Prediction and Time Series Computer Exercise Week 5

Name: Nguyen Xuan Binh Student ID: 887799

12/02/2021

Exercise 5.5: The file t38.txt contains three quarterly time series. The time series start from the first quarter of the year 1953 and the corresponding time series are, CONS = total consumption (billions), INC = income (billions), INFLAT = inflation (%)

The time series CONS and INC represent the observed total consumption and income in an imaginary country. The time series INFLAT represents inflation. The goal is to estimate a so-called consumption function that explains the time series CONS with the time series INC and INFLAT. The conventional linear regression model for the response variable CONS is $CONS_t = \beta0 + \beta1 * INC_t + \beta2 * INFLAT_t + \varepsilon_t$ (5)

a) Estimate model (5) and study the goodness of fit.

```
library(car)

## Loading required package: carData

library(forecast)

## Warning: package 'forecast' was built under R version 4.1.2

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

library(lmtest)

## Warning: package 'lmtest' was built under R version 4.1.2

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.1.2

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

t38<-read.table("t38.txt",header=T,sep=" ")
cons<-t38$CONS
consTS <- ts(cons, start = 1953, frequency = 3)
inc<-t38$INC
incTS <- ts(inc, start = 1953, frequency = 3)
```

```
inflat<-t38$INFLAT
inflatTS <- ts(inflat, start = 1953, frequency = 3)
model5 <- lm (formula = cons ~ inc + inflat)
summary(model5)

##
## Call:
## lm(formula = cons ~ inc + inflat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.0491 -2.1273  0.4948  2.3026  8.6025
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -147.38977   21.71937  -6.786 2.26e-10 ***
## inc            1.15263    0.02431  47.420  < 2e-16 ***
## inflat        -2.47468    0.20268 -12.210  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.267 on 156 degrees of freedom
## Multiple R-squared:  0.9425, Adjusted R-squared:  0.9418
## F-statistic:  1279 on 2 and 156 DF,  p-value: < 2.2e-16
```

Model study: The regression coefficients corresponding to the income variable INC and inflation variable INFLAT are statistically significant with 5% level of significance.

The coefficient of the income is positive and the coefficient of the inflation is negative, indicating direction proportional relationship between consumption and inverse proportional relationship between consumption and inflation.

The coefficient of determination is 0.9425, high enough to be considered as a good model

```
par(mfrow = c(3,2),mar = c(3, 3, 3, 3))
# Q-Q plot of the residuals of model
qqnorm(model5$residuals,pch=16, main="Fig 1: Q-Q plot of the residuals of mod
el 5")
qqline(model5$residuals,col="red",lwd=2)

#fitted model and original time series
fit <- ts(predict(model5), start = 1953, frequency = 3)
plot(consTS,col="red",main = "Fig 2: fitted model and original time series",x
lab="Time",ylab="")
lines(fit,col="blue")
legend("topright", legend=c("cons", "fit"),
       col=c("red","blue"),lty=c(1,1),cex=0.5)
# Estimated residuals of model 5
plot(model5$residuals,type="p",main="Fig 3: Estimated residuals of model 5",y
lab="Residuals",xlab="Year",pch=16,xaxt="n")
axis(1,at=seq(from=0,to=159,by=3),labels=seq(from=1953,to=2006,by=1))
```

```
abline(0,0)
# Cook’s distances of model 5
plot(cooks.distance(model5),main="Fig 4: Cook’s distances of model 5",ylab="C
ook's distances",xlab="Year",pch=16,xaxt="n")
axis(1,at=seq(from=0,to=159,by=3),labels=seq(from=1953,to=2006,by=1))
# Estimated residuals of model 5
hist(model5$residuals,xlab="Residuals",ylab="Frequency",main="Fig 5: Estimate
d residuals of model 5")
# ACF of the estimated residuals of model 5
acf(model5$residuals, main = "Fig 6: ACF of the estimated residuals of model
5")
```

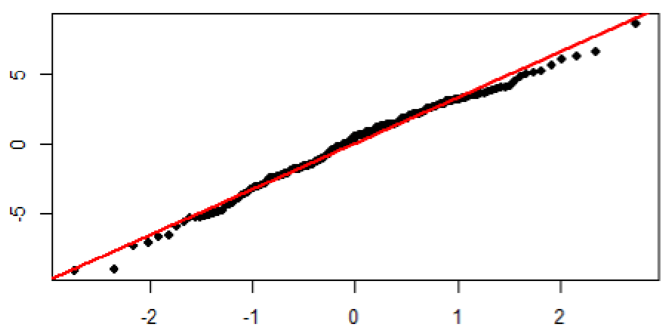


Fig 1: Q-Q plot of the residuals of model 5

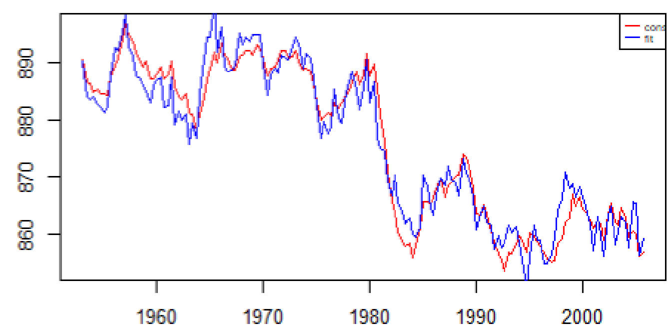


Fig 2: fitted model and original time series

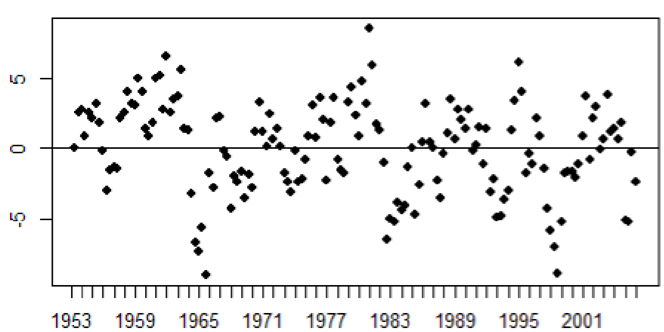


Fig 3: Estimated residuals of model 5

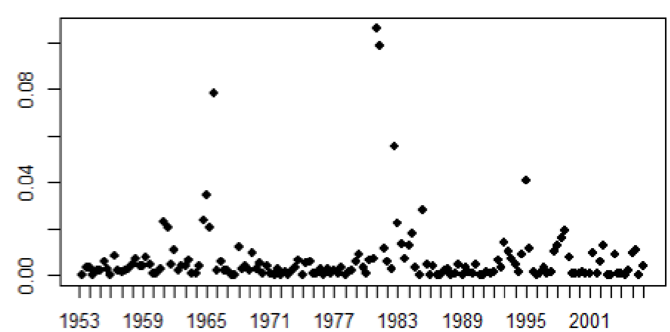


Fig 4: Cook’s distances of model 5

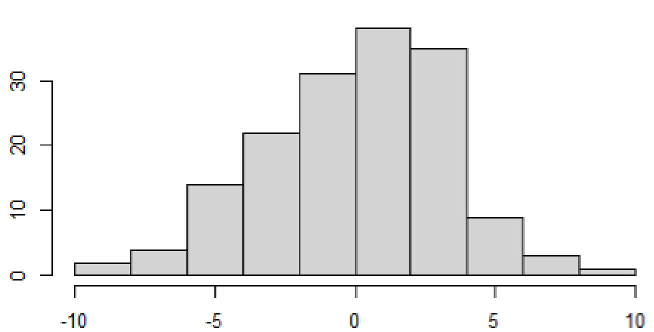


Fig 5: Estimated residuals of model 5

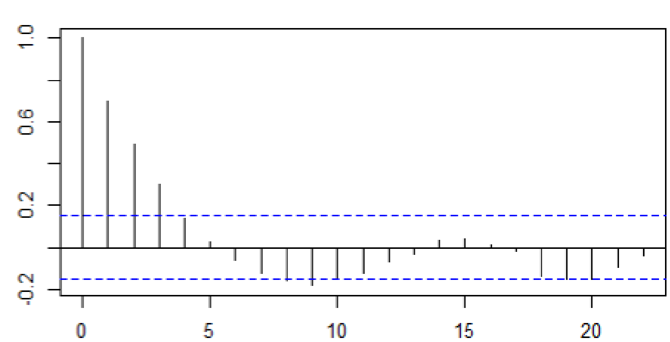


Fig 6: ACF of the estimated residuals of model 5

```r
# VIF of model 5
vif(model5)

##     inc  inflat
## 1.01252 1.01252
```

Comments for the model 5: • By Figures 5, the residuals seem to be normally distributed.

• By Figures 1 and 6, the residuals seem to be heavily correlated.

• By the variance inflation factor VIF, multicollinearity between CONS and INFLAT is unlikely as VIF is very close to 1.

• The reason for the correlatedness of the residuals can be seen from Figure 2, where the fitted curve stays above and below the response variable CONS for long time periods.

• There are many outliers in the cook's distance in figure 4 => Model 5 can be considered to be inefficient in describing the model

b) Estimate the difference model corresponding to (5) and study the goodness of fit

```r
consD <- diff(cons)
incD <- diff(inc)
inflatD <- diff(inflat)
model5D<-lm(consD ~ incD + inflatD)
summary(model5D)

##
## Call:
## lm(formula = consD ~ incD + inflatD)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6284 -0.8637  0.0631  0.9223  3.9466
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.11968    0.11453  -1.045  0.29764
## incD         0.51830    0.03527  14.696  < 2e-16 ***
## inflatD     -0.71594    0.23934  -2.991  0.00323 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.437 on 155 degrees of freedom
## Multiple R-squared:  0.5826, Adjusted R-squared:  0.5772
## F-statistic: 108.2 on 2 and 155 DF,  p-value: < 2.2e-16
```

Model study: The regression coefficients corresponding to the differenced income variable incD and differenced inflation variable inflatD are statistically significant with 5% level of significance. The coefficient of the differenced income is positive and the coefficient of the differenced inflation is negative, still indicating their relationship with consumption found in (a)

The coefficient of determination is 0.5826, suggesting that the differenced model may not be the correct model

```
par(mfrow = c(3,2),mar = c(3, 3, 3, 3))
# Q-Q plot of the residuals of model
qqnorm(model5D$residuals,pch=16, main="Fig 1: Q-Q plot of the residuals of di
fferenced model 5")
qqline(model5D$residuals,col="red",lwd=2)

#fitted model and original time series
fit <- ts(predict(model5D, start = 1953, frequency = 3))
plot(ts(consD),col="red",main = "Fig 2: differenced fitted model and differen
ced original time series",xlab="Time",ylab="")
lines(fit,col="blue")
legend("topright", legend=c("differenced cons", "differenced fit"),
       col=c("red","blue"),lty=c(1,1),cex=0.3)
# Estimated residuals of model 5
plot(model5D$residuals,type="p",main="Fig 3: Estimated residuals of differenc
ed model 5",ylab="Residuals",xlab="Year",pch=16,xaxt="n")
axis(1,at=seq(from=0,to=159,by=3),labels=seq(from=1953,to=2006,by=1))
abline(0,0)
# Cook’s distances of model 5
plot(cooks.distance(model5D),main="Fig 4: Cook’s distances of differenced mod
el 5",ylab="Cook's distances",xlab="Year",pch=16,xaxt="n")
axis(1,at=seq(from=0,to=159,by=3),labels=seq(from=1953,to=2006,by=1))
# Estimated residuals of model 5
hist(model5D$residuals,xlab="Residuals",ylab="Frequency",main="Fig 5: Estimat
ed residuals of differenced model 5")
# ACF of the estimated residuals of model 5
acf(model5D$residuals, main = "Fig 6: ACF of estimated residuals of differenc
ed model 5")
```

Fig 1: Q-Q plot of the residuals of differenced model 5

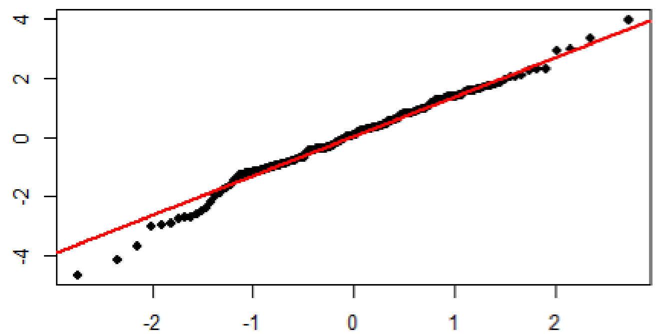


Fig 2: differenced fit and differenced original time series

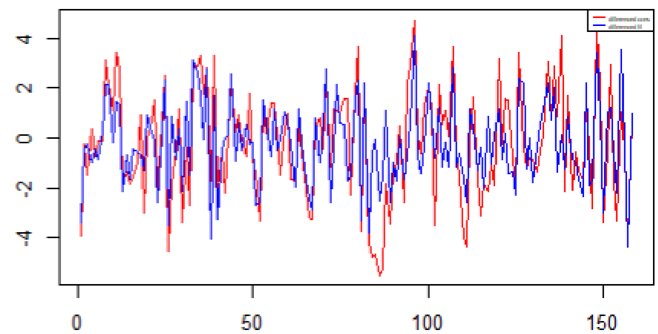


Fig 3: Estimated residuals of differenced model 5

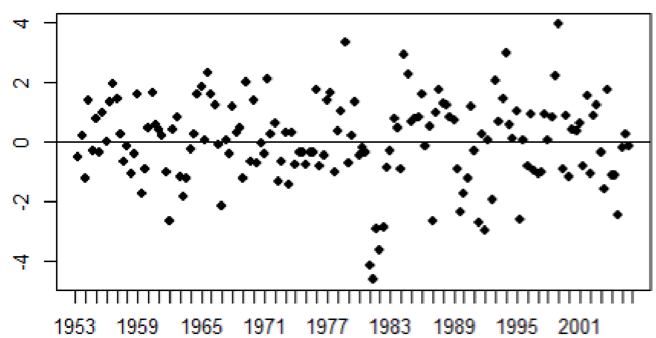


Fig 4: Cook's distances of differenced model 5

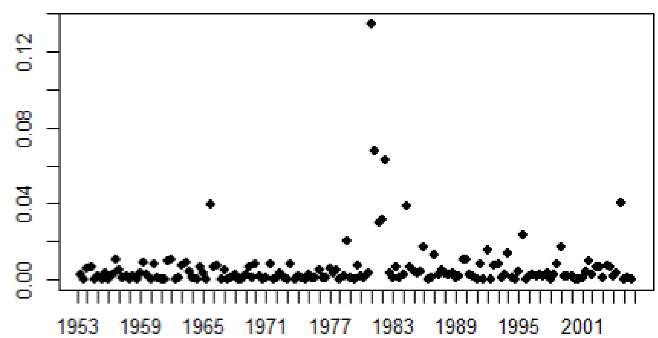


Fig 5: Estimated residuals of differenced model 5

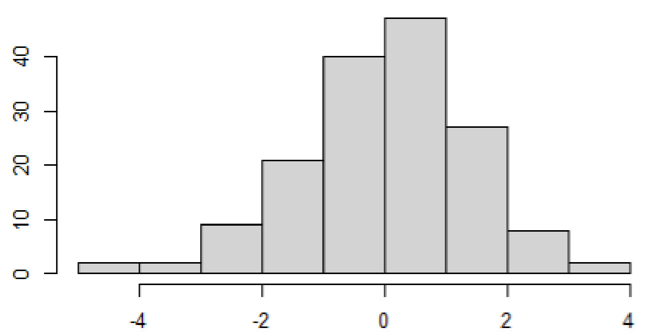
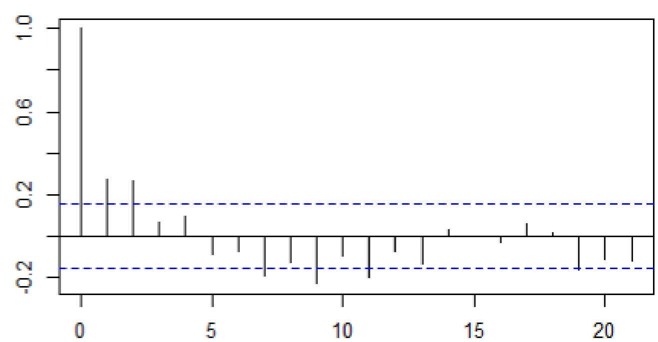


Fig 6: ACF of estimated residuals of differenced model 5




```
vif(model5D)

##      incD  inflatD
## 1.064351 1.064351
```

Comments for the model 5 with differencing:
• By Figures 5, the residuals could be normally distributed.
• By Figures 1 and 6, the residuals are no longer heavily correlated.
• By the variance inflation factor VIF, multicollinearity between CONS diff and INFLAT diff is unlikely as VIF is very close to 1.
• There are many fewer outliers in the cook's distance in figure 4 compared to the undifferenced model

• In figure 2, the fit line does not match the differenced time series of consumption, suggesting that the differencing model is incorrect

```
install.packages("lmtest")

## Warning: package 'lmtest' is in use and will not be installed

library(lmtest)
model5_bg <- rep(NA,155)
# Breusch-Godfrey can be performed up to order:
# (sample size) - (number of estimated parameters) = 158-3 = 155
for (i in 1:155) {
  model5_bg[i]= bgtest(model5D, order=i)$p.value
}
which(model5_bg > 0.05)

##  [1]  77  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99 100 1
01 102
## [20] 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 1
20 121
## [39] 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 1
39 140
## [58] 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155

# Null hypothesis of no autocorrelation accepted with all lags
```

Judging by the Breusch-Godfrey test, there are many residuals with p_value > 0.05, proving to be correlated

=> the null hypothesis of non-correlation is rejected

=> The differenced model 5 is inefficient to explain the relationship between response variable consumption and explanatory variables income and inflation

c) Estimate dynamic regression model:

CONS_t = β0 + β1 * CONS_t-1 + β2 * INC_t + β3 * INC_t-1 + β4 * INFLAT_t + β5 * INFLAT_t-1 + ε_t and study the goodness of fit

```
n <- nrow(t38)
model5dyna <- lm(consTS[-1] ~ consTS[-n] + incTS[-1] + incTS[-n] +
inflatTS[-1] + inflatTS[-n])
summary(model5dyna)

##
## Call:
## lm(formula = consTS[-1] ~ consTS[-n] + incTS[-1] + incTS[-n] +
##     inflatTS[-1] + inflatTS[-n])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.49953 -0.76349 -0.04695  0.62801  3.15931
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  -20.26950    8.52628  -2.377   0.0187 *  
## consTS[-n]     0.79831    0.02716  29.393  < 2e-16 ***
## incTS[-1]      0.49894    0.02833  17.611  < 2e-16 ***
## incTS[-n]     -0.27611    0.03788  -7.290 1.59e-11 ***
## inflatTS[-1]  -0.79309    0.18395  -4.311 2.90e-05 ***
## inflatTS[-n]  -0.25061    0.20310  -1.234   0.2191    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.078 on 152 degrees of freedom
## Multiple R-squared:  0.9939, Adjusted R-squared:  0.9937
## F-statistic:  4915 on 5 and 152 DF,  p-value: < 2.2e-16
```

- All regression coefficients are statistically significant except inflat_t-1 variable

- The coefficient of the variable cons_t with lag 1 is 0.798, which implies that the adjustment to changes in income and inflation is moderately fast.

- The signs of the coefficients of the income and inflation variables with lag 0 are as expected: the coefficient -0.79 of the inflation variable is negative and the coefficient +0.49 of income variable is positive. These coefficients describe the instant effects of changes in income and inflation.

- The signs of the coefficients of the income and inflation variables with lag 1 are also as expected.

- Interpretations of the regression coefficients of income and inflation variables with lag 0:
– If the income goes up by 1%, then total consumptions are instantly increased by (without a lag) 0.498%.
– If the inflation is increased by 1%, then total consumptions are reduced by 0.793%.

- Interpretations of the long term elasticities of income and inflation variables:
– If the income goes up by 1%, then the total consumptions are reduced by 0.27% in the long term.
– If the inflation is increased by 1%, then the total consumptions are reduced by 0.25% in the long term.

- However, it is not possible to draw conclusions from the coefficient of the determination.

```
par(mfrow = c(3,2),mar = c(3, 3, 3, 3))
# Q-Q plot of the residuals of model
qqnorm(model5dyna$residuals,pch=16, main="Fig 1: Q-Q plot of the residuals of differenced model 5")
qqline(model5dyna$residuals,col="red",lwd=2)
#fitted model and original time series
fit <- ts(predict(model5dyna, start = 1953, frequency = 3))
plot(ts(cons),col="red",main = "Fig 2: differenced fitted model and differenced original time series",xlab="Time",ylab="")
lines(fit,col="blue")
legend("topright", legend=c("differenced cons", "differenced fit"),
```

```
       col=c("red","blue"),lty=c(1,1),cex=0.3)
# Estimated residuals of model 5
plot(model5dyna$residuals,type="p",main="Fig 3: Estimated residuals of differ
enced model 5",ylab="Residuals",xlab="Year",pch=16,xaxt="n")
axis(1,at=seq(from=0,to=159,by=3),labels=seq(from=1953,to=2006,by=1))
abline(0,0)
# Cook's distances of model 5
plot(cooks.distance(model5dyna),main="Fig 4: Cook's distances of differenced
model 5",ylab="Cook's distances",xlab="Year",pch=16,xaxt="n")
axis(1,at=seq(from=0,to=159,by=3),labels=seq(from=1953,to=2006,by=1))
# Estimated residuals of model 5
hist(model5dyna$residuals,xlab="Residuals",ylab="Frequency",main="Fig 5: Esti
mated residuals of differenced model 5")
# ACF of the estimated residuals of model 5
acf(model5dyna$residuals, main = "Fig 6: ACF of estimated residuals of differ
enced model 5")
```

Fig 1: Q-Q plot of the residuals of dynamic regression model

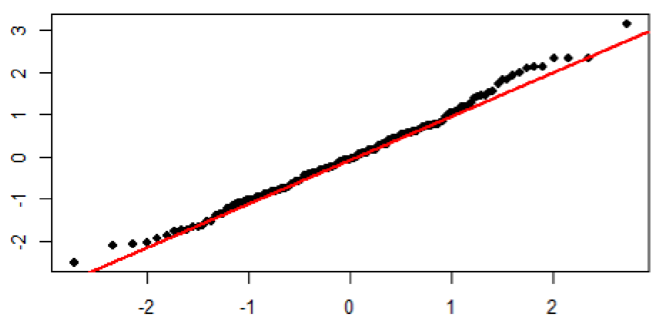


Fig 2: dynamic fitted model and dynamic original time series

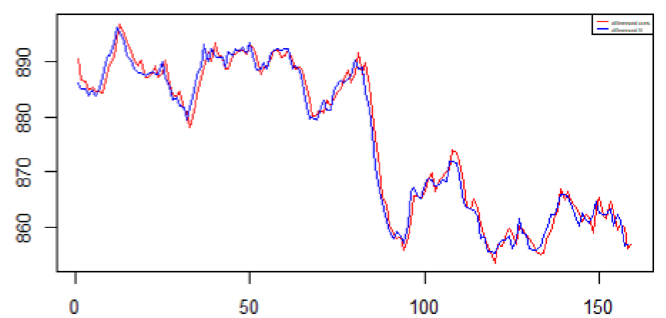


Fig 3: Estimated residuals of dynamic regression model

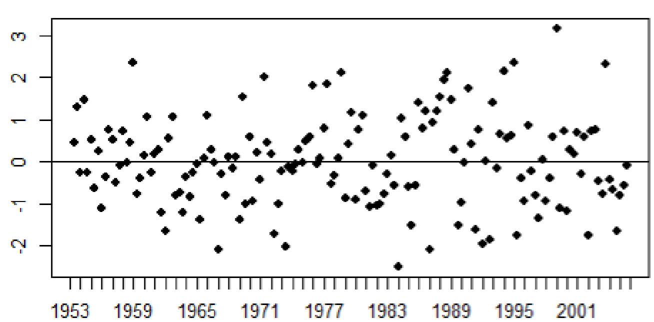


Fig 4: Cook's distances of dynamic regression model

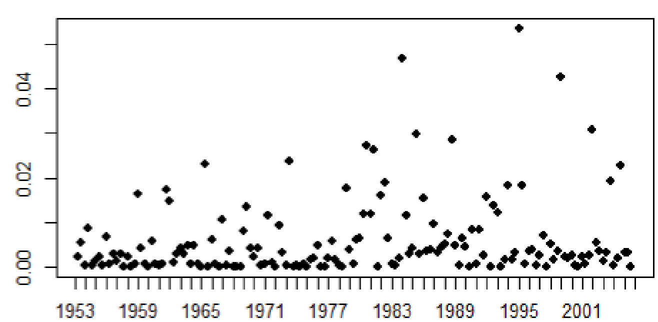


Fig 5: Estimated residuals of dynamic regression model

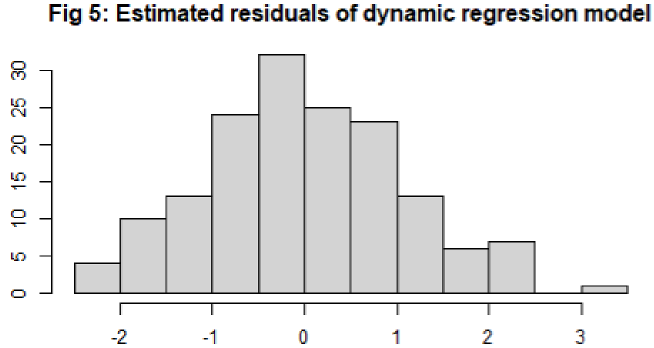
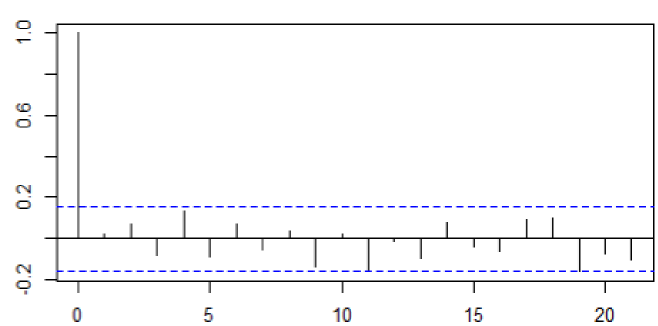


Fig 6: ACF of estimated residuals of dynamic regression model



Comments related to the dynamic regression of model 5:
• By Figures 1 and 5, the residuals could be normally distributed.
• By Figures 3 and 6, the residuals are not correlated.
• By the residual diagrams in figure 3, there is no evidence of heteroscedasticity.
• By Figure 2, the fitted model coincides better with the original time series than the fits of original model 5 found in (a) and differenced model 5 found in (b).

Now we test Breusch-Godfrey and VIF test

```
library(lmtest)
model5dyna_bg <- rep(NA,152)
# Breusch-Godfrey can be performed up to order:
# (sample size) - (number of estimated parameters) = 158-6 = 152
for (i in 1:152) {
  model5dyna_bg[i]= bgtest(model5dyna, order=i)$p.value
}
which(model5dyna_bg < 0.05)

## integer(0)

model5dyna_bg

##   [1] 0.7404511 0.6045556 0.5650183 0.2787527 0.3042654 0.3551158 0.454795
0
##   [8] 0.5631049 0.4735877 0.5690298 0.4061289 0.4731610 0.5252836 0.578153
9
##  [15] 0.6520422 0.5894026 0.3956821 0.4355298 0.2071538 0.2106953 0.156750
3
##  [22] 0.1534830 0.1858849 0.2243955 0.2578920 0.2124698 0.2438560 0.265090
5
##  [29] 0.3049527 0.3230335 0.3143937 0.3220023 0.3133261 0.3383738 0.340553
2
##  [36] 0.3833834 0.2832329 0.3059693 0.3015777 0.3016666 0.2474447 0.190445
0
##  [43] 0.1997935 0.2274582 0.2592239 0.2802037 0.3031284 0.3299429 0.336761
7
##  [50] 0.3662879 0.3262469 0.3513529 0.3804923 0.3600569 0.3943330 0.408330
1
##  [57] 0.4052707 0.3568331 0.3836353 0.4141415 0.4498262 0.2846277 0.315215
0
##  [64] 0.2920044 0.3111106 0.2746370 0.2839722 0.3105616 0.3010305 0.326604
5
##  [71] 0.3440329 0.3242977 0.3514303 0.3808195 0.4123068 0.4152464 0.425857
2
##  [78] 0.3463123 0.3760194 0.3869832 0.4172575 0.3888215 0.4049522 0.427185
9
##  [85] 0.4504424 0.4378654 0.4266472 0.4468663 0.4414324 0.4709179 0.498522
4
##  [92] 0.5109369 0.5256059 0.5159122 0.4443017 0.4474901 0.4747258 0.383956
6
```

```
##  [99] 0.3469387 0.3438840 0.3571655 0.3698975 0.3952581 0.4077615 0.417802
5
## [106] 0.4302655 0.4564772 0.4684763 0.4728930 0.4867985 0.5076234 0.334416
1
## [113] 0.3588214 0.2854976 0.3008617 0.2785145 0.2904988 0.1866446 0.204275
0
## [120] 0.2159708 0.2037159 0.2217452 0.2394409 0.2586381 0.2714962 0.235172
2
## [127] 0.2028571 0.2156090 0.2337641 0.2278207 0.2445561 0.2436024 0.253283
4
## [134] 0.2700678 0.2904124 0.2889899 0.3083564 0.2653938 0.2709980 0.246039
3
## [141] 0.2633544 0.2775652 0.2967588 0.2927641 0.3092860 0.3047361 0.311479
9
## [148] 0.3285512 0.3216544 0.3286712 0.3403599 0.3527818
```

```
vif(model5dyna)
```

```
##   consTS[-n]    incTS[-1]    incTS[-n] inflatTS[-1] inflatTS[-n]
##    18.137266    12.433573    22.292466     7.552629     9.329719
```

• By the Breusch-Godfrey test, the residuals are not correlated. The null hypothesis is accepted with 5% level of significance for all lags: By VIF, there is strong multicollinearity in the model as all of the VIF values of the explanatory variables are bigger than 10 because the model involves same variables with different lags.

=> We consider this model to be sufficient in explaining the consumption variable

d) Which of the previous models are sufficient in explaining the behavior of the response variable CONS?

The dynamic regression model found in (c) is an efficient model for the relationship between the response variable consumption and the explanatory variables income and inflation, as proven in (c)