

sold_units_ml

2/21/2022

Analysis of the factors related with the number of units sold per year

```
#Importing the packages
```

```
library(readr)
```

```
library(car)
```

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

```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-3
```

```
library(leaps)
```

Importing the data

```
file_path<-"../raw/sold_units_final.csv"
```

```
sold_units<-read_csv(file_path)
```

```
##
## -- Column specification -----
## cols(
##   Año = col_double(),
##   'Unidades Vendidas' = col_double(),
##   'ITCRB Estados Unidos Promedio' = col_double(),
##   'Restriccion de importaciones' = col_double(),
##   'Crisis Semiconductores' = col_double(),
##   'Brecha Cambiaria' = col_double(),
##   'Reservas Internacionales' = col_double(),
##   'PIB (Millones de US$ a precios actuales)' = col_double(),
##   'PIB/reservas' = col_double(),
##   'Devaluacion Interanual' = col_double(),
##   Inflacion = col_double()
## )
```

```

#Dropping the year column, inflation and devaluación.
sold_units<-sold_units[,2:9]

#Centering the variables to reduce structural multicollinearity
sold_units[,6]<-scale(sold_units[,6],scale=FALSE)
sold_units[,7]<-scale(sold_units[,7],scale=FALSE)
sold_units[,8]<-scale(sold_units[,8],scale=FALSE)

#Renaming the columns
my_names<-c("num_units","itcrb","import_restriction","semiconductor_crisis",
            "brecha_cambiaria", "reserves", "PIB", "PIB_over_reserves")
names(sold_units)<-my_names

```

Building the model

```

sold_units_selected<-lm(sold_units)
summary(sold_units_selected)

##
## Call:
## lm(formula = sold_units)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -90321 -33475   5143  29988  89886
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.817e+06  2.244e+05   8.096 3.33e-06 ***
## itcrb          -1.035e+04  1.833e+03  -5.648 0.000108 ***
## import_restriction -2.416e+05  1.130e+05  -2.138 0.053764 .
## semiconductor_crisis -2.884e+05  9.195e+04  -3.136 0.008592 **
## brecha_cambiaria    4.615e+05  2.466e+05   1.871 0.085877 .
## reserves        -3.415e+01  5.531e+00  -6.174 4.77e-05 ***
## PIB              2.683e+00  5.903e-01   4.546 0.000671 ***
## PIB_over_reserves  -1.312e+05  2.039e+04  -6.434 3.23e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59680 on 12 degrees of freedom
## Multiple R-squared:  0.9623, Adjusted R-squared:  0.9404
## F-statistic: 43.81 on 7 and 12 DF,  p-value: 1.309e-07

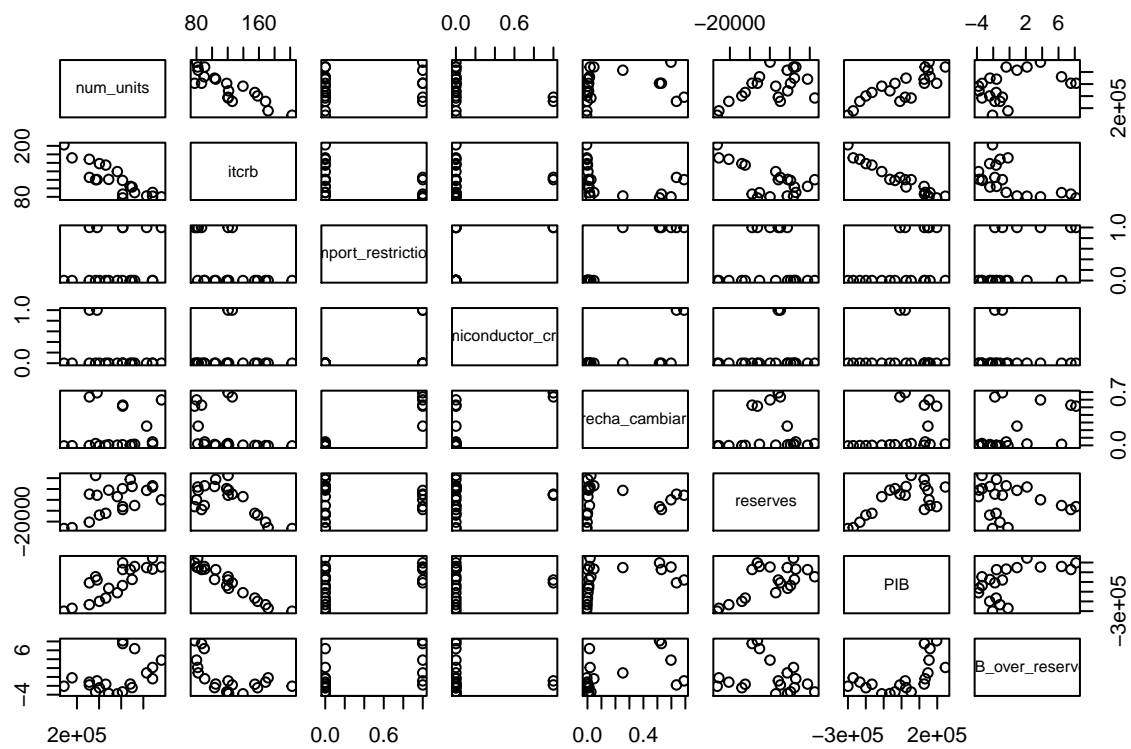
```

Pairwise plots of the features

```

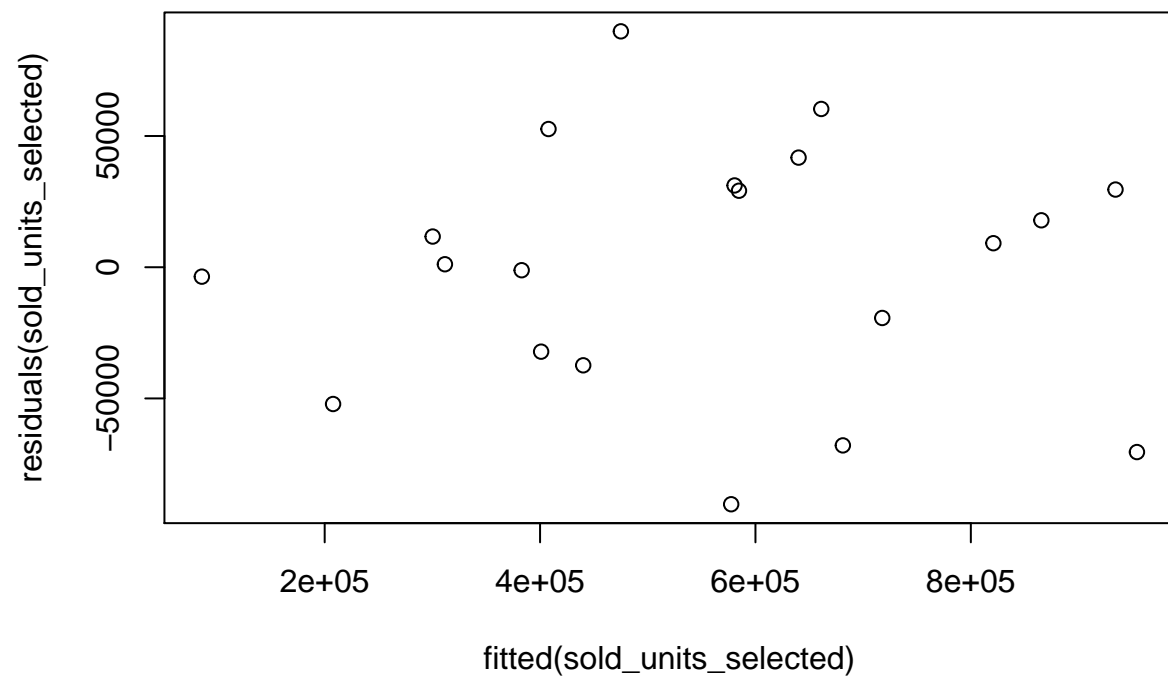
pairs(sold_units)

```

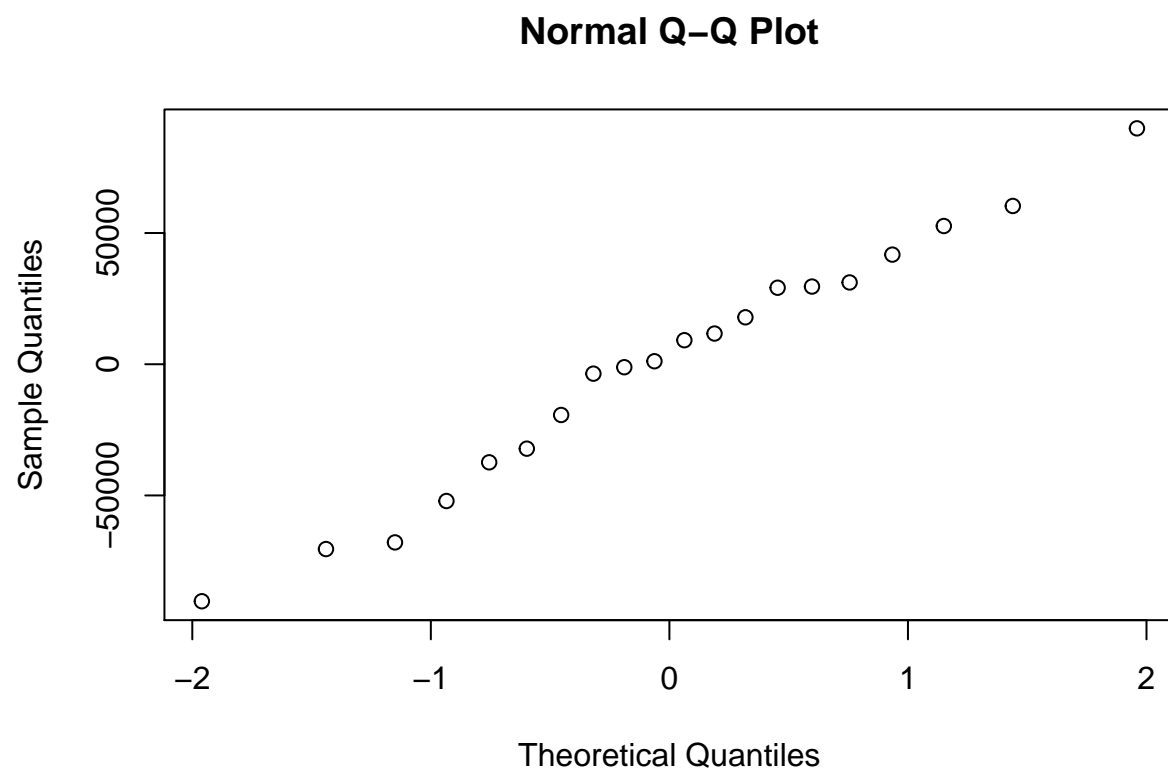


Analyzing the residuals

```
plot(fitted(sold_units_selected),residuals(sold_units_selected))
```

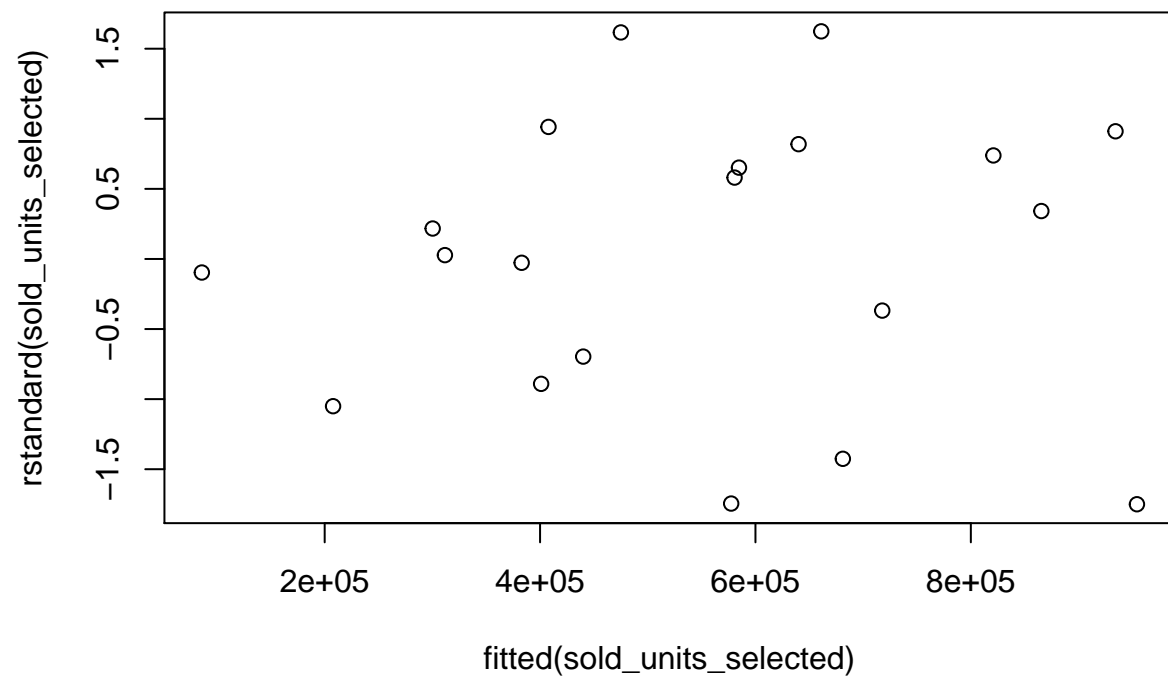


```
qqnorm(residuals(sold_units_selected))
```

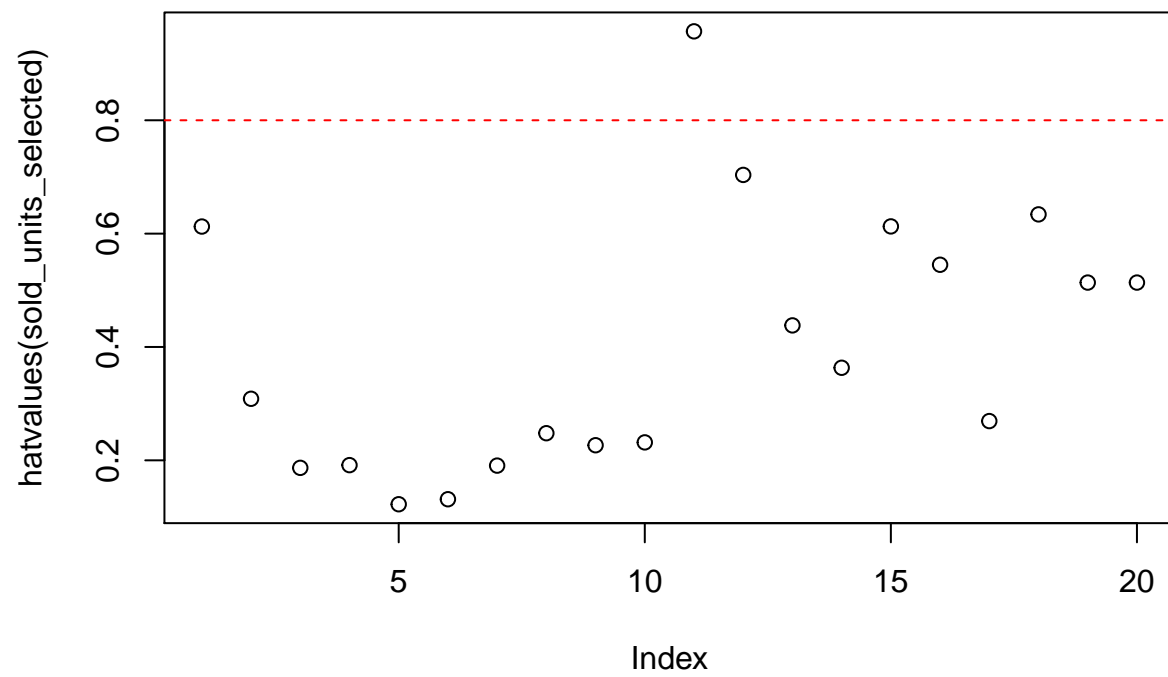


Looking for outliers and high leverage points

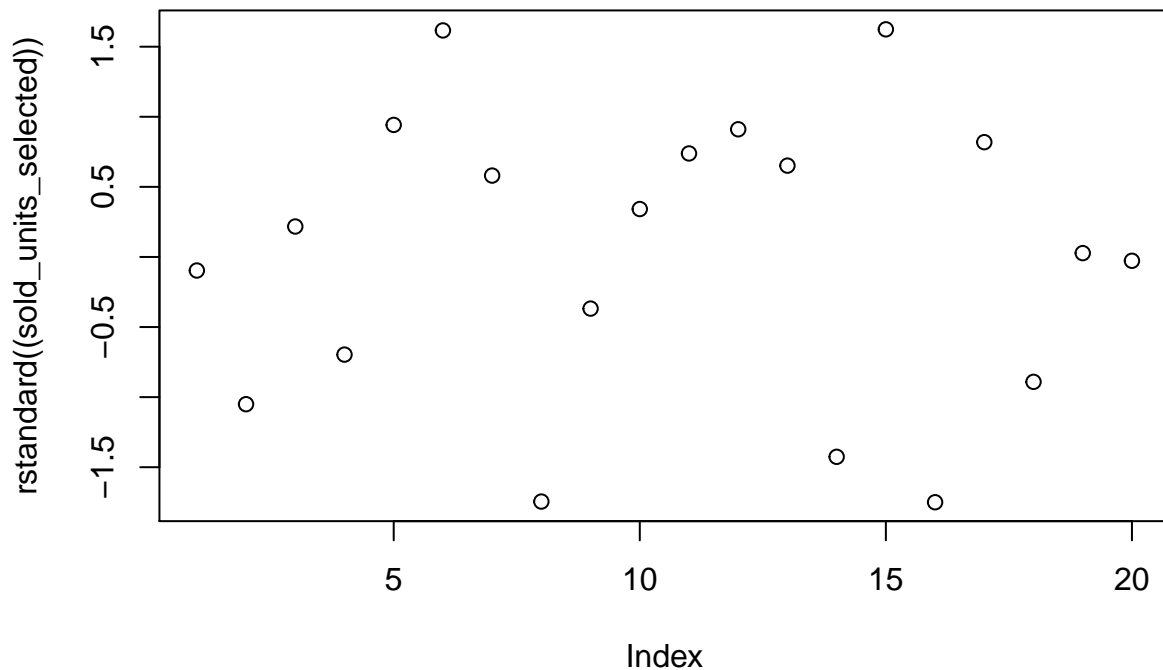
```
plot(fitted(sold_units_selected),rstandard(sold_units_selected))
```



```
plot(hatvalues(sold_units_selected))  
abline(h=length(coef(sold_units_selected))/nrow(sold_units)*2,  
       col = "red",lty = 2)
```



```
plot(rstandard((sold_units_selected)))
```



Looking for colinearity Correlation matrix

```
cor(sold_units[, -1])
```

```
##               itcrb import_restriction semiconductor_crisis
## itcrb          1.00000000      -0.45615253      0.03102529
## import_restriction -0.45615253      1.00000000      0.50917508
## semiconductor_crisis 0.03102529      0.50917508      1.00000000
## brecha_cambiaria    -0.40454741      0.95207008      0.65090261
## reserves            -0.64341540      0.02503357      0.11349808
## PIB                 -0.97038025      0.42912174      0.01189703
## PIB_over_reserves   -0.57391965      0.53395398     -0.12356516
##               brecha_cambiaria reserves      PIB PIB_over_reserves
## itcrb          -0.40454741 -0.64341540 -0.97038025     -0.5739197
## import_restriction 0.95207008 0.02503357 0.42912174      0.5339540
## semiconductor_crisis 0.65090261 0.11349808 0.01189703     -0.1235652
## brecha_cambiaria    1.00000000 0.03724469 0.39340556      0.4822820
## reserves            0.03724469 1.00000000 0.65862991     -0.2101472
## PIB                 0.39340556 0.65862991 1.00000000      0.5813857
## PIB_over_reserves   0.48228198 -0.21014720 0.58138567      1.0000000
```

Variance inflation factors

```
vif(sold_units_selected)
```

```
##               itcrb  import_restriction semiconductor_crisis
```



```
##          23.242185          15.049932          4.273157
## brecha_cambiaria          reserves          PIB
##          22.091141          33.508869          51.541641
## PIB_over_reserves
##          30.529316
```

Eigenvalues of the correlation matrix

```
eigen(cor(sold_units[,-1]))
```

```
## eigen() decomposition
## $values
## [1] 3.529110025 1.866345590 1.306811850 0.228993645 0.031741984 0.027613064
## [7] 0.009383842
##
## $vectors
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,]  0.4572047 -0.36047310 -0.025077008  0.0007034938  0.41810586  0.696628890
## [2,] -0.4352454 -0.36576393  0.034867055  0.5500205953 -0.46305960  0.378068571
## [3,] -0.1868202 -0.47317764 -0.528121845 -0.6364679031 -0.22893922 -0.001509614
## [4,] -0.4269210 -0.41634948 -0.056986209  0.2240359298  0.66397305 -0.340543219
## [5,] -0.2247470  0.45641815 -0.567917927  0.1494362930  0.29069066  0.201333815
## [6,] -0.4565077  0.36102603  0.003989135 -0.2287408918  0.05575892  0.437617276
## [7,] -0.3541913 -0.01935454  0.627254841 -0.4093162092  0.17285778  0.154130451
##
##          [,7]
## [1,] -0.01700982
## [2,]  0.12521901
## [3,]  0.06914914
## [4,] -0.18472178
## [5,]  0.52083866
## [6,] -0.64366503
## [7,]  0.50946005
```

Feature selection

Applying best subset selection

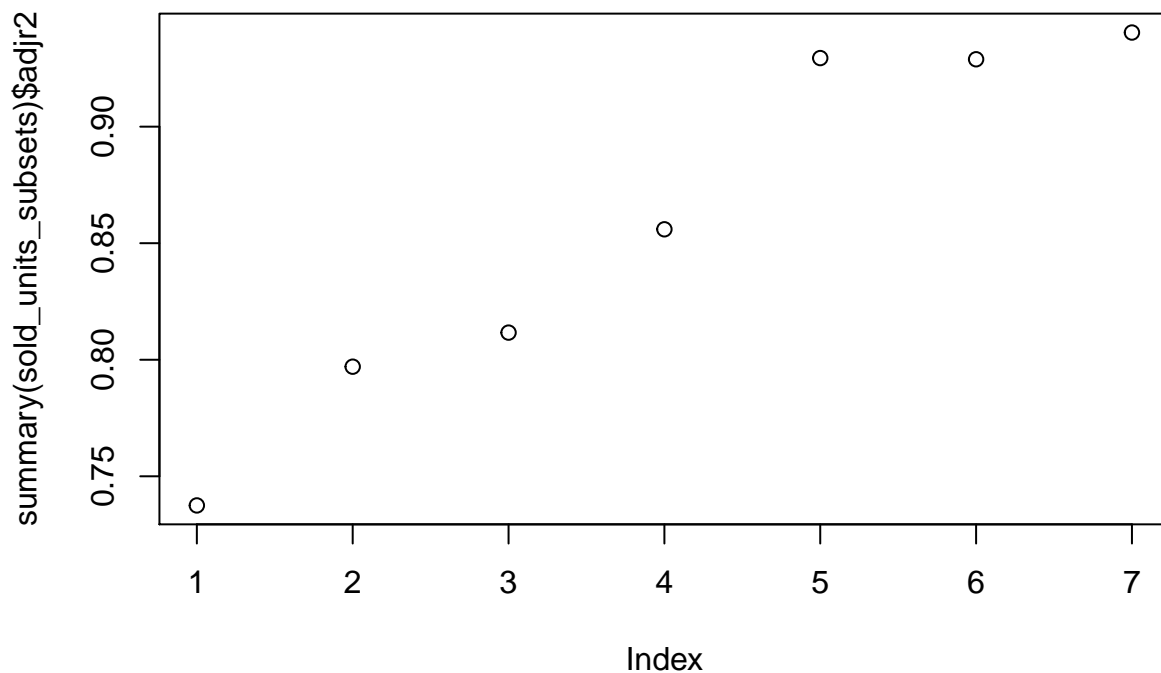
```
sold_units_subsets<-regsubsets(sold_units$num_units~.,sold_units,nvmax = 10)
summary(sold_units_subsets)
```

```
## Subset selection object
## Call: regsubsets.formula(sold_units$num_units ~ ., sold_units, nvmax = 10)
## 7 Variables (and intercept)
##
##          Forced in Forced out
## itcrb          FALSE      FALSE
## import_restriction  FALSE      FALSE
## semiconductor_crisis  FALSE      FALSE
## brecha_cambiaria     FALSE      FALSE
## reserves             FALSE      FALSE
## PIB                 FALSE      FALSE
## PIB_over_reserves    FALSE      FALSE
```

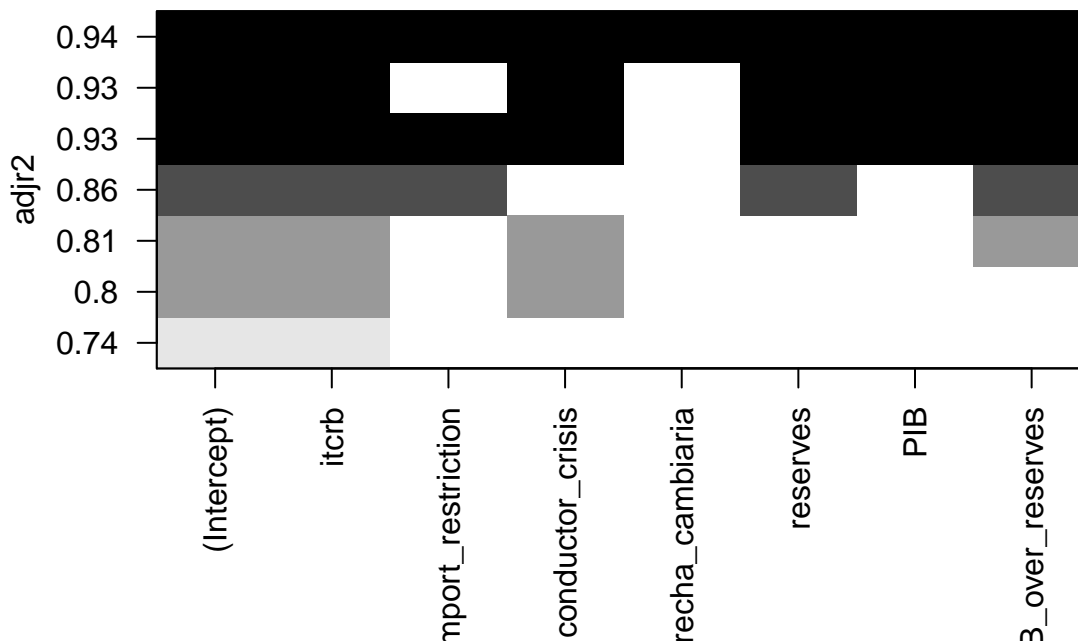
```
## 1 subsets of each size up to 7
## Selection Algorithm: exhaustive
##           itcrb import_restriction semiconductor_crisis brecha_cambiaria
## 1 ( 1 ) "*" " " " " " "
## 2 ( 1 ) "*" " " "*" " "
## 3 ( 1 ) "*" " " "*" " "
## 4 ( 1 ) "*" "*" " " " "
## 5 ( 1 ) "*" " " "*" " "
## 6 ( 1 ) "*" "*" "*" " "
## 7 ( 1 ) "*" "*" "*" "*"
##           reserves PIB PIB_over_reserves
## 1 ( 1 ) " " " " " "
## 2 ( 1 ) " " " " " "
## 3 ( 1 ) " " " " "*"
## 4 ( 1 ) "*" " " "*"
## 5 ( 1 ) "*" "*" "*"
## 6 ( 1 ) "*" "*" "*"
## 7 ( 1 ) "*" "*" "*"

```

```
plot(summary(sold_units_subsets)$adjr2)
```



```
plot (sold_units_subsets, scale = "adjr2")
```



```
which.max(summary(sold_units_subsets)$adjr2)
```

```
## [1] 7
```

```
coef(sold_units_subsets, which.max(summary(sold_units_subsets)$adjr2))
```

```
##      (Intercept)      itcrb  import_restriction
##      1816868.49156      -10353.09949      -241555.11874
## semiconductor_crisis brecha_cambiaría      reserves
##      -288384.74282      461498.42292      -34.14516
##      PIB      PIB_over_reserves
##      2.68351      -131198.09571
```

Building the selected model

```
sold_units_selected<-
  lm(sold_units[,names(sold_units)%in%
    c("num_units",names(coef(sold_units_subsets,
      which.max(summary(sold_units_subsets)$adjr2)))))]
summary(sold_units_selected)
```

```
##
## Call:
## lm(formula = sold_units[, names(sold_units) %in% c("num_units",
```

```
##      names(coef(sold_units_subsets, which.max(summary(sold_units_subsets)$adjr2))))))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -90321 -33475   5143  29988  89886
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.817e+06  2.244e+05   8.096 3.33e-06 ***
## itcrb          -1.035e+04  1.833e+03  -5.648 0.000108 ***
## import_restriction -2.416e+05  1.130e+05  -2.138 0.053764 .
## semiconductor_crisis -2.884e+05  9.195e+04  -3.136 0.008592 **
## brecha_cambiaria    4.615e+05  2.466e+05   1.871 0.085877 .
## reserves         -3.415e+01  5.531e+00  -6.174 4.77e-05 ***
## PIB              2.683e+00  5.903e-01   4.546 0.000671 ***
## PIB_over_reserves  -1.312e+05  2.039e+04  -6.434 3.23e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59680 on 12 degrees of freedom
## Multiple R-squared:  0.9623, Adjusted R-squared:  0.9404
## F-statistic: 43.81 on 7 and 12 DF,  p-value: 1.309e-07
```

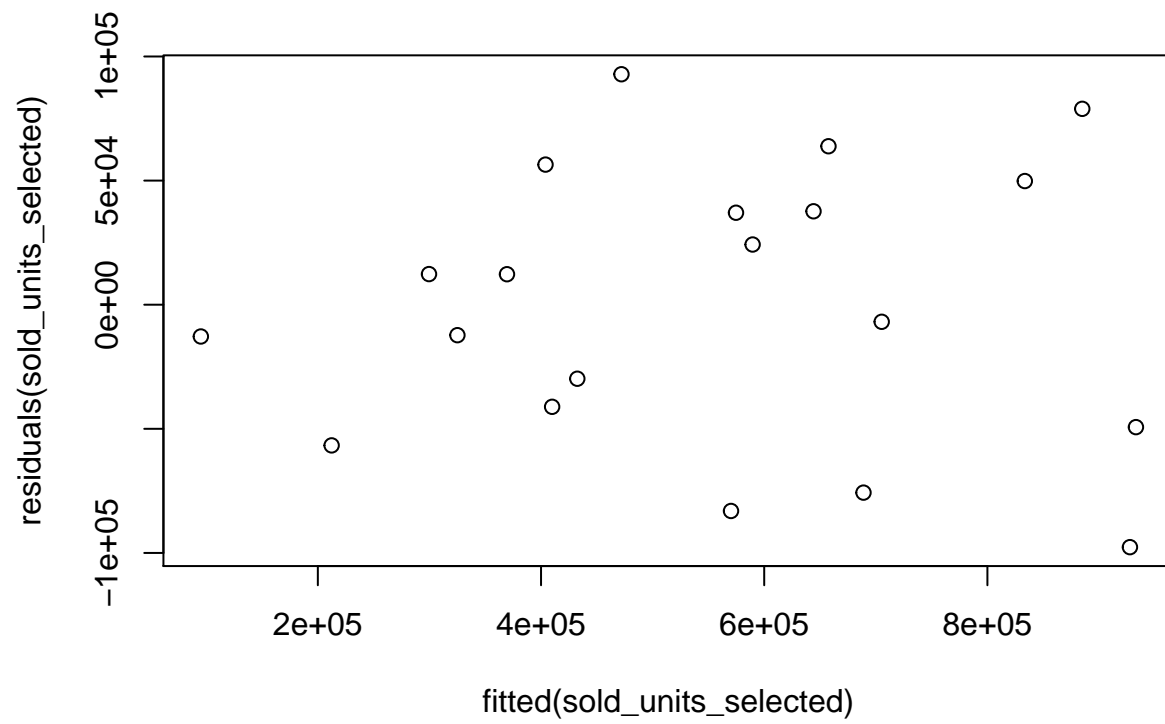
Building the model with 5 predictors

```
sold_units_selected<-
  lm(sold_units[,names(sold_units)%in%
      c("num_units",names(coef(sold_units_subsets,5)))]
      summary(sold_units_selected)
```

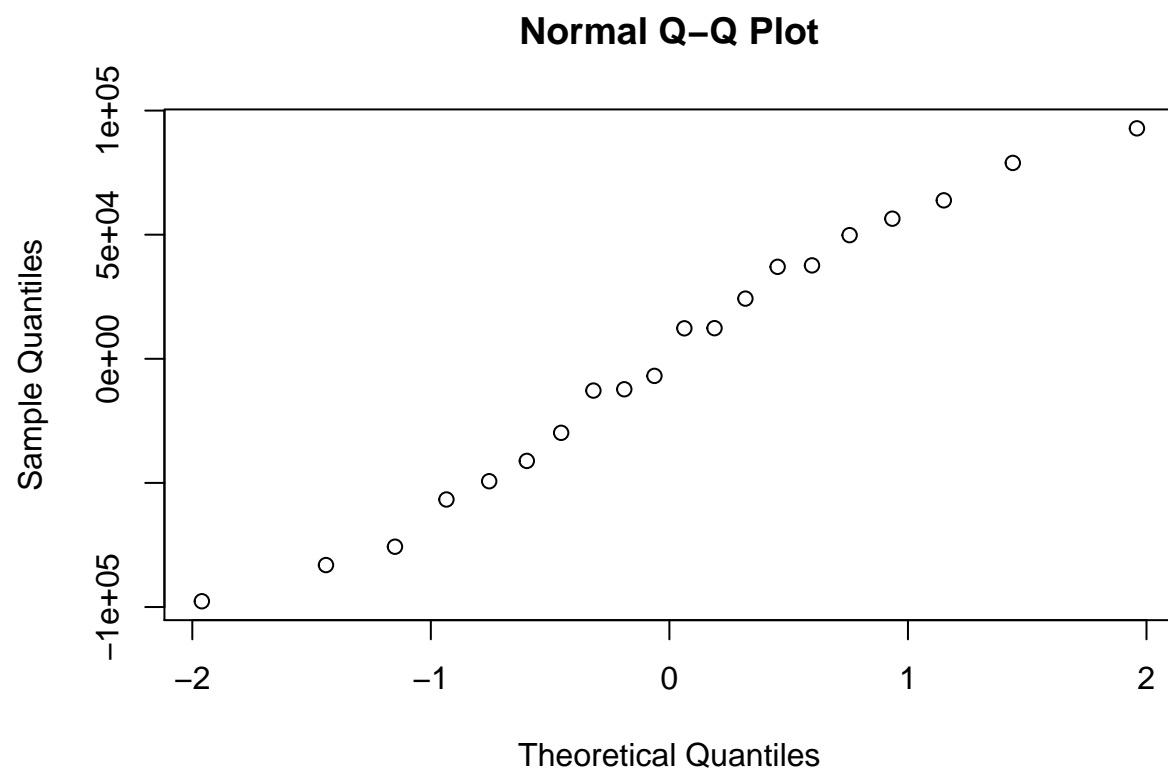
```
##
## Call:
## lm(formula = sold_units[, names(sold_units) %in% c("num_units",
##      names(coef(sold_units_subsets, 5)))]
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -97670 -43182   2699  40677  92839
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.730e+06  2.173e+05   7.962 1.45e-06 ***
## itcrb          -9.645e+03  1.815e+03  -5.315 0.000109 ***
## semiconductor_crisis -2.228e+05  4.977e+04  -4.476 0.000523 ***
## reserves         -3.113e+01  5.814e+00  -5.353 0.000102 ***
## PIB              2.510e+00  6.151e-01   4.081 0.001122 **
## PIB_over_reserves  -1.203e+05  2.081e+04  -5.780 4.77e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 64930 on 14 degrees of freedom
## Multiple R-squared:  0.948, Adjusted R-squared:  0.9294
## F-statistic: 51.04 on 5 and 14 DF,  p-value: 1.708e-08
```

Analyzing the residuals

```
plot(fitted(sold_units_selected),residuals(sold_units_selected))
```

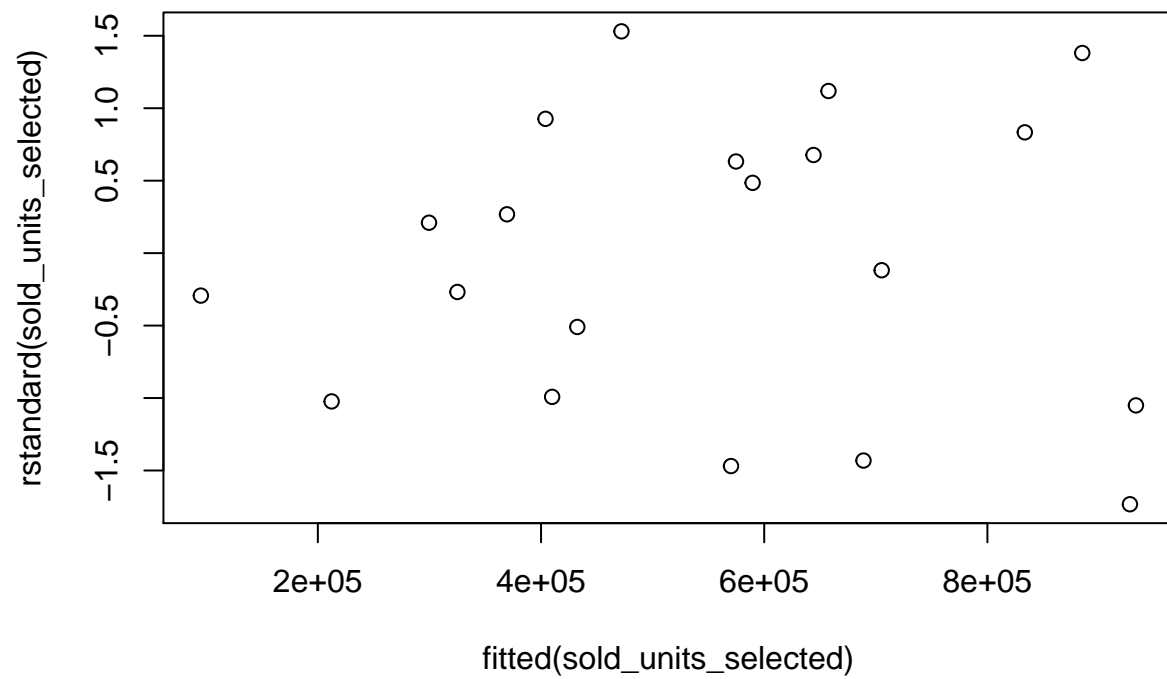


```
qqnorm(residuals(sold_units_selected))
```

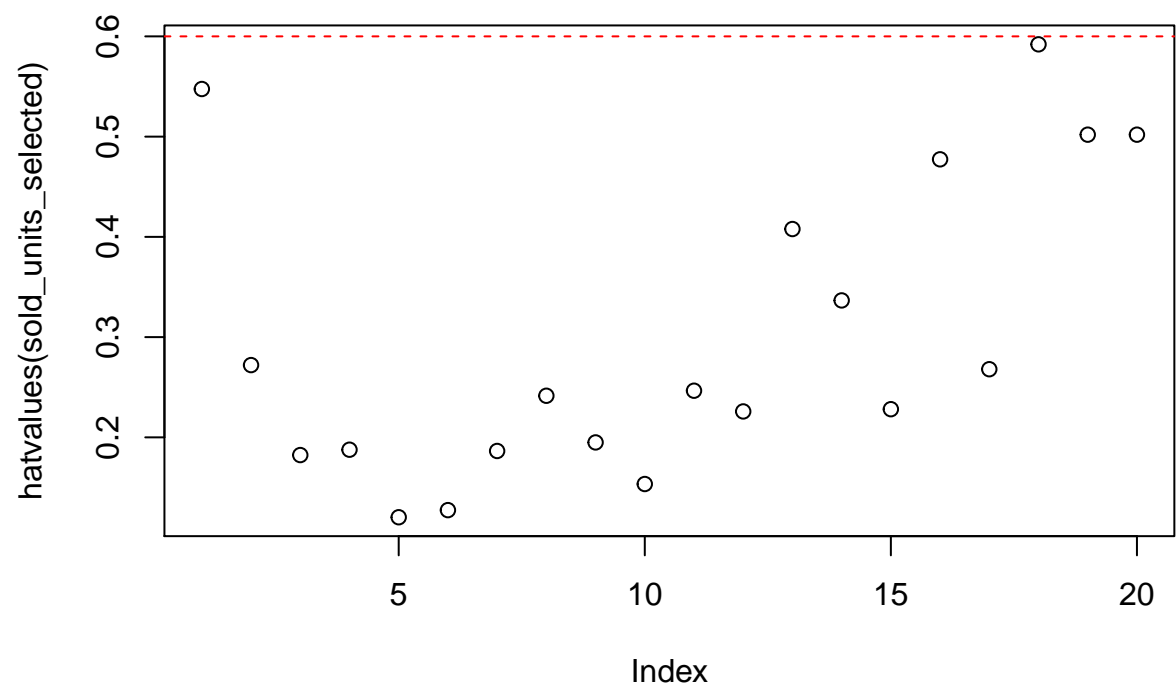


Looking for outliers and high leverage points

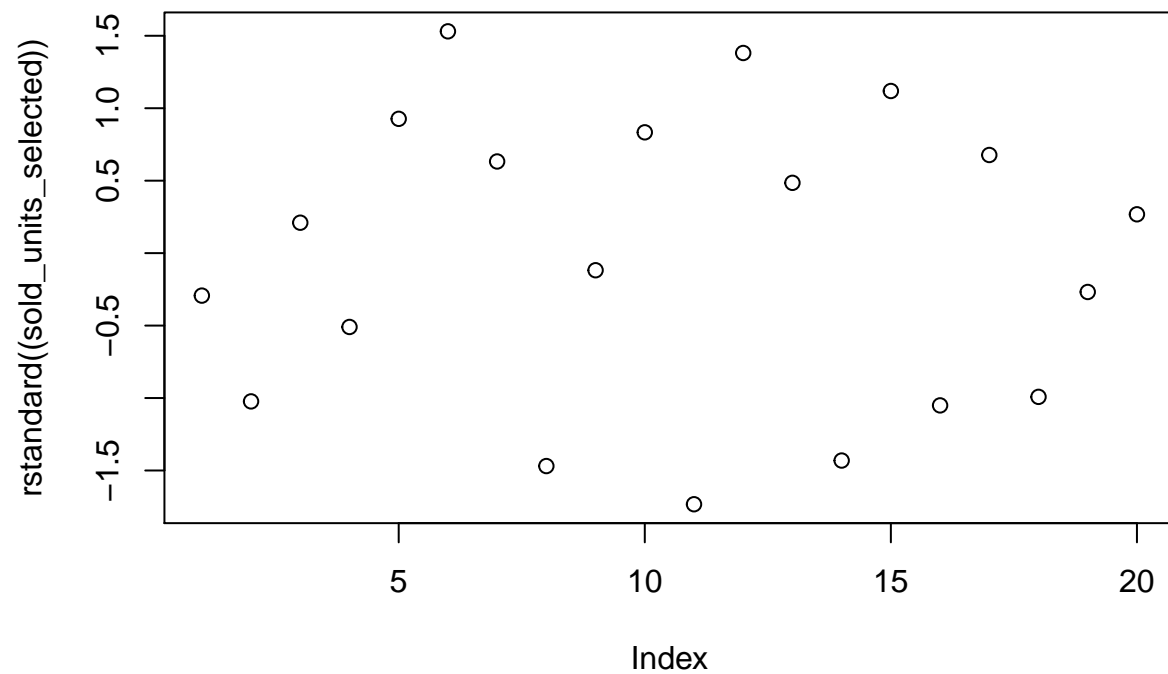
```
plot(fitted(sold_units_selected),rstandard(sold_units_selected))
```



```
plot(hatvalues(sold_units_selected))
abline(h=length(coef(sold_units_selected))/nrow(sold_units)*2,
       col = "red", lty = 2)
```



```
plot(rstandard((sold_units_selected)))
```

Looking for colinearity Variance inflation factors

```
vif(sold_units_selected)
```

```
##          itcrb semiconductor_crisis          reserves
##      19.23772          1.05751          31.28108
##          PIB      PIB_over_reserves
##      47.26227          26.86648
```