ISyE 6414 Project 2

Xiaochen Yan

2023-11-23

We eliminated the non-positive values of the response variables (Dividend Payout Ratio)

```
data1<-read.csv("pos_data.csv",sep=',',header=TRUE)[-1]
dim(data1)
```

```
## [1] 2399    6
```

————-Part 3: Data Description————

3.1 Descriptive statistics

The maximum values of the five variables show that there are some unusual observations.

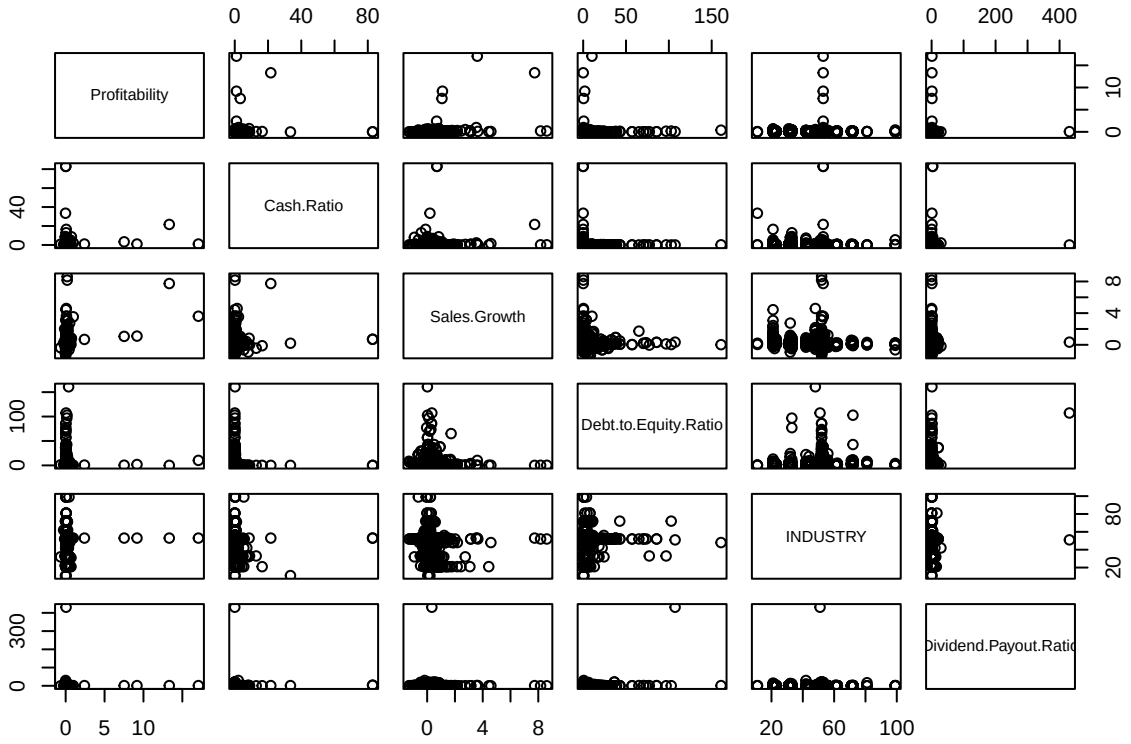
```
summary(data1)
```

```
##  Profitability        Cash.Ratio        Sales.Growth      Debt.to.Equity.Ratio
##  Min.   :-0.65671   Min.   : 0.00000   Min.   :-1.31015   Min.   :  0.00078
##  1st Qu.: 0.02001   1st Qu.: 0.02373   1st Qu.: 0.03135   1st Qu.:  1.13818
##  Median : 0.03816   Median : 0.06300   Median : 0.11222   Median :  2.76868
##  Mean   : 0.08882   Mean   : 0.30706   Mean   : 0.17826   Mean   :  5.98880
##  3rd Qu.: 0.08773   3rd Qu.: 0.16554   3rd Qu.: 0.22982   3rd Qu.:  9.32147
##  Max.   :17.06412   Max.   :82.88839   Max.   : 8.61472   Max.   :160.89600
##     INDUSTRY     Dividend.Payout.Ratio
##  Min.   :11.00   Min.   :  0.0002
##  1st Qu.:33.00   1st Qu.:  0.2052
##  Median :52.00   Median :  0.3596
##  Mean   :45.47   Mean   :  0.9336
##  3rd Qu.:52.00   3rd Qu.:  0.7111
##  Max.   :99.00   Max.   :430.8850
```

3.2 Scatter Plot

The scatter plot also shows unusual observations. Besides, due to unusual values, the relationship between response variable and independent variables is not clear.

```
plot(data1[,-ncol(data1)-1])
```



3.3 Process Unusual Observations

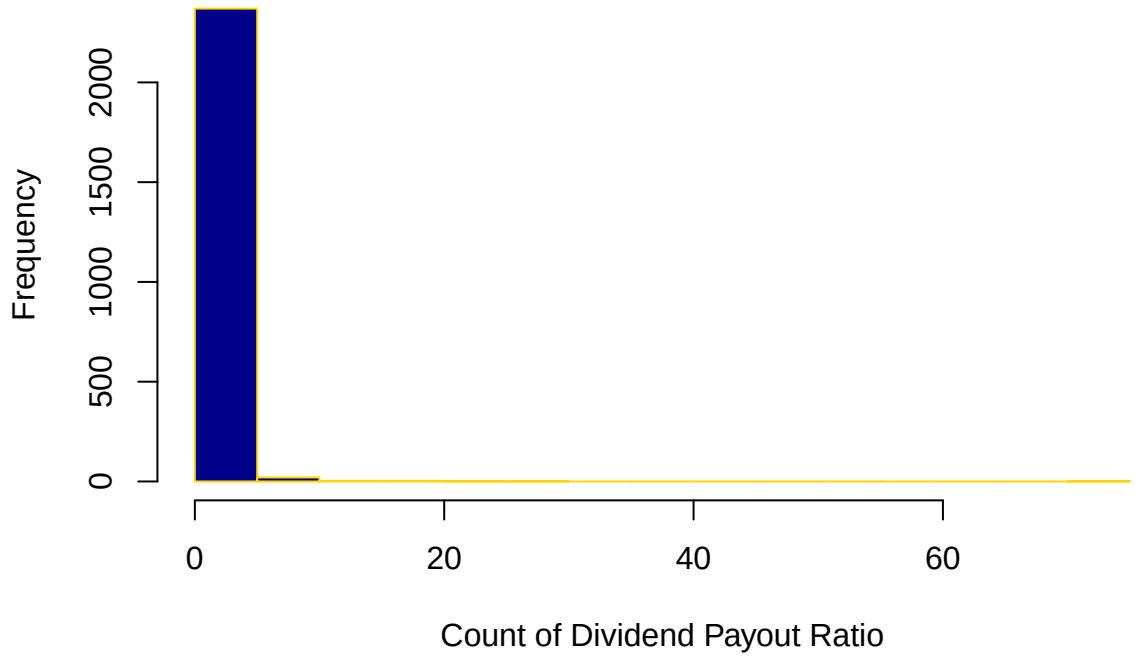
```
# Replace values greater than the 99% quantile with the 99% quantile value, excluding "Industry"
data_99 <- data1  # Copy the original data frame

quantile_99 <- apply(data_99, 2, function(x) quantile(x, probs = 0.99))
data_99["Profitability"] <- apply(data_99["Profitability"], 2, function(x, q) ifelse(x > q, q, x), quant
data_99["Cash.Ratio"] <- apply(data_99["Cash.Ratio"], 2, function(x, q) ifelse(x > q, q, x), quantile_99
data_99["Sales.Growth"] <- apply(data_99["Sales.Growth"], 2, function(x, q) ifelse(x > q, q, x), quantil
data_99["Debt.to.Equity.Ratio"] <- apply(data_99["Debt.to.Equity.Ratio"], 2, function(x, q) ifelse(x > q
data_99["Dividend.Payout.Ratio"] <- apply(data_99["Dividend.Payout.Ratio"], 2, function(x, q) ifelse(x >

# Print the modified data frame
summary(data_99)
```

```
##  Profitability.Profitability Cash.Ratio.Cash.Ratio Sales.Growth.Sales.Growth
##  Min.   :-0.656706           Min.   : 0.00000      Min.   :-1.310151
##  1st Qu.: 0.020008           1st Qu.: 0.02373      1st Qu.: 0.031351
##  Median : 0.038162           Median : 0.06300      Median : 0.112217
##  Mean   : 0.075430           Mean   : 0.19274      Mean   : 0.166009
##  3rd Qu.: 0.087729           3rd Qu.: 0.16554      3rd Qu.: 0.229823
##  Max.   : 8.668683           Max.   :33.49268      Max.   : 8.614715
##  Debt.to.Equity.Ratio.Debt.to.Equity.Ratio    INDUSTRY
##  Min.   : 0.00078                          Min.   :11.00
##  1st Qu.: 0.65206                          1st Qu.:33.00
##  Median : 1.60591                          Median :52.00
```

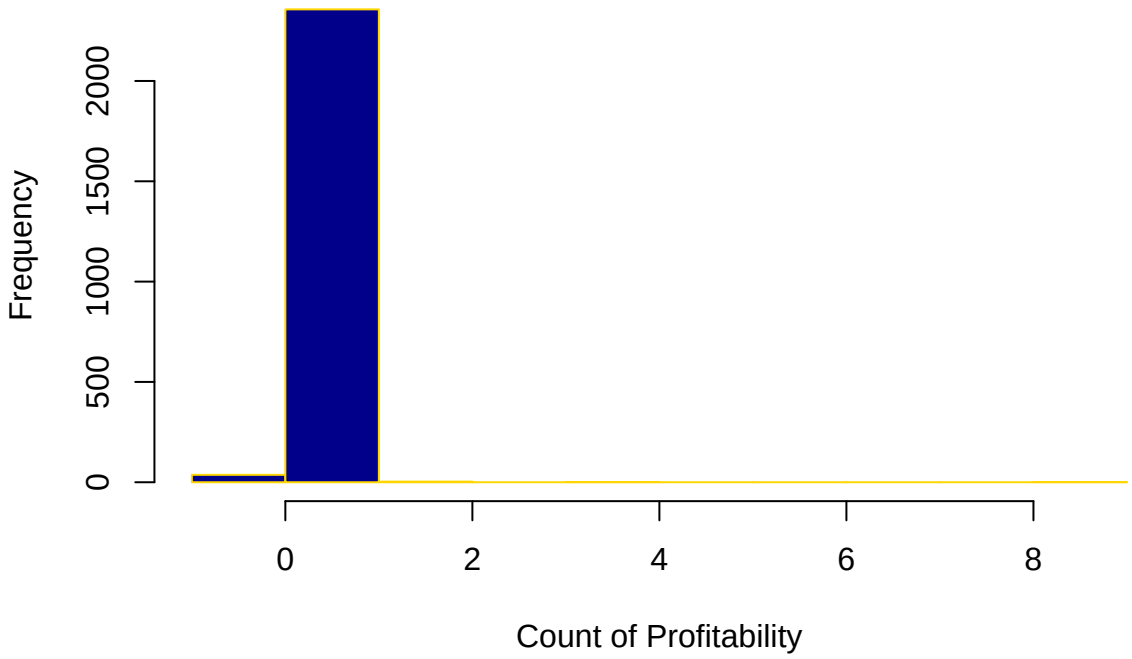
```
##  Mean   : 3.30944                          Mean   :45.47
##  3rd Qu.: 3.89000                          3rd Qu.:52.00
##  Max.   :72.00000                          Max.   :99.00
##  Dividend.Payout.Ratio.Dividend.Payout.Ratio
##  Min.   : 0.00017
##  1st Qu.: 0.20524
##  Median : 0.35956
##  Mean   : 0.65949
##  3rd Qu.: 0.60034
##  Max.   :72.00000
```

3.4 Data Histogram

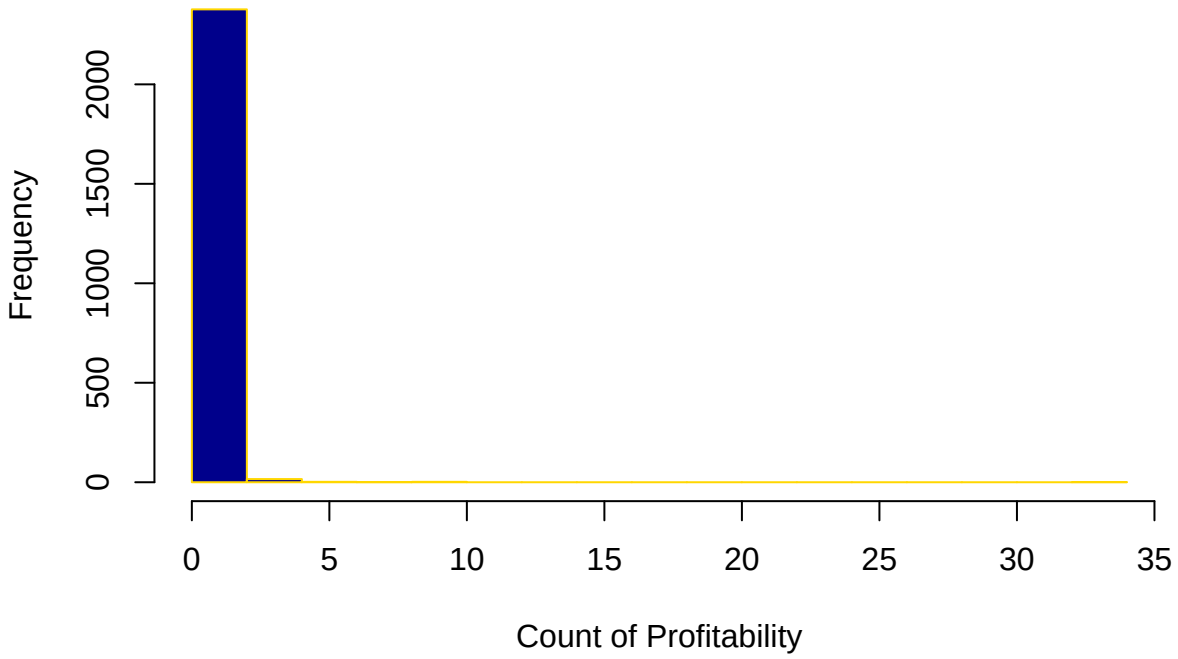
```
hist(data_99$Dividend.Payout.Ratio,main="",xlab="Count of Dividend Payout Ratio", border="gold",col="da
```



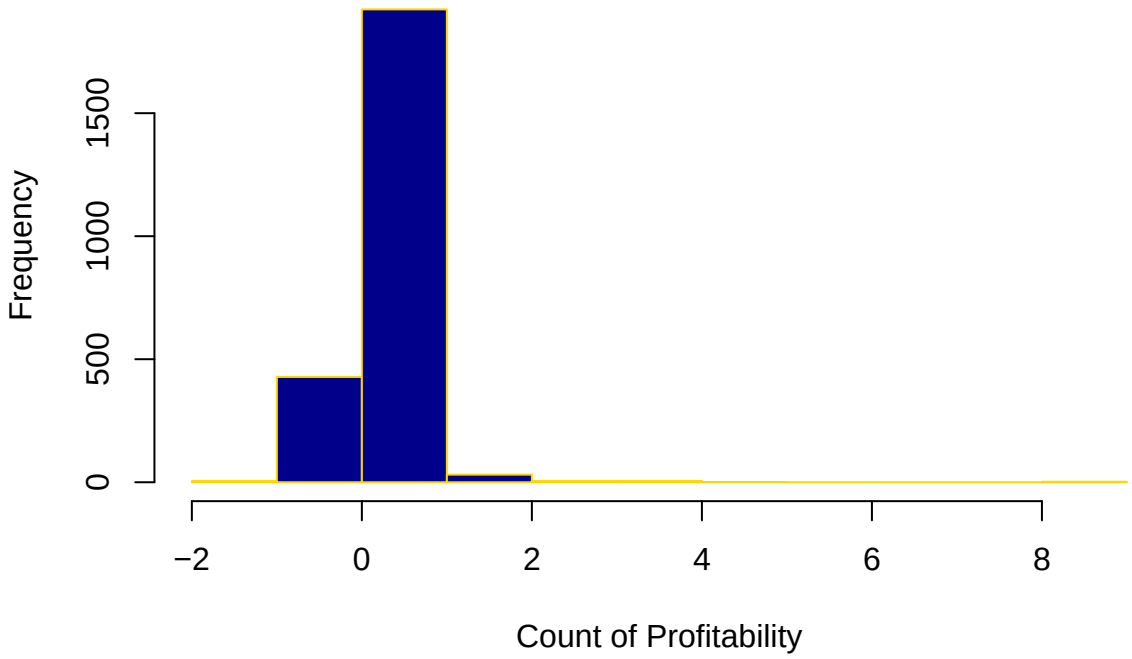

```
hist(data_99$Profitability,main="",xlab="Count of Profitability", border="gold",col="darkblue")
```



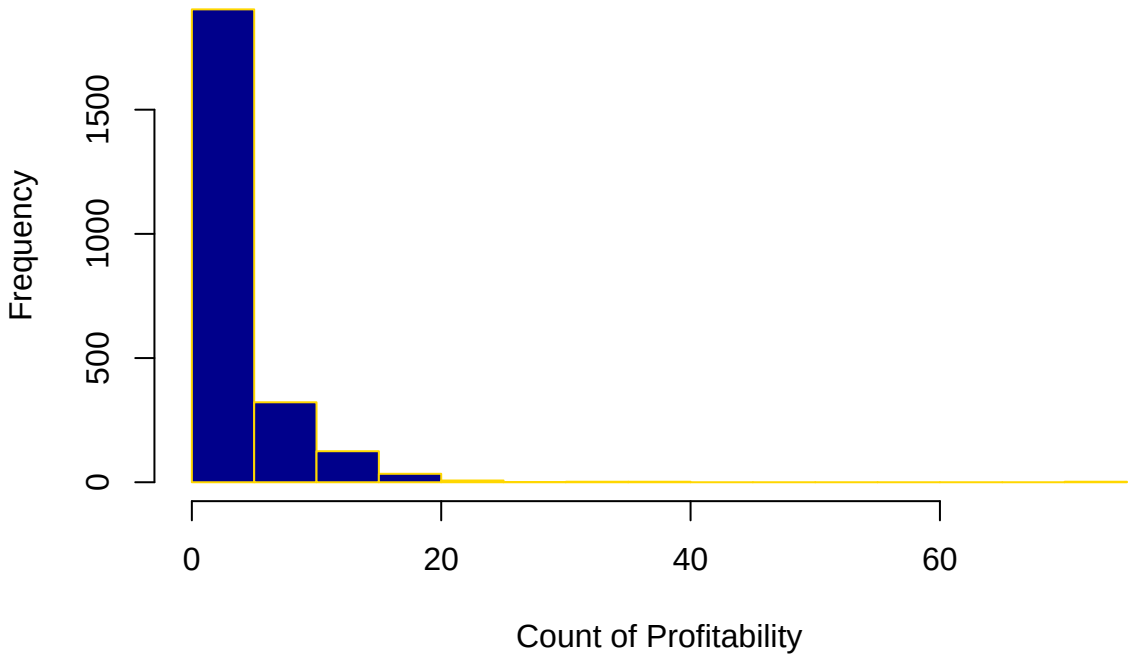

```
hist(data_99$Cash.Ratio,main="",xlab="Count of Profitability", border="gold",col="darkblue")
```




```
hist(data_99$Sales.Growth,main="",xlab="Count of Profitability", border="gold",col="darkblue")
```




```
hist(data_99$Debt.to.Equity.Ratio,main="",xlab="Count of Profitability", border="gold",col="darkblue")
```



———-Part 4: Analysis————————-

4.1 Split data set and correlation analysis

```
set.seed(1234)

sample_size = floor(0.8*nrow(data_99))
picked = sample(seq_len(nrow(data_99)),size = sample_size)

data = data_99[picked,]
y <- data$Dividend.Payout.Ratio

attach(data)
```

```
summary(y)
```

```
##  Dividend.Payout.Ratio
##  Min.   : 0.00017
##  1st Qu.: 0.20522
##  Median : 0.36500
##  Mean   : 0.65083
##  3rd Qu.: 0.60730
##  Max.   :72.00000
```

```
print(length(y))
```

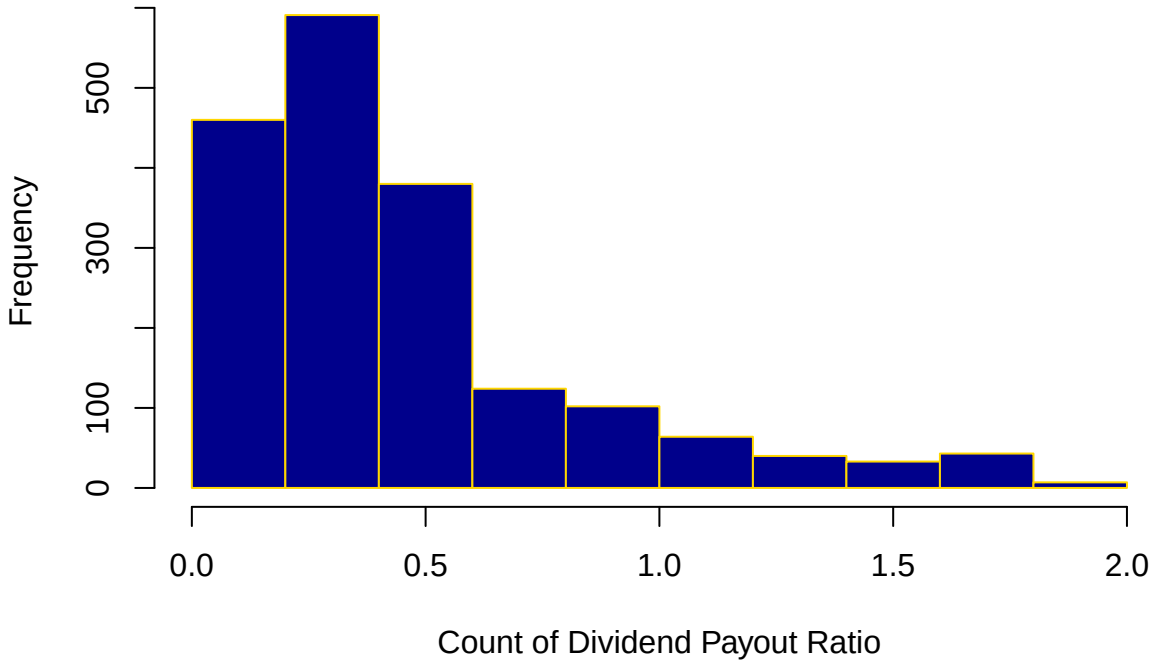
```
## [1] 1919
```

Since the response value is a ratio, we choose the majority of data (less than 2) for histogram.

```
adj_y = y[y<2]
length(adj_y)
```

```
## [1] 1844
```

```
hist(adj_y,main="",xlab="Count of Dividend Payout Ratio",breaks = 10,xlim = c(0, 2), border="gold",col=
```




```
corr = cor(data)
print(corr)
```

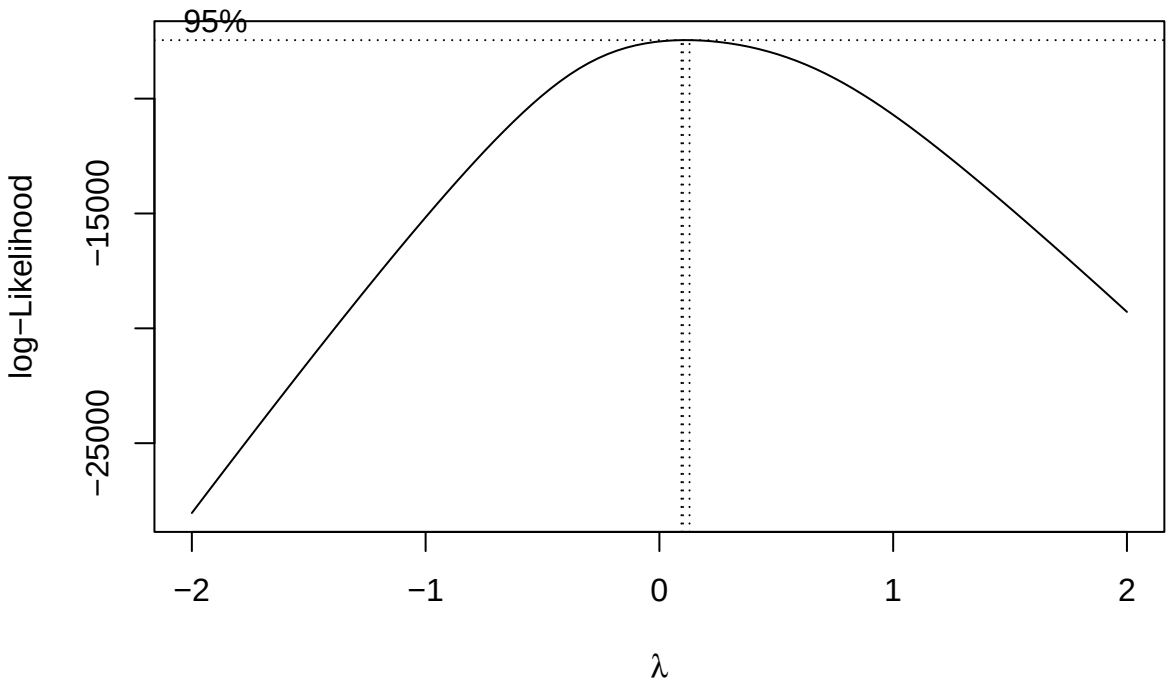
```
##                       Profitability  Cash.Ratio Sales.Growth
## Profitability           1.000000000  0.10510905  0.246097638
## Cash.Ratio              0.105109053  1.00000000  0.018436108
## Sales.Growth            0.246097638  0.01843611  1.000000000
## Debt.to.Equity.Ratio   -0.077978232 -0.08983835 -0.034586555
## INDUSTRY               -0.087992911 -0.09427238 -0.054092359
## Dividend.Payout.Ratio  -0.009517243  0.01424959 -0.008897196
##                       Debt.to.Equity.Ratio    INDUSTRY Dividend.Payout.Ratio
## Profitability                  -0.07797823 -0.08799291          -0.009517243
## Cash.Ratio                     -0.08983835 -0.09427238           0.014249588
## Sales.Growth                   -0.03458655 -0.05409236          -0.008897196
## Debt.to.Equity.Ratio            1.00000000  0.24609586           0.261383200
```

```
## INDUSTRY                    0.24609586  1.00000000          0.011138242
## Dividend.Payout.Ratio       0.26138320  0.01113824          1.000000000
```

4.2 Box-Cox Transformation

Apply the box-cox and the result shows that we need to apply natural logarithm transformation to the response variable

```
library(MASS)
result <- boxcox(y ~ 1, data = data)
```




```
result
```

```
## $x
##   [1] -2.00000000 -1.95959596 -1.91919192 -1.87878788 -1.83838384 -1.79797980
##   [7] -1.75757576 -1.71717172 -1.67676768 -1.63636364 -1.59595960 -1.55555556
##  [13] -1.51515152 -1.47474747 -1.43434343 -1.39393939 -1.35353535 -1.31313131
##  [19] -1.27272727 -1.23232323 -1.19191919 -1.15151515 -1.11111111 -1.07070707
##  [25] -1.03030303 -0.98989899 -0.94949495 -0.90909091 -0.86868687 -0.82828283
##  [31] -0.78787879 -0.74747475 -0.70707071 -0.66666667 -0.62626263 -0.58585859
##  [37] -0.54545455 -0.50505051 -0.46464646 -0.42424242 -0.38383838 -0.34343434
##  [43] -0.30303030 -0.26262626 -0.22222222 -0.18181818 -0.14141414 -0.10101010
##  [49] -0.06060606 -0.02020202  0.02020202  0.06060606  0.10101010  0.14141414
##  [55]  0.18181818  0.22222222  0.26262626  0.30303030  0.34343434  0.38383838
##  [61]  0.42424242  0.46464646  0.50505051  0.54545455  0.58585859  0.62626263
```

```
##  [67]  0.66666667  0.70707071  0.74747475  0.78787879  0.82828283  0.86868687
##  [73]  0.90909091  0.94949495  0.98989899  1.03030303  1.07070707  1.11111111
##  [79]  1.15151515  1.19191919  1.23232323  1.27272727  1.31313131  1.35353535
##  [85]  1.39393939  1.43434343  1.47474747  1.51515152  1.55555556  1.59595960
##  [91]  1.63636364  1.67676768  1.71717172  1.75757576  1.79797980  1.83838384
##  [97]  1.87878788  1.91919192  1.95959596  2.00000000
##
## $y
##   [1] -28039.463 -27498.138 -26957.898 -26418.807 -25880.925 -25344.314
##   [7] -24809.035 -24275.162 -23742.774 -23211.956 -22682.800 -22155.406
##  [13] -21629.887 -21106.359 -20584.958 -20065.828 -19549.127 -19035.037
##  [19] -18523.748 -18015.480 -17510.476 -17008.998 -16511.358 -16017.879
##  [25] -15528.952 -15045.001 -14566.500 -14094.018 -13628.154 -13169.649
##  [31] -12719.315 -12278.082 -11847.083 -11427.513 -11020.861 -10628.746
##  [37] -10252.958  -9895.547  -9558.582  -9244.125  -8954.228  -8690.510
##  [43]  -8454.029  -8245.578  -8064.801  -7910.965  -7782.828  -7678.565
##  [49]  -7596.319  -7534.229  -7490.448  -7463.415  -7451.859  -7454.640
##  [55]  -7471.059  -7500.538  -7542.764  -7597.665  -7665.247  -7745.747
##  [61]  -7839.453  -7946.722  -8067.964  -8203.554  -8353.777  -8518.883
##  [67]  -8698.909  -8893.746  -9103.167  -9326.665  -9563.676  -9813.462
##  [73] -10075.184 -10347.979 -10630.942 -10923.168 -11223.811 -11532.053
##  [79] -11847.132 -12168.381 -12495.156 -12826.916 -13163.163 -13503.449
##  [85] -13847.401 -14194.663 -14544.941 -14897.969 -15253.506 -15611.351
##  [91] -15971.314 -16333.234 -16696.967 -17062.378 -17429.354 -17797.785
##  [97] -18167.583 -18538.666 -18910.951 -19284.348
```

4.3 First-order model 4.3.1 Regression

F-statistic shows the model as a whole is significant. The coefficient of Sales.Growth and Debt.to.Equity.Ratio is significant.

First -order model

```
model1 = lm(log(y)~Profitability+Cash.Ratio+Sales.Growth+Debt.to.Equity.Ratio+factor(INDUSTRY),data = da
summary(model1)
```

```
##
## Call:
## lm(formula = log(y) ~ Profitability + Cash.Ratio + Sales.Growth +
##     Debt.to.Equity.Ratio + factor(INDUSTRY), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6802 -0.4756  0.0782  0.5037  4.7345
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -1.230141   0.957679  -1.285    0.199
## Profitability         0.033066   0.109945   0.301    0.764
## Cash.Ratio            0.045189   0.034352   1.315    0.189
## Sales.Growth         -0.329543   0.062089  -5.308 1.24e-07 ***
## Debt.to.Equity.Ratio  0.006291   0.005748   1.094    0.274
```

```
## factor(INDUSTRY)21   -0.293469   0.957721  -0.306    0.759
## factor(INDUSTRY)22    0.518739   0.961200   0.540    0.589
## factor(INDUSTRY)23   -0.637049   0.996337  -0.639    0.523
## factor(INDUSTRY)31    0.424131   0.969999   0.437    0.662
## factor(INDUSTRY)32    0.180924   0.960615   0.188    0.851
## factor(INDUSTRY)33   -0.086561   0.954582  -0.091    0.928
## factor(INDUSTRY)42    0.160074   0.969042   0.165    0.869
## factor(INDUSTRY)44   -0.553825   1.003224  -0.552    0.581
## factor(INDUSTRY)45    0.410634   1.070759   0.383    0.701
## factor(INDUSTRY)48   -0.147012   0.962983  -0.153    0.879
## factor(INDUSTRY)49    0.444044   1.441313   0.308    0.758
## factor(INDUSTRY)51    0.427972   0.961860   0.445    0.656
## factor(INDUSTRY)52    0.038705   0.955216   0.041    0.968
## factor(INDUSTRY)53    1.078123   0.957111   1.126    0.260
## factor(INDUSTRY)54    0.351270   1.004342   0.350    0.727
## factor(INDUSTRY)56    0.230762   0.981052   0.235    0.814
## factor(INDUSTRY)61    0.289823   1.049300   0.276    0.782
## factor(INDUSTRY)62   -0.359721   1.003698  -0.358    0.720
## factor(INDUSTRY)71   -1.252867   1.066540  -1.175    0.240
## factor(INDUSTRY)72    0.108796   0.990319   0.110    0.913
## factor(INDUSTRY)81   -0.146302   1.442266  -0.101    0.919
## factor(INDUSTRY)99   -0.881648   1.137723  -0.775    0.438
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.079 on 1892 degrees of freedom
## Multiple R-squared:  0.122,  Adjusted R-squared:   0.11
## F-statistic: 10.11 on 26 and 1892 DF,  p-value: < 2.2e-16
```

test for subset of coefficient

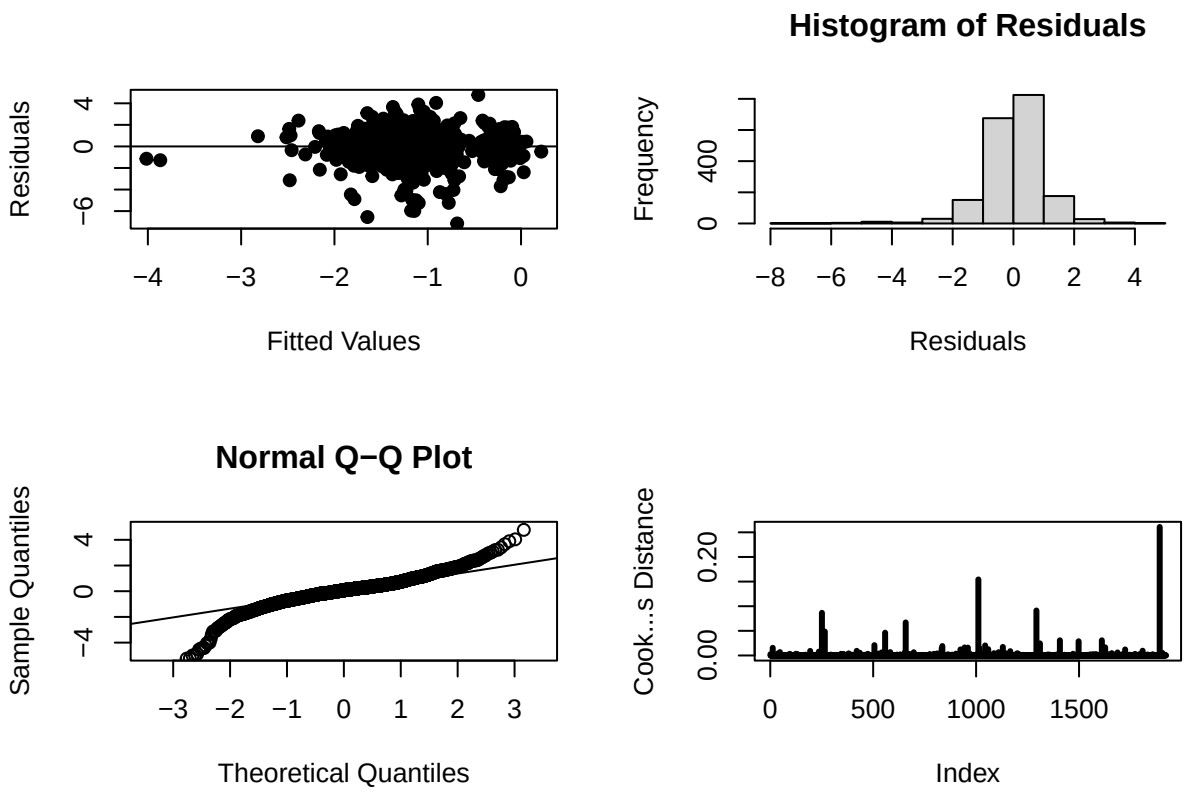
```
reduced_model1 = lm(log(y) ~ Cash.Ratio + Sales.Growth,data=data)
anova(reduced_model1, model1)
```

```
## Analysis of Variance Table
##
## Model 1: log(y) ~ Cash.Ratio + Sales.Growth
## Model 2: log(y) ~ Profitability + Cash.Ratio + Sales.Growth + Debt.to.Equity.Ratio +
##     factor(INDUSTRY)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   1916 2469.1
## 2   1892 2201.8 24    267.25 9.5687 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4.3.2 Model Diagnostics Analysis of residuals:

```
res = stdres(model1)


par(mfrow = c(2,2))
plot(model1$fitted.values, res, xlab = "Fitted Values", ylab = "Residuals",pch = 19)
abline(h = 0)
```

```
# plot(Profitability, res, xlab = "Profitability", ylab = "Residuals",xlim=c(0,1), pch = 19)
# abline(h = 0)
# plot(Cash.Ratio, res, xlab = "Cash.Ratio", ylab = "Residuals",xlim=c(0,5), pch = 19)
# abline(h = 0)
# plot(Sales.Growth, res, xlab = "Sales.Growth", ylab = "Residuals",xlim=c(0,3), pch = 19)
# abline(h = 0)
# plot(Debt.to.Equity.Ratio, res, xlab = "Debt.to.Equity.Ratio", ylab = "Residuals",xlim=c(0,50), pch =
# abline(h = 0)
# plot(INDUSTRY, res, xlab = "INDUSTRY", ylab = "Residuals", pch = 19)
# abline(h = 0)


hist(res, xlab="Residuals", main= "Histogram of Residuals")

qqnorm(res,ylim=c(-5,5))
qqline(res)

cook = cooks.distance(model1)
plot(cook,type="h",lwd=3, ylab = "Cook's Distance")
```

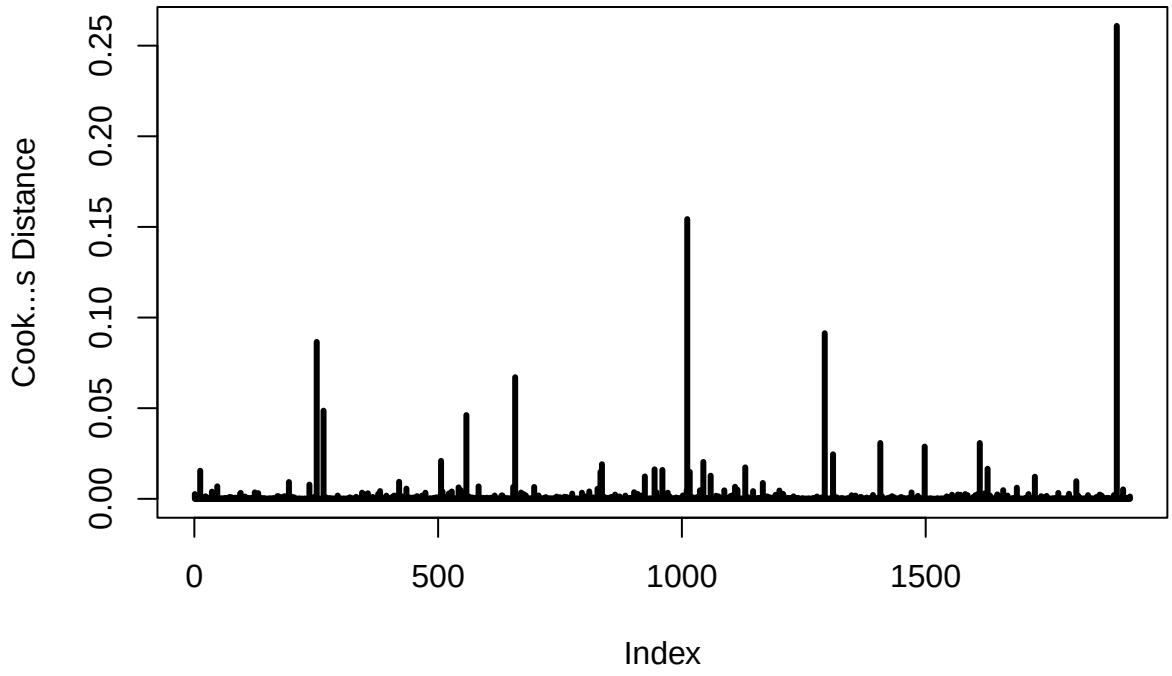


Use Cook's distance to check if there are any outliers.

```
cook = cooks.distance(model1)
plot(cook,type="h",lwd=3, ylab = "Cook's Distance")
```

```
## Warning in title(...): conversion failure on 'Cook's Distance' in 'mbcsToSbcs':
```

```
## dot substituted for <e2>

## Warning in title(...): conversion failure on 'Cook's Distance' in 'mbcsToSbcs':
## dot substituted for <80>

## Warning in title(...): conversion failure on 'Cook's Distance' in 'mbcsToSbcs':
## dot substituted for <99>
```




```
hurdle = 4/length(data)

indices <- which(cook > hurdle)
if(length(indices) > 0) {
    print(cook[indices])
} else {
    print("No Outliers")
}
```

```
## [1] "No Outliers"
```

Multicollinearity:

```
library(car)

vif(model1)
```

```
##                          GVIF Df GVIF^(1/(2*Df))
## Profitability        1.120496  1        1.058535
## Cash.Ratio           1.601131  1        1.265358
## Sales.Growth         1.098022  1        1.047865
## Debt.to.Equity.Ratio 1.209605  1        1.099820
## factor(INDUSTRY)     2.027163 22        1.016190
```

```
max(10,1/(1-summary(model1)$r.squared))
```

```
## [1] 10
```

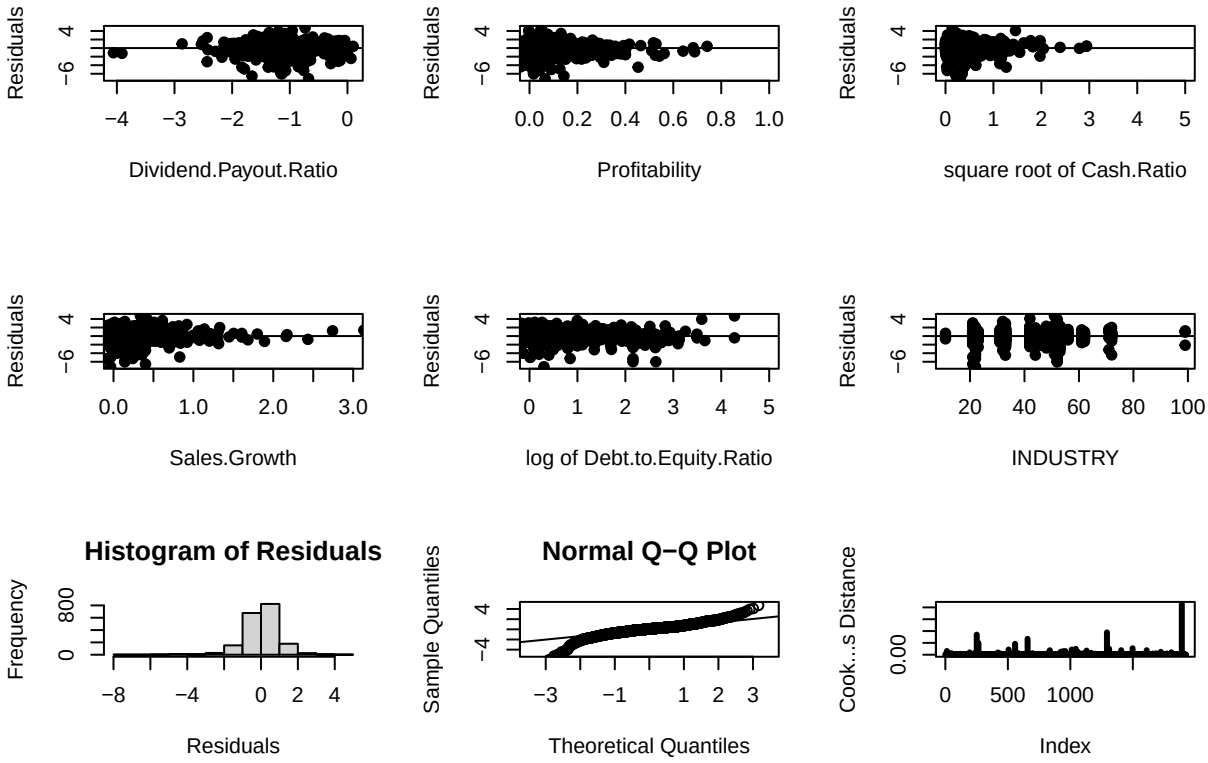
Apply transformation to the predict variables. Only Debt.to.Equity.Ratio is positive so we will use log transformation to it.

```
model1_trans = lm(log(y)~Profitability+sqrt(Cash.Ratio)+Sales.Growth+log(Debt.to.Equity.Ratio)+factor(I
summary(model1_trans)
```

```
##
## Call:
## lm(formula = log(y) ~ Profitability + sqrt(Cash.Ratio) + Sales.Growth +
##     log(Debt.to.Equity.Ratio) + factor(INDUSTRY), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6902 -0.4795  0.0748  0.5008  5.0105
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -0.51435    0.81633  -0.630   0.5287
## Profitability              0.04186    0.11193   0.374   0.7085
## sqrt(Cash.Ratio)           0.04822    0.09885   0.488   0.6257
## Sales.Growth              -0.32813    0.06212  -5.282 1.42e-07 ***
## log(Debt.to.Equity.Ratio)  0.03885    0.02416   1.608   0.1080
## factor(INDUSTRY)21        -0.99029    0.81183  -1.220   0.2227
## factor(INDUSTRY)22        -0.20132    0.81744  -0.246   0.8055
## factor(INDUSTRY)23        -1.35223    0.85335  -1.585   0.1132
## factor(INDUSTRY)31        -0.30426    0.82273  -0.370   0.7116
## factor(INDUSTRY)32        -0.53610    0.81242  -0.660   0.5094
## factor(INDUSTRY)33        -0.80031    0.80617  -0.993   0.3210
## factor(INDUSTRY)42        -0.55588    0.82526  -0.674   0.5007
## factor(INDUSTRY)44        -1.29373    0.86230  -1.500   0.1337
## factor(INDUSTRY)45        -0.32686    0.93981  -0.348   0.7280
## factor(INDUSTRY)48        -0.86268    0.81549  -1.058   0.2902
## factor(INDUSTRY)49        -0.30606    1.34597  -0.227   0.8201
## factor(INDUSTRY)51        -0.28387    0.81448  -0.349   0.7275
## factor(INDUSTRY)52        -0.69276    0.80714  -0.858   0.3908
## factor(INDUSTRY)53         0.37038    0.81170   0.456   0.6482
## factor(INDUSTRY)54        -0.37884    0.86354  -0.439   0.6609
## factor(INDUSTRY)56        -0.50039    0.83687  -0.598   0.5500
## factor(INDUSTRY)61        -0.42647    0.91412  -0.467   0.6409
## factor(INDUSTRY)62        -1.08274    0.86328  -1.254   0.2099
## factor(INDUSTRY)71        -1.96450    0.93541  -2.100   0.0358 *
```

```
## factor(INDUSTRY)72        -0.62545    0.84746  -0.738   0.4606
## factor(INDUSTRY)81        -0.87927    1.35043  -0.651   0.5151
## factor(INDUSTRY)99        -1.60546    1.01467  -1.582   0.1138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.079 on 1892 degrees of freedom
## Multiple R-squared:  0.122,  Adjusted R-squared:  0.1099
## F-statistic: 10.11 on 26 and 1892 DF,  p-value: < 2.2e-16
```

```
res = stdres(model1_trans)
cook = cooks.distance(model1_trans)

par(mfrow = c(3,3))
plot(model1_trans$fitted.values, res, xlab = "Dividend.Payout.Ratio", ylab = "Residuals",pch = 19)
abline(h = 0)
plot(Profitability, res, xlab = "Profitability", ylab = "Residuals",xlim=c(0,1), pch = 19)
abline(h = 0)
plot(sqrt(Cash.Ratio), res, xlab = "square root of Cash.Ratio", ylab = "Residuals",xlim=c(0,5), pch = 19
abline(h = 0)
plot(Sales.Growth, res, xlab = "Sales.Growth", ylab = "Residuals",xlim=c(0,3), pch = 19)
abline(h = 0)
plot(log(Debt.to.Equity.Ratio), res, xlab = "log of Debt.to.Equity.Ratio", ylab = "Residuals",xlim=c(0,5
abline(h = 0)
plot(INDUSTRY, res, xlab = "INDUSTRY", ylab = "Residuals", pch = 19)
abline(h = 0)


hist(res, xlab="Residuals", main= "Histogram of Residuals")

qqnorm(res,ylim=c(-5,5))
qqline(res)

plot(cook,type="h",lwd=3, ylab = "Cook's Distance")
```




```
vif(model1_trans)
```

```
##                              GVIF Df GVIF^(1/(2*Df))
## Profitability            1.161313  1        1.077643
## sqrt(Cash.Ratio)         1.497474  1        1.223713
## Sales.Growth             1.098953  1        1.048310
## log(Debt.to.Equity.Ratio) 1.468457  1        1.211799
## factor(INDUSTRY)         1.615700 22        1.010963
```

```
max(10,1/(1-summary(model1_trans)$r.squared))
```

```
## [1] 10
```

Poisson

```
model3 = glm(round(y*100) ~ Profitability+Cash.Ratio+Sales.Growth+Debt.to.Equity.Ratio+factor(INDUSTRY)

summary(model3)
```

```
##
## Call:
## glm(formula = round(y * 100) ~ Profitability + Cash.Ratio + Sales.Growth +
```

```
##     Debt.to.Equity.Ratio + factor(INDUSTRY), family = "poisson",
##     data = data)
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            0.5993287  0.1684481   3.558 0.000374 ***
## Profitability         -0.3215725  0.0218707 -14.703  < 2e-16 ***
## Cash.Ratio             0.1283283  0.0043687  29.374  < 2e-16 ***
## Sales.Growth          -0.1423145  0.0100964 -14.096  < 2e-16 ***
## Debt.to.Equity.Ratio   0.0538307  0.0002137 251.850  < 2e-16 ***
## factor(INDUSTRY)21     3.2145586  0.1675220  19.189  < 2e-16 ***
## factor(INDUSTRY)22     3.6258856  0.1687412  21.488  < 2e-16 ***
## factor(INDUSTRY)23     2.4705238  0.1774524  13.922  < 2e-16 ***
## factor(INDUSTRY)31     3.3547216  0.1697223  19.766  < 2e-16 ***
## factor(INDUSTRY)32     3.6787161  0.1687335  21.802  < 2e-16 ***
## factor(INDUSTRY)33     3.1801217  0.1677159  18.961  < 2e-16 ***
## factor(INDUSTRY)42     4.2426302  0.1681453  25.232  < 2e-16 ***
## factor(INDUSTRY)44     2.8906439  0.1744285  16.572  < 2e-16 ***
## factor(INDUSTRY)45     3.4346147  0.1777167  19.326  < 2e-16 ***
## factor(INDUSTRY)48     3.1824307  0.1692167  18.807  < 2e-16 ***
## factor(INDUSTRY)49     3.1380246  0.2241657  13.999  < 2e-16 ***
## factor(INDUSTRY)51     4.1063552  0.1678590  24.463  < 2e-16 ***
## factor(INDUSTRY)52     2.9569225  0.1680970  17.591  < 2e-16 ***
## factor(INDUSTRY)53     4.1752231  0.1679790  24.856  < 2e-16 ***
## factor(INDUSTRY)54     3.1901018  0.1734171  18.396  < 2e-16 ***
## factor(INDUSTRY)56     3.0297853  0.1712444  17.693  < 2e-16 ***
## factor(INDUSTRY)61     3.3839534  0.1762256  19.202  < 2e-16 ***
## factor(INDUSTRY)62     2.5863062  0.1773039  14.587  < 2e-16 ***
## factor(INDUSTRY)71     2.0828035  0.1991155  10.460  < 2e-16 ***
## factor(INDUSTRY)72     3.2221030  0.1714697  18.791  < 2e-16 ***
## factor(INDUSTRY)81     2.4944524  0.2681271   9.303  < 2e-16 ***
## factor(INDUSTRY)99     2.3802245  0.2085522  11.413  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 213187  on 1918  degrees of freedom
## Residual deviance: 144853  on 1892  degrees of freedom
## AIC: 155208
##
## Number of Fisher Scoring iterations: 6
```

Test for overall regression

```
1-pchisq((213187- 144853), 1918-1892)
```

```
## [1] 0
```

Goodness of Fit Test - Deviance Test

```
with(model3, cbind(res.deviance = deviance, df = df.residual,
                           p = 1- pchisq(deviance, df.residual)))
```

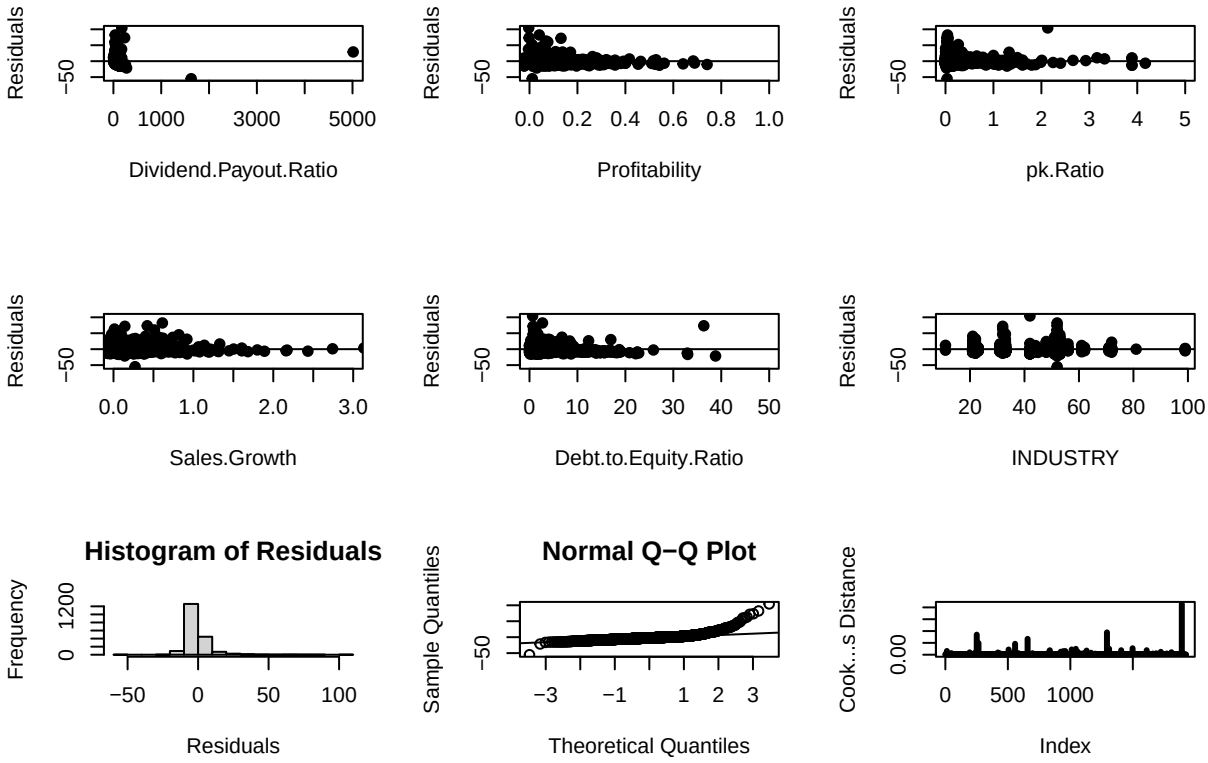
```
##      res.deviance   df p
## [1,]     144852.8 1892 0
```

p-value = 0. Reject the null hypothesis of good fit.

Residual Analysis

```
res <- resid(model3,type="deviance")

par(mfrow = c(3,3))
plot(model3$fitted.values, res, xlab = "Dividend.Payout.Ratio", ylab = "Residuals", pch = 19)
abline(h = 0)
plot(Profitability, res, xlab = "Profitability", ylab = "Residuals",xlim=c(0,1), pch = 19)
abline(h = 0)
plot(Cash.Ratio, res, xlab = "pk.Ratio", ylab = "Residuals",xlim=c(0,5), pch = 19)
abline(h = 0)
plot(Sales.Growth, res, xlab = "Sales.Growth", ylab = "Residuals",xlim=c(0,3), pch = 19)
abline(h = 0)
plot(Debt.to.Equity.Ratio, res, xlab = "Debt.to.Equity.Ratio", ylab = "Residuals",xlim=c(0,50), pch = 19
abline(h = 0)
plot(INDUSTRY, res, xlab = "INDUSTRY", ylab = "Residuals", pch = 19)
abline(h = 0)


hist(res, xlab="Residuals", main= "Histogram of Residuals")

qqnorm(res)
qqline(res)

plot(cook,type="h",lwd=3, ylab = "Cook's Distance")
```

```
## Warning in title(...): conversion failure on 'Cook's Distance' in 'mbcsToSbcs':
## dot substituted for <e2>

## Warning in title(...): conversion failure on 'Cook's Distance' in 'mbcsToSbcs':
## dot substituted for <80>

## Warning in title(...): conversion failure on 'Cook's Distance' in 'mbcsToSbcs':
## dot substituted for <99>
```




```
## Do we have overdispersion?
wdf <- model3$df.residual # n-p-1
dev <- model3$deviance
overdisp <- dev/wdf
print(overdisp)
```

```
## [1] 76.56068
```

```
## Assuming GOF: Deviance ~ N(0,1)
##  Evaluate Deviance residuals vs the Normal quantile with prob=0.99995
dev_residuals <- residuals(model3, type = "deviance")
outliers <- which(abs(dev_residuals) > qnorm(.99995))
length(outliers)
```

```
## [1] 959
```

959 outliers, indicating overdispersion.

Apply an overdispersed Poisson Regression

```
quasipoisson_model <- glm(round(y*100) ~ Profitability+Cash.Ratio+Sales.Growth+Debt.to.Equity.Ratio+fac
summary(quasipoisson_model)
```

```
##
```

```
## Call:
## glm(formula = round(y * 100) ~ Profitability + Cash.Ratio + Sales.Growth +
##     Debt.to.Equity.Ratio + factor(INDUSTRY), family = "quasipoisson",
##     data = data)
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.599329   2.235421   0.268   0.7886
## Profitability        -0.321572   0.290239  -1.108   0.2680
## Cash.Ratio            0.128328   0.057976   2.213   0.0270 *
## Sales.Growth         -0.142315   0.133986  -1.062   0.2883
## Debt.to.Equity.Ratio  0.053831   0.002836  18.978   <2e-16 ***
## factor(INDUSTRY)21    3.214559   2.223131   1.446   0.1484
## factor(INDUSTRY)22    3.625886   2.239311   1.619   0.1056
## factor(INDUSTRY)23    2.470524   2.354914   1.049   0.2943
## factor(INDUSTRY)31    3.354722   2.252330   1.489   0.1365
## factor(INDUSTRY)32    3.678716   2.239207   1.643   0.1006
## factor(INDUSTRY)33    3.180122   2.225704   1.429   0.1532
## factor(INDUSTRY)42    4.242630   2.231403   1.901   0.0574 .
## factor(INDUSTRY)44    2.890644   2.314784   1.249   0.2119
## factor(INDUSTRY)45    3.434615   2.358422   1.456   0.1455
## factor(INDUSTRY)48    3.182431   2.245620   1.417   0.1566
## factor(INDUSTRY)49    3.138025   2.974831   1.055   0.2916
## factor(INDUSTRY)51    4.106355   2.227602   1.843   0.0654 .
## factor(INDUSTRY)52    2.956922   2.230761   1.326   0.1852
## factor(INDUSTRY)53    4.175223   2.229195   1.873   0.0612 .
## factor(INDUSTRY)54    3.190102   2.301363   1.386   0.1659
## factor(INDUSTRY)56    3.029785   2.272529   1.333   0.1826
## factor(INDUSTRY)61    3.383953   2.338633   1.447   0.1481
## factor(INDUSTRY)62    2.586306   2.352943   1.099   0.2718
## factor(INDUSTRY)71    2.082803   2.642398   0.788   0.4307
## factor(INDUSTRY)72    3.222103   2.275519   1.416   0.1569
## factor(INDUSTRY)81    2.494452   3.558229   0.701   0.4834
## factor(INDUSTRY)99    2.380224   2.767630   0.860   0.3899
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 176.1109)
##
##     Null deviance: 213187  on 1918  degrees of freedom
## Residual deviance: 144853  on 1892  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

Second -order model

```
model2.1 = lm(log(y)~(Profitability+sqrt(Cash.Ratio)+Sales.Growth+log(Debt.to.Equity.Ratio))**2+factor(
summary(model2.1)
```

```
##
## Call:
```

```
## lm(formula = log(y) ~ (Profitability + sqrt(Cash.Ratio) + Sales.Growth +
##     log(Debt.to.Equity.Ratio))^2 + factor(INDUSTRY) + I(Profitability^2) +
##     I(sqrt(Cash.Ratio)^2) + I(Sales.Growth^2) + I(log(Debt.to.Equity.Ratio^2)),
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6282 -0.4738  0.0624  0.5120  4.8325
##
## Coefficients: (1 not defined because of singularities)
##                                          Estimate Std. Error t value
## (Intercept)                             -1.138969   1.031651  -1.104
## Profitability                           -1.222296   0.593369  -2.060
## sqrt(Cash.Ratio)                        -0.232307   0.169069  -1.374
## Sales.Growth                            -0.454667   0.156952  -2.897
## log(Debt.to.Equity.Ratio)                0.038076   0.035886   1.061
## factor(INDUSTRY)21                      -0.111983   1.042641  -0.107
## factor(INDUSTRY)22                       0.518102   1.036602   0.500
## factor(INDUSTRY)23                      -0.523859   1.075841  -0.487
## factor(INDUSTRY)31                       0.531247   1.050386   0.506
## factor(INDUSTRY)32                       0.313406   1.041872   0.301
## factor(INDUSTRY)33                       0.023652   1.038884   0.023
## factor(INDUSTRY)42                       0.264303   1.048385   0.252
## factor(INDUSTRY)44                      -0.481892   1.080709  -0.446
## factor(INDUSTRY)45                       0.491468   1.143927   0.430
## factor(INDUSTRY)48                      -0.045409   1.043780  -0.044
## factor(INDUSTRY)49                       0.590198   1.494471   0.395
## factor(INDUSTRY)51                       0.528711   1.046656   0.505
## factor(INDUSTRY)52                       0.057592   1.036860   0.056
## factor(INDUSTRY)53                       1.100379   1.037633   1.060
## factor(INDUSTRY)54                       0.462408   1.080529   0.428
## factor(INDUSTRY)56                       0.330195   1.058071   0.312
## factor(INDUSTRY)61                       0.384885   1.128621   0.341
## factor(INDUSTRY)62                      -0.280425   1.079627  -0.260
## factor(INDUSTRY)71                      -1.140843   1.144220  -0.997
## factor(INDUSTRY)72                       0.162186   1.068483   0.152
## factor(INDUSTRY)81                      -0.161568   1.492430  -0.108
## factor(INDUSTRY)99                      -0.824351   1.211309  -0.681
## I(Profitability^2)                      -0.117070   0.240342  -0.487
## I(sqrt(Cash.Ratio)^2)                   -0.019948   0.097674  -0.204
## I(Sales.Growth^2)                        0.001771   0.023414   0.076
## I(log(Debt.to.Equity.Ratio^2))                 NA         NA      NA
## Profitability:sqrt(Cash.Ratio)           1.002507   0.613846   1.633
## Profitability:Sales.Growth               0.167795   0.471711   0.356
## Profitability:log(Debt.to.Equity.Ratio)  0.305960   0.251895   1.215
## sqrt(Cash.Ratio):Sales.Growth            0.342717   0.222875   1.538
## sqrt(Cash.Ratio):log(Debt.to.Equity.Ratio) -0.122094   0.079150  -1.543
## Sales.Growth:log(Debt.to.Equity.Ratio)   0.084404   0.074354   1.135
##                                         Pr(>|t|)
## (Intercept)                              0.26972
## Profitability                            0.03954 *
## sqrt(Cash.Ratio)                         0.16959
## Sales.Growth                             0.00381 **
## log(Debt.to.Equity.Ratio)                0.28882
```

```
## factor(INDUSTRY)21                          0.91448
## factor(INDUSTRY)22                          0.61727
## factor(INDUSTRY)23                          0.62636
## factor(INDUSTRY)31                          0.61308
## factor(INDUSTRY)32                          0.76359
## factor(INDUSTRY)33                          0.98184
## factor(INDUSTRY)42                          0.80099
## factor(INDUSTRY)44                          0.65572
## factor(INDUSTRY)45                          0.66751
## factor(INDUSTRY)48                          0.96530
## factor(INDUSTRY)49                          0.69295
## factor(INDUSTRY)51                          0.61352
## factor(INDUSTRY)52                          0.95571
## factor(INDUSTRY)53                          0.28907
## factor(INDUSTRY)54                          0.66874
## factor(INDUSTRY)56                          0.75502
## factor(INDUSTRY)61                          0.73312
## factor(INDUSTRY)62                          0.79509
## factor(INDUSTRY)71                          0.31887
## factor(INDUSTRY)72                          0.87937
## factor(INDUSTRY)81                          0.91380
## factor(INDUSTRY)99                          0.49624
## I(Profitability^2)                          0.62624
## I(sqrt(Cash.Ratio)^2)                       0.83820
## I(Sales.Growth^2)                           0.93971
## I(log(Debt.to.Equity.Ratio^2))                   NA
## Profitability:sqrt(Cash.Ratio)              0.10260
## Profitability:Sales.Growth                  0.72209
## Profitability:log(Debt.to.Equity.Ratio)     0.22466
## sqrt(Cash.Ratio):Sales.Growth               0.12429
## sqrt(Cash.Ratio):log(Debt.to.Equity.Ratio)  0.12310
## Sales.Growth:log(Debt.to.Equity.Ratio)      0.25645
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.076 on 1883 degrees of freedom
## Multiple R-squared:  0.1303, Adjusted R-squared:  0.1141
## F-statistic:  8.06 on 35 and 1883 DF,  p-value: < 2.2e-16
```

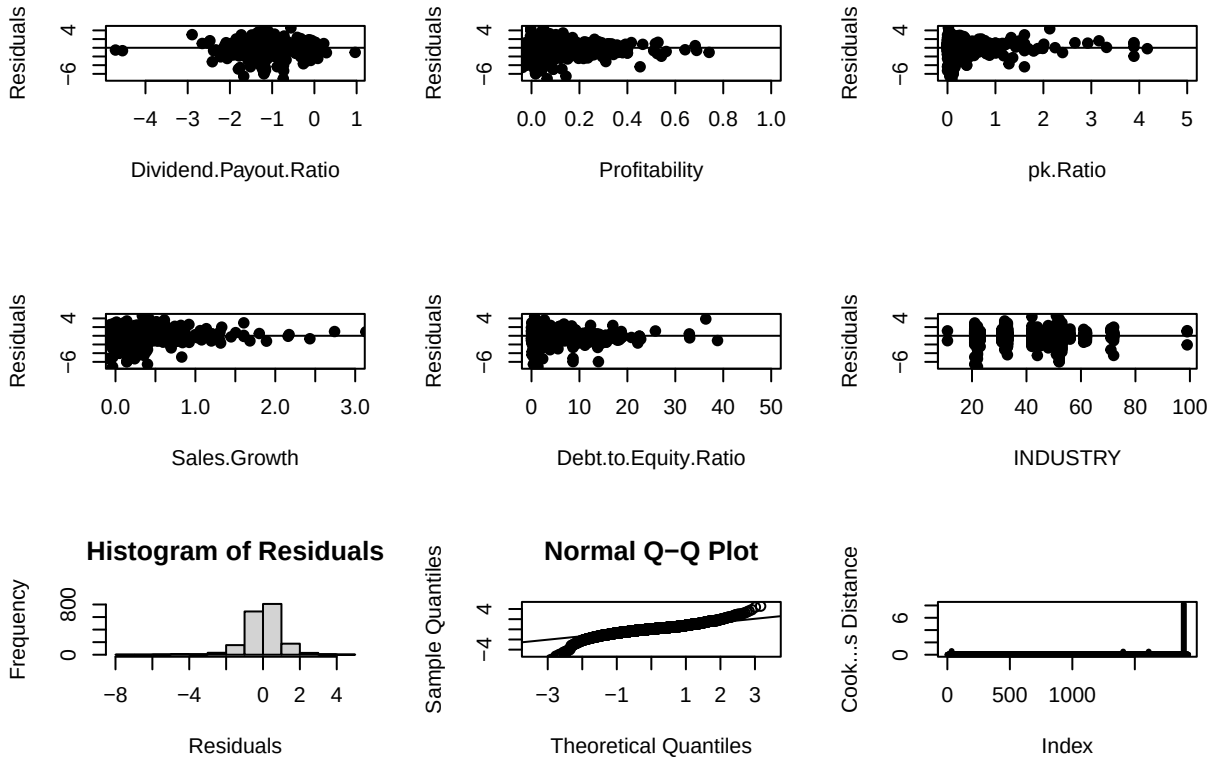
```
model2.2 = lm(log(y)~(Profitability+Cash.Ratio+Sales.Growth+log(Debt.to.Equity.Ratio))**2+factor(INDUSTR
summary(model2.2)
```

```
##
## Call:
## lm(formula = log(y) ~ (Profitability + Cash.Ratio + Sales.Growth +
##     log(Debt.to.Equity.Ratio))^2 + factor(INDUSTRY) + I(Profitability^2) +
##     I(Cash.Ratio^2) + I(Sales.Growth^2) + I(log(Debt.to.Equity.Ratio^2)),
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6709 -0.4743  0.0769  0.5023  4.9131
##
## Coefficients: (1 not defined because of singularities)
```

```
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             -0.878965   1.076474  -0.817   0.4143
## Profitability                           -0.590647   0.413248  -1.429   0.1531
## Cash.Ratio                               0.059527   0.119617   0.498   0.6188
## Sales.Growth                            -0.353226   0.125699  -2.810   0.0050
## log(Debt.to.Equity.Ratio)                0.035205   0.028405   1.239   0.2154
## factor(INDUSTRY)21                      -0.500182   1.084326  -0.461   0.6446
## factor(INDUSTRY)22                       0.196076   1.080242   0.182   0.8560
## factor(INDUSTRY)23                      -0.918071   1.113853  -0.824   0.4099
## factor(INDUSTRY)31                       0.145181   1.090190   0.133   0.8941
## factor(INDUSTRY)32                      -0.068771   1.082337  -0.064   0.9493
## factor(INDUSTRY)33                      -0.367897   1.079263  -0.341   0.7332
## factor(INDUSTRY)42                      -0.116828   1.091767  -0.107   0.9148
## factor(INDUSTRY)44                      -0.868980   1.120699  -0.775   0.4382
## factor(INDUSTRY)45                       0.133712   1.180596   0.113   0.9098
## factor(INDUSTRY)48                      -0.435818   1.084790  -0.402   0.6879
## factor(INDUSTRY)49                       0.193114   1.524452   0.127   0.8992
## factor(INDUSTRY)51                       0.132548   1.086396   0.122   0.9029
## factor(INDUSTRY)52                      -0.304843   1.077209  -0.283   0.7772
## factor(INDUSTRY)53                       0.766412   1.078858   0.710   0.4775
## factor(INDUSTRY)54                       0.071371   1.120418   0.064   0.9492
## factor(INDUSTRY)56                      -0.069882   1.100539  -0.063   0.9494
## factor(INDUSTRY)61                      -0.012243   1.162617  -0.011   0.9916
## factor(INDUSTRY)62                      -0.674341   1.119953  -0.602   0.5472
## factor(INDUSTRY)71                      -1.534057   1.179430  -1.301   0.1935
## factor(INDUSTRY)72                      -0.202620   1.107891  -0.183   0.8549
## factor(INDUSTRY)81                      -0.482570   1.522778  -0.317   0.7514
## factor(INDUSTRY)99                      -1.213792   1.243070  -0.976   0.3290
## I(Profitability^2)                       0.197161   0.223894   0.881   0.3786
## I(Cash.Ratio^2)                         -0.005039   0.004182  -1.205   0.2284
## I(Sales.Growth^2)                        0.008424   0.024197   0.348   0.7278
## I(log(Debt.to.Equity.Ratio^2))                 NA         NA      NA       NA
## Profitability:Cash.Ratio                -0.052908   0.223831  -0.236   0.8132
## Profitability:Sales.Growth              -0.294849   0.470851  -0.626   0.5313
## Profitability:log(Debt.to.Equity.Ratio) -0.012874   0.226524  -0.057   0.9547
## Cash.Ratio:Sales.Growth                  0.278354   0.135701   2.051   0.0404
## Cash.Ratio:log(Debt.to.Equity.Ratio)    -0.017412   0.038863  -0.448   0.6542
## Sales.Growth:log(Debt.to.Equity.Ratio)   0.078216   0.074899   1.044   0.2965
##
## (Intercept)
## Profitability
## Cash.Ratio
## Sales.Growth                            **
## log(Debt.to.Equity.Ratio)
## factor(INDUSTRY)21
## factor(INDUSTRY)22
## factor(INDUSTRY)23
## factor(INDUSTRY)31
## factor(INDUSTRY)32
## factor(INDUSTRY)33
## factor(INDUSTRY)42
## factor(INDUSTRY)44
## factor(INDUSTRY)45
## factor(INDUSTRY)48
```

```
## factor(INDUSTRY)49
## factor(INDUSTRY)51
## factor(INDUSTRY)52
## factor(INDUSTRY)53
## factor(INDUSTRY)54
## factor(INDUSTRY)56
## factor(INDUSTRY)61
## factor(INDUSTRY)62
## factor(INDUSTRY)71
## factor(INDUSTRY)72
## factor(INDUSTRY)81
## factor(INDUSTRY)99
## I(Profitability^2)
## I(Cash.Ratio^2)
## I(Sales.Growth^2)
## I(log(Debt.to.Equity.Ratio^2))
## Profitability:Cash.Ratio
## Profitability:Sales.Growth
## Profitability:log(Debt.to.Equity.Ratio)
## Cash.Ratio:Sales.Growth                  *
## Cash.Ratio:log(Debt.to.Equity.Ratio)
## Sales.Growth:log(Debt.to.Equity.Ratio)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.077 on 1883 degrees of freedom
## Multiple R-squared:  0.129,  Adjusted R-squared:  0.1128
## F-statistic:  7.97 on 35 and 1883 DF,  p-value: < 2.2e-16
```

```
res = stdres(model2.1)
cook = cooks.distance(model2.1)

par(mfrow = c(3,3))
plot(model2.1$fitted.values, res, xlab = "Dividend.Payout.Ratio", ylab = "Residuals", pch = 19)
abline(h = 0)
plot(Profitability, res, xlab = "Profitability", ylab = "Residuals",xlim=c(0,1), pch = 19)
abline(h = 0)
plot(Cash.Ratio, res, xlab = "pk.Ratio", ylab = "Residuals",xlim=c(0,5), pch = 19)
abline(h = 0)
plot(Sales.Growth, res, xlab = "Sales.Growth", ylab = "Residuals",xlim=c(0,3), pch = 19)
abline(h = 0)
plot(Debt.to.Equity.Ratio, res, xlab = "Debt.to.Equity.Ratio", ylab = "Residuals",xlim=c(0,50), pch = 19
abline(h = 0)
plot(INDUSTRY, res, xlab = "INDUSTRY", ylab = "Residuals", pch = 19)
abline(h = 0)


hist(res, xlab="Residuals", main= "Histogram of Residuals")

qqnorm(res,ylim=c(-5,5))
qqline(res)

plot(cook,type="h",lwd=3, ylab = "Cook's Distance")
```




```
model2 = lm(log(y)~(Profitability+sqrt(Cash.Ratio)+Sales.Growth+log(Debt.to.Equity.Ratio))**2+factor(IN
summary(model2)
```

```
##
## Call:
## lm(formula = log(y) ~ (Profitability + sqrt(Cash.Ratio) + Sales.Growth +
##     log(Debt.to.Equity.Ratio))^2 + factor(INDUSTRY), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6257 -0.4730  0.0605  0.5129  4.8469
##
## Coefficients:
##                                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)                            -1.29583    0.89802  -1.443 0.149190
## Profitability                          -1.09990    0.52784  -2.084 0.037313
## sqrt(Cash.Ratio)                       -0.23165    0.13811  -1.677 0.093645
## Sales.Growth                           -0.42498    0.11499  -3.696 0.000225
## log(Debt.to.Equity.Ratio)               0.03946    0.02941   1.341 0.179920
## factor(INDUSTRY)21                      0.03683    0.89562   0.041 0.967200
## factor(INDUSTRY)22                      0.66696    0.89883   0.742 0.458159
## factor(INDUSTRY)23                     -0.37816    0.94083  -0.402 0.687774
## factor(INDUSTRY)31                      0.67962    0.91010   0.747 0.455304
## factor(INDUSTRY)32                      0.46279    0.89955   0.514 0.606988
## factor(INDUSTRY)33                      0.17240    0.89570   0.192 0.847391
```

```
## factor(INDUSTRY)42                            0.41068    0.90794   0.452 0.651088
## factor(INDUSTRY)44                           -0.33193    0.94623  -0.351 0.725782
## factor(INDUSTRY)45                            0.64702    1.01175   0.640 0.522571
## factor(INDUSTRY)48                            0.09942    0.90298   0.110 0.912342
## factor(INDUSTRY)49                            0.74226    1.39856   0.531 0.595664
## factor(INDUSTRY)51                            0.67944    0.90547   0.750 0.453125
## factor(INDUSTRY)52                            0.20650    0.89494   0.231 0.817542
## factor(INDUSTRY)53                            1.25113    0.89762   1.394 0.163533
## factor(INDUSTRY)54                            0.60878    0.94619   0.643 0.520044
## factor(INDUSTRY)56                            0.47626    0.92415   0.515 0.606373
## factor(INDUSTRY)61                            0.53590    0.99721   0.537 0.591054
## factor(INDUSTRY)62                           -0.13718    0.94682  -0.145 0.884816
## factor(INDUSTRY)71                           -0.99198    1.01233  -0.980 0.327263
## factor(INDUSTRY)72                            0.31302    0.93063   0.336 0.736642
## factor(INDUSTRY)81                           -0.01310    1.39776  -0.009 0.992526
## factor(INDUSTRY)99                           -0.67408    1.08941  -0.619 0.536149
## Profitability:sqrt(Cash.Ratio)                0.77117    0.39775   1.939 0.052673
## Profitability:Sales.Growth                   -0.03521    0.18910  -0.186 0.852293
## Profitability:log(Debt.to.Equity.Ratio)       0.22438    0.19192   1.169 0.242491
## sqrt(Cash.Ratio):Sales.Growth                 0.34447    0.21590   1.596 0.110767
## sqrt(Cash.Ratio):log(Debt.to.Equity.Ratio) -0.10816    0.04560  -2.372 0.017785
## Sales.Growth:log(Debt.to.Equity.Ratio)        0.07268    0.05496   1.322 0.186190
##
## (Intercept)
## Profitability                                *
## sqrt(Cash.Ratio)                             .
## Sales.Growth                                 ***
## log(Debt.to.Equity.Ratio)
## factor(INDUSTRY)21
## factor(INDUSTRY)22
## factor(INDUSTRY)23
## factor(INDUSTRY)31
## factor(INDUSTRY)32
## factor(INDUSTRY)33
## factor(INDUSTRY)42
## factor(INDUSTRY)44
## factor(INDUSTRY)45
## factor(INDUSTRY)48
## factor(INDUSTRY)49
## factor(INDUSTRY)51
## factor(INDUSTRY)52
## factor(INDUSTRY)53
## factor(INDUSTRY)54
## factor(INDUSTRY)56
## factor(INDUSTRY)61
## factor(INDUSTRY)62
## factor(INDUSTRY)71
## factor(INDUSTRY)72
## factor(INDUSTRY)81
## factor(INDUSTRY)99
## Profitability:sqrt(Cash.Ratio)               .
## Profitability:Sales.Growth
## Profitability:log(Debt.to.Equity.Ratio)
## sqrt(Cash.Ratio):Sales.Growth
```

```
## sqrt(Cash.Ratio):log(Debt.to.Equity.Ratio) *
## Sales.Growth:log(Debt.to.Equity.Ratio)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.075 on 1886 degrees of freedom
## Multiple R-squared:  0.1302, Adjusted R-squared:  0.1154
## F-statistic: 8.819 on 32 and 1886 DF,  p-value: < 2.2e-16
```

```
model2_adj = lm(log(y)~Profitability+sqrt(Cash.Ratio)+Sales.Growth+log(Debt.to.Equity.Ratio)+factor(INDU
summary(model2_adj)
```

```
##
## Call:
## lm(formula = log(y) ~ Profitability + sqrt(Cash.Ratio) + Sales.Growth +
##     log(Debt.to.Equity.Ratio) + factor(INDUSTRY) + Profitability *
##     sqrt(Cash.Ratio) + Profitability * Sales.Growth + Profitability *
##     log(Debt.to.Equity.Ratio) + sqrt(Cash.Ratio) * log(Debt.to.Equity.Ratio),
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6278 -0.4731  0.0652  0.5107  4.8965
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                              -1.27485    0.89825  -1.419   0.1560
## Profitability                            -1.14424    0.52671  -2.172   0.0299
## sqrt(Cash.Ratio)                         -0.18188    0.13495  -1.348   0.1779
## Sales.Growth                             -0.30846    0.06870  -4.490 7.55e-06
## log(Debt.to.Equity.Ratio)                 0.05032    0.02876   1.749   0.0804
## factor(INDUSTRY)21                       -0.01269    0.89555  -0.014   0.9887
## factor(INDUSTRY)22                        0.62927    0.89896   0.700   0.4840
## factor(INDUSTRY)23                       -0.41393    0.94101  -0.440   0.6601
## factor(INDUSTRY)31                        0.63912    0.91017   0.702   0.4826
## factor(INDUSTRY)32                        0.42965    0.89974   0.478   0.6330
## factor(INDUSTRY)33                        0.12986    0.89576   0.145   0.8848
## factor(INDUSTRY)42                        0.36525    0.90797   0.402   0.6875
## factor(INDUSTRY)44                       -0.37321    0.94633  -0.394   0.6933
## factor(INDUSTRY)45                        0.61076    1.01197   0.604   0.5462
## factor(INDUSTRY)48                        0.08440    0.90326   0.093   0.9256
## factor(INDUSTRY)49                        0.68904    1.39864   0.493   0.6223
## factor(INDUSTRY)51                        0.63224    0.90547   0.698   0.4851
## factor(INDUSTRY)52                        0.16577    0.89503   0.185   0.8531
## factor(INDUSTRY)53                        1.21679    0.89778   1.355   0.1755
## factor(INDUSTRY)54                        0.56327    0.94623   0.595   0.5517
## factor(INDUSTRY)56                        0.43828    0.92430   0.474   0.6354
## factor(INDUSTRY)61                        0.49071    0.99730   0.492   0.6227
## factor(INDUSTRY)62                       -0.17058    0.94698  -0.180   0.8571
## factor(INDUSTRY)71                       -1.03124    1.01251  -1.018   0.3086
## factor(INDUSTRY)72                        0.26921    0.93068   0.289   0.7724
## factor(INDUSTRY)81                       -0.05621    1.39809  -0.040   0.9679
## factor(INDUSTRY)99                       -0.71212    1.08966  -0.654   0.5135
## Profitability:sqrt(Cash.Ratio)            0.88232    0.39336   2.243   0.0250
```

```
## Profitability:Sales.Growth                 -0.03120    0.18506  -0.169   0.8661
## Profitability:log(Debt.to.Equity.Ratio)     0.25640    0.18349   1.397   0.1625
## sqrt(Cash.Ratio):log(Debt.to.Equity.Ratio) -0.10882    0.04551  -2.391   0.0169
##
## (Intercept)
## Profitability                              *
## sqrt(Cash.Ratio)
## Sales.Growth                               ***
## log(Debt.to.Equity.Ratio)                  .
## factor(INDUSTRY)21
## factor(INDUSTRY)22
## factor(INDUSTRY)23
## factor(INDUSTRY)31
## factor(INDUSTRY)32
## factor(INDUSTRY)33
## factor(INDUSTRY)42
## factor(INDUSTRY)44
## factor(INDUSTRY)45
## factor(INDUSTRY)48
## factor(INDUSTRY)49
## factor(INDUSTRY)51
## factor(INDUSTRY)52
## factor(INDUSTRY)53
## factor(INDUSTRY)54
## factor(INDUSTRY)56
## factor(INDUSTRY)61
## factor(INDUSTRY)62
## factor(INDUSTRY)71
## factor(INDUSTRY)72
## factor(INDUSTRY)81
## factor(INDUSTRY)99
## Profitability:sqrt(Cash.Ratio)             *
## Profitability:Sales.Growth
## Profitability:log(Debt.to.Equity.Ratio)
## sqrt(Cash.Ratio):log(Debt.to.Equity.Ratio) *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.076 on 1888 degrees of freedom
## Multiple R-squared:  0.1285, Adjusted R-squared:  0.1147
## F-statistic: 9.282 on 30 and 1888 DF,  p-value: < 2.2e-16
```

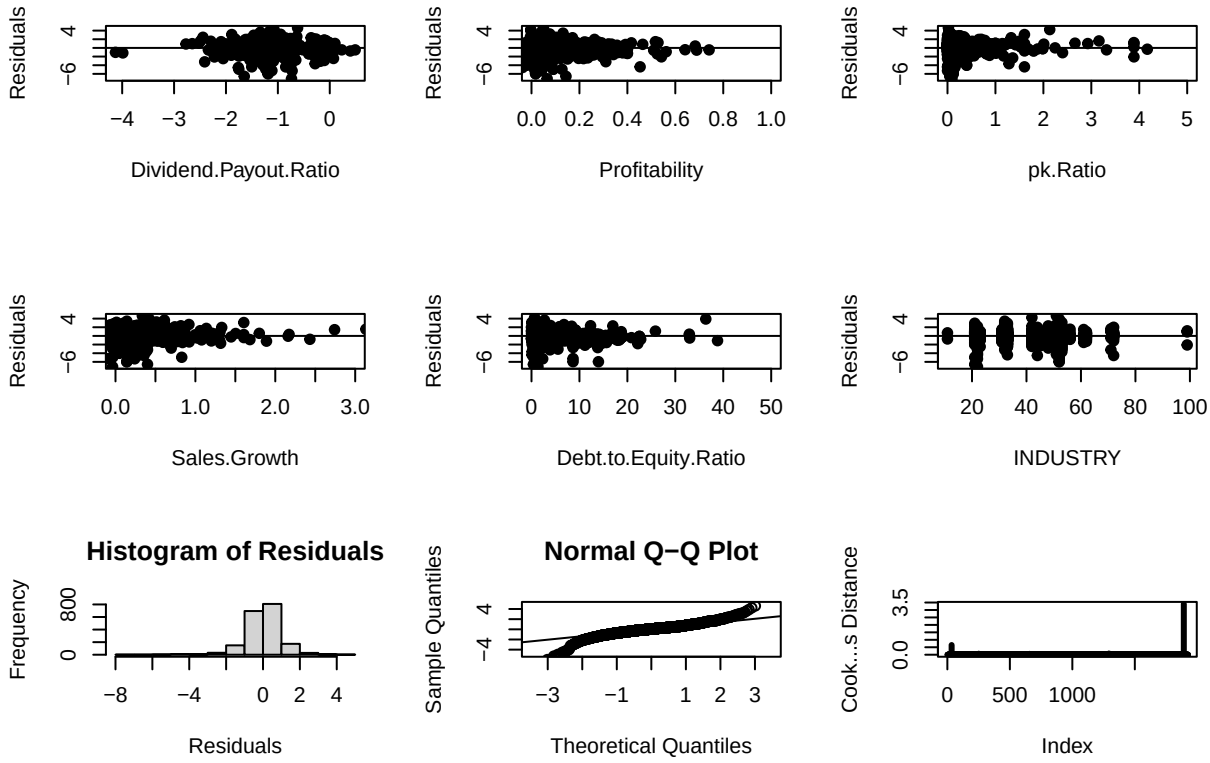
```
anova(model2_adj, model2)
```

```
## Analysis of Variance Table
##
## Model 1: log(y) ~ Profitability + sqrt(Cash.Ratio) + Sales.Growth + log(Debt.to.Equity.Ratio) +
##     factor(INDUSTRY) + Profitability * sqrt(Cash.Ratio) + Profitability *
##     Sales.Growth + Profitability * log(Debt.to.Equity.Ratio) +
##     sqrt(Cash.Ratio) * log(Debt.to.Equity.Ratio)
## Model 2: log(y) ~ (Profitability + sqrt(Cash.Ratio) + Sales.Growth + log(Debt.to.Equity.Ratio))^2 +
##     factor(INDUSTRY)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   1888 2185.5
```

```
## 2   1886 2181.4  2    4.0665 1.7579 0.1727
```

We fail to reject Null hypothesis! so we can choose the model with subset of coefficients

```
res = stdres(model2_adj)
cook = cooks.distance(model2_adj)

par(mfrow = c(3,3))
plot(model2_adj$fitted.values, res, xlab = "Dividend.Payout.Ratio", ylab = "Residuals", pch = 19)
abline(h = 0)
plot(Profitability, res, xlab = "Profitability", ylab = "Residuals",xlim=c(0,1), pch = 19)
abline(h = 0)
plot(Cash.Ratio, res, xlab = "pk.Ratio", ylab = "Residuals",xlim=c(0,5), pch = 19)
abline(h = 0)
plot(Sales.Growth, res, xlab = "Sales.Growth", ylab = "Residuals",xlim=c(0,3), pch = 19)
abline(h = 0)
plot(Debt.to.Equity.Ratio, res, xlab = "Debt.to.Equity.Ratio", ylab = "Residuals",xlim=c(0,50), pch = 19
abline(h = 0)
plot(INDUSTRY, res, xlab = "INDUSTRY", ylab = "Residuals", pch = 19)
abline(h = 0)


hist(res, xlab="Residuals", main= "Histogram of Residuals")

qqnorm(res,ylim=c(-5,5))
qqline(res)

plot(cook,type="h",lwd=3, ylab = "Cook's Distance")
```




```
outliers = which(cook>1)
length(outliers)
```

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

```
data_without_outliers <- data[-outliers, ]
y_without_outliers <- y[-outliers, ]
```

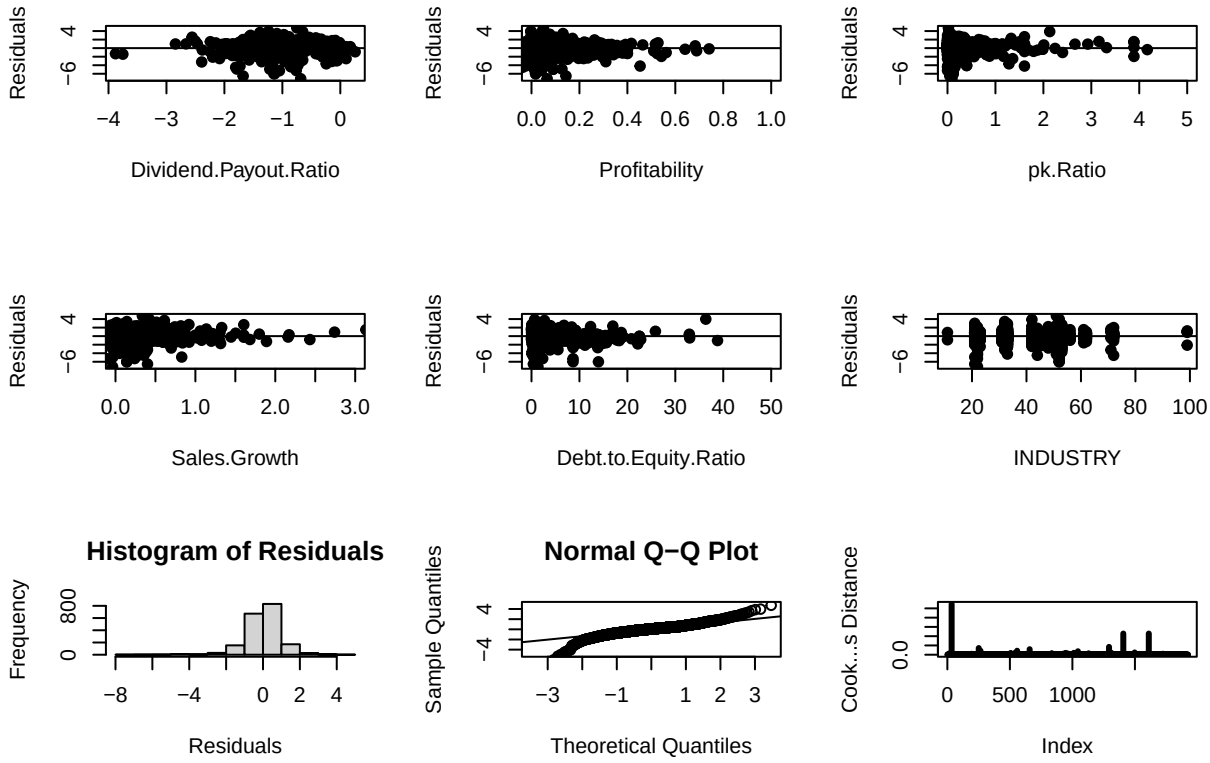
```
model2_no_outlier = lm(log(y_without_outliers)~Profitability+Cash.Ratio+Sales.Growth+log(Debt.to.Equity
summary(model2_no_outlier)
```

```
##
## Call:
## lm(formula = log(y_without_outliers) ~ Profitability + Cash.Ratio +
##     Sales.Growth + log(Debt.to.Equity.Ratio) + factor(INDUSTRY) +
##     Profitability * Cash.Ratio + Profitability * Sales.Growth +
##     Profitability * log(Debt.to.Equity.Ratio) + Cash.Ratio *
##     log(Debt.to.Equity.Ratio), data = data_without_outliers)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6810 -0.4724  0.0710  0.5010  5.0264
##
## Coefficients:
```

```
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             -1.215934   1.024795  -1.187   0.2356
## Profitability                           -0.973699   0.429688  -2.266   0.0236
## Cash.Ratio                               0.123145   0.118029   1.043   0.2969
## Sales.Growth                            -0.403704   0.089343  -4.519 6.61e-06
## log(Debt.to.Equity.Ratio)                0.023514   0.026365   0.892   0.3726
## factor(INDUSTRY)21                      -0.141585   1.026244  -0.138   0.8903
## factor(INDUSTRY)22                       0.557013   1.028010   0.542   0.5880
## factor(INDUSTRY)23                      -0.555429   1.064518  -0.522   0.6019
## factor(INDUSTRY)31                       0.512179   1.037743   0.494   0.6217
## factor(INDUSTRY)32                       0.304133   1.029017   0.296   0.7676
## factor(INDUSTRY)33                      -0.003287   1.026560  -0.003   0.9974
## factor(INDUSTRY)42                       0.251003   1.038201   0.242   0.8090
## factor(INDUSTRY)44                      -0.515907   1.069711  -0.482   0.6297
## factor(INDUSTRY)45                       0.490714   1.128984   0.435   0.6639
## factor(INDUSTRY)48                      -0.038151   1.032897  -0.037   0.9705
## factor(INDUSTRY)49                       0.548673   1.484986   0.369   0.7118
## factor(INDUSTRY)51                       0.477380   1.034785   0.461   0.6446
## factor(INDUSTRY)52                       0.060252   1.024556   0.059   0.9531
## factor(INDUSTRY)53                       1.123363   1.026511   1.094   0.2739
## factor(INDUSTRY)54                       0.446583   1.069588   0.418   0.6763
## factor(INDUSTRY)56                       0.297977   1.049888   0.284   0.7766
## factor(INDUSTRY)61                       0.330825   1.115205   0.297   0.7668
## factor(INDUSTRY)62                      -0.306744   1.070476  -0.287   0.7745
## factor(INDUSTRY)71                      -1.155651   1.131273  -1.022   0.3071
## factor(INDUSTRY)72                       0.146782   1.055597   0.139   0.8894
## factor(INDUSTRY)81                      -0.111165   1.485992  -0.075   0.9404
## factor(INDUSTRY)99                      -0.861155   1.199076  -0.718   0.4727
## Profitability:Cash.Ratio                 0.183212   0.148007   1.238   0.2159
## Profitability:Sales.Growth               0.599749   0.443662   1.352   0.1766
## Profitability:log(Debt.to.Equity.Ratio)  0.276150   0.212495   1.300   0.1939
## Cash.Ratio:log(Debt.to.Equity.Ratio)     0.014456   0.026234   0.551   0.5817
##
## (Intercept)
## Profitability                           *
## Cash.Ratio
## Sales.Growth                            ***
## log(Debt.to.Equity.Ratio)
## factor(INDUSTRY)21
## factor(INDUSTRY)22
## factor(INDUSTRY)23
## factor(INDUSTRY)31
## factor(INDUSTRY)32
## factor(INDUSTRY)33
## factor(INDUSTRY)42
## factor(INDUSTRY)44
## factor(INDUSTRY)45
## factor(INDUSTRY)48
## factor(INDUSTRY)49
## factor(INDUSTRY)51
## factor(INDUSTRY)52
## factor(INDUSTRY)53
## factor(INDUSTRY)54
## factor(INDUSTRY)56
```

```
## factor(INDUSTRY)61
## factor(INDUSTRY)62
## factor(INDUSTRY)71
## factor(INDUSTRY)72
## factor(INDUSTRY)81
## factor(INDUSTRY)99
## Profitability:Cash.Ratio
## Profitability:Sales.Growth
## Profitability:log(Debt.to.Equity.Ratio)
## Cash.Ratio:log(Debt.to.Equity.Ratio)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.077 on 1887 degrees of freedom
## Multiple R-squared:  0.1264, Adjusted R-squared:  0.1125
## F-statistic:   9.1 on 30 and 1887 DF,  p-value: < 2.2e-16
```

```
res = stdres(model2_no_outlier)
cook = cooks.distance(model2_no_outlier)

par(mfrow = c(3,3))
plot(model2_no_outlier$fitted.values, res, xlab = "Dividend.Payout.Ratio", ylab = "Residuals", pch = 19)
abline(h = 0)
plot(Profitability[-outliers, ], res, xlab = "Profitability", ylab = "Residuals",xlim=c(0,1), pch = 19)
abline(h = 0)
plot(Cash.Ratio[-outliers, ], res, xlab = "pk.Ratio", ylab = "Residuals",xlim=c(0,5), pch = 19)
abline(h = 0)
plot(Sales.Growth[-outliers, ], res, xlab = "Sales.Growth", ylab = "Residuals",xlim=c(0,3), pch = 19)
abline(h = 0)
plot(Debt.to.Equity.Ratio[-outliers, ], res, xlab = "Debt.to.Equity.Ratio", ylab = "Residuals",xlim=c(0
abline(h = 0)
plot(INDUSTRY[-outliers], res, xlab = "INDUSTRY", ylab = "Residuals", pch = 19)
abline(h = 0)


hist(res, xlab="Residuals", main= "Histogram of Residuals")

qqnorm(res,ylim=c(-5,5))
qqline(res)

plot(cook,type="h",lwd=3, ylab = "Cook's Distance")
```



4.5 Model Selection

Model selection

```
detach(data)

test=data_99[-picked,]
```

```
pred1 <- predict(model1_trans, test, interval = 'prediction')
test.pred1 <- pred1[,1]
test.lwr1 <- pred1[,2]
test.upr1 <- pred1[,3]
# Mean Squared Prediction Error (MSPE)
mean((test.pred1-test$Dividend.Payout.Ratio)^2)
```

```
## [1] 5.875187
```

```
# Mean Absolute Prediction Error (MAE)
mean(abs(test.pred1-test$Dividend.Payout.Ratio))
```

```
## [1] 1.778934
```

```r
# Mean Absolute Percentage Error (MAPE)
mean(abs(test.pred1-test$Dividend.Payout.Ratio)/test$Dividend.Payout.Ratio)
```

```
## [1] 10.80136
```

```r
# Precision Measure (PM)
sum((test.pred1-test$Dividend.Payout.Ratio)^2)/sum((test$Dividend.Payout.Ratio-mean(test$Dividend.Payou
```

```
## [1] 2.109635
```

```r
# CI Measure (CIM)
(sum(test$Dividend.Payout.Ratio<test.lwr1)+sum(test$Dividend.Payout.Ratio>test.upr1))/nrow(test)
```

```
## [1] 0.11875
```

```r
pred1 <- predict(model2_adj, test, interval = 'prediction')
test.pred1 <- pred1[,1]
test.lwr1 <- pred1[,2]
test.upr1 <- pred1[,3]
# Mean Squared Prediction Error (MSPE)
mean((test.pred1-test$Dividend.Payout.Ratio)^2)
```

```
## [1] 5.87487
```

```r
# Mean Absolute Prediction Error (MAE)
mean(abs(test.pred1-test$Dividend.Payout.Ratio))
```

```
## [1] 1.779311
```

```r
# Mean Absolute Percentage Error (MAPE)
mean(abs(test.pred1-test$Dividend.Payout.Ratio)/test$Dividend.Payout.Ratio)
```

```
## [1] 10.66672
```

```r
# Precision Measure (PM)
sum((test.pred1-test$Dividend.Payout.Ratio)^2)/sum((test$Dividend.Payout.Ratio-mean(test$Dividend.Payou
```

```
## [1] 2.109521
```

```r
# CI Measure (CIM)
(sum(test$Dividend.Payout.Ratio<test.lwr1)+sum(test$Dividend.Payout.Ratio>test.upr1))/nrow(test)
```

```
## [1] 0.11875
```

```r
pred1 <- predict(model2_no_outlier, test, interval = 'prediction')
test.pred1 <- pred1[,1]
test.lwr1 <- pred1[,2]
test.upr1 <- pred1[,3]
# Mean Squared Prediction Error (MSPE)
mean((test.pred1-test$Dividend.Payout.Ratio)^2)
```

```
## [1] 5.864595
```

```
# Mean Absolute Prediction Error (MAE)
mean(abs(test.pred1-test$Dividend.Payout.Ratio))
```

```
## [1] 1.773047
```

```
# Mean Absolute Percentage Error (MAPE)
mean(abs(test.pred1-test$Dividend.Payout.Ratio)/test$Dividend.Payout.Ratio)
```

```
## [1] 10.21335
```

```
# Precision Measure (PM)
sum((test.pred1-test$Dividend.Payout.Ratio)^2)/sum((test$Dividend.Payout.Ratio-mean(test$Dividend.Payou
```

```
## [1] 2.105832
```

```
# CI Measure (CIM)
(sum(test$Dividend.Payout.Ratio<test.lwr1)+sum(test$Dividend.Payout.Ratio>test.upr1))/nrow(test)
```

```
## [1] 0.1104167
```