Tracking Returns

Code

Casey Tirshfield

The file `d_nasdaq_82stocks.txt` contains the daily log returns of the NASDAQ Composite Index and 82 stocks from January 3, 1990 to December 29, 2006. We want to track the returns of NASDAQ by using a small number of stocks from the given 82 stocks.

Hide

```
# import .txt file as data frame
df <- read.table('d_nasdaq_82stocks.txt', header=TRUE)
# in order to avoid issues with naming, we change the ticker for Ford Motor Company f
rom F to FORD, and the ticker for Brown-Forman Corporation Class B from BF-B to BFB
names(df)[23] <- 'BFB'
names(df)[51] <- 'FORD'
# we make the headers of our data frame callable objects
attach(df)
```

(a) Construct a full regression model.

Hide

```
# this question's code is inspired by the sample code found at the following link htt
ps://web.stanford.edu/~xing/statfinbook/_BookFun/chap1_ana_data.txt
fit_full <- lm(Nasdaq ~., data=df[,-1])
summary(fit_full)
```

```

Call:
lm(formula = Nasdaq ~ ., data = df[, -1])

Residuals:
       Min         1Q     Median         3Q        Max
-0.0288872 -0.0014119  0.0000462  0.0015263  0.0173120

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.597e-05  4.454e-05  -1.930 0.053679 .
```

```
AA           1.824e-03  2.745e-03   0.665 0.506407
AAPL         1.474e-02  1.540e-03   9.570  < 2e-16 ***
AET          2.254e-03  2.371e-03   0.951 0.341774
ALTR         2.266e-02  1.455e-03  15.570  < 2e-16 ***
AMAT         2.231e-02  1.726e-03  12.929  < 2e-16 ***
AMD          7.580e-03  1.318e-03   5.753 9.40e-09 ***
AMGN         2.978e-02  2.027e-03  14.689  < 2e-16 ***
AOC          5.162e-03  2.478e-03   2.083 0.037294 *
APA          2.686e-03  2.493e-03   1.078 0.281210
APD         -8.368e-03  3.233e-03  -2.588 0.009681 **
ASH          6.134e-03  3.328e-03   1.843 0.065348 .
AT           2.123e-02  3.170e-03   6.697 2.41e-11 ***
AVY         -1.905e-04  2.989e-03  -0.064 0.949174
AXP          7.856e-03  3.004e-03   2.615 0.008950 **
AYE          1.275e-04  2.315e-03   0.055 0.956061
BA           5.296e-03  2.725e-03   1.943 0.052075 .
BAC         -4.788e-03  3.297e-03  -1.452 0.146551
BAX         -3.604e-03  2.745e-03  -1.313 0.189264
BBY          7.881e-03  1.338e-03   5.889 4.19e-09 ***
BC           5.916e-03  2.332e-03   2.537 0.011208 *
BFB         -3.074e-03  3.644e-03  -0.844 0.398978
BLL          4.434e-03  2.824e-03   1.570 0.116560
BMY          1.811e-03  3.069e-03   0.590 0.555232
BNI          3.887e-04  3.173e-03   0.122 0.902524
C            1.082e-02  3.044e-03   3.555 0.000382 ***
CAH          3.707e-03  2.390e-03   1.551 0.120954
CAT         -4.740e-03  2.837e-03  -1.671 0.094887 .
CCE          3.300e-03  2.322e-03   1.421 0.155258
CCL          7.846e-03  2.267e-03   3.461 0.000544 ***
CEG         -8.797e-03  4.161e-03  -2.114 0.034558 *
CFC          6.246e-03  2.115e-03   2.952 0.003170 **
CLX          1.114e-03  3.002e-03   0.371 0.710563
CMI          2.473e-03  2.533e-03   0.976 0.328942
COP         -3.952e-03  3.548e-03  -1.114 0.265414
COST         1.202e-02  2.235e-03   5.379 7.89e-08 ***
CSX         -1.959e-03  3.120e-03  -0.628 0.530244
CVX         -1.141e-03  4.255e-03  -0.268 0.788613
DD          -1.688e-02  3.387e-03  -4.983 6.51e-07 ***
DELL         1.511e-02  1.517e-03   9.964  < 2e-16 ***
DIS          1.248e-02  2.744e-03   4.549 5.55e-06 ***
DOV          5.097e-03  3.388e-03   1.504 0.132596
DUK         -2.919e-03  4.001e-03  -0.730 0.465736
ED          -1.407e-02  5.246e-03  -2.683 0.007336 **
EDS          6.362e-03  1.883e-03   3.379 0.000734 ***
EIX         -4.434e-03  2.348e-03  -1.888 0.059040 .
EK           2.272e-03  2.532e-03   0.897 0.369664
EOG          2.406e-03  2.435e-03   0.988 0.323128
EXC         -3.029e-03  3.872e-03  -0.782 0.434071
FORD        -5.682e-03  2.752e-03  -2.065 0.038997 *
FPL         -4.521e-04  5.072e-03  -0.089 0.928981
```

```
FRX          8.976e-03  2.183e-03   4.111 4.01e-05 ***
GE           3.303e-02  3.949e-03   8.365  < 2e-16 ***
GM           4.229e-03  2.829e-03   1.495 0.135017
GPS          3.683e-03  1.935e-03   1.903 0.057133 .
GT          -4.592e-04  2.135e-03  -0.215 0.829681
HD          -1.197e-03  2.813e-03  -0.425 0.670507
HLT          3.403e-04  2.250e-03   0.151 0.879795
HON          1.036e-02  2.667e-03   3.885 0.000104 ***
HPC          4.338e-03  2.625e-03   1.653 0.098506 .
HPQ          2.027e-02  2.152e-03   9.417  < 2e-16 ***
HRB          2.853e-03  2.566e-03   1.112 0.266295
HSY         -1.075e-02  3.469e-03  -3.100 0.001945 **
HUM          3.524e-03  1.730e-03   2.037 0.041757 *
IBM          1.947e-02  2.833e-03   6.872 7.24e-12 ***
INTC         4.606e-02  2.336e-03  19.717  < 2e-16 ***
IPG          1.324e-02  2.339e-03   5.659 1.62e-08 ***
JNJ         -1.286e-02  3.724e-03  -3.454 0.000558 ***
JPM          1.274e-02  2.913e-03   4.374 1.25e-05 ***
JWN          2.014e-03  1.949e-03   1.033 0.301561
K           -4.565e-03  3.274e-03  -1.394 0.163361
KBH          4.928e-03  2.065e-03   2.387 0.017041 *
KMI          7.191e-03  2.459e-03   2.924 0.003471 **
KO          -1.361e-02  3.609e-03  -3.770 0.000165 ***
KR          -5.139e-03  2.270e-03  -2.264 0.023637 *
LEG          9.504e-03  2.865e-03   3.317 0.000917 ***
LEN          9.544e-03  2.287e-03   4.173 3.07e-05 ***
LLY         -4.022e-03  2.931e-03  -1.372 0.170060
LM           2.092e-02  2.592e-03   8.072 8.93e-16 ***
LMT         -1.871e-03  2.705e-03  -0.692 0.489179
LOW          3.257e-03  2.318e-03   1.405 0.160068
LUV          5.386e-03  2.128e-03   2.531 0.011411 *
MAS         -6.793e-03  2.719e-03  -2.498 0.012535 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0029 on 4203 degrees of freedom
Multiple R-squared:  0.8065,    Adjusted R-squared:  0.8027
F-statistic: 213.6 on 82 and 4203 DF,  p-value: < 2.2e-16
```

(b) Use partial F-statistics and backward elimination to select variables form the full regression model in (a). Write down the selected model.

Hide

```
# tries all one-term deletions from the model and removes the least significant expla
natory variable of each iteration until all explanatory variables have a p-value of l
ess than alpha
backward_elimination <- function(data, alpha) {
  while (1) {
    model <- lm(Nasdaq ~ ., data=data)
    temp <- dropterm(model, test='F')
    stock_removed = which.min(temp$F)
    p_value <- temp$`Pr(F)`[stock_removed]
    if (p_value < alpha) {
      break;
    }
    data = data[, -stock_removed];
  }
  data
}
selected <- backward_elimination(df[,-1], 0.05);
summary(selected)
```

```
     Nasdaq                AAPL                 ALTR                 AMAT
AMD                   AMGN
 Min.   :-0.0441609   Min.   :-0.7308588   Min.   :-0.3157970   Min.   :-0.1880522
Min.   :-0.4769241   Min.   :-0.2240090
 1st Qu.:-0.0026999   1st Qu.:-0.0166505   1st Qu.:-0.0202560   1st Qu.:-0.0201772
1st Qu.:-0.0198026   1st Qu.:-0.0125642
 Median : 0.0005034   Median : 0.0000000   Median : 0.0000000   Median : 0.0000000
Median : 0.0000000   Median : 0.0000000
 Mean   : 0.0001682   Mean   : 0.0005314   Mean   : 0.0008815   Mean   : 0.0008658
Mean   : 0.0003727   Mean   : 0.0009698
 3rd Qu.: 0.0032769   3rd Qu.: 0.0171961   3rd Qu.: 0.0220693   3rd Qu.: 0.0213710
3rd Qu.: 0.0205771   3rd Qu.: 0.0142315
 Max.   : 0.0575642   Max.   : 0.2866694   Max.   : 0.2247073   Max.   : 0.2281850
Max.   : 0.2381447   Max.   : 0.1406494
      AOC                  APD                  ASH                   AT
AXP                    BA
 Min.   :-0.3611045   Min.   :-0.1118734   Min.   :-0.1854172   Min.   :-0.1251912
Min.   :-0.1461158   Min.   :-0.1938306
 1st Qu.:-0.0078238   1st Qu.:-0.0090580   1st Qu.:-0.0079640   1st Qu.:-0.0080953
1st Qu.:-0.0103951   1st Qu.:-0.0100884
 Median : 0.0000000   Median : 0.0000000   Median : 0.0000000   Median : 0.0000000
Median : 0.0000000   Median : 0.0000000
 Mean   : 0.0003561   Mean   : 0.0004858   Mean   : 0.0003012   Mean   : 0.0004307
Mean   : 0.0005206   Mean   : 0.0004106
 3rd Qu.: 0.0091244   3rd Qu.: 0.0099811   3rd Qu.: 0.0092309   3rd Qu.: 0.0086336
3rd Qu.: 0.0113347   3rd Qu.: 0.0107299
 Max.   : 0.1997797   Max.   : 0.1122120   Max.   : 0.0821425   Max.   : 0.1177031
Max.   : 0.1198440   Max.   : 0.1098945
      BBY                   BC                    C                   CCL
CEG                   CFC
```

```
 Min.   :-0.490829   Min.   :-0.1563542   Min.   :-0.1708447   Min.   :-0.3838365   M
in.   :-0.2394297   Min.   :-0.1271005
 1st Qu.:-0.016000   1st Qu.:-0.0109104   1st Qu.:-0.0102296   1st Qu.:-0.0110978   1
st Qu.:-0.0067715   1st Qu.:-0.0126284
 Median : 0.000000   Median : 0.0000000   Median : 0.0000000   Median : 0.0000000   M
edian : 0.0000000   Median : 0.0000000
 Mean   : 0.001262   Mean   : 0.0002761   Mean   : 0.0008202   Mean   : 0.0005948   M
ean   : 0.0004482   Mean   : 0.0009504
 3rd Qu.: 0.018479   3rd Qu.: 0.0113409   3rd Qu.: 0.0117187   3rd Qu.: 0.0116921   3
rd Qu.: 0.0080498   3rd Qu.: 0.0139302
 Max.   : 0.208311   Max.   : 0.1753633   Max.   : 0.1680757   Max.   : 0.1527422   M
ax.   : 0.1228797   Max.   : 0.1978257
      COST                 DD                  DELL                DIS
ED                  EDS
 Min.   :-0.2677794   Min.   :-0.1093209   Min.   :-0.277632   Min.   :-0.2031775   M
in.   :-0.0695699   Min.   :-0.7510589
 1st Qu.:-0.0127526   1st Qu.:-0.0095107   1st Qu.:-0.015614   1st Qu.:-0.0104395   1
st Qu.:-0.0067014   1st Qu.:-0.0104599
 Median : 0.0000000   Median : 0.0000000   Median : 0.000000   Median : 0.0000000   M
edian : 0.0000000   Median : 0.0000000
 Mean   : 0.0003597   Mean   : 0.0003229   Mean   : 0.001408   Mean   : 0.0003275   M
ean   : 0.0003557   Mean   : 0.0002211
 3rd Qu.: 0.0132804   3rd Qu.: 0.0097128   3rd Qu.: 0.018962   3rd Qu.: 0.0107027   3
rd Qu.: 0.0072938   3rd Qu.: 0.0115512
 Max.   : 0.1774138   Max.   : 0.0939340   Max.   : 0.187742   Max.   : 0.1423899   M
ax.   : 0.0904406   Max.   : 0.1491984
      EIX                 FRX                  GE                  GPS
HON                  HPQ
 Min.   :-0.4273862   Min.   :-0.2767231   Min.   :-0.1128923   Min.   :-0.2366094
Min.   :-0.1909042   Min.   :-0.200151
 1st Qu.:-0.0076689   1st Qu.:-0.0110103   1st Qu.:-0.0080678   1st Qu.:-0.0134817
1st Qu.:-0.0098796   1st Qu.:-0.012980
 Median : 0.0000000   Median : 0.0000000   Median : 0.0000000   Median : 0.0000000
Median : 0.0000000   Median : 0.000000
 Mean   : 0.0003543   Mean   : 0.0006935   Mean   : 0.0005367   Mean   : 0.0005706
Mean   : 0.0004709   Mean   : 0.000547
 3rd Qu.: 0.0086233   3rd Qu.: 0.0122489   3rd Qu.: 0.0093230   3rd Qu.: 0.0146702
3rd Qu.: 0.0106013   3rd Qu.: 0.013908
 Max.   : 0.3024843   Max.   : 0.1603890   Max.   : 0.1174122   Max.   : 0.1751452
Max.   : 0.2484743   Max.   : 0.159253
      HSY                 HUM                  IBM                 INTC
IPG                  JNJ
 Min.   :-0.1283812   Min.   :-0.2976997   Min.   :-0.1688993   Min.   :-0.2487877
Min.   :-0.3534155   Min.   :-0.1725264
 1st Qu.:-0.0071985   1st Qu.:-0.0134747   1st Qu.:-0.0095253   1st Qu.:-0.0144195
1st Qu.:-0.0096853   1st Qu.:-0.0080216
 Median : 0.0000000   Median : 0.0000000   Median : 0.0000000   Median : 0.0000000
Median : 0.0000000   Median : 0.0000000
 Mean   : 0.0004786   Mean   : 0.0004477   Mean   : 0.0003897   Mean   : 0.0006914
Mean   : 0.0002314   Mean   : 0.0005772
```

```
 3rd Qu.: 0.0079745   3rd Qu.: 0.0144927   3rd Qu.: 0.0097648   3rd Qu.: 0.0158764
3rd Qu.: 0.0101508   3rd Qu.: 0.0086888
 Max.   : 0.2255167   Max.   : 0.1469095   Max.   : 0.1236629   Max.   : 0.1832673
Max.   : 0.1867144   Max.   : 0.0789695
      JPM                 KBH                 KMI                 KO
KR                  LEG
 Min.   :-0.199319   Min.   :-0.1916917   Min.   :-0.3323115   Min.   :-0.1108306   M
in.   :-0.2949261   Min.   :-0.216096
 1st Qu.:-0.010987   1st Qu.:-0.0148953   1st Qu.:-0.0077043   1st Qu.:-0.0077254   1
st Qu.:-0.0106225   1st Qu.:-0.008459
 Median : 0.000000   Median : 0.0000000   Median : 0.0000000   Median : 0.0000000   M
edian : 0.0000000   Median : 0.000000
 Mean   : 0.000517   Mean   : 0.0005156   Mean   : 0.0006464   Mean   : 0.0004365   M
ean   : 0.0004202   Mean   : 0.000514
 3rd Qu.: 0.011055   3rd Qu.: 0.0152638   3rd Qu.: 0.0088541   3rd Qu.: 0.0088047   3
rd Qu.: 0.0109494   3rd Qu.: 0.009559
 Max.   : 0.148359   Max.   : 0.1745596   Max.   : 0.4372542   Max.   : 0.0935826   M
ax.   : 0.1307342   Max.   : 0.159386
      LEN                 LM                  LUV                 MAS
 Min.   :-0.2809803   Min.   :-0.1892698   Min.   :-0.2752111   Min.   :-0.1740038
 1st Qu.:-0.0121708   1st Qu.:-0.0096546   1st Qu.:-0.0132730   1st Qu.:-0.0096001
 Median : 0.0000000   Median : 0.0000000   Median : 0.0000000   Median : 0.0000000
 Mean   : 0.0009141   Mean   : 0.0008879   Mean   : 0.0006299   Mean   : 0.0002951
 3rd Qu.: 0.0128193   3rd Qu.: 0.0112263   3rd Qu.: 0.0139167   3rd Qu.: 0.0101298
 Max.   : 0.1495949   Max.   : 0.1425047   Max.   : 0.1436243   Max.   : 0.1128037
```

Hide

```
# now we write down the selected model, which contains only significant explanatory v
ariables
fit_reduce <- lm(Nasdaq ~ AAPL + ALTR + AMAT + AMD + AMGN + AOC + APD + ASH + AT + AX
P + BA + BBY + BC + C + CCL + CEG + CFC + COST + DD + DELL + DIS + ED + EDS + EIX + F
RX + GE + GPS + HON + HPQ + HSY + HUM + IBM + INTC + IPG + JNJ + JPM + KBH + KMI + KO
+ KR + LEG + LEN + LM + LUV + MAS)
summary(fit_reduce)
```

```

Call:
lm(formula = Nasdaq ~ AAPL + ALTR + AMAT + AMD + AMGN + AOC +
    APD + ASH + AT + AXP + BA + BBY + BC + C + CCL + CEG + CFC +
    COST + DD + DELL + DIS + ED + EDS + EIX + FRX + GE + GPS +
    HON + HPQ + HSY + HUM + IBM + INTC + IPG + JNJ + JPM + KBH +
    KMI + KO + KR + LEG + LEN + LM + LUV + MAS)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0289356 -0.0014230  0.0000468  0.0015105  0.0177293

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) -8.623e-05  4.450e-05  -1.938 0.052718 .
AAPL         1.471e-02  1.527e-03   9.632  < 2e-16 ***
ALTR         2.273e-02  1.448e-03  15.698  < 2e-16 ***
AMAT         2.242e-02  1.719e-03  13.041  < 2e-16 ***
AMD          7.755e-03  1.310e-03   5.920 3.48e-09 ***
AMGN         2.948e-02  2.001e-03  14.735  < 2e-16 ***
AOC          5.209e-03  2.451e-03   2.125 0.033624 *
APD         -6.824e-03  3.110e-03  -2.194 0.028265 *
ASH          7.142e-03  3.183e-03   2.244 0.024900 *
AT           2.156e-02  3.144e-03   6.858 7.98e-12 ***
AXP          7.837e-03  2.970e-03   2.639 0.008354 **
BA           5.389e-03  2.674e-03   2.015 0.043965 *
BBY          8.543e-03  1.312e-03   6.510 8.40e-11 ***
BC           6.401e-03  2.296e-03   2.788 0.005325 **
C            9.739e-03  2.968e-03   3.281 0.001041 **
CCL          8.636e-03  2.232e-03   3.870 0.000111 ***
CEG         -1.043e-02  3.827e-03  -2.724 0.006470 **
CFC          5.482e-03  2.082e-03   2.633 0.008497 **
COST         1.271e-02  2.174e-03   5.848 5.36e-09 ***
DD          -1.631e-02  3.217e-03  -5.070 4.14e-07 ***
DELL         1.511e-02  1.508e-03  10.023  < 2e-16 ***
DIS          1.287e-02  2.711e-03   4.748 2.12e-06 ***
ED          -1.700e-02  4.693e-03  -3.622 0.000296 ***
EDS          6.869e-03  1.860e-03   3.693 0.000225 ***
EIX         -4.514e-03  2.283e-03  -1.977 0.048073 *
FRX          8.639e-03  2.102e-03   4.109 4.05e-05 ***
GE           3.180e-02  3.897e-03   8.162 4.32e-16 ***
GPS          4.265e-03  1.868e-03   2.283 0.022454 *
HON          1.124e-02  2.608e-03   4.307 1.69e-05 ***
HPQ          2.064e-02  2.139e-03   9.652  < 2e-16 ***
HSY         -1.201e-02  3.341e-03  -3.593 0.000330 ***
HUM          4.184e-03  1.667e-03   2.510 0.012103 *
IBM          1.965e-02  2.816e-03   6.979 3.42e-12 ***
INTC         4.640e-02  2.314e-03  20.049  < 2e-16 ***
IPG          1.363e-02  2.322e-03   5.873 4.60e-09 ***
JNJ         -1.448e-02  3.438e-03  -4.212 2.58e-05 ***
JPM          1.159e-02  2.725e-03   4.253 2.15e-05 ***
KBH          5.539e-03  2.049e-03   2.703 0.006897 **
KMI          7.713e-03  2.414e-03   3.196 0.001406 **
KO          -1.408e-02  3.420e-03  -4.118 3.90e-05 ***
KR          -5.119e-03  2.245e-03  -2.280 0.022665 *
LEG          1.113e-02  2.824e-03   3.940 8.29e-05 ***
LEN          1.007e-02  2.269e-03   4.439 9.28e-06 ***
LM           2.214e-02  2.566e-03   8.627  < 2e-16 ***
LUV          5.576e-03  2.084e-03   2.675 0.007496 **
MAS         -6.272e-03  2.661e-03  -2.357 0.018472 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.002902 on 4240 degrees of freedom
```

```
Multiple R-squared:  0.8045,    Adjusted R-squared:  0.8024
F-statistic: 387.6 on 45 and 4240 DF,  p-value: < 2.2e-16
```

(c) Compare the full and selected models. Summarize your comparison in an ANOVA table.

Hide

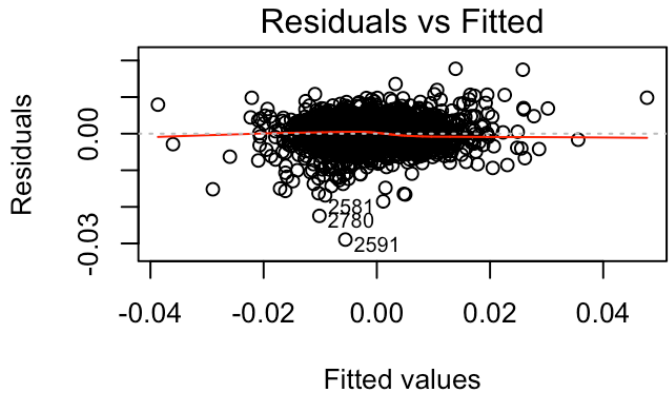
```
anova(fit_reduce, fit_full)
```

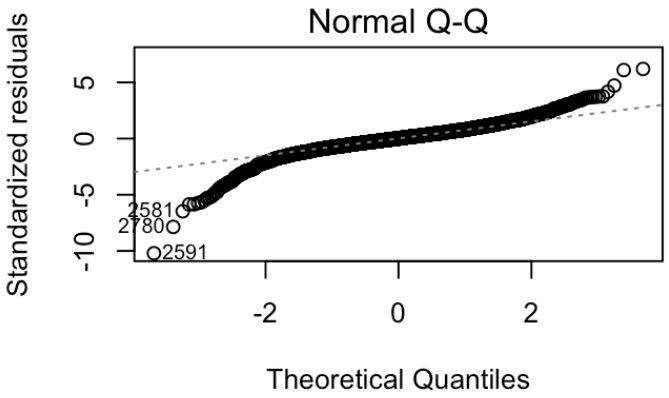
```
Analysis of Variance Table

Model 1: Nasdaq ~ AAPL + ALTR + AMAT + AMD + AMGN + AOC + APD + ASH +
    AT + AXP + BA + BBY + BC + C + CCL + CEG + CFC + COST + DD +
    DELL + DIS + ED + EDS + EIX + FRX + GE + GPS + HON + HPQ +
    HSY + HUM + IBM + INTC + IPG + JNJ + JPM + KBH + KMI + KO +
    KR + LEG + LEN + LM + LUV + MAS
Model 2: Nasdaq ~ AA + AAPL + AET + ALTR + AMAT + AMD + AMGN + AOC + APA +
    APD + ASH + AT + AVY + AXP + AYE + BA + BAC + BAX + BBY +
    BC + BFB + BLL + BMY + BNI + C + CAH + CAT + CCE + CCL +
    CEG + CFC + CLX + CMI + COP + COST + CSX + CVX + DD + DELL +
    DIS + DOV + DUK + ED + EDS + EIX + EK + EOG + EXC + FORD +
    FPL + FRX + GE + GM + GPS + GT + HD + HLT + HON + HPC + HPQ +
    HRB + HSY + HUM + IBM + INTC + IPG + JNJ + JPM + JWN + K +
    KBH + KMI + KO + KR + LEG + LEN + LLY + LM + LMT + LOW +
    LUV + MAS
  Res.Df      RSS Df  Sum of Sq      F Pr(>F)
1   4240 0.035718
2   4203 0.035353 37 0.00036456 1.1714 0.2204
```

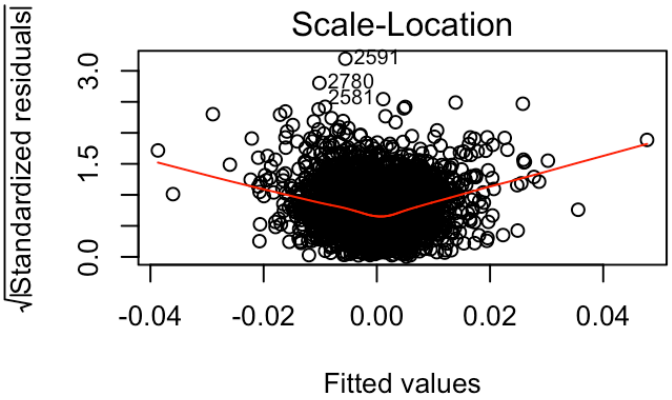
(d) For the selected regression model in (b), perform residual diagnostics.

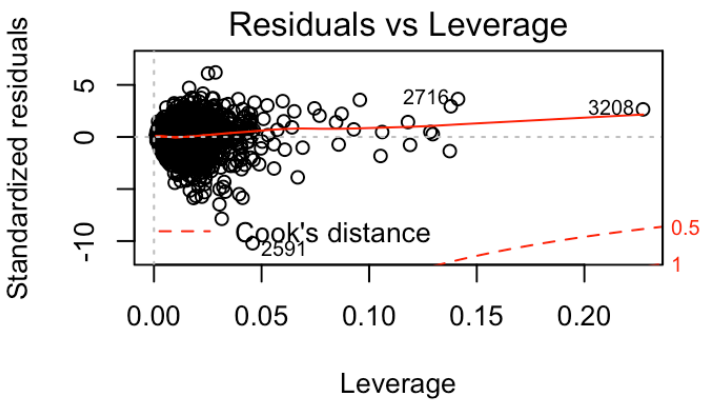
Hide

```
par(mfrow=c(2, 2))
plot(fit_reduce)
# plot studentized residuals
par(mfrow=c(1, 2))
```







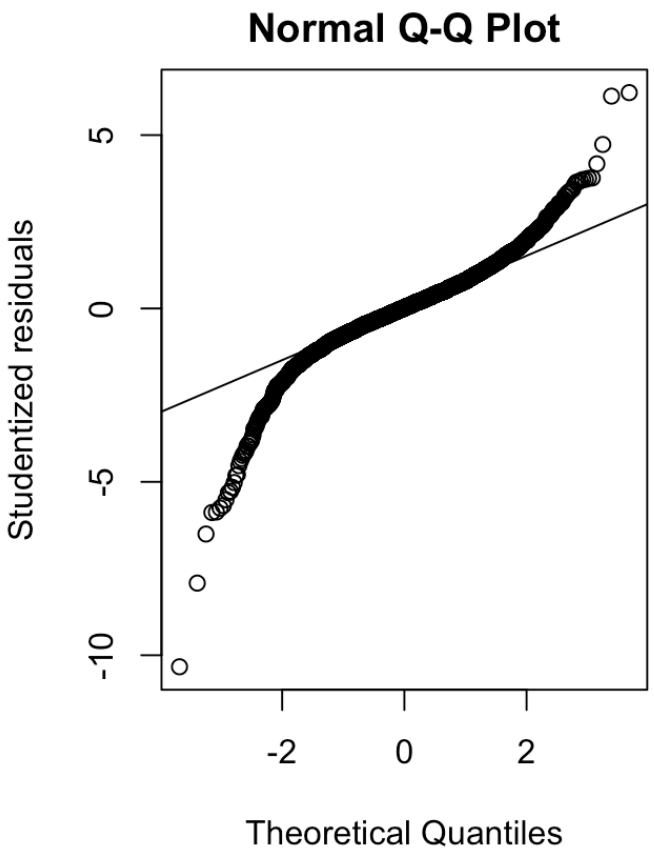


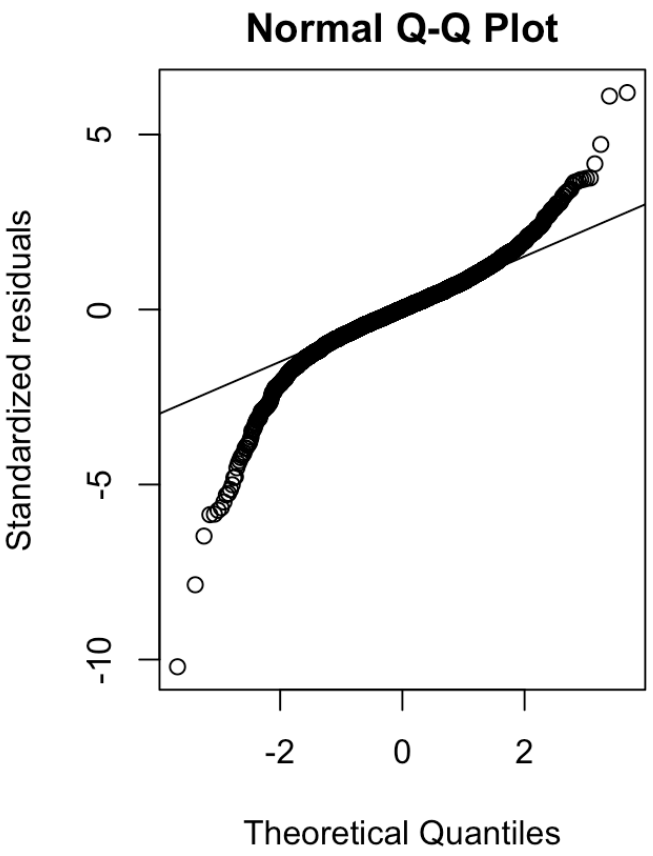
Hide

```
studresid <- as.numeric(studres(fit_reduce))
qqnorm(studresid, ylab='Studentized residuals')
qqline(studresid)
```

Hide

```
# plot standardized residuals
standresid <- as.numeric(stdres(fit_reduce))
qqnorm(standresid, ylab='Standardized residuals')
qqline(standresid)
```





(e) If you can only use at most five stocks to track the daily NASDAQ log returns, decribe your model selection procedure and your constructed model.

Hide

```
# we employ forward selection
# first we initialize the null model consisting of no predictors, just the intercept
null <- lm(Nasdaq ~ 1, data=df[,-1])
# we then add terms one by one
stepAIC(null, direction='forward', steps=5, trace=0,
        scope = list(upper=fit_full, lower=null))
```

```

Call:
lm(formula = Nasdaq ~ INTC + GE + ALTR + LM + AMAT, data = df[,
    -1])

Coefficients:
(Intercept)         INTC           GE         ALTR           LM         AMAT
 -5.105e-05    7.688e-02    9.438e-02    3.544e-02    5.636e-02    3.944e-02
```

Hide

```
# now we write down the constructed model
fit_five <- lm(Nasdaq ~ INTC + GE + ALTR + LM + AMAT)
summary(fit_five)
```

```
Call:
lm(formula = Nasdaq ~ INTC + GE + ALTR + LM + AMAT)

Residuals:
      Min        1Q    Median        3Q       Max
-0.034947 -0.001797  0.000065  0.001923  0.022637

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.105e-05  5.463e-05  -0.935     0.35
INTC         7.688e-02  2.634e-03  29.193   <2e-16 ***
GE           9.438e-02  3.858e-03  24.465   <2e-16 ***
ALTR         3.544e-02  1.735e-03  20.429   <2e-16 ***
LM           5.636e-02  2.871e-03  19.629   <2e-16 ***
AMAT         3.944e-02  2.016e-03  19.559   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.003572 on 4280 degrees of freedom
Multiple R-squared:  0.701, Adjusted R-squared:  0.7006
F-statistic:  2007 on 5 and 4280 DF,  p-value: < 2.2e-16
```

Hide

```
# we compare the full and constructed models
anova(fit_five, fit_full)
```

```
Analysis of Variance Table

Model 1: Nasdaq ~ INTC + GE + ALTR + LM + AMAT
Model 2: Nasdaq ~ AA + AAPL + AET + ALTR + AMAT + AMD + AMGN + AOC + APA +
    APD + ASH + AT + AVY + AXP + AYE + BA + BAC + BAX + BBY +
    BC + BFB + BLL + BMY + BNI + C + CAH + CAT + CCE + CCL +
    CEG + CFC + CLX + CMI + COP + COST + CSX + CVX + DD + DELL +
    DIS + DOV + DUK + ED + EDS + EIX + EK + EOG + EXC + FORD +
    FPL + FRX + GE + GM + GPS + GT + HD + HLT + HON + HPC + HPQ +
    HRB + HSY + HUM + IBM + INTC + IPG + JNJ + JPM + JWN + K +
    KBH + KMI + KO + KR + LEG + LEN + LLY + LM + LMT + LOW +
    LUV + MAS
  Res.Df      RSS Df Sum of Sq      F    Pr(>F)
1   4280 0.054619
2   4203 0.035353 77  0.019265 29.745 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```