NHUT CAO 906939

COMPUTER EXERCISE 5

5.5, a)

First load necessary libraries, read the data and examine the model

```
library(car)
```

```
## Loading required package: carData
```

```
library(forecast)
```

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

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

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

```
# 5.4 a, Model 1
data <- read.table("t38.txt",header=T)
CONS <- ts(data$CONS)
INC <- ts(data$INC)
INFLAT <- ts(data$INFLAT)

model1 <-lm(CONS~INC+INFLAT)
summary(model1)
```

Base on the summary table below:

- All the coefficients of the difference model (1) are statistically significant with 5% level of significance.
- The signs of the regression coefficients of the income and inflation variables are as expected: the coefficient of the inflation variable is negative and the coefficient of the income variable is positive.
- The coefficient of determination of the model is 94.25%
- Interpretation of the regression coefficients as elasticities:
  If the income increases by 1%, then the total consumption increases by 1.152%
  If the inflation increases by 1%, then the total consumption is reduced by 2.475%

```
##
## 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
```

```
# the residuals are almost normally distributed
par(mfrow=c(3,2))
hist(model1$residuals,xlab="Residuals",ylab="Frequency",main=" ")

acf(model1$residuals,main="")
plot(model1$residuals,type="p",ylab="Residuals",xlab="Time",pch=16,xaxt="n")
abline(0,0)

qqnorm(model1$residuals, pch=16)
qqline(model1$residuals,col="red",lwd=2)

plot(model1$fitted.values, ylab="Model 1 Fitted")

plot(cooks.distance(model1),ylab="Cook's distances",xlab="Time",pch=16,xaxt="n")
axis(1)
```

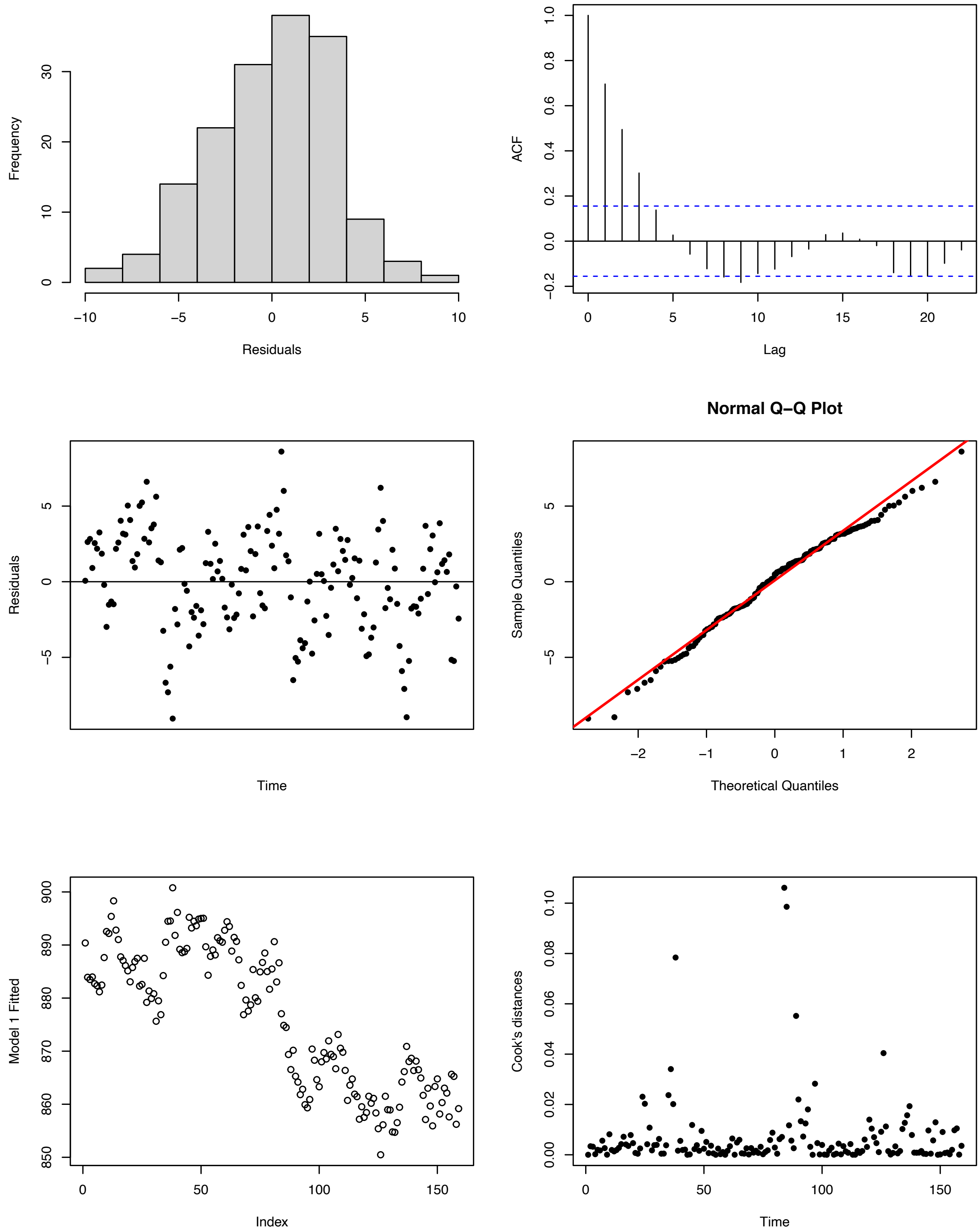
Below we study the normality of the residuals, Cook's distances and compare the fitted model with the original time series.

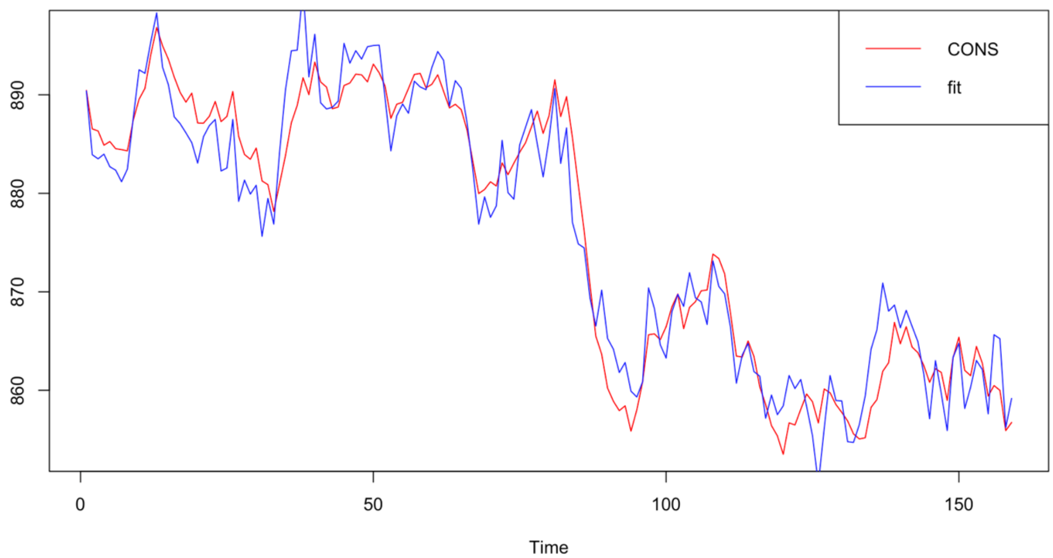
- Based on the histogram and the Q-Q plot, we can see that the residuals are mostly normally distributed
- Based on ACF and the estimated residuals plot, we can say that the residuals are not correlated.

```
plot(CONS,col="red",xlab="Time",ylab="")
fit <- ts(predict(model1))
lines(fit,col="blue")
legend("topright", legend=c("CONS", "fit"),
       col=c("red","blue"),lty=c(1,1),cex=1)

vif(model1)
```

```
##      INC   INFLAT
## 1.01252 1.01252
```

The VIF shows there is no multicollinearity.


Frequency
Residuals
ACF
Lag
Normal Q−Q Plot
Residuals
Time
Sample Quantiles
Theoretical Quantiles
Model 1 Fitted
Index
Cook's distances
Time



Compare the fitted model and the original time series, we see that there are still errors in predicting the time series, but mostly correctly follow the original plot, therefore this model is satisfactory.

b) Study the data, here the parameters are differenced.

```
# b Model 2
DCONS <- diff(CONS)
DINC <- diff(INC)
DINFLAT <- diff(INFLAT)

model2 <- lm(DCONS~DINC+DINFLAT)
summary(model2)
```

```
##
## Call:
## lm(formula = DCONS ~ DINC + DINFLAT)
##
## 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
## DINC         0.51830    0.03527  14.696  < 2e-16 ***
## DINFLAT     -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
```

```
vif(model2)
```

```
##     DINC  DINFLAT
## 1.064351 1.064351
```

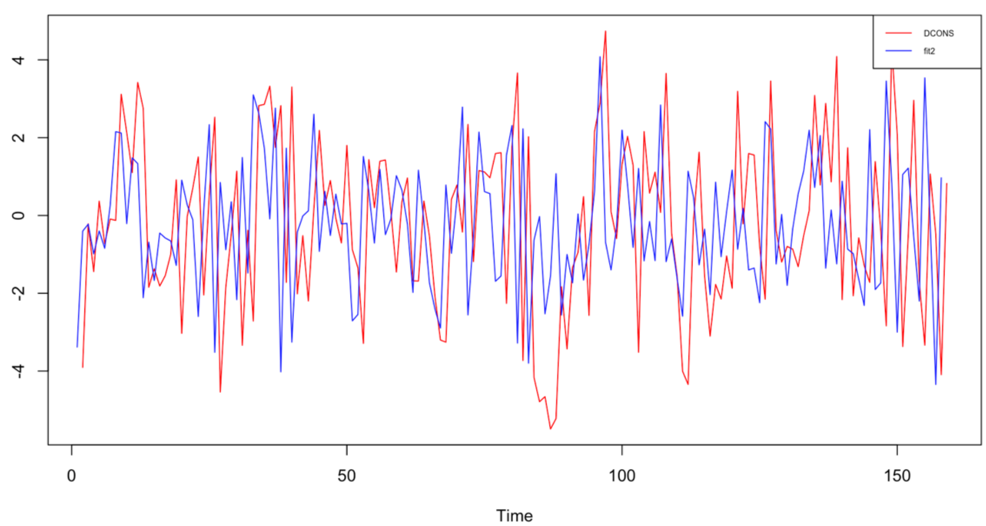
- Based on the VIF, the multicollinearity is not a problem

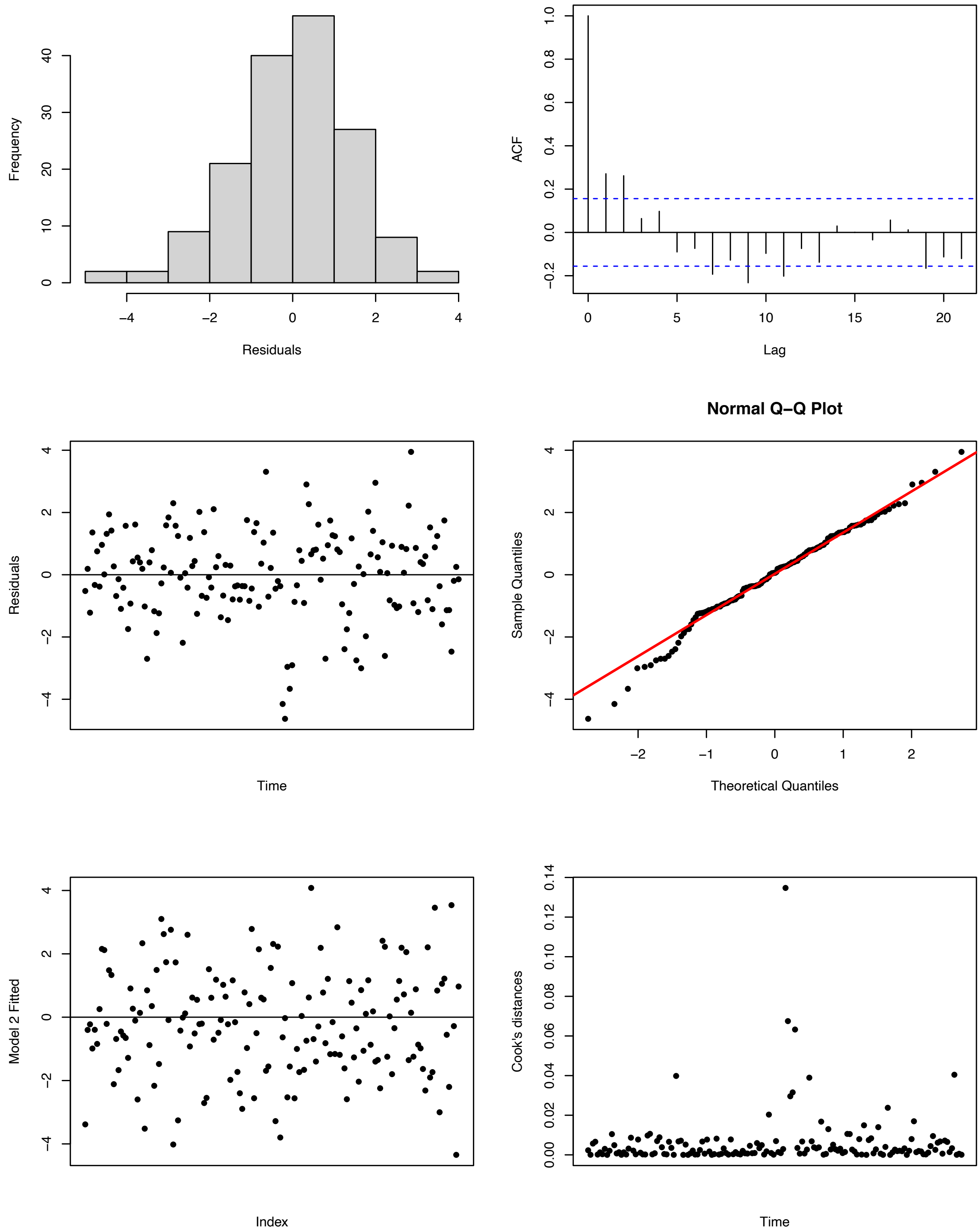
- The coefficient of determination is 58.26%
- All the coefficients of the difference model (1) are statistically significant with 5% level of significance. Except the constant term
- The signs of the regression coefficients of the income and inflation variables are as expected: the coefficient of the inflation variable is negative and the coefficient of the income variable is positive.
- Interpretation of the regression coefficients as elasticities:
  If the income increases by 1%, then the total consumption increases by 0.518%
  If the inflation increases by 1%, then the total consumption is reduced by 0.716%

Next, we study the normality of the residuals, Cook's distances and compare the fitted model with the original time series.

```
par(mfrow=c(3,2))
hist(model2$residuals,xlab="Residuals",ylab="Frequency",main=" ")
# the residuals are normally distributed
acf(model2$residuals,main="")
plot(model2$residuals,type="p",ylab="Residuals",xlab="Time",pch=16,xaxt="n")
abline(0,0)

qqnorm(model2$residuals, pch=16)
qqline(model2$residuals,col="red",lwd=2)

plot(model2$fitted.values, ylab="Model 2 Fitted",xlab="Index",pch=16,xaxt="n")
abline(0,0)

plot(cooks.distance(model2),ylab="Cook's distances",xlab="Time",pch=16,xaxt="n")
axis(1)
```

- The residuals seem to be normally distributed and mostly correlated
- Plotting the fitted model against the original time series shows that the model doesn't provide good prediction.




Frequency
Residuals
ACF
Lag
Normal Q–Q Plot
Residuals
Time
Sample Quantiles
Theoretical Quantiles
Model 2 Fitted
Index
Cook's distances
Time

```r
plot(DCONS,col="red",xlab="Time",ylab="")
fit2 <- ts(predict(model2))
lines(fit2, col="blue")
legend("topright", legend=c("DCONS", "fit2"),
       col=c("red","blue"),lty=c(1,1),cex=0.5)

model2_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)
{
  model2_bg[i]= bgtest(model2, order=i)$p.value
}
model2_bg
```

```
##   [1] 0.0005879214 0.0000898417 0.0003011303 0.0006804155 0.0006436374
##   [6] 0.0012665766 0.0011012242 0.0020677546 0.0016284475 0.0029820923
##  [11] 0.0025777788 0.0044030950 0.0055857041 0.0080918554 0.0126024357
##  [16] 0.0079146566 0.0098757348 0.0108479960 0.0009305510 0.0010668814
##  [21] 0.0012893710 0.0009040996 0.0014001155 0.0019421903 0.0027881008
##  [26] 0.0023499849 0.0032680136 0.0039262270 0.0054085488 0.0073176619
##  [31] 0.0067465407 0.0078259304 0.0106504684 0.0115743999 0.0095570203
##  [36] 0.0127563195 0.0127330943 0.0105265427 0.0104830044 0.0137299864
##  [41] 0.0176307860 0.0220569107 0.0069454054 0.0085195284 0.0110696094
##  [46] 0.0093935767 0.0114792144 0.0121402940 0.0098726367 0.0098528288
##  [51] 0.0087503452 0.0104241885 0.0132362326 0.0115905922 0.0121763606
##  [56] 0.0113020775 0.0107743901 0.0135346550 0.0167910738 0.0200937260
##  [61] 0.0198467699 0.0173314586 0.0198385338 0.0219926726 0.0260803011
##  [66] 0.0278098892 0.0334604329 0.0394848049 0.0412136608 0.0317942461
##  [71] 0.0274585904 0.0294305814 0.0320445785 0.0342573986 0.0388645952
##  [76] 0.0453800346 0.0526029259 0.0355265556 0.0394876225 0.0376074805
##  [81] 0.0414412854 0.0370228312 0.0385571361 0.0449953808 0.0521744571
##  [86] 0.0590129858 0.0657735609 0.0725000429 0.0812868637 0.0859076637
##  [91] 0.0816653186 0.0930229144 0.1047383729 0.1023596953 0.1105456974
##  [96] 0.1220355647 0.1336573677 0.1455528671 0.1160496962 0.1227290774
## [101] 0.1373088823 0.1403937676 0.1557276741 0.1725201693 0.1900028889
## [106] 0.1829241124 0.1978893044 0.2138772170 0.2325418902 0.2520935907
## [111] 0.2382338479 0.2310304879 0.1583447894 0.1651373624 0.1802140738
## [116] 0.1690933336 0.1566682891 0.1713082941 0.1842977026 0.2017405254
## [121] 0.2097228614 0.2275503949 0.2452815463 0.2609908994 0.2815875581
## [126] 0.2853384486 0.3039901570 0.2476193236 0.2635699444 0.2654411527
## [131] 0.2582737757 0.2729706494 0.2931702984 0.3132976526 0.3153110305
## [136] 0.3135490303 0.3143163674 0.3318457259 0.2752869555 0.2788686100
## [141] 0.2990825519 0.2302010773 0.2321796457 0.2458011649 0.2635265517
## [146] 0.2828006276 0.2902583696 0.3100558972 0.3306708698 0.3510609149
## [151] 0.3612834804 0.3731048768 0.3949052003 0.4169910456 0.4392933282
```

```r
which(model2_bg > 0.05)
```

```
##  [1]  77  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99 100 101 102
## [20] 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121
## [39] 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
## [58] 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155
```

- By the Breusch-Godfrey test, the residuals are correlated, the null hypothesis is rejected with the 5% of level of significance for lags 77-155.
- By the regression diagnostics, the model is not satisfactory.

c) Now we have a dynamic regression model

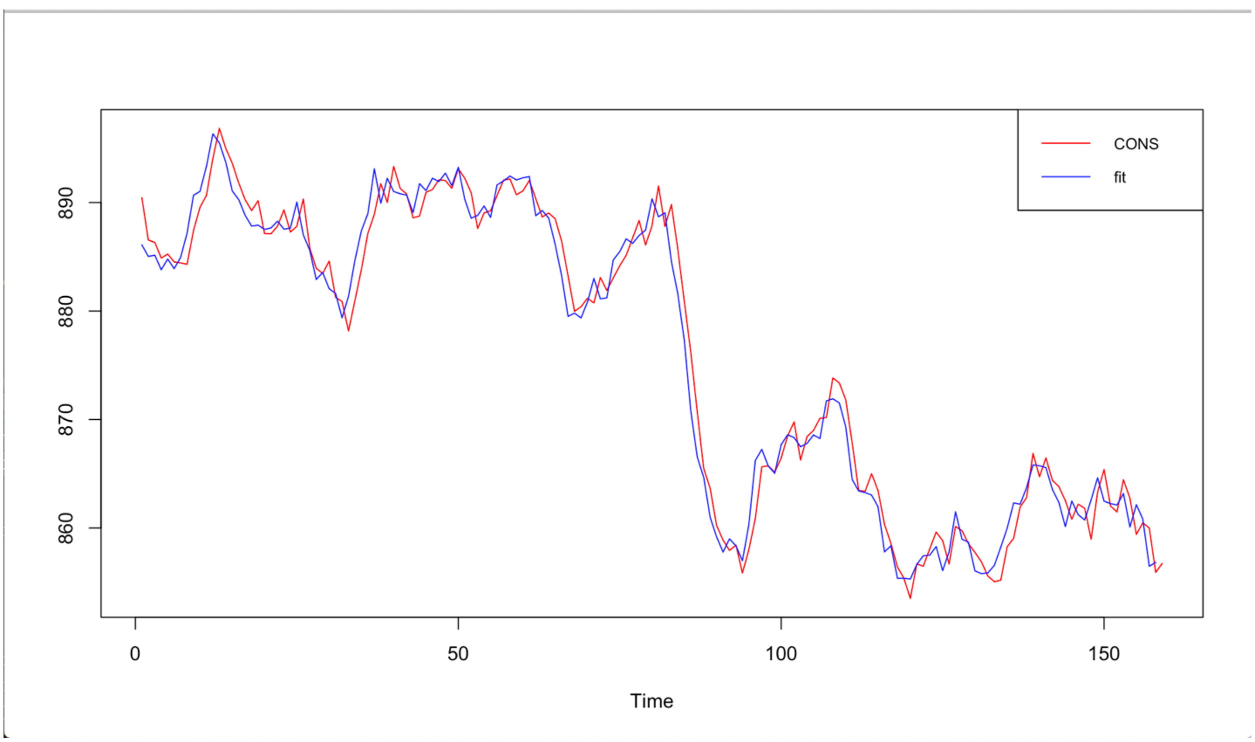
```
# c Model 3

n <- nrow(data)
model3 <- lm(CONS[-1]~CONS[-n]+INC[-1]+INC[-n]+INFLAT[-1]+INFLAT[-n])
summary(model3)
```

```
##
## Call:
## lm(formula = CONS[-1] ~ CONS[-n] + INC[-1] + INC[-n] + INFLAT[-1] +
##     INFLAT[-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 *
## CONS[-n]      0.79831    0.02716  29.393  < 2e-16 ***
## INC[-1]       0.49894    0.02833  17.611  < 2e-16 ***
## INC[-n]      -0.27611    0.03788  -7.290 1.59e-11 ***
## INFLAT[-1]   -0.79309    0.18395  -4.311 2.90e-05 ***
## INFLAT[-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
```

- It is not possible to draw direct conclusions regarding the significance of the regression coefficients based on the t-tests. However, the results give some general direction, and the results indicate that all regression coefficients would be statistically significant with the exception of INF_{t-1}
- 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 the income variable is positive. These coefficients describe the instant effects of changes in income and inflation
- The coefficient of the variable CONS with lag 1 is 0.79, which implies that the adjustment to changes in income and inflation is not too fast.
- Interpretations of the regression coefficients of income and inflation variables with lag 0:
    - If the income goes up by 1%, then the consumptions are increased by (without a lag) 0.49%.
    - If the inflations are increased by 1%, then the consumptions are reduced by 0.79%.
- Interpretations of the long term elasticities of income and inflation variables:
    - If the income goes up by 1%, then the consumptions are increased by 0.27% in the long term.
    - If the inflation are increased by 1%, then the consumptions are reduced by 0.25% in the long term.
- The coefficient of determination is 99.39%, but it’s impossible to draw any conclusions from it.

```r
# ,breaks=seq(from=-0.1,to=0.1,by=0.02
par(mfrow=c(3,2))
hist(model3$residuals,xlab="Residuals",ylab="Frequency",main=" ")

acf(model3$residuals,main="")

qqnorm(model3$residuals,pch=16)
qqline(model3$residuals)

plot(model3$residuals,type="p",ylab="Residuals",xlab="Time",pch=16,xaxt="n")
abline(0,0)

plot(model3$fitted.values, model3$residuals,type="p",ylab="Residuals", xlab = "Fitted values", pch=16)
abline(0,0)
plot(cooks.distance(model3),ylab="Cook's distances",xlab="Index",pch=16,xaxt="n")
axis(1)

dev.off()
```

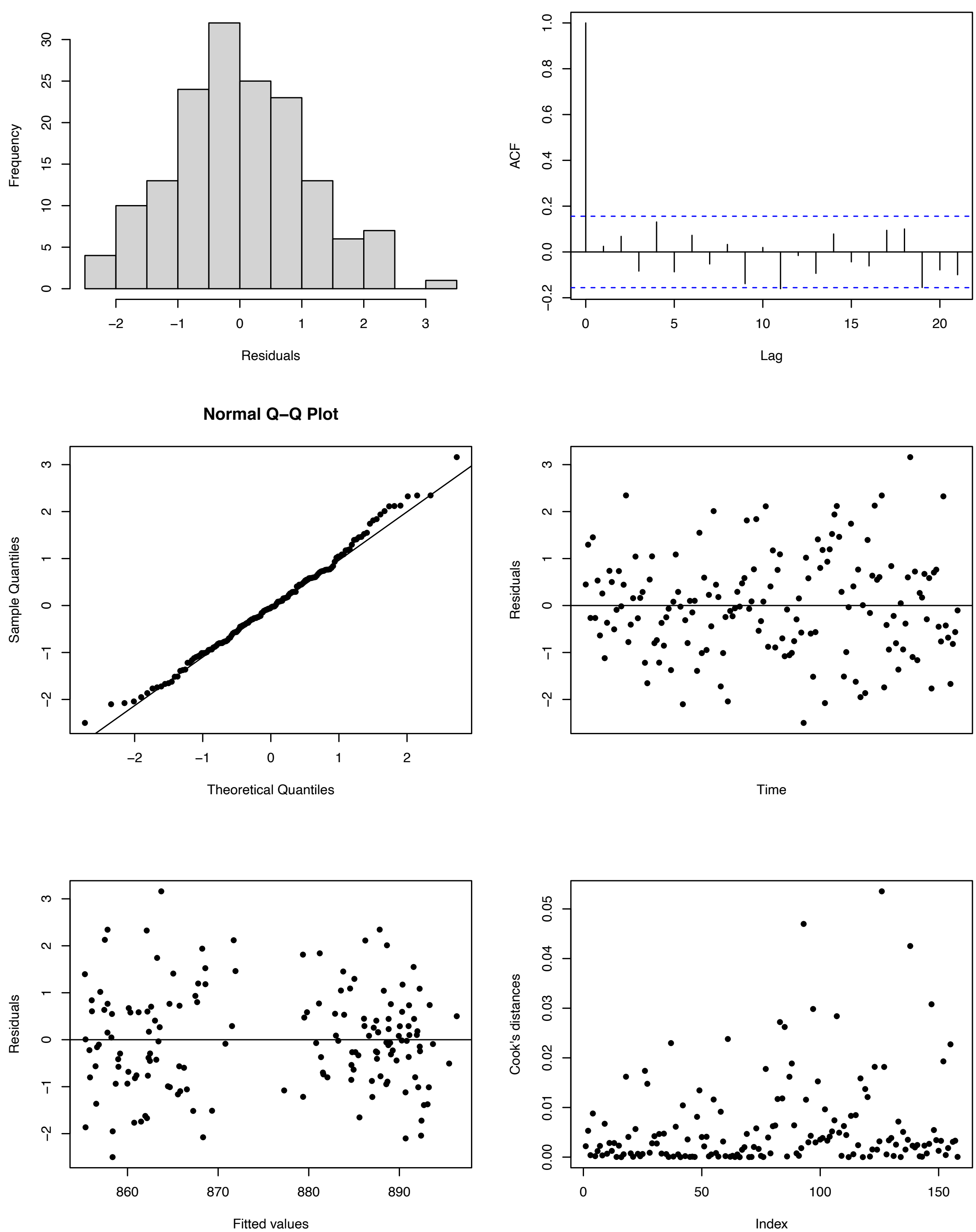
```r
plot(ts(CONS),col="red",xlab="Time",ylab="",type="l")
fit3 <- ts(predict(model3))
lines(fit3,col="blue")
legend("topright", legend=c("CONS", "fit"),
       col=c("red","blue"),lty=c(1,1),cex=0.8)#dev.off()
```



We can see that this fitted model coincides better with the original time series than the other two.

Based on the plots below:

- The residuals are normally distributed and not correlated
- By the residual diagrams, there is no evidence of heteroscedasticity.


Frequency
Residuals
ACF
Lag
Normal Q–Q Plot
Sample Quantiles
Theoretical Quantiles
Residuals
Time
Residuals
Fitted values
Cook's distances
Index

```
model3_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)
{
  model3_bg[i]= bgtest(model3, order=i)$p.value
}
model3_bg
```

```
##   [1] 0.7404511 0.6045556 0.5650183 0.2787527 0.3042654 0.3551158 0.4547950
##   [8] 0.5631049 0.4735877 0.5690298 0.4061289 0.4731610 0.5252836 0.5781539
##  [15] 0.6520422 0.5894026 0.3956821 0.4355298 0.2071538 0.2106953 0.1567503
##  [22] 0.1534830 0.1858849 0.2243955 0.2578920 0.2124698 0.2438560 0.2650905
##  [29] 0.3049527 0.3230335 0.3143937 0.3220023 0.3133261 0.3383738 0.3405532
##  [36] 0.3833834 0.2832329 0.3059693 0.3015777 0.3016666 0.2474447 0.1904450
##  [43] 0.1997935 0.2274582 0.2592239 0.2802037 0.3031284 0.3299429 0.3367617
##  [50] 0.3662879 0.3262469 0.3513529 0.3804923 0.3600569 0.3943330 0.4083301
##  [57] 0.4052707 0.3568331 0.3836353 0.4141415 0.4498262 0.2846277 0.3152150
##  [64] 0.2920044 0.3111106 0.2746370 0.2839722 0.3105616 0.3010305 0.3266045
##  [71] 0.3440329 0.3242977 0.3514303 0.3808195 0.4123068 0.4152464 0.4258572
##  [78] 0.3463123 0.3760194 0.3869832 0.4172575 0.3888215 0.4049522 0.4271859
##  [85] 0.4504424 0.4378654 0.4266472 0.4468663 0.4414324 0.4709179 0.4985224
##  [92] 0.5109369 0.5256059 0.5159122 0.4443017 0.4474901 0.4747258 0.3839566
##  [99] 0.3469387 0.3438840 0.3571655 0.3698975 0.3952581 0.4077615 0.4178025
## [106] 0.4302655 0.4564772 0.4684763 0.4728930 0.4867985 0.5076234 0.3344161
## [113] 0.3588214 0.2854976 0.3008617 0.2785145 0.2904988 0.1866446 0.2042750
## [120] 0.2159708 0.2037159 0.2217452 0.2394409 0.2586381 0.2714962 0.2351722
## [127] 0.2028571 0.2156090 0.2337641 0.2278207 0.2445561 0.2436024 0.2532834
## [134] 0.2700678 0.2904124 0.2889899 0.3083564 0.2653938 0.2709980 0.2460393
## [141] 0.2633544 0.2775652 0.2967588 0.2927641 0.3092860 0.3047361 0.3114799
## [148] 0.3285512 0.3216544 0.3286712 0.3403599 0.3527818
```

```
which(model3_bg < 0.05)
```

```
## integer(0)
```

With the Breusch-Godfrey test, the residuals are not correlated, and the null hypothesis is accepted for all lags with 5% level of significance.

```
vif(model3)
```

```
##   CONS[-n]    INC[-1]    INC[-n] INFLAT[-1] INFLAT[-n]
##  18.137266  12.433573  22.292466   7.552629   9.329719
```

There is multicollinearity in this model, but it's not unusual as the model has same variables with different lags.

⇨ We consider this model is sufficient to explain the total consumptions.

d) Model 3 (dynamic regression) is the most suitable one to explain the behavior of response variable CONS.