma1 -0.990483 0.059627 -16.611 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## ARIMA(1,1,1)

model2.da=Arima(train.diff, order=c(1,1,1),method="ML")  
summary(model2.da) #AIC=1893.47

## Series: train.diff   
## ARIMA(1,1,1)   
##   
## Coefficients:  
## ar1 ma1  
## -0.1356 -0.9795  
## s.e. 0.1017 0.0359  
##   
## sigma^2 = 13355537: log likelihood = -943.74  
## AIC=1893.47 AICc=1893.73 BIC=1901.23  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 343.0162 3598.725 2456.658 NaN Inf 0.6590975 -0.01416801

lmtest::coeftest(model2.da) #terdapat parameter tidak signifikan

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 -0.13559 0.10171 -1.3331 0.1825   
## ma1 -0.97946 0.03589 -27.2902 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## ARIMA(2,1,1)

model3.da=Arima(train.diff, order=c(2,1,1),method="ML")  
summary(model3.da) #AIC=1894.42

## Series: train.diff   
## ARIMA(2,1,1)   
##   
## Coefficients:  
## ar1 ar2 ma1  
## -0.1577 -0.1131 -0.9695  
## s.e. 0.1032 0.1095 0.0338  
##   
## sigma^2 = 13370521: log likelihood = -943.21  
## AIC=1894.42 AICc=1894.85 BIC=1904.76  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 364.5922 3581.94 2455.869 NaN Inf 0.6588859 -0.01036391

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

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 -0.157726 0.103161 -1.5289 0.1263   
## ar2 -0.113097 0.109523 -1.0326 0.3018   
## ma1 -0.969481 0.033827 -28.6604 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## ARIMA(2,1,2)

model4.da=Arima(train.diff, order=c(2,1,2),method="ML")  
summary(model4.da) #AIC=1896.25

## Series: train.diff   
## ARIMA(2,1,2)   
##   
## Coefficients:  
## ar1 ar2 ma1 ma2  
## 0.1605 -0.0729 -1.2928 0.3174  
## s.e. 0.6603 0.1569 0.6596 0.6438  
##   
## sigma^2 = 13495968: log likelihood = -943.13  
## AIC=1896.25 AICc=1896.91 BIC=1909.18  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 374.697 3579.714 2455.343 NaN Inf 0.6587447 -0.006826188

lmtest::coeftest(model4.da) #terdapat parameter tidak signifikan

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 0.160517 0.660305 0.2431 0.80793   
## ar2 -0.072945 0.156907 -0.4649 0.64201   
## ma1 -1.292837 0.659575 -1.9601 0.04998 \*  
## ma2 0.317363 0.643813 0.4929 0.62205   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## ARIMA(3,1,0)

model5.da=Arima(train.diff, order=c(3,1,0),method="ML")  
summary(model5.da) #AIC=1913.93

## Series: train.diff   
## ARIMA(3,1,0)   
##   
## Coefficients:  
## ar1 ar2 ar3  
## -0.8654 -0.6392 -0.3215  
## s.e. 0.0958 0.1164 0.1002  
##   
## sigma^2 = 16704933: log likelihood = -952.96  
## AIC=1913.93 AICc=1914.36 BIC=1924.27  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -124.9609 4003.747 2617.632 NaN Inf 0.7022852 -0.05952906

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

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 -0.865410 0.095808 -9.0327 < 2.2e-16 \*\*\*  
## ar2 -0.639174 0.116409 -5.4908 4.002e-08 \*\*\*  
## ar3 -0.321480 0.100169 -3.2094 0.00133 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## ARIMA(3,1,1)

model6.da=Arima(train.diff, order=c(3,1,1),method="ML")  
summary(model6.da) #AIC=1896.25

## Series: train.diff   
## ARIMA(3,1,1)   
##   
## Coefficients:  
## ar1 ar2 ar3 ma1  
## -0.1676 -0.1248 -0.0482 -0.9647  
## s.e. 0.1056 0.1129 0.1150 0.0359  
##   
## sigma^2 = 13493593: log likelihood = -943.12  
## AIC=1896.25 AICc=1896.9 BIC=1909.17  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 372.2676 3579.399 2457.509 NaN Inf 0.6593258 -0.007046909

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

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 -0.167592 0.105639 -1.5865 0.1126   
## ar2 -0.124750 0.112910 -1.1049 0.2692   
## ar3 -0.048208 0.115022 -0.4191 0.6751   
## ma1 -0.964706 0.035916 -26.8600 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Penentuan Model Terbaik

perbandingan <- matrix(c("1893.21","Semua\_signifikan",  
 "1893.47","ar1\_tidak\_signifikan",  
 "1894.42","ar1\_ar2\_tidak\_signifikan",  
 "1896.25","ar1\_ar2\_ma2\_tidak\_signifikan",  
 "1913.93","Semua\_signifikan",  
 "1896.25","ar1\_ar2\_ar3\_tidak\_signifikan"),nrow=6,ncol=2,byrow = T)  
colnames(perbandingan) <- c("AIC", "Parameter")  
row.names(perbandingan) <- c("ARIMA(0,1,1)","ARIMA(1,1,1)","ARIMA(2,1,1)","ARIMA(2,1,2)","ARIMA(3,1,0)","ARIMA(3,1,1)")  
perbandingan

## AIC Parameter   
## ARIMA(0,1,1) "1893.21" "Semua\_signifikan"   
## ARIMA(1,1,1) "1893.47" "ar1\_tidak\_signifikan"   
## ARIMA(2,1,1) "1894.42" "ar1\_ar2\_tidak\_signifikan"   
## ARIMA(2,1,2) "1896.25" "ar1\_ar2\_ma2\_tidak\_signifikan"  
## ARIMA(3,1,0) "1913.93" "Semua\_signifikan"   
## ARIMA(3,1,1) "1896.25" "ar1\_ar2\_ar3\_tidak\_signifikan"

Berdasarkan pendugaan parameter di atas, nilai AIC terkecil dimiliki oleh model ARIMA(0,1,1) dan parameter model ARIMA(0,1,1) juga seluruhnya signifikan sehingga model yang dipilih adalah model ARIMA(0,1,1).