Week 1 Questions

Dr M L Mackenzie

January 2018

Feedback questions

Introduction

Tobin's Q is the ratio between a physical assets market value and its replacement value and is an interesting quantity in the markets.

If the market value reflected solely the recorded assets of a company, Tobin's Q would be 1.0 while if Tobin's Q is greater than 1.0, then the market value is greater than the value of the company's recorded assets. This suggests that the market value reflects some unmeasured or unrecorded assets of the company. This is useful information:

- High Tobin's Q values encourage companies to invest more in capital because they are "worth" more than the price they paid for them
- Low Tobin's Q values (e.g. less than 1) means the market value is less than the recorded value of the assets of the company which suggests that the market may be undervaluing the company or that the company is in trouble

Tobin's Q is far from perfect and amongst other things ignores:

- market hype and speculation, reflecting, for example, analysts' views of the prospects for companies, or speculation such as bid rumors.
- the "intellectual capital" of corporations, that is, the unmeasured contribution of knowledge, goodwill, technology and other intangible assets that a company may have but which aren't recorded by accountants.

We will look at how Tobin's Q is related to various aspects of US-based companies and the data set ($N = 13539$) we will use has been collected over 25 years from 1240 American companies which are sourced from 41 industry classes. The data set contains the following variables:

- id: Company identifier (anonymised at source)
- year: financial year
- assets: value of assets
- capex: capital expenditure
- ltd: long-term debt
- ebitda: operating profits
- ppe: value of the property and plant equipment
- sales: value of sales
- ads: cost of advertising
- rd: research & development expenses
- mv: market value externally assessed
- indclass: class of industry

Research questions

The main questions for this data set are:

1. What are the relationships between market value and each covariate?

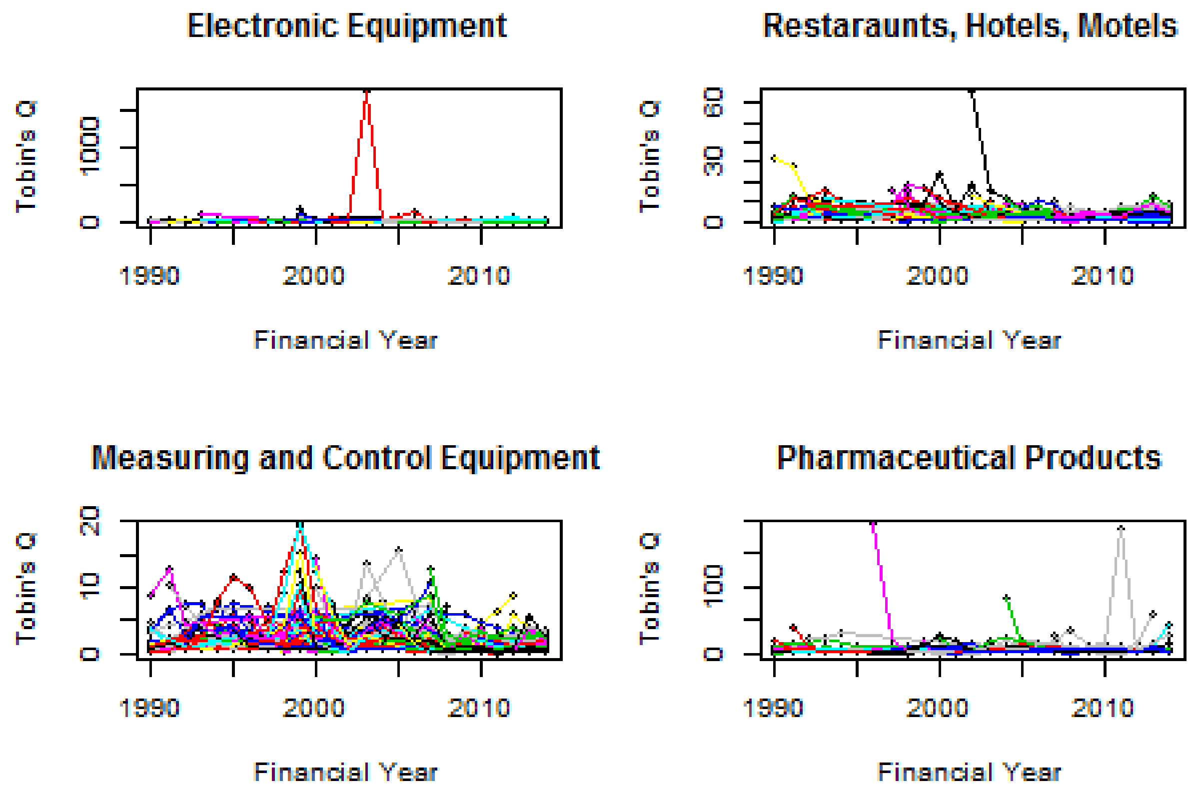
2. What are the main drivers of market value for these companies? Do these drivers differ across industry classes.
3. How well can we predict market value based on the information available and can we maximise market value based on what we can see? e.g. if the effect of advertising on market value plateaus at some point but there is a persistently linear trend between research and development spending and market value then it might be wiser to invest in R&D than concentrate on advertising.

Exploratory Data Analysis

The first thing to notice is that the response varies over time within company and across companies within industry class. Some industry classes vary less than others across time and across companies within industry classes.

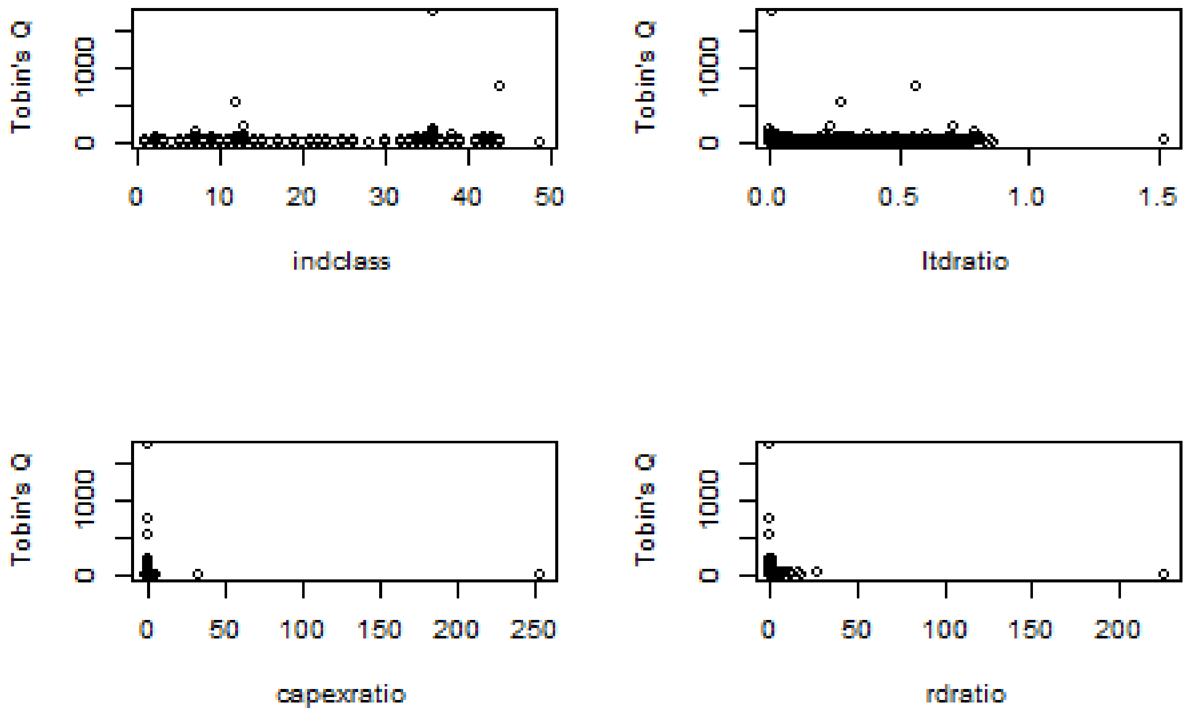
```
x <- as.numeric(rownames(sort(table(dat$indclass), decreasing = TRUE)[1:4]))

# or using tidyverse: require(tidyverse) x<- group_by(dat,
# indclass) %>% summarise(n=n()) %>% arrange(desc(n)) %>%
# top_n(4) %>% pull(indclass)

nams <- c("Electronic Equipment", "Restaraunts, Hotels, Motels",
    "Measuring and Control Equipment", "Pharmaceutical Products")
```

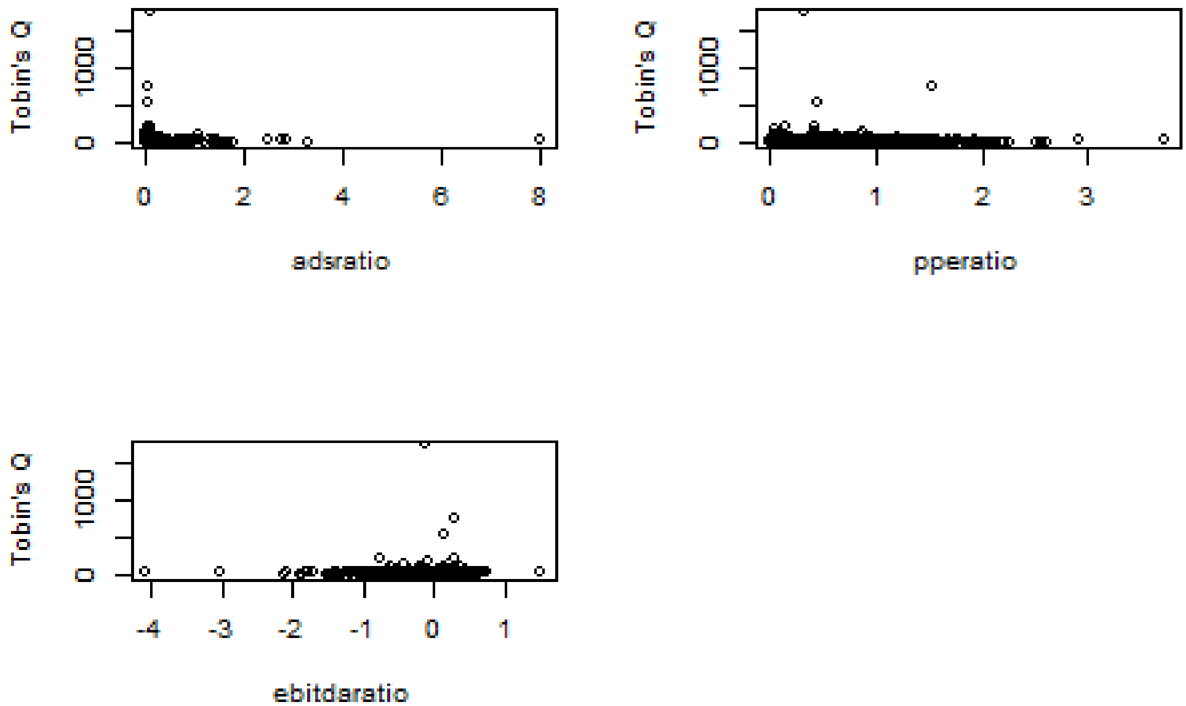
```
par(mfrow = c(2, 2))
for (j in 1:4) {
    datuse <- dat[dat$indclass == x[j], ]
    plot(datuse$year, datuse$tobinsQ, xlab = "Financial Year",
        ylab = "Tobin's Q", pch = 20, main = nams[j])
    for (k in unique(dat$id)) {
        lines(datuse$year[datuse$id == k], datuse$tobinsQ[datuse$id ==
            k], col = k)
    }
}
```



Initial inspection of the data reveals some very large response scores and some large gaps in the data:

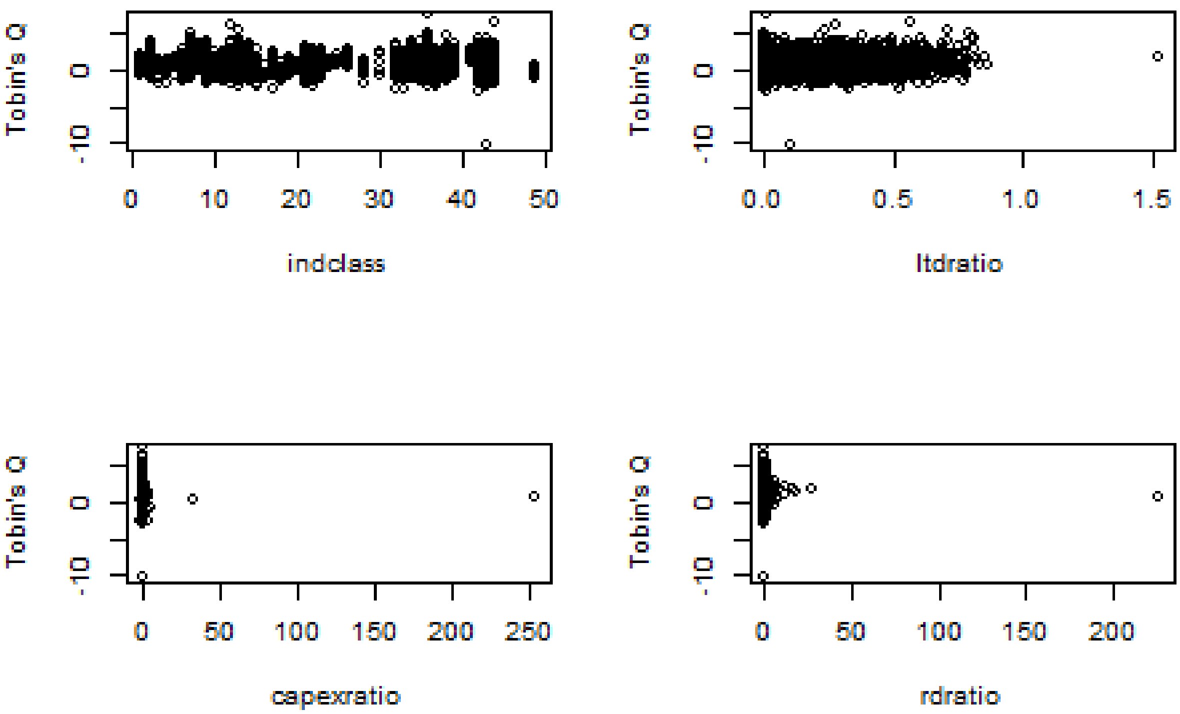
```
par(mfrow = c(2, 2))
for (i in c(13, 15:20)) {
    plot(dat[, i], (dat[, 14]), xlab = names(dat)[i], ylab = "Tobin's Q")
}
```

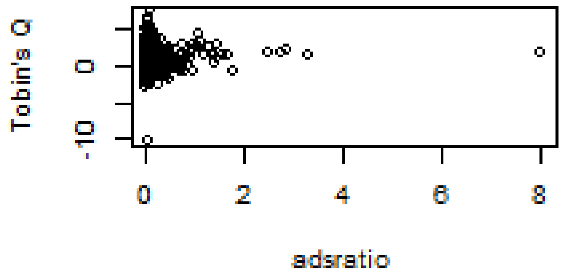

Tobin's Q
0
1000
0
10
20
30
40
50
indclass
0.0
0.5
1.0
1.5
ltdratio
50
100
150
200
250
capexratio
rdratio

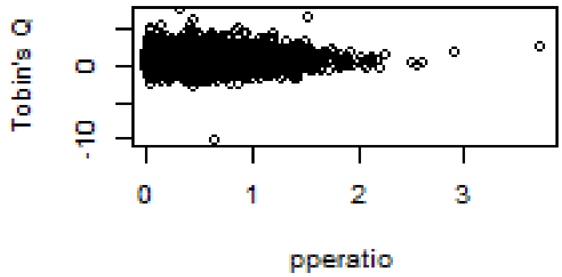


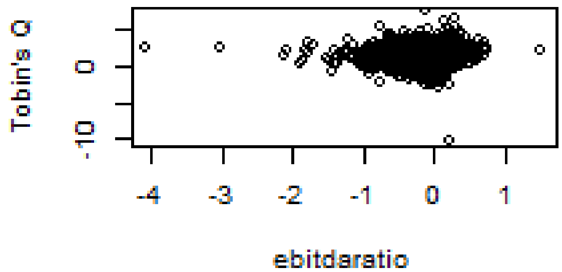
and these gaps are still evident (naturally) when the response scores are shown on the log scale:

```
par(mfrow = c(2, 2))
for (i in c(13, 15:20)) {
    plot(dat[, i], log(dat[, 14]), xlab = names(dat)[i], ylab = "Tobin's Q")
}
```


Tobin's Q
-10
0
0
10
20
30
40
50
indclass
Tobin's Q
-10
0
0.0
0.5
1.0
1.5
ltdratio
Tobin's Q
-10
0
0
50
100
150
200
250
capexratio
Tobin's Q
-10
0
0
50
100
150
200
rdratio







We can reduce the data set to remove the very large values and produce a summary to confirm the new ranges:

```
# newdat<- dat[dat$ltdratio<1,] newdat<-
# newdat[newdat$capexratio<20,] newdat<-
# newdat[newdat$rdratio<10,] newdat<-
# newdat[newdat$adsratio<2,] newdat<-
# newdat[newdat$tobinsQ<200,] summary(newdat)
```

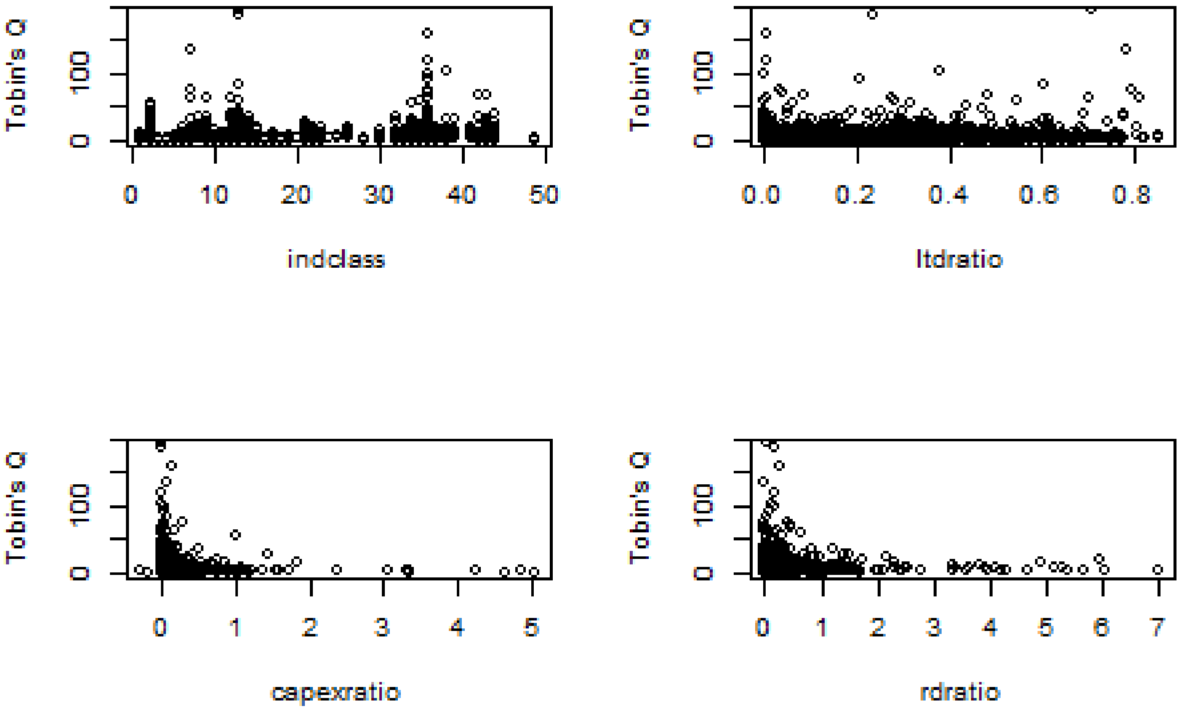
```
newdat <- filter(dat, ltdratio < 1, capexratio < 20, rdratio <
    10, adsratio < 2, tobinsQ < 200)

summary(newdat)
```

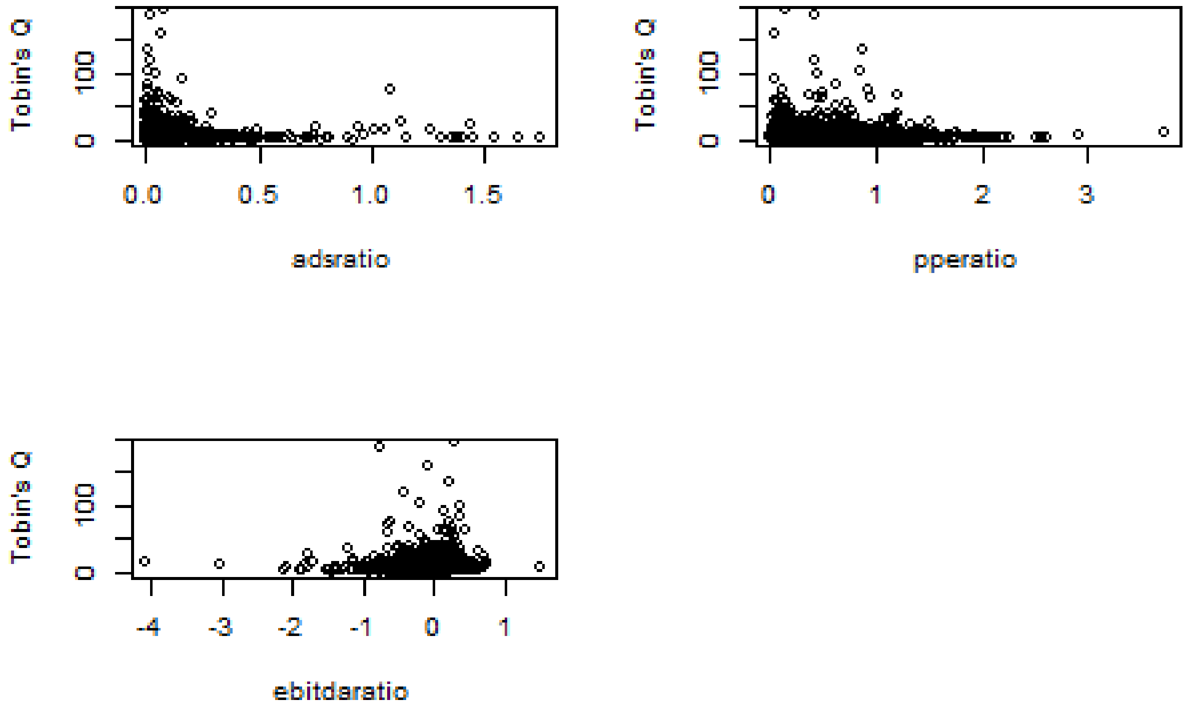
```
       id              year          assets              capex
 Min.   :  1050   Min.   :1990   Min.   :    0.72   Min.   :-401.609
 1st Qu.: 10391   1st Qu.:1999   1st Qu.:   59.32   1st Qu.:   1.572
 Median : 25390   Median :2004   Median :  216.73   Median :   8.203
 Mean   : 54440   Mean   :2004   Mean   : 1434.82   Mean   :  74.055
 3rd Qu.:104598   3rd Qu.:2009   3rd Qu.:  914.16   3rd Qu.:  40.154
 Max.   :270705   Max.   :2014   Max.   :80033.55   Max.   :7150.000
      ltd                ebitda                ppe
 Min.   :    0.000   Min.   :-1134.000   Min.   :    0.00
 1st Qu.:    0.000   1st Qu.:    2.293   1st Qu.:   15.13
 Median :    3.129   Median :   21.395   Median :   67.45
 Mean   :  293.976   Mean   :  199.175   Mean   :  768.24
 3rd Qu.:  100.000   3rd Qu.:  123.826   3rd Qu.:  341.80
```

```
 Max.   :24380.700   Max.   : 7723.890   Max.   :59762.83  
     sales                ads                  rd          
 Min.   :     0.05   Min.   :   0.000   Min.   :   0.000  
 1st Qu.:    58.35   1st Qu.:   0.537   1st Qu.:   0.000  
 Median :   229.86   Median :   3.055   Median :   3.996  
 Mean   :  1453.51   Mean   :  40.312   Mean   :  43.057  
 3rd Qu.:   992.47   3rd Qu.:  18.500   3rd Qu.:  23.247  
 Max.   :108465.00   Max.   :2840.000   Max.   :3146.829  
    bookval               mv               indclass        tobinsQ         
 Min.   :    0.04   Min.   :    0.10   Min.   : 1.00   Min.   :  0.00004  
 1st Qu.:   34.52   1st Qu.:   60.49   1st Qu.:15.00   1st Qu.:  1.23689  
 Median :  128.56   Median :  275.95   Median :36.00   Median :  2.10815  
 Mean   :  654.89   Mean   : 1759.15   Mean   :29.35   Mean   :  3.21056  
 3rd Qu.:  482.49   3rd Qu.: 1231.72   3rd Qu.:38.00   3rd Qu.:  3.58202  
 Max.   :41466.76   Max.   :50174.89   Max.   :49.00   Max.   :193.71598  
    ltdratio          capexratio          rdratio           adsratio       
 Min.   :0.00000   Min.   :-0.25973   Min.   :0.0000   Min.   :0.000000  
 1st Qu.:0.00000   1st Qu.: 0.01781   1st Qu.:0.0000   1st Qu.:0.005603  
 Median :0.03931   Median : 0.03282   Median :0.0290   Median :0.014364  
 Mean   :0.11473   Mean   : 0.05643   Mean   :0.0908   Mean   :0.032936  
 3rd Qu.:0.19905   3rd Qu.: 0.05886   3rd Qu.:0.1159   3rd Qu.:0.035927  
 Max.   :0.85654   Max.   : 5.04205   Max.   :7.0207   Max.   :1.757624  
    pperatio        ebitdaratio      
 Min.   :0.0000   Min.   :-4.06708  
 1st Qu.:0.1830   1st Qu.: 0.04678  
 Median :0.3484   Median : 0.11329  
 Mean   :0.4367   Mean   : 0.08921  
 3rd Qu.:0.6103   3rd Qu.: 0.17485  
 Max.   :3.7330   Max.   : 1.50841  
```

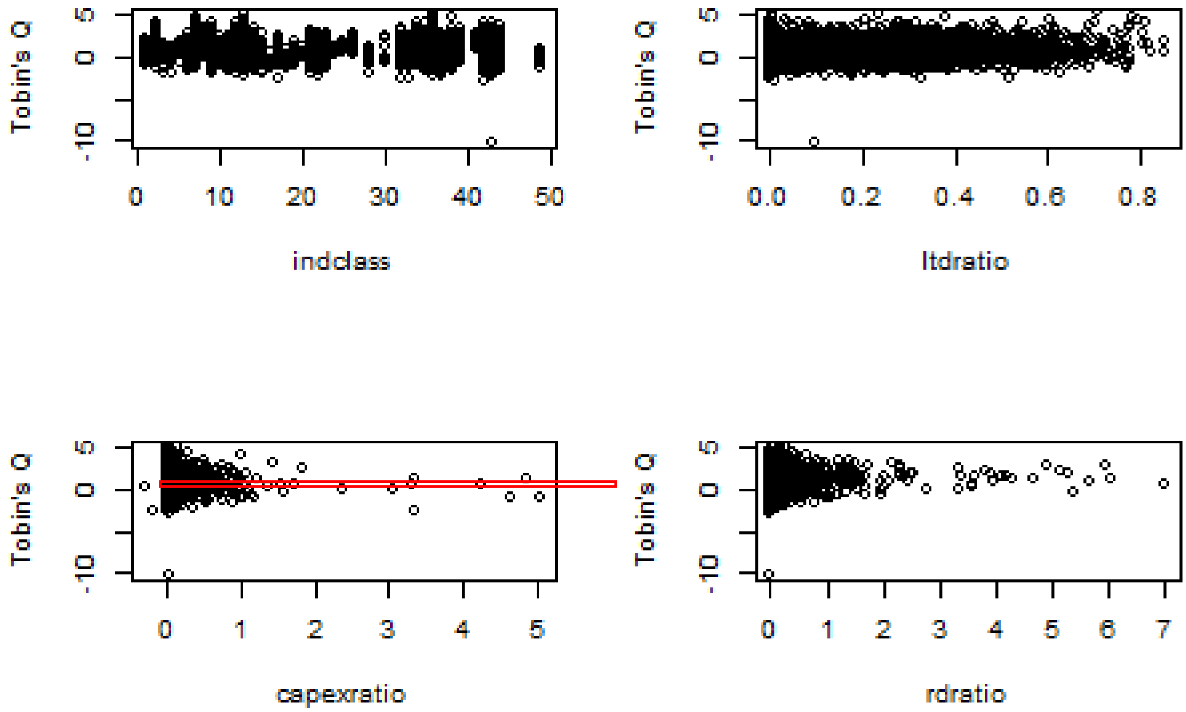
which results in a new set of plots on either the raw or log scales:

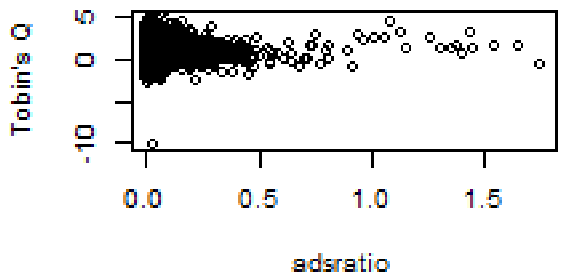
```r
par(mfrow = c(2, 2))
for (i in c(13, 15:20)) {
    plot(newdat[, i], (newdat[, 14]), xlab = names(dat)[i], ylab = "Tobin's Q")
}
```

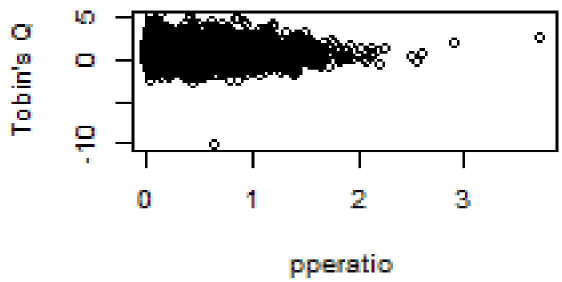

Tobin's Q
0
100
0
10
20
30
40
50
indclass
Tobin's Q
0
100
0.0
0.2
0.4
0.6
0.8
ltdratio
Tobin's Q
0
100
0
1
2
3
4
5
capexratio
Tobin's Q
0
100
0
1
2
3
4
5
6
7
rdratio

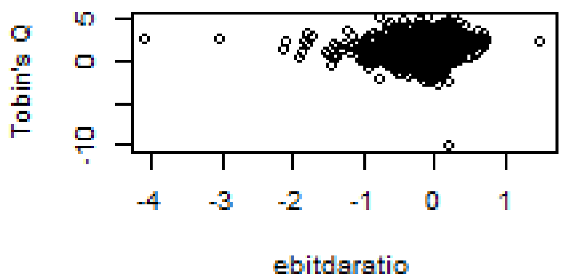



```r
par(mfrow = c(2, 2))
for (i in c(13, 15:20)) {
    plot(newdat[, i], log(newdat[, 14]), xlab = names(dat)[i],
        ylab = "Tobin's Q")
}
```


Tobin's Q
5
0
-10
0
10
20
30
40
50
indclass
Tobin's Q
0.0
0.2
0.4
0.6
0.8
ltdratio
Tobin's Q
0
1
2
3
4
5
capexratio
Tobin's Q
0
1
2
3
4
5
6
7
rdratio







Model Specification and Fitting

Model 1

An initial model (with all candidate variables included) reveals not all covariates are significantly related to the response:

```
fit <- glm(tobinsQ ~ ltdratio + capexratio + rdratio + adsratio +
    pperatio + ebitdaratio + as.factor(year) + as.factor(indclass),
    data = newdat)
```

```
require(car)
vif(fit)
```

```
                        GVIF Df GVIF^(1/(2*Df))
ltdratio            1.232321  1        1.110099
capexratio          1.213780  1        1.101717
rdratio             1.395413  1        1.181276
adsratio            1.182557  1        1.087454
pperatio            1.679157  1        1.295823
ebitdaratio         1.287513  1        1.134686
as.factor(year)     1.140738 24        1.002747
as.factor(indclass) 2.534756 40        1.011694
```

```
Anova(fit)
```

```
Analysis of Deviance Table (Type II tests)
```

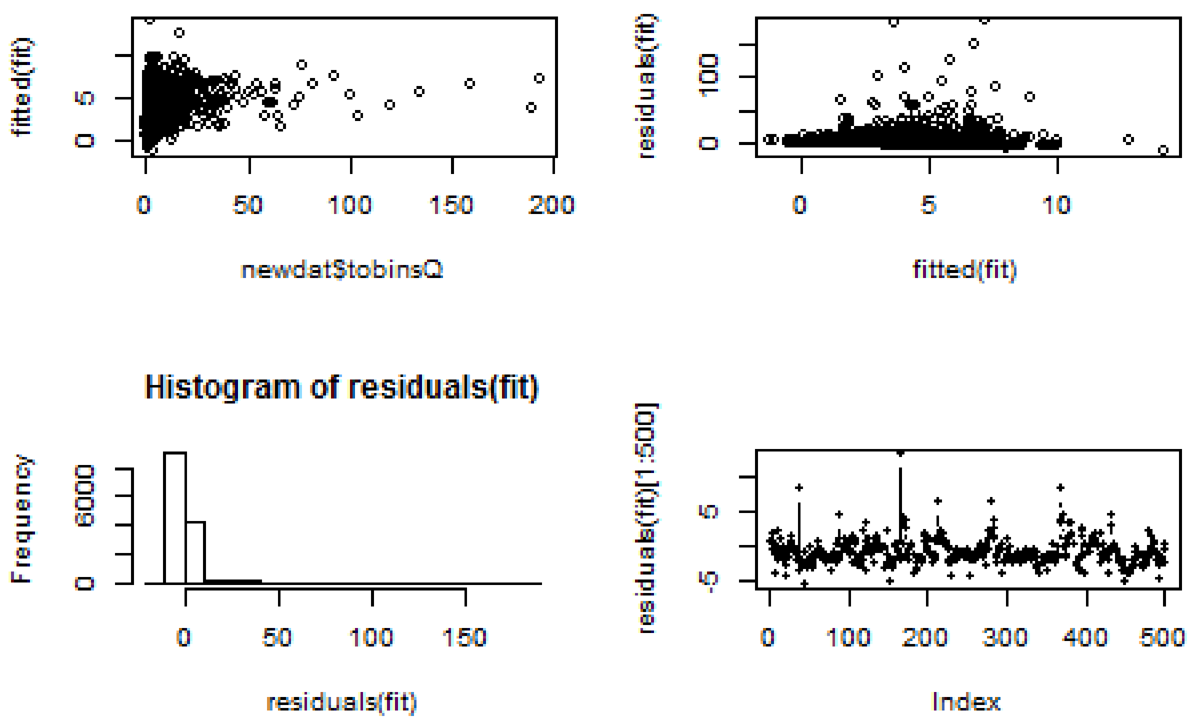
```
Response: tobinsQ
                     LR Chisq Df Pr(>Chisq)
ltdratio                58.87  1  1.688e-14 ***
capexratio               0.10  1     0.7464
rdratio                 24.84  1  6.215e-07 ***
adsratio                15.47  1  8.385e-05 ***
pperatio                 0.41  1     0.5239
ebitdaratio             23.54  1  1.223e-06 ***
as.factor(year)        231.40 24  < 2.2e-16 ***
as.factor(indclass)    450.32 40  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit)
```

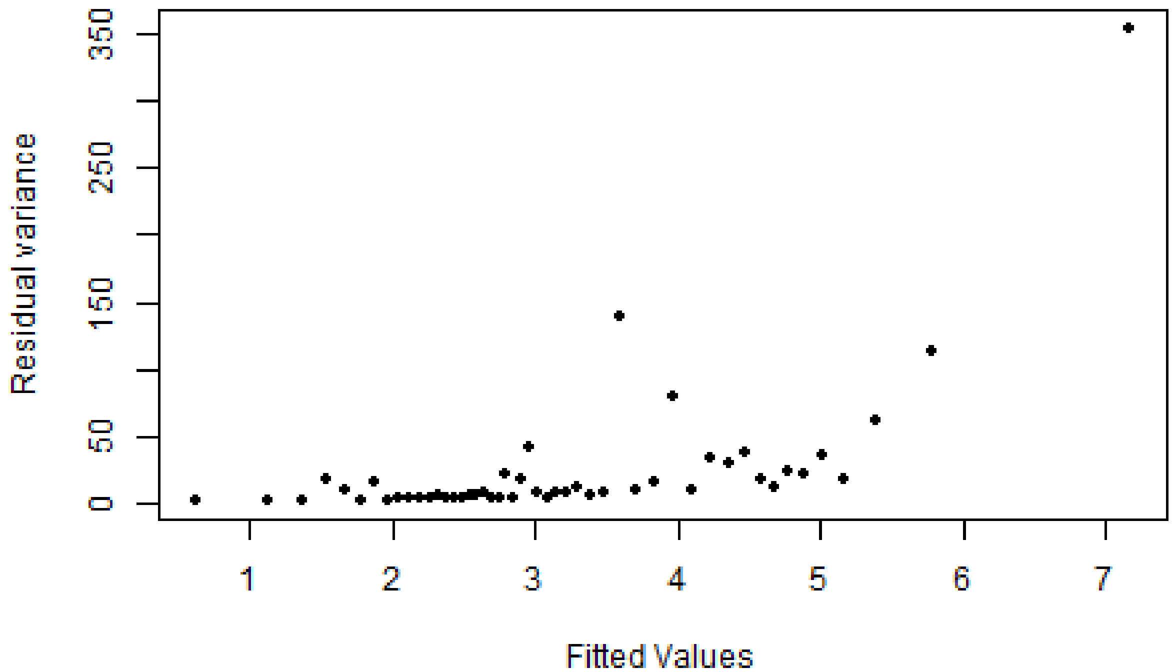
```
Call:
glm(formula = tobinsQ ~ ltdratio + capexratio + rdratio + adsratio +
    pperatio + ebitdaratio + as.factor(year) + as.factor(indclass),
    data = newdat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-11.725   -1.751   -0.731    0.518  186.556

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            1.10905    1.11026   0.999 0.317858
ltdratio               2.50805    0.32689   7.672 1.80e-14 ***
capexratio             0.12032    0.37208   0.323 0.746425
rdratio                1.07732    0.21614   4.984 6.29e-07 ***
adsratio               2.64664    0.67291   3.933 8.43e-05 ***
pperatio              -0.10843    0.17013  -0.637 0.523906
ebitdaratio            1.35502    0.27928   4.852 1.24e-06 ***
as.factor(year)1991    1.11195    0.41995   2.648 0.008110 **
as.factor(year)1992    1.33977    0.40981   3.269 0.001081 **
as.factor(year)1993    1.52078    0.39984   3.803 0.000143 ***
as.factor(year)1994    0.67524    0.43682   1.546 0.122169
as.factor(year)1995    0.81439    0.42328   1.924 0.054377 .
as.factor(year)1996    1.34930    0.40487   3.333 0.000863 ***
as.factor(year)1997    1.00575    0.39661   2.536 0.011227 *
as.factor(year)1998    0.92173    0.39222   2.350 0.018786 *
as.factor(year)1999    3.09607    0.38266   8.091 6.42e-16 ***
as.factor(year)2000    0.95711    0.37241   2.570 0.010179 *
as.factor(year)2001    0.75316    0.36933   2.039 0.041443 *
as.factor(year)2002    0.18347    0.36446   0.503 0.614695
as.factor(year)2003    1.12172    0.36206   3.098 0.001951 **
as.factor(year)2004    1.16917    0.35783   3.267 0.001088 **
as.factor(year)2005    1.01845    0.35680   2.854 0.004319 **
as.factor(year)2006    1.17432    0.35730   3.287 0.001016 **
as.factor(year)2007    0.79956    0.35870   2.229 0.025828 *
as.factor(year)2008   -0.56420    0.35864  -1.573 0.115701
as.factor(year)2009    0.06232    0.36017   0.173 0.862622
as.factor(year)2010    0.38601    0.36319   1.063 0.287870
as.factor(year)2011    0.48276    0.36729   1.314 0.188741
```

```
as.factor(year)2012    0.50495    0.37239   1.356 0.175133
as.factor(year)2013    1.36800    0.37643   3.634 0.000280 ***
as.factor(year)2014    1.39209    0.38388   3.626 0.000289 ***
as.factor(indclass)2   2.09192    1.10514   1.893 0.058394 .
as.factor(indclass)3  -0.73503    1.77702  -0.414 0.679150
as.factor(indclass)4  -0.94130    1.29173  -0.729 0.466191
as.factor(indclass)5  -0.09236    1.69906  -0.054 0.956648
as.factor(indclass)6  -0.49397    1.11166  -0.444 0.656795
as.factor(indclass)7   1.17056    1.14407   1.023 0.306255
as.factor(indclass)8   4.43601    1.34011   3.310 0.000935 ***
as.factor(indclass)9   0.25356    1.10428   0.230 0.818392
as.factor(indclass)10 -0.28114    1.13888  -0.247 0.805020
as.factor(indclass)11 -0.13318    1.16676  -0.114 0.909123
as.factor(indclass)12  0.91036    1.08347   0.840 0.400799
as.factor(indclass)13  2.34357    1.08253   2.165 0.030413 *
as.factor(indclass)14  0.33799    1.12026   0.302 0.762879
as.factor(indclass)15  0.08343    1.21011   0.069 0.945032
as.factor(indclass)16 -1.79135    2.01433  -0.889 0.373856
as.factor(indclass)17 -0.42883    1.12354  -0.382 0.702709
as.factor(indclass)18 -2.00792    1.61473  -1.243 0.213705
as.factor(indclass)19 -1.08432    1.27581  -0.850 0.395391
as.factor(indclass)20 -0.82631    2.10508  -0.393 0.694672
as.factor(indclass)21  0.26428    1.08839   0.243 0.808150
as.factor(indclass)22 -0.29389    1.09709  -0.268 0.788798
as.factor(indclass)23 -0.07082    1.11455  -0.064 0.949336
as.factor(indclass)24  0.55182    1.73504   0.318 0.750457
as.factor(indclass)25  1.63593    1.82283   0.897 0.369485
as.factor(indclass)26  1.86536    1.28282   1.454 0.145939
as.factor(indclass)28 -0.54985    1.51186  -0.364 0.716096
as.factor(indclass)30  2.59045    2.35255   1.101 0.270862
as.factor(indclass)32  0.24347    1.13418   0.215 0.830030
as.factor(indclass)33 -0.52966    1.26963  -0.417 0.676555
as.factor(indclass)34  1.58045    1.11623   1.416 0.156834
as.factor(indclass)35  0.41276    1.08636   0.380 0.703991
as.factor(indclass)36  2.11349    1.07336   1.969 0.048970 *
as.factor(indclass)37  0.02149    1.07761   0.020 0.984086
as.factor(indclass)38  0.37124    1.08947   0.341 0.733293
as.factor(indclass)39  1.08465    1.17290   0.925 0.355107
as.factor(indclass)41  5.90316    1.87708   3.145 0.001665 **
as.factor(indclass)42  0.11785    1.10177   0.107 0.914822
as.factor(indclass)43 -0.13223    1.07406  -0.123 0.902023
as.factor(indclass)44  0.21422    1.08986   0.197 0.844173
as.factor(indclass)49 -1.04508    1.69889  -0.615 0.538463
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 26.16942)

    Null deviance: 375420  on 13524  degrees of freedom
Residual deviance: 352083  on 13454  degrees of freedom
AIC: 82609

Number of Fisher Scoring iterations: 2
```

```
par(mfrow = c(2, 2))
plot(newdat$tobinsQ, fitted(fit))
plot(fitted(fit), residuals(fit))
hist(residuals(fit))
plot(residuals(fit)[1:500], type = "b", pch = 20)
```




```
par(mfrow = c(1, 1))
cut.x <- cut(fitted(fit), quantile(fitted(fit), probs = seq(0,
    1, length = 50)))
means <- tapply(fitted(fit), cut.x, mean)
vars <- tapply(residuals(fit), cut.x, var)
plot(means, vars, pch = 20, xlab = "Fitted Values", ylab = "Residual variance")
```



Model 2

```r
newdat$mvmil <- newdat$mv * 1e+06
newdat$bkmil <- newdat$bookval * 1e+06
```

```r
fit_pois <- glm(mvmil ~ ltdratio + capexratio + rdratio + adsratio +
    pperatio + ebitdaratio + as.factor(year) + as.factor(indclass) +
    offset(log(bkmil)), data = newdat, family = poisson)
```

```r
Anova(fit_pois)
```

```
Analysis of Deviance Table (Type II tests)

Response: mvmil
                    LR Chisq Df Pr(>Chisq)
ltdratio           2.0656e+10  1  < 2.2e-16 ***
capexratio         8.3793e+09  1  < 2.2e-16 ***
rdratio            3.0448e+11  1  < 2.2e-16 ***
adsratio           1.0340e+10  1  < 2.2e-16 ***
pperatio           1.0951e+11  1  < 2.2e-16 ***
ebitdaratio        2.6136e+12  1  < 2.2e-16 ***
as.factor(year)    7.7533e+11 24  < 2.2e-16 ***
as.factor(indclass) 1.0698e+12 40  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit_pois)
```

```

Call:
glm(formula = mvmil ~ ltdratio + capexratio + rdratio + adsratio +
    pperatio + ebitdaratio + as.factor(year) + as.factor(indclass) +
    offset(log(bkmil)), family = poisson, data = newdat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-298313    -7649    -1404     6123   368953

Coefficients:
                        Estimate Std. Error   z value Pr(>|z|)
(Intercept)           -1.627e-02  1.964e-05    -828.3   <2e-16 ***
ltdratio               2.594e-01  1.799e-06  144167.4   <2e-16 ***
capexratio             1.917e-01  1.941e-06   98777.5   <2e-16 ***
rdratio                7.717e-01  1.035e-06  745765.0   <2e-16 ***
adsratio               5.612e-01  5.438e-06  103199.5   <2e-16 ***
pperatio              -2.787e-01  8.484e-07 -328520.1   <2e-16 ***
ebitdaratio            4.140e+00  2.185e-06 1894718.7   <2e-16 ***
as.factor(year)1991    3.034e-01  3.066e-06   98942.8   <2e-16 ***
as.factor(year)1992    3.646e-01  2.993e-06  121803.1   <2e-16 ***
as.factor(year)1993    4.519e-01  2.946e-06  153427.8   <2e-16 ***
as.factor(year)1994    1.567e-01  3.099e-06   50565.9   <2e-16 ***
as.factor(year)1995    2.774e-01  2.949e-06   94077.7   <2e-16 ***
as.factor(year)1996    3.546e-01  2.892e-06  122610.5   <2e-16 ***
as.factor(year)1997    5.299e-01  2.780e-06  190620.9   <2e-16 ***
as.factor(year)1998    6.138e-01  2.763e-06  222179.7   <2e-16 ***
as.factor(year)1999    9.092e-01  2.627e-06  346162.2   <2e-16 ***
as.factor(year)2000    6.791e-01  2.622e-06  259018.2   <2e-16 ***
as.factor(year)2001    5.497e-01  2.663e-06  206437.5   <2e-16 ***
as.factor(year)2002    3.288e-01  2.678e-06  122784.1   <2e-16 ***
as.factor(year)2003    5.181e-01  2.589e-06  200122.7   <2e-16 ***
as.factor(year)2004    4.576e-01  2.560e-06  178709.7   <2e-16 ***
as.factor(year)2005    4.315e-01  2.555e-06  168871.9   <2e-16 ***
as.factor(year)2006    4.561e-01  2.540e-06  179556.7   <2e-16 ***
as.factor(year)2007    4.838e-01  2.536e-06  190809.5   <2e-16 ***
as.factor(year)2008   -2.624e-02  2.633e-06   -9967.3   <2e-16 ***
as.factor(year)2009    2.643e-01  2.562e-06  103166.3   <2e-16 ***
as.factor(year)2010    3.033e-01  2.534e-06  119692.1   <2e-16 ***
as.factor(year)2011    2.380e-01  2.549e-06   93355.3   <2e-16 ***
as.factor(year)2012    2.883e-01  2.543e-06  113358.0   <2e-16 ***
as.factor(year)2013    5.362e-01  2.509e-06  213673.4   <2e-16 ***
as.factor(year)2014    5.988e-01  2.508e-06  238796.0   <2e-16 ***
as.factor(indclass)2   1.988e-01  1.953e-05   10179.7   <2e-16 ***
as.factor(indclass)3  -5.618e-02  1.994e-05   -2817.0   <2e-16 ***
as.factor(indclass)4  -2.439e-01  1.960e-05  -12443.4   <2e-16 ***
as.factor(indclass)5  -1.767e-01  1.963e-05   -9003.4   <2e-16 ***
as.factor(indclass)6  -2.187e-01  1.962e-05  -11145.2   <2e-16 ***
as.factor(indclass)7   1.390e-01  1.958e-05    7099.5   <2e-16 ***
as.factor(indclass)8   7.695e-01  1.965e-05   39165.7   <2e-16 ***
as.factor(indclass)9   1.382e-01  1.956e-05    7063.1   <2e-16 ***
as.factor(indclass)10 -3.159e-01  1.968e-05  -16053.3   <2e-16 ***
```

```
as.factor(indclass)11 -1.847e-01  2.004e-05   -9217.3   <2e-16 ***
as.factor(indclass)12  1.628e-01  1.955e-05    8325.6   <2e-16 ***
as.factor(indclass)13  7.766e-02  1.954e-05    3975.1   <2e-16 ***
as.factor(indclass)14  1.294e-02  1.956e-05     661.8   <2e-16 ***
as.factor(indclass)15  3.947e-02  1.971e-05    2002.5   <2e-16 ***
as.factor(indclass)16 -1.094e+00  7.832e-05  -13965.0   <2e-16 ***
as.factor(indclass)17 -2.463e-01  1.975e-05  -12470.6   <2e-16 ***
as.factor(indclass)18 -4.689e-01  2.722e-05  -17227.5   <2e-16 ***
as.factor(indclass)19 -8.485e-01  1.959e-05  -43314.3   <2e-16 ***
as.factor(indclass)20 -3.865e-01  8.672e-05   -4457.2   <2e-16 ***
as.factor(indclass)21 -6.444e-02  1.954e-05   -3298.2   <2e-16 ***
as.factor(indclass)22 -8.837e-02  1.962e-05   -4504.3   <2e-16 ***
as.factor(indclass)23 -2.173e-01  1.956e-05  -11112.5   <2e-16 ***
as.factor(indclass)24  3.110e-01  2.021e-05   15390.5   <2e-16 ***
as.factor(indclass)25  1.758e-01  2.534e-05    6936.2   <2e-16 ***
as.factor(indclass)26 -2.442e-01  2.069e-05  -11802.5   <2e-16 ***
as.factor(indclass)28 -2.605e-01  2.017e-05  -12912.7   <2e-16 ***
as.factor(indclass)30  5.497e-01  7.309e-05    7520.4   <2e-16 ***
as.factor(indclass)32 -4.410e-01  1.956e-05  -22545.1   <2e-16 ***
as.factor(indclass)33 -7.288e-01  2.074e-05  -35138.6   <2e-16 ***
as.factor(indclass)34  1.010e-01  1.966e-05    5138.6   <2e-16 ***
as.factor(indclass)35 -1.181e-01  1.955e-05   -6040.8   <2e-16 ***
as.factor(indclass)36  1.978e-01  1.953e-05   10125.7   <2e-16 ***
as.factor(indclass)37 -1.611e-01  1.953e-05   -8246.8   <2e-16 ***
as.factor(indclass)38 -5.569e-02  1.955e-05   -2848.6   <2e-16 ***
as.factor(indclass)39  3.524e-01  1.956e-05   18021.0   <2e-16 ***
as.factor(indclass)41  6.869e-01  2.017e-05   34051.0   <2e-16 ***
as.factor(indclass)42 -3.443e-02  1.958e-05   -1757.9   <2e-16 ***
as.factor(indclass)43 -8.237e-02  1.953e-05   -4218.3   <2e-16 ***
as.factor(indclass)44  2.210e-01  1.956e-05   11302.1   <2e-16 ***
as.factor(indclass)49 -1.398e-01  3.742e-05   -3735.0   <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: 1.2915e+13  on 13524  degrees of freedom
Residual deviance: 7.2149e+12  on 13454  degrees of freedom
AIC: 7.2149e+12

Number of Fisher Scoring iterations: 9
```

```r
fit_qpois <- glm(mvmil ~ ltdratio + capexratio + rdratio + adsratio +
    pperatio + ebitdaratio + as.factor(year) + as.factor(indclass) +
    offset(log(bkmil)), data = newdat, family = quasipoisson)
```

```r
logLik(fit_pois)
```

```
'log Lik.' -3.607428e+12 (df=71)
```

```r
Anova(fit_qpois, test = "F")
```

```
Analysis of Deviance Table (Type II tests)

Response: mvmil
Error estimate based on Pearson residuals
```

```
                         SS    Df       F    Pr(>F)    
ltdratio            2.0656e+10     1  0.3099    0.5778    
capexratio          8.3793e+09     1  0.1257    0.7229    
rdratio             3.0448e+11     1  4.5674    0.0326 *  
adsratio            1.0340e+10     1  0.1551    0.6937    
pperatio            1.0951e+11     1  1.6427    0.2000    
ebitdaratio         2.6136e+12     1 39.2063 3.929e-10 ***
as.factor(year)     7.7533e+11    24  0.4846    0.9837    
as.factor(indclass) 1.0698e+12    40  0.4012    0.9997    
Residuals           8.9689e+14 13454                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit_qpois)
```

```
Call:
glm(formula = mvmil ~ ltdratio + capexratio + rdratio + adsratio + 
    pperatio + ebitdaratio + as.factor(year) + as.factor(indclass) + 
    offset(log(bkmil)), family = quasipoisson, data = newdat)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-298313    -7649    -1404     6123   368953  

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)           -0.01627    5.07170  -0.003  0.99744    
ltdratio               0.25938    0.46452   0.558  0.57660    
capexratio             0.19173    0.50114   0.383  0.70204    
rdratio                0.77168    0.26716   2.888  0.00388 ** 
adsratio               0.56115    1.40393   0.400  0.68938    
pperatio              -0.27873    0.21906  -1.272  0.20326    
ebitdaratio            4.14031    0.56420   7.338 2.29e-13 ***
as.factor(year)1991    0.30336    0.79161   0.383  0.70157    
as.factor(year)1992    0.36461    0.77287   0.472  0.63711    
as.factor(year)1993    0.45193    0.76051   0.594  0.55236    
as.factor(year)1994    0.15669    0.80006   0.196  0.84473    
as.factor(year)1995    0.27745    0.76145   0.364  0.71559    
as.factor(year)1996    0.35457    0.74666   0.475  0.63488    
as.factor(year)1997    0.52990    0.71774   0.738  0.46035    
as.factor(year)1998    0.61382    0.71331   0.861  0.38952    
as.factor(year)1999    0.90922    0.67816   1.341  0.18003    
as.factor(year)2000    0.67905    0.67689   1.003  0.31578    
as.factor(year)2001    0.54965    0.68745   0.800  0.42399    
as.factor(year)2002    0.32882    0.69144   0.476  0.63440    
as.factor(year)2003    0.51805    0.66837   0.775  0.43830    
as.factor(year)2004    0.45757    0.66108   0.692  0.48885    
as.factor(year)2005    0.43152    0.65977   0.654  0.51309    
as.factor(year)2006    0.45605    0.65577   0.695  0.48679    
as.factor(year)2007    0.48383    0.65469   0.739  0.45991    
as.factor(year)2008   -0.02624    0.67972  -0.039  0.96921    
as.factor(year)2009    0.26434    0.66155   0.400  0.68948    
as.factor(year)2010    0.30334    0.65434   0.464  0.64296    
```

```
as.factor(year)2011    0.23797    0.65816   0.362  0.71768
as.factor(year)2012    0.28827    0.65657   0.439  0.66064
as.factor(year)2013    0.53615    0.64786   0.828  0.40793
as.factor(year)2014    0.59884    0.64748   0.925  0.35505
as.factor(indclass)2   0.19881    5.04249   0.039  0.96855
as.factor(indclass)3  -0.05618    5.14940  -0.011  0.99129
as.factor(indclass)4  -0.24387    5.06016  -0.048  0.96156
as.factor(indclass)5  -0.17673    5.06807  -0.035  0.97218
as.factor(indclass)6  -0.21872    5.06694  -0.043  0.96557
as.factor(indclass)7   0.13899    5.05464   0.027  0.97806
as.factor(indclass)8   0.76953    5.07297   0.152  0.87943
as.factor(indclass)9   0.13819    5.05142   0.027  0.97818
as.factor(indclass)10 -0.31590    5.08068  -0.062  0.95042
as.factor(indclass)11 -0.18474    5.17501  -0.036  0.97152
as.factor(indclass)12  0.16277    5.04780   0.032  0.97428
as.factor(indclass)13  0.07766    5.04405   0.015  0.98772
as.factor(indclass)14  0.01294    5.05010   0.003  0.99795
as.factor(indclass)15  0.03947    5.08928   0.008  0.99381
as.factor(indclass)16 -1.09371   20.22112  -0.054  0.95687
as.factor(indclass)17 -0.24635    5.10042  -0.048  0.96148
as.factor(indclass)18 -0.46888    7.02718  -0.067  0.94680
as.factor(indclass)19 -0.84852    5.05796  -0.168  0.86677
as.factor(indclass)20 -0.38654   22.39129  -0.017  0.98623
as.factor(indclass)21 -0.06444    5.04430  -0.013  0.98981
as.factor(indclass)22 -0.08837    5.06563  -0.017  0.98608
as.factor(indclass)23 -0.21732    5.04930  -0.043  0.96567
as.factor(indclass)24  0.31102    5.21777   0.060  0.95247
as.factor(indclass)25  0.17576    6.54256   0.027  0.97857
as.factor(indclass)26 -0.24422    5.34257  -0.046  0.96354
as.factor(indclass)28 -0.26046    5.20786  -0.050  0.96011
as.factor(indclass)30  0.54969   18.87199   0.029  0.97676
as.factor(indclass)32 -0.44098    5.05024  -0.087  0.93042
as.factor(indclass)33 -0.72876    5.35481  -0.136  0.89175
as.factor(indclass)34  0.10102    5.07586   0.020  0.98412
as.factor(indclass)35 -0.11812    5.04871  -0.023  0.98133
as.factor(indclass)36  0.19776    5.04255   0.039  0.96872
as.factor(indclass)37 -0.16110    5.04365  -0.032  0.97452
as.factor(indclass)38 -0.05569    5.04784  -0.011  0.99120
as.factor(indclass)39  0.35240    5.04896   0.070  0.94436
as.factor(indclass)41  0.68686    5.20809   0.132  0.89508
as.factor(indclass)42 -0.03443    5.05656  -0.007  0.99457
as.factor(indclass)43 -0.08237    5.04170  -0.016  0.98696
as.factor(indclass)44  0.22102    5.04917   0.044  0.96509
as.factor(indclass)49 -0.13977    9.66196  -0.014  0.98846
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 66662993274)

    Null deviance: 1.2915e+13  on 13524  degrees of freedom
Residual deviance: 7.2149e+12  on 13454  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 9
```

Model 3

```
fit_int <- glm(mvmil ~ ltdratio * indclass + capexratio * indclass +
    rdratio * indclass + adsratio * indclass + pperatio * indclass +
    ebitdaratio * indclass + as.factor(year) + as.factor(indclass) +
    offset(log(bkmil)), data = newdat, family = quasipoisson)
```

```
Anova(fit_int, test = "F")
```

```
Analysis of Deviance Table (Type II tests)

Response: mvmil
Error estimate based on Pearson residuals

                            SS    Df      F   Pr(>F)
ltdratio            3.3809e+10     1 0.1127 0.737075
indclass                           0
capexratio          6.7843e+09     1 0.0226 0.880455
rdratio             3.1265e+11     1 1.0424 0.307291
adsratio            2.1042e+09     1 0.0070 0.933250
pperatio            1.2300e+11     1 0.4101 0.521938
ebitdaratio         2.4157e+12     1 8.0538 0.004548 **
as.factor(year)     7.5237e+11    24 0.1045 1.000000
as.factor(indclass) 1.0818e+12    39 0.0925 1.000000
ltdratio:indclass   1.0372e+11     1 0.3458 0.556520
indclass:capexratio 1.2573e+10     1 0.0419 0.837782
indclass:rdratio    3.4774e+10     1 0.1159 0.733491
indclass:adsratio   9.0119e+10     1 0.3005 0.583607
indclass:pperatio   6.8379e+10     1 0.2280 0.633040
indclass:ebitdaratio 3.3348e+10    1 0.1112 0.738810
Residuals           4.0336e+15 13448
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_int)
```

```
Call:
glm(formula = mvmil ~ ltdratio * indclass + capexratio * indclass +
    rdratio * indclass + adsratio * indclass + pperatio * indclass +
    ebitdaratio * indclass + as.factor(year) + as.factor(indclass) +
    offset(log(bkmil)), family = quasipoisson, data = newdat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-230064    -7668    -1314     6077   362649

Coefficients: (1 not defined because of singularities)
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           -0.63271   11.01236  -0.057    0.954
ltdratio               1.36575    2.00648   0.681    0.496
indclass               0.01221    0.42746   0.029    0.977
capexratio            -0.56723    3.93501  -0.144    0.885
rdratio                0.37030    1.49639   0.247    0.805
adsratio               2.76156    5.42572   0.509    0.611
pperatio               0.26479    1.27019   0.208    0.835
```

ebitdaratio	5.23448	3.57436	1.464	0.143
as.factor(year)1991	0.30295	1.67913	0.180	0.857
as.factor(year)1992	0.36334	1.63945	0.222	0.825
as.factor(year)1993	0.45517	1.61319	0.282	0.778
as.factor(year)1994	0.17041	1.69783	0.100	0.920
as.factor(year)1995	0.28223	1.61601	0.175	0.861
as.factor(year)1996	0.37081	1.58508	0.234	0.815
as.factor(year)1997	0.54665	1.52382	0.359	0.720
as.factor(year)1998	0.64208	1.51533	0.424	0.672
as.factor(year)1999	0.92943	1.44193	0.645	0.519
as.factor(year)2000	0.71532	1.44019	0.497	0.619
as.factor(year)2001	0.57263	1.46313	0.391	0.696
as.factor(year)2002	0.35698	1.47097	0.243	0.808
as.factor(year)2003	0.54750	1.42198	0.385	0.700
as.factor(year)2004	0.49677	1.40767	0.353	0.724
as.factor(year)2005	0.47219	1.40447	0.336	0.737
as.factor(year)2006	0.51218	1.39624	0.367	0.714
as.factor(year)2007	0.54228	1.39421	0.389	0.697
as.factor(year)2008	0.02573	1.44782	0.018	0.986
as.factor(year)2009	0.31362	1.40877	0.223	0.824
as.factor(year)2010	0.35478	1.39403	0.255	0.799
as.factor(year)2011	0.29278	1.40227	0.209	0.835
as.factor(year)2012	0.34364	1.39905	0.246	0.806
as.factor(year)2013	0.58177	1.37995	0.422	0.673
as.factor(year)2014	0.64788	1.37920	0.470	0.639
as.factor(indclass)2	-0.01959	10.48681	-0.002	0.999
as.factor(indclass)3	-0.19573	10.52515	-0.019	0.985
as.factor(indclass)4	-0.27676	10.13284	-0.027	0.978
as.factor(indclass)5	-0.05919	9.98400	-0.006	0.995
as.factor(indclass)6	-0.26137	9.82801	-0.027	0.979
as.factor(indclass)7	-0.16979	9.65351	-0.018	0.986
as.factor(indclass)8	0.74659	9.57425	0.078	0.938
as.factor(indclass)9	0.05544	9.41097	0.006	0.995
as.factor(indclass)10	-0.20469	9.39338	-0.022	0.983
as.factor(indclass)11	-0.13878	9.54831	-0.015	0.988
as.factor(indclass)12	0.30402	9.19170	0.033	0.974
as.factor(indclass)13	0.23072	9.15005	0.025	0.980
as.factor(indclass)14	0.06553	9.14890	0.007	0.994
as.factor(indclass)15	0.06807	9.25568	0.007	0.994
as.factor(indclass)16	-0.99674	42.53537	-0.023	0.981
as.factor(indclass)17	-0.12894	9.35931	-0.014	0.989
as.factor(indclass)18	-0.26670	13.92953	-0.019	0.985
as.factor(indclass)19	-0.78896	9.40893	-0.084	0.933
as.factor(indclass)20	-0.25820	47.23408	-0.005	0.996
as.factor(indclass)21	0.10483	9.60418	0.011	0.991
as.factor(indclass)22	0.08799	9.79568	0.009	0.993
as.factor(indclass)23	-0.02537	9.91323	-0.003	0.998
as.factor(indclass)24	0.51905	10.46328	0.050	0.960
as.factor(indclass)25	0.37983	13.54169	0.028	0.978
as.factor(indclass)26	-0.08786	11.10893	-0.008	0.994
as.factor(indclass)28	-0.06172	11.24521	-0.005	0.996
as.factor(indclass)30	0.96149	40.22048	0.024	0.981
as.factor(indclass)32	-0.15523	11.93419	-0.013	0.990
as.factor(indclass)33	-0.42239	12.78404	-0.033	0.974

```
as.factor(indclass)34  0.34957   12.54477   0.028    0.978
as.factor(indclass)35  0.06435   12.79039   0.005    0.996
as.factor(indclass)36  0.40417   13.08230   0.031    0.975
as.factor(indclass)37  0.03766   13.39400   0.003    0.998
as.factor(indclass)38  0.12255   13.71462   0.009    0.993
as.factor(indclass)39  0.80774   14.04299   0.058    0.954
as.factor(indclass)41  0.89597   14.94985   0.060    0.952
as.factor(indclass)42  0.25956   15.04337   0.017    0.986
as.factor(indclass)43  0.37702   15.36549   0.025    0.980
as.factor(indclass)44  0.75342   15.72398   0.048    0.962
as.factor(indclass)49       NA         NA      NA       NA
ltdratio:indclass     -0.03915    0.06673  -0.587    0.557
indclass:capexratio    0.02645    0.13016   0.203    0.839
indclass:rdratio       0.02599    0.07073   0.367    0.713
indclass:adsratio     -0.12568    0.23435  -0.536    0.592
indclass:pperatio     -0.01824    0.03806  -0.479    0.632
indclass:ebitdaratio  -0.03621    0.10834  -0.334    0.738

(Dispersion parameter for quasipoisson family taken to be 2.99912e+11)

    Null deviance: 1.2915e+13  on 13524  degrees of freedom
Residual deviance: 6.8156e+12  on 13448  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 9
```

Questions

1. Which of the following is FALSE?

    - It is always wise to mention the removal of any values as part of any reporting to ensure the reader understands the extent of the data being modelled.
    - The variance inflation factors show no cause for concern once the very large covariate values are removed.
    - A histogram of the Tobin's Q values would be useful before any modelling is carried out since this will tell us definitively if a Normal errors model would be appropriate. caution: if few data
    - It is often wise to remove extreme values for one or more covariates; there is little support otherwise for the modelled relationship assumed to hold for the entire covariate range.
    - It is very difficult to see the shape of the relationships in this case between each covariate and the response due to over-plotting.

2. Which of the following is FALSE?

    - It is difficult to tell from the scatterplots alone if there are any genuine relationships between the covariates and the response, or if any visible patterns are simply due to chance.
    - The average Tobin's Q score appears to decrease as capexratio and/or rdratio increases. P11
    - The variability in the Tobin's Q scores appears to decrease as capexratio and/or rdratio increases.
    - While there is a great deal of variability in the data, it is still clear from the scatterplots that we will be able to predict Tobin's Q with a great deal of precision
    - The pearson correlation coefficients between each covariate and the response wouldnt be very informative here since the relationships are not likely to be linear.

3. Which of the following about the results from Model 1 is FALSE?

    - If the logLik for Model 1 is -1000, it should be approximately -890 for the model fitted with year as a linear term (rather than a factor) to obtain the same BIC score for the two models.
    - All covariates show significantly non-zero (specifically linear) relationships at the 5% level with Tobin's Q except the 'capexratio' and 'pperatio' variables
    - We can confirm from the Anova tables (for all models) that by removing these extra large covariate values this has not reduced the number of years of data available or reduced the levels of 'indclass' (as stated in the Introduction)
    - The average value of Tobin's Q in year 1990 (all other things held constant) was not significantly different (at the 5% level) to the average value in 8 of the subsequent years.
    - Tobin's Q was significantly higher (at the 5% level) in 16 of the years subsequent to the average response value in 1990 (all other things held constant).

4. Which of the following about the results from Model 1 is FALSE?

    - The collection of diagnostics presented here indicates that inference is not safe using this model
    - The model over-predicts the very large Tobin's Q scores; the fitted values do not exceed 20 while the observed values are as high as approximately 200.
    - The residuals are heavily right skewed and this skewness is necessarily caused by the very large observed values in the response.
    - The residuals appear to be correlated through time; there seem to be systematic patterns in the residuals when plotted in order.
    - The residual variance appears to increase sharply when the mean of the fitted values exceeds approximately 5.

5. Which of the following about the results from Models 2 and 3 is FALSE?

    - The response was converted into millions of USD in order to return a ratio based on whole numbers for the ratio of the Tobin's Q score; this respects the requirement for integer only data for Poisson based models.

- The Poisson based model returns significant relationships for all covariates because it unrealistically assumes a dispersion parameter of 1.
- When a quasi-Poisson model is fitted (without interactions) only two of the covariates are significant at the 5% level - fewer than the model assuming the dispersion parameter is equal to 1.
- Based on the quasi-Poisson model, there is no evidence that any of the covariate relationships trialled vary with industry classes ('indclass'). however these interaction terms would be need to removed one by one to ensure we do not erroneously remove any interaction terms as part of this process.
- Under the quasi-Poisson-based model, Tobin's Q significantly increases with expenditure on research and development (as a ratio of total sales) and also significantly increases (more quickly) with operating profits (as a ratio of total assets). The value of the coefficients can vary sometimes between models when a dispersion parameter is estimated however (as compared with assuming this is equal to 1) and so the analyst should inspect both sets of output to be sure. 43.7711965

6. Calculate the QAIC score for Model 3 based on the output associated with Model 2 and Model 3.
178.0565766
7. TRUE or FALSE? If the relationships between covariates are nonlinear (but strongly related) it is possible for a VIF to return a relatively low value.
8. Which of the following is FALSE?
   - Models which are more complicated than the underlying function tend to exhibit high variance since they give predictions which tend to vary greatly across data sets generated from the same underlying function.
   - Models which are more simplistic than the underlying function tend to exhibit high bias since they give predictions dont model the data very closely.
   - Models which are more simplistic than the underlying function tend to exhibit low variance since they tend to give predictions which are similar regardless of the particular sample of data generated from the same underlying function.+ We seek models which are neither overfitted or underfitted and in particular we seek models with low bias and high variance.
   - Models which are more complicated than the underlying function tend to exhibit low bias since they give predictions which are very close to the observed data.
9. TRUE or FALSE? One remedy for collinearity is to remove one or more of the collinear variables from the model and re-fit the model.
10. Which of the following is FALSE?
    - Ridge regression and lasso regression are both specials cases of the Elastic net.
    - When the penalty is zero (regardless of the method), the coefficients are identical to those obtained using maximum/quasi-likelihood.
    - Penalty based methods are likely to give strikingly different predictions compared with the results from Models 1, 2 and 3 since some of the covariates appear to be highly collinear.
    - Ridge regression is one method we can use to model correlated covariates simultaneously in a model which reduces the size of the coefficients (so they are closer to zero) but this means a bias is introduced (the predictions tend to be systematically too large or systematically too small).
    - The Lasso method is also a penalty based method but this permits some coefficents to be exactly zero and therefore effectively removes these coefficients from the model.