liquid<-read.csv(file.choose(), header=TRUE)

#file name is Liquidity

attach(liquid)

summary(liquid)

AVGT VOLUME NTRAN PRICE

Min. : 0.590 Min. : 0.658 Min. : 999 Min. : 9.125

1st Qu.: 2.916 1st Qu.: 5.841 1st Qu.: 3090 1st Qu.: 24.562

Median : 4.284 Median :11.556 Median : 5071 Median : 34.375

Mean : 5.441 Mean :13.423 Mean : 6436 Mean : 38.796

3rd Qu.: 7.010 3rd Qu.:17.509 3rd Qu.: 7414 3rd Qu.: 46.312

Max. :20.772 Max. :64.572 Max. :36420 Max. :122.375

SHARE VALUE DEBEQ TIC

Min. : 6.736 Min. : 0.1150 Min. : 0.1850 AA : 1

1st Qu.: 28.402 1st Qu.: 0.9975 1st Qu.: 0.8005 AAL : 1

Median : 53.827 Median : 2.0650 Median : 1.1050 AEP : 1

Mean : 94.731 Mean : 4.1160 Mean : 2.6973 AGE : 1

3rd Qu.:112.028 3rd Qu.: 3.9755 3rd Qu.: 1.7440 AHS : 1

Max. :783.051 Max. :75.4370 Max. :53.6280 ALN : 1

(Other):117

COMPANY

ALEXANDER & ALEX SVCS: 1

ALLEN GROUP INC : 1

ALUMINUM CO AMER : 1

AMERICAN ELEC PWR INC: 1

AMERICAN HOSP SUPPLY : 1

AMOCO CORP : 1

(Other) :117

> model1 =lm(VOLUME ~ PRICE + SHARE +VALUE)

> CooksD = cooks.distance(model1)

> CooksD [order(CooksD)]

86 117 7 97 99 87

8.512740e-07 1.997072e-06 4.437384e-06 4.438645e-06 6.056272e-06 1.411033e-05

25 29 16 33 90 17

1.640481e-05 3.240051e-05 3.502571e-05 4.177816e-05 4.381435e-05 5.168936e-05

18 100 68 115 19 83

6.743720e-05 7.593679e-05 7.667741e-05 8.290429e-05 8.312731e-05 1.054776e-04

5 94 49 92 20 79

1.141920e-04 1.155801e-04 1.176477e-04 1.225693e-04 2.309967e-04 2.333564e-04

98 73 108 28 67 12

2.395078e-04 2.681925e-04 2.837674e-04 3.271342e-04 3.364762e-04 3.580106e-04

24 112 80 120 41 72

3.987282e-04 4.322680e-04 4.347117e-04 4.439034e-04 4.833492e-04 5.697235e-04

50 78 65 1 43 55

6.052000e-04 6.197630e-04 6.675922e-04 6.881457e-04 6.922516e-04 6.953259e-04

93 84 58 119 95 121

7.202983e-04 7.462715e-04 7.643830e-04 8.310743e-04 8.558718e-04 9.042452e-04

56 88 89 40 15 101

9.166063e-04 9.418038e-04 9.608331e-04 9.764616e-04 9.836350e-04 1.072595e-03

118 106 2 44 14 23

1.076330e-03 1.103148e-03 1.125888e-03 1.155296e-03 1.194263e-03 1.352292e-03

96 62 26 54 46 21

1.419620e-03 1.605116e-03 1.612270e-03 1.641489e-03 1.677308e-03 1.704214e-03

70 4 51 52 116 34

1.780044e-03 1.979036e-03 2.000667e-03 2.085893e-03 2.146823e-03 2.217030e-03

76 81 123 114 74 38

2.303865e-03 2.308306e-03 2.335830e-03 2.591559e-03 2.685451e-03 2.701212e-03

109 91 30 107 63 10

2.717886e-03 2.743880e-03 2.845332e-03 2.913305e-03 3.006660e-03 3.013045e-03

35 61 69 57 31 110

3.106542e-03 3.170492e-03 3.171394e-03 3.211063e-03 3.353191e-03 3.493573e-03

39 111 82 9 22 27

3.516210e-03 3.653256e-03 3.720325e-03 3.853619e-03 4.291861e-03 4.731934e-03

66 75 32 113 47 104

5.570043e-03 5.990115e-03 6.412144e-03 6.610613e-03 6.789477e-03 7.129163e-03

102 42 3 11 6 8

7.193953e-03 7.198898e-03 7.697297e-03 7.849408e-03 8.186637e-03 8.871407e-03

36 105 48 59 13 64

1.328547e-02 1.519727e-02 1.952481e-02 1.985911e-02 2.217454e-02 2.421815e-02

77 71 103 85 37 45

2.422240e-02 2.446598e-02 4.136422e-02 5.701303e-02 5.980794e-02 6.633302e-02

53 122 60

1.060768e-01 1.598960e+00 2.595438e+00

> dim(liquid)

[1] 123 9

>

> #n=123,k = # X's in model=3

##df1=k+1; df2=n-(k+1)

> qf(.95,df1=4,df2=119)

[1] 2.447881

> # Only obs 60 above has exceeded the cutoff point,

> #so according to Cook's D,it has an unusual high value

\*\*\*\*

> model1 =lm(VOLUME ~ PRICE + SHARE +VALUE)

> summary(model1)

Call:

lm(formula = VOLUME ~ PRICE + SHARE + VALUE)

Residuals:

Min 1Q Median 3Q Max

-20.708 -4.179 -1.091 3.108 28.230

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.90916 1.60140 4.939 2.59e-06 \*\*\*

PRICE -0.02224 0.03503 -0.635 0.5267

SHARE 0.05372 0.01034 5.194 8.63e-07 \*\*\*

VALUE 0.31271 0.16150 1.936 0.0552 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.722 on 119 degrees of freedom

Multiple R-squared: 0.6101, Adjusted R-squared: 0.6003

F-statistic: 62.07 on 3 and 119 DF, p-value: < 2.2e-16

\*\*

> model1b =lm(VOLUME ~ PRICE + SHARE +VALUE,subset = -c(60))

> summary(model1b)

Call:

lm(formula = VOLUME ~ PRICE + SHARE + VALUE, subset = -c(60))

Residuals:

Min 1Q Median 3Q Max

-21.233 -4.026 -1.256 3.004 28.531

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.899332 1.800752 3.831 0.000206 \*\*\*

PRICE 0.002701 0.040519 0.067 0.946974

SHARE 0.071199 0.017687 4.026 0.000101 \*\*\*

VALUE -0.124973 0.394114 -0.317 0.751730

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.709 on 118 degrees of freedom

Multiple R-squared: 0.5239, Adjusted R-squared: 0.5118

F-statistic: 43.28 on 3 and 118 DF, p-value: < 2.2e-16

> #R -sqr both got worse, p-val is the same val as it was

##The bill Gate VS the norm scenario

> # VALUE variable was borderline significant, now it is definitely not significant, by the p-value

> #s went down and got a little better.

> model1 =lm(VOLUME ~ PRICE + SHARE +VALUE)

> outlier<-read.csv(file.choose(), header=TRUE)

> attach(outlier)

> outlier

X Y CODES

1 1.5 3.0 0

2 1.7 2.5 0

3 2.0 3.5 0

4 2.2 3.0 0

5 2.5 3.1 0

6 2.5 3.6 0

7 2.7 3.2 0

8 2.9 3.9 0

9 3.0 4.0 0

10 3.5 4.0 0

11 3.8 4.2 0

12 4.2 4.1 0

13 4.3 4.8 0

14 4.6 4.2 0

15 4.9 5.1 0

16 5.0 5.1 0

17 5.1 5.1 0

18 5.2 4.8 0

19 5.5 5.3 0

20 4.3 8.0 1

21 9.5 8.0 2

22 9.5 2.5 3

> model2 = lm(Y~X)

> rstandard = rstandard(model2)

> rstandard [order(rstandard)]

22 2 4 5 7 1

-3.510341763 -0.808881406 -0.534226891 -0.533343503 -0.506477473 -0.347210514

14 12 6 10 3 11

-0.226416006 -0.196765362 -0.133073721 -0.086378873 -0.076639711 -0.009597529

8 9 18 13 17 16

-0.002735013 0.049461176 0.085473019 0.328564021 0.350587017 0.377404358

19 15 21 20

0.402545525 0.404210905 1.754151769 2.853960893

> # Outliers are outside of +-2,so obs 22 and 20 are outliers

>

> leverages = hatvalues(model2)

> leverages [order(leverages)]

12 13 20 11 14 10 15

0.04554154 0.04583820 0.04583820 0.04646022 0.04799135 0.04935980 0.05203928

16 17 18 9 8 19 7

0.05380964 0.05579054 0.05798197 0.05840303 0.06084327 0.06581944 0.06635534

5 6 4 3 2 1 21

0.07270953 0.07270953 0.08381978 0.09227928 0.10654749 0.11711229 0.35137515

22

0.35137515

> #Cutoff = 3(k+1)/n = 3\*2/22=.2727

> #Obs 21 and 22 are high leverage points

>

> CooksD = cooks.distance(model2)

> CooksD [order(CooksD)]

8 11 9 10 18 3

2.423055e-07 2.244044e-06 7.586985e-05 1.937056e-04 2.248339e-04 2.985587e-04

6 12 14 13 17 16

6.942728e-04 9.236725e-04 1.292130e-03 2.593078e-03 3.631231e-03 4.050097e-03

15 19 1 7 5 4

4.484633e-03 5.708518e-03 7.995631e-03 9.115584e-03 1.115217e-02 1.305531e-02

2 20 21 22

3.901319e-02 1.956463e-01 8.334543e-01 3.337692e+00

> qf(.95,2 20)

Error: unexpected numeric constant in "qf(.95,2 20"

> qf(.95,2 ,20)

[1] 3.492828

> #No obs are high enough to exceed