Project\_1

2024-06-16

# Library setup

library(tseries)

## Warning: package 'tseries' was built under R version 4.3.3

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

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(leaps)

## Warning: package 'leaps' was built under R version 4.3.3

# data inspection

fl <- read.csv('Data\_files\\Business\_data.csv')  
head(fl)

## DATE N\_employees N\_applications Inflation BUSINV Loan\_tight Loan\_Demand  
## 1 1/1/2006 135732.7 56229.17 2.504368 1330287 -10.7 16.1  
## 2 4/1/2006 136261.7 55211.54 2.827660 1362519 -12.3 3.5  
## 3 7/1/2006 136678.3 47388.57 3.172047 1389556 -8.9 -1.9  
## 4 10/1/2006 137056.0 45140.77 3.181040 1405920 0.0 -3.6  
## 5 1/1/2007 137604.3 54681.54 3.129683 1417512 0.0 -1.8  
## 6 4/1/2007 137969.0 53343.85 2.784459 1437327 -3.8 -22.6  
## Profit Sales Quarter Year  
## 1 880.676 1058566 First 2006  
## 2 902.575 1069808 Second 2006  
## 3 969.334 1074451 Third 2006  
## 4 905.291 1076956 Fourth 2006  
## 5 874.867 1095281 First 2007  
## 6 897.436 1122932 Second 2007

# Converting data to time series

N\_employees <- ts(fl$N\_employees,start=c(2006,1), end=c(2024,1), frequency=4)  
N\_applications <- ts(fl$N\_applications,start=c(2006,1), end=c(2024,1), frequency=4)  
Inflation <- ts(fl$Inflation,start=c(2006,1), end=c(2024,1), frequency=4)  
BUSINV <- ts(fl$BUSINV,start=c(2006,1), end=c(2024,1), frequency=4)  
Loan\_tight <- ts(fl$Loan\_tight,start=c(2006,1), end=c(2024,1), frequency=4)  
Loan\_demand <- ts(fl$Loan\_Demand,start=c(2006,1), end=c(2024,1), frequency=4)  
Profit <- ts(fl$Profit,start=c(2006,1), end=c(2024,1), frequency=4)  
Sales <- ts(fl$Sales,start=c(2006,1), end=c(2024,1), frequency=4)

# Stationary Test(ADF)

adf.test(N\_employees)

##   
## Augmented Dickey-Fuller Test  
##   
## data: N\_employees  
## Dickey-Fuller = -2.5078, Lag order = 4, p-value = 0.3685  
## alternative hypothesis: stationary

adf.test(N\_applications)

##   
## Augmented Dickey-Fuller Test  
##   
## data: N\_applications  
## Dickey-Fuller = -1.8567, Lag order = 4, p-value = 0.6338  
## alternative hypothesis: stationary

adf.test(Inflation)

##   
## Augmented Dickey-Fuller Test  
##   
## data: Inflation  
## Dickey-Fuller = -0.83189, Lag order = 4, p-value = 0.9545  
## alternative hypothesis: stationary

adf.test(BUSINV)

##   
## Augmented Dickey-Fuller Test  
##   
## data: BUSINV  
## Dickey-Fuller = -2.3643, Lag order = 4, p-value = 0.4269  
## alternative hypothesis: stationary

adf.test(Loan\_tight)

##   
## Augmented Dickey-Fuller Test  
##   
## data: Loan\_tight  
## Dickey-Fuller = -2.8991, Lag order = 4, p-value = 0.2091  
## alternative hypothesis: stationary

adf.test(Loan\_demand)

##   
## Augmented Dickey-Fuller Test  
##   
## data: Loan\_demand  
## Dickey-Fuller = -3.1794, Lag order = 4, p-value = 0.09797  
## alternative hypothesis: stationary

adf.test(Profit)

##   
## Augmented Dickey-Fuller Test  
##   
## data: Profit  
## Dickey-Fuller = -1.8556, Lag order = 4, p-value = 0.6342  
## alternative hypothesis: stationary

adf.test(Sales)

##   
## Augmented Dickey-Fuller Test  
##   
## data: Sales  
## Dickey-Fuller = -2.2142, Lag order = 4, p-value = 0.4881  
## alternative hypothesis: stationary

All columns of data are non-stationary, therefore we need to differentiate our data.

# Data Differentiation

## Making it stationary

diff\_emp <- N\_employees %>% diff()%>% diff()  
diff\_app <- N\_applications %>% diff()%>% diff()  
diff\_Infl <- Inflation %>% diff()  
diff\_BUSI <- BUSINV %>% diff()  
diff\_Ltig <- Loan\_tight %>% diff()  
diff\_Ldem <- Loan\_demand %>% diff()  
diff\_Profit <- Profit %>% diff()%>% diff()  
diff\_Sales <- Sales %>% diff()%>% diff()  
  
adf.test(diff\_emp)

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

##   
## Augmented Dickey-Fuller Test  
##   
## data: diff\_emp  
## Dickey-Fuller = -5.8692, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary

adf.test(diff\_app)

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

##   
## Augmented Dickey-Fuller Test  
##   
## data: diff\_app  
## Dickey-Fuller = -6.441, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary

adf.test(diff\_Infl)

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

##   
## Augmented Dickey-Fuller Test  
##   
## data: diff\_Infl  
## Dickey-Fuller = -4.6087, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary

adf.test(diff\_BUSI)

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

##   
## Augmented Dickey-Fuller Test  
##   
## data: diff\_BUSI  
## Dickey-Fuller = -4.1289, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary

adf.test(diff\_Ltig)

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

##   
## Augmented Dickey-Fuller Test  
##   
## data: diff\_Ltig  
## Dickey-Fuller = -4.1349, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary

adf.test(diff\_Ldem)

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

##   
## Augmented Dickey-Fuller Test  
##   
## data: diff\_Ldem  
## Dickey-Fuller = -5.0973, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary

adf.test(diff\_Profit)

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

##   
## Augmented Dickey-Fuller Test  
##   
## data: diff\_Profit  
## Dickey-Fuller = -5.3632, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary

adf.test(diff\_Sales)

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

##   
## Augmented Dickey-Fuller Test  
##   
## data: diff\_Sales  
## Dickey-Fuller = -5.401, Lag order = 4, p-value = 0.01  
## alternative hypothesis: stationary

# Raw Data Model

diff\_Infl <- diff\_Infl[-1]  
diff\_BUSI <- diff\_BUSI[-1]  
diff\_Ltig <- diff\_Ltig[-1]  
diff\_Ldem <- diff\_Ldem[-1]  
  
diff\_data <- data.frame(diff\_emp,diff\_app,diff\_Infl,diff\_BUSI,diff\_Ltig,diff\_Ldem,diff\_Sales)  
lm\_diff <- lm(diff\_Profit ~ . , data = diff\_data)  
summary(lm\_diff)

##   
## Call:  
## lm(formula = diff\_Profit ~ ., data = diff\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -234.53 -60.77 -11.06 70.17 325.28   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.2234535 17.6080562 0.524 0.6022   
## diff\_emp -0.0148291 0.0065782 -2.254 0.0277 \*   
## diff\_app 0.0020714 0.0008843 2.342 0.0223 \*   
## diff\_Infl 80.0242332 58.4831272 1.368 0.1761   
## diff\_BUSI -0.0006958 0.0006369 -1.092 0.2788   
## diff\_Ltig 0.9239153 1.1626523 0.795 0.4298   
## diff\_Ldem 0.3655919 0.9978832 0.366 0.7153   
## diff\_Sales 0.0030057 0.0004833 6.219 4.49e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 122.8 on 63 degrees of freedom  
## Multiple R-squared: 0.6097, Adjusted R-squared: 0.5663   
## F-statistic: 14.06 on 7 and 63 DF, p-value: 7.881e-11

# Finding Best Model

results <- leaps(diff\_data,diff\_Profit,int=TRUE,method=c("Cp", "adjr2", "r2"),nbest=5,names=NULL, df=NROW(diff\_data))  
results

## $which  
## 1 2 3 4 5 6 7  
## 1 FALSE FALSE FALSE FALSE FALSE FALSE TRUE  
## 1 FALSE TRUE FALSE FALSE FALSE FALSE FALSE  
## 1 TRUE FALSE FALSE FALSE FALSE FALSE FALSE  
## 1 FALSE FALSE FALSE FALSE TRUE FALSE FALSE  
## 1 FALSE FALSE TRUE FALSE FALSE FALSE FALSE  
## 2 FALSE TRUE FALSE FALSE FALSE FALSE TRUE  
## 2 TRUE FALSE FALSE FALSE FALSE FALSE TRUE  
## 2 FALSE FALSE FALSE FALSE TRUE FALSE TRUE  
## 2 FALSE FALSE FALSE FALSE FALSE TRUE TRUE  
## 2 FALSE FALSE TRUE FALSE FALSE FALSE TRUE  
## 3 TRUE TRUE FALSE FALSE FALSE FALSE TRUE  
## 3 FALSE TRUE FALSE FALSE TRUE FALSE TRUE  
## 3 TRUE FALSE FALSE FALSE TRUE FALSE TRUE  
## 3 FALSE TRUE FALSE TRUE FALSE FALSE TRUE  
## 3 FALSE TRUE TRUE FALSE FALSE FALSE TRUE  
## 4 TRUE TRUE FALSE FALSE TRUE FALSE TRUE  
## 4 TRUE TRUE TRUE FALSE FALSE FALSE TRUE  
## 4 TRUE TRUE FALSE TRUE FALSE FALSE TRUE  
## 4 TRUE TRUE FALSE FALSE FALSE TRUE TRUE  
## 4 FALSE TRUE TRUE TRUE FALSE FALSE TRUE  
## 5 TRUE TRUE TRUE TRUE FALSE FALSE TRUE  
## 5 TRUE TRUE TRUE FALSE TRUE FALSE TRUE  
## 5 TRUE TRUE FALSE TRUE TRUE FALSE TRUE  
## 5 TRUE TRUE FALSE FALSE TRUE TRUE TRUE  
## 5 TRUE TRUE TRUE FALSE FALSE TRUE TRUE  
## 6 TRUE TRUE TRUE TRUE TRUE FALSE TRUE  
## 6 TRUE TRUE TRUE TRUE FALSE TRUE TRUE  
## 6 TRUE TRUE TRUE FALSE TRUE TRUE TRUE  
## 6 TRUE TRUE FALSE TRUE TRUE TRUE TRUE  
## 6 FALSE TRUE TRUE TRUE TRUE TRUE TRUE  
## 7 TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
##   
## $label  
## [1] "(Intercept)" "1" "2" "3" "4"   
## [6] "5" "6" "7"   
##   
## $size  
## [1] 2 2 2 2 2 3 3 3 3 3 4 4 4 4 4 5 5 5 5 5 6 6 6 6 6 7 7 7 7 7 8  
##   
## $Cp  
## [1] 12.743224 57.326982 63.318503 90.571069 92.165041 7.195719 8.807025  
## [8] 12.288307 13.885919 14.090863 3.584919 7.572609 8.398470 8.518280  
## [15] 8.765768 3.982304 4.243639 5.514644 5.584749 8.221494 4.720105  
## [22] 5.212643 5.923496 5.951289 6.243637 6.134225 6.631487 7.193328  
## [29] 7.872328 11.081772 8.000000

colnames(diff\_data)

## [1] "diff\_emp" "diff\_app" "diff\_Infl" "diff\_BUSI" "diff\_Ltig"   
## [6] "diff\_Ldem" "diff\_Sales"

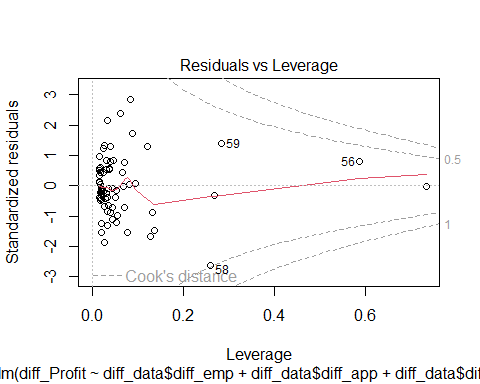
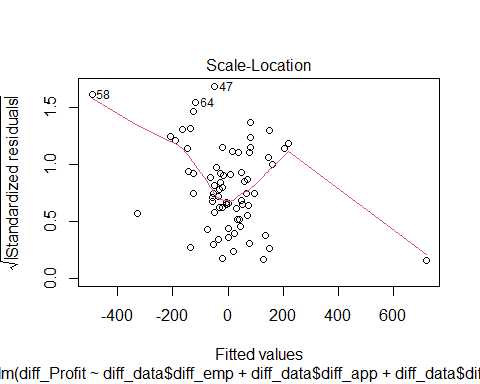
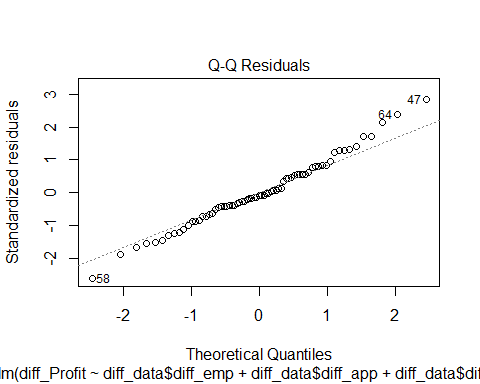
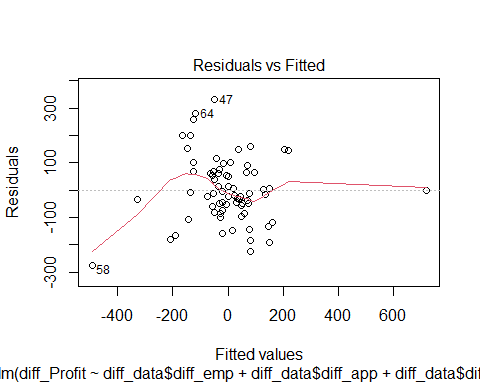
# Best model

lm\_1 <- lm(diff\_Profit ~ diff\_data$diff\_emp+diff\_data$diff\_app+diff\_data$diff\_Ltig+diff\_data$diff\_Sales)  
summary(lm\_1)

##   
## Call:  
## lm(formula = diff\_Profit ~ diff\_data$diff\_emp + diff\_data$diff\_app +   
## diff\_data$diff\_Ltig + diff\_data$diff\_Sales)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -273.82 -67.93 -11.37 66.25 330.62   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.7539633 14.4650079 -0.052 0.9586   
## diff\_data$diff\_emp -0.0149178 0.0062606 -2.383 0.0201 \*   
## diff\_data$diff\_app 0.0021612 0.0008466 2.553 0.0130 \*   
## diff\_data$diff\_Ltig 1.3883907 1.0882352 1.276 0.2065   
## diff\_data$diff\_Sales 0.0030387 0.0004647 6.539 1.06e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 121.8 on 66 degrees of freedom  
## Multiple R-squared: 0.5974, Adjusted R-squared: 0.573   
## F-statistic: 24.48 on 4 and 66 DF, p-value: 1.893e-12

# Model Analyses

plot(lm\_1)



We have some problem points which are point 58, 64 and 47. Point 58 was derived from the 2020 pandemic, therefore this point is an outlier and shouldn’t be considered in our model.