

MT5764: Advanced Data Analysis

Weekly Questions One

Rachel Sippy

20 January 2022

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 (i.e. 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. The dataset ($N = 13539$), has been collected over 25 years from 1240 American companies which are sourced from 41 industry classes.

Note: Don't forget to answer the weekly questions at the end of this document (and every weekly questions in the future) on [Moodle](#).

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.

Data description

The data set contains the following variables:

```
# Libraries
library(tidyverse) # ggplot()

# Read dataset
df <- read.csv("TobinsQ.csv")
df <- na.omit(df) # remove rows with NA

# Set ID as factor
df$ID <- as.factor(df$ID)

dim(df)

[1] 13539     20

head(df)

  ID Year Assets CapEx    LTD EBITDA     PPE   Sales   Ads     RD BookValue
1 1050 1994  7.231 0.156  1.399  1.297  3.101 10.327 0.072 0.044      3.17
2 1050 1995  8.645 0.168  1.253  0.486  3.289  8.435 0.139 0.017      4.36
3 1050 1996  9.201 0.079  1.142  1.135  3.360  9.848 0.131 0.117      5.11
4 1050 1997 13.961 0.210  1.737  0.860  3.701 14.531 0.117 0.092      6.74
5 1050 1998 15.475 0.241  1.570  1.755  4.035 26.382 0.158 0.097      7.56
6 1050 1999 51.073 0.165 28.290  1.171 16.542 23.862 0.087 0.033      9.04
  MarketValue IndClass TobinsQ      rLTD      rCapEx      rRD      rAds
1       11.42 Machinery 3.602524 0.1934725 0.015106033 0.004260676 0.006972015
2       20.88 Machinery 4.788991 0.1449393 0.019917012 0.002015412 0.016478957
3       14.40 Machinery 2.818004 0.1241169 0.008021933 0.011880585 0.013302193
4       24.40 Machinery 3.620178 0.1244180 0.014451862 0.006331292 0.008051751
5       24.75 Machinery 3.273810 0.1014540 0.009135016 0.003676749 0.005988932
6       20.63 Machinery 2.282080 0.5539130 0.006914760 0.001382952 0.003645964
  rPPE      rEBITDA
1 0.4288480 0.17936662
2 0.3804511 0.05621747
3 0.3651777 0.12335616
4 0.2650956 0.06160017
5 0.2607431 0.11340872
6 0.3238893 0.02292797
```

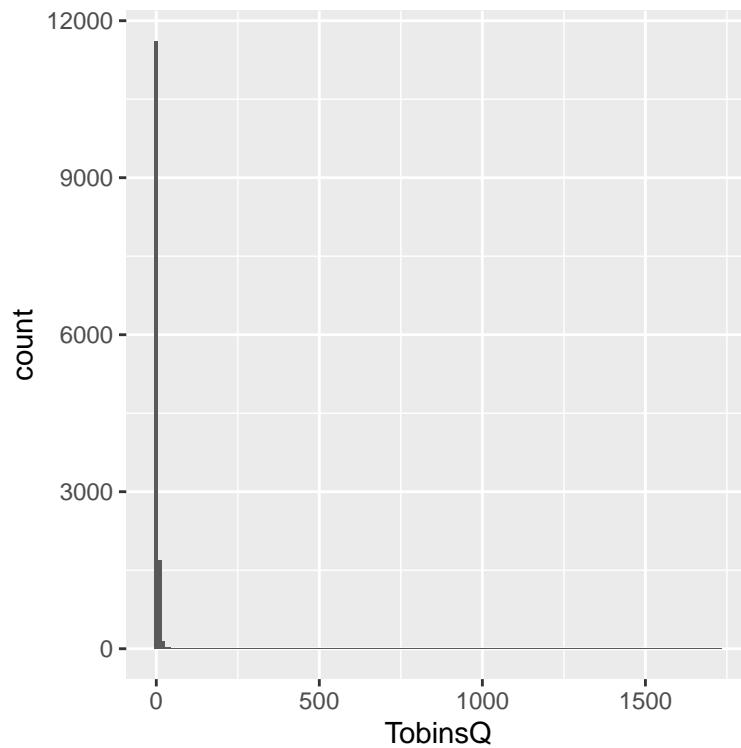
- **TobinsQ** (response variable): the ratio between a physical assets market value and its replacement value (i.e. **MarketValue/BookValue**).
- **ID**: Company identifier (anonymised at source).
- **Year**: financial year.
- **Assets**: value of assets.
- **CapEx**: **capital expenditure**.
- **LTD**: long-term debt.
- **EBITDA**: **Earnings Before Interest, Taxes, Depreciation and Amortization** (i.e. operating profits).
- **PPE**: value of the **property and plant equipment**.
- **Sales**: value of sales.

- **Ads**: cost of advertising.
- **RD**: research and development expenses.
- **BookValue**: book value.
- **MarketValue**: market value externally assessed.
- **IndClass**: type of industry.
- **rLTD**: long-term debt to total asset ratio (i.e. LTD/Assets).
- **rCapEx**: CapEx to Sales ratio.
- **rRD**: RD to Sales ratio.
- **rAds**: Ads to Sales ratio.
- **rPPE**: PPE to Assets ratio.
- **rEBITDA**: EBITDA to Assets ratio.

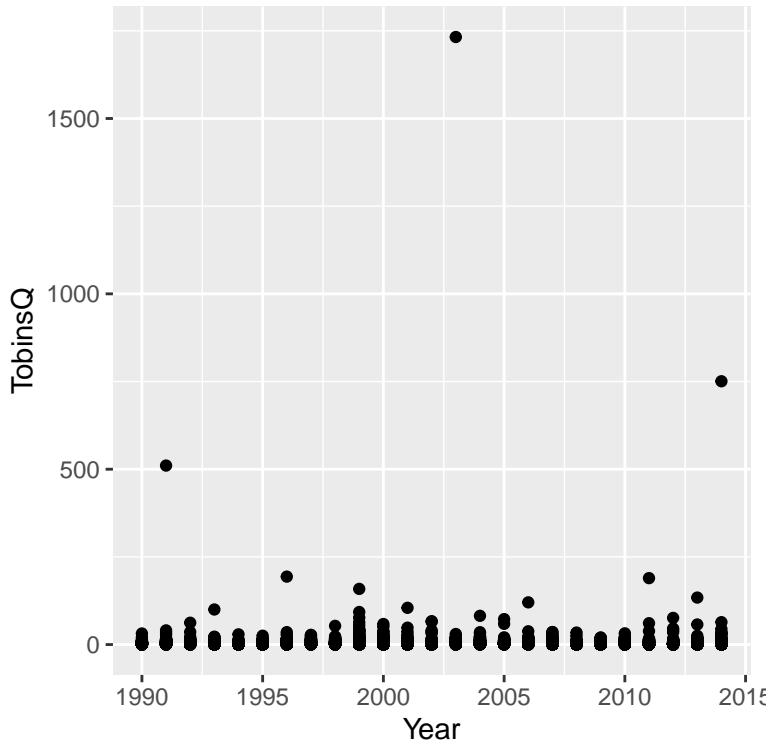
Exploratory analysis

Let's start by looking at a histogram of our response and then plot the whole dataset over time:

```
# Histogram
ggplot(df, aes(x=TobinsQ)) +
  geom_histogram(binwidth=10)
```



```
# TobinsQ over time
ggplot(df, aes(x=Year, y=TobinsQ)) +
  geom_point()
```



Initial inspection reveals some extremely large TobinsQ values.

Let's pick 9 industry types at random and plot TobinsQ over time to get an idea of the within and across companies variation (within and across different industry classes).

Note: change in y-axis scaling.

```
# Number of companies per industry type
table(df$IndClass)
```

Agriculture	Aircraft	Banking
23	14	757
BeerLiquor	BusinessServices	BusinessSupplies
50	246	546
CandySoda	CarsTrucks	Chemicals
13	255	229
Clothing	Computers	Construction
168	644	18
ConstructionMaterials	ConsumerGoods	Defense
215	335	52
ElectricalEquipment	ElectronicEquipment	Entertainment
407	2266	175
FabricatedProducts	FoodProducts	Healthcare
8	327	118
LabEquipment	Machinery	MedicalEquipment
1192	574	751
Mining	Oil	Other
23	6	15
PersonalServices	Pharmaceutical	Publishing

	56	838	40
Recreation		Restaurants	Retail
	274	2066	352
RubberPlastic		Shipbuilding	ShippingContainers
	82	12	113
SteelWorks		Telecom	Textiles
	54	190	9
Tobacco		Wholesale	
	15	11	

```
# Extract 9 industry types at random
set.seed(70456)
myIndClass <- sample(unique(df$IndClass), 9)
```

```
# Plot TobinsQ over time for 9 industry types at random
ggplot(subset(df, IndClass %in% myIndClass),
       aes(x=Year, y=TobinsQ, colour=ID)) +
  geom_point() +
  geom_line() +
  facet_wrap(~IndClass, scales="free_y") +
  theme(legend.position="none")
```



The extremely large TobinsQ values swamp the overall scale. Let's plot on the log-scale instead.

```
# Plot TobinsQ over time for 9 industry types at random
ggplot(subset(df, IndClass %in% myIndClass),
       aes(x=Year, y=log(TobinsQ), colour=ID)) +
  geom_point() +
  geom_line() +
  facet_wrap(~IndClass, scales="free_y") +
  theme(legend.position="none")
```



The dataset is very heterogeneous; there's a lot of variation over time, both within and across companies, and within and across different industry types.

Let's plot $\log(\text{TobinsQ})$ against some of the covariates of interest (mainly the ratio quantities):

```
# Libraries
library(gridExtra) # to arrange ggplots in a grid

# Create list of plots of log(TobinsQ) against other covariates
p <- list()
for (covariate in c("IndClass", "rLTD", "rCapEx", "rRD",
                    "rAds", "rPPE", "rEBITDA"))
{
  p[[covariate]] <- ggplot(df, aes_string(x=covariate,
```

```

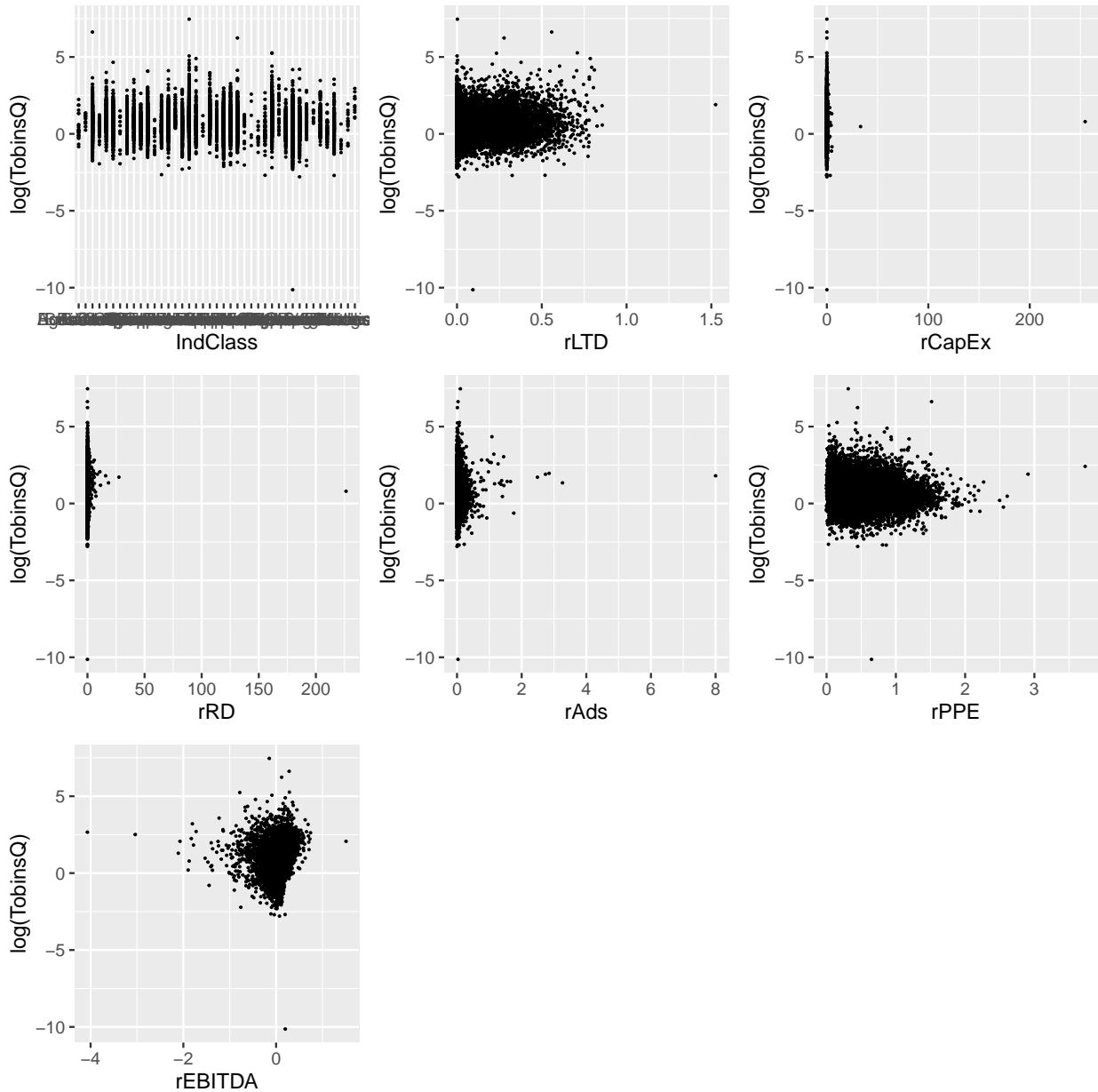
        y="log(TobinsQ))" )+
  geom_point(size=0.2)
}

```

```

# Arrange ggplots in a grid
grid.arrange(grobs=p)

```



As noted earlier we observe some extremely large TobinsQ values, but also some covariate values that are way outside their “normal” range.

Upon consulting with field experts, it was decided that it is sensible to work on a filtered version of the dataset.

Let's filter out the extremely large values for both response and covariates.

```
newDf <- df %>%
  filter(rLTD<1 & rCapEx<20 & rRD<10 & rAds<2 & TobinsQ<200)
dim(df)
```

```
[1] 13539     20
```

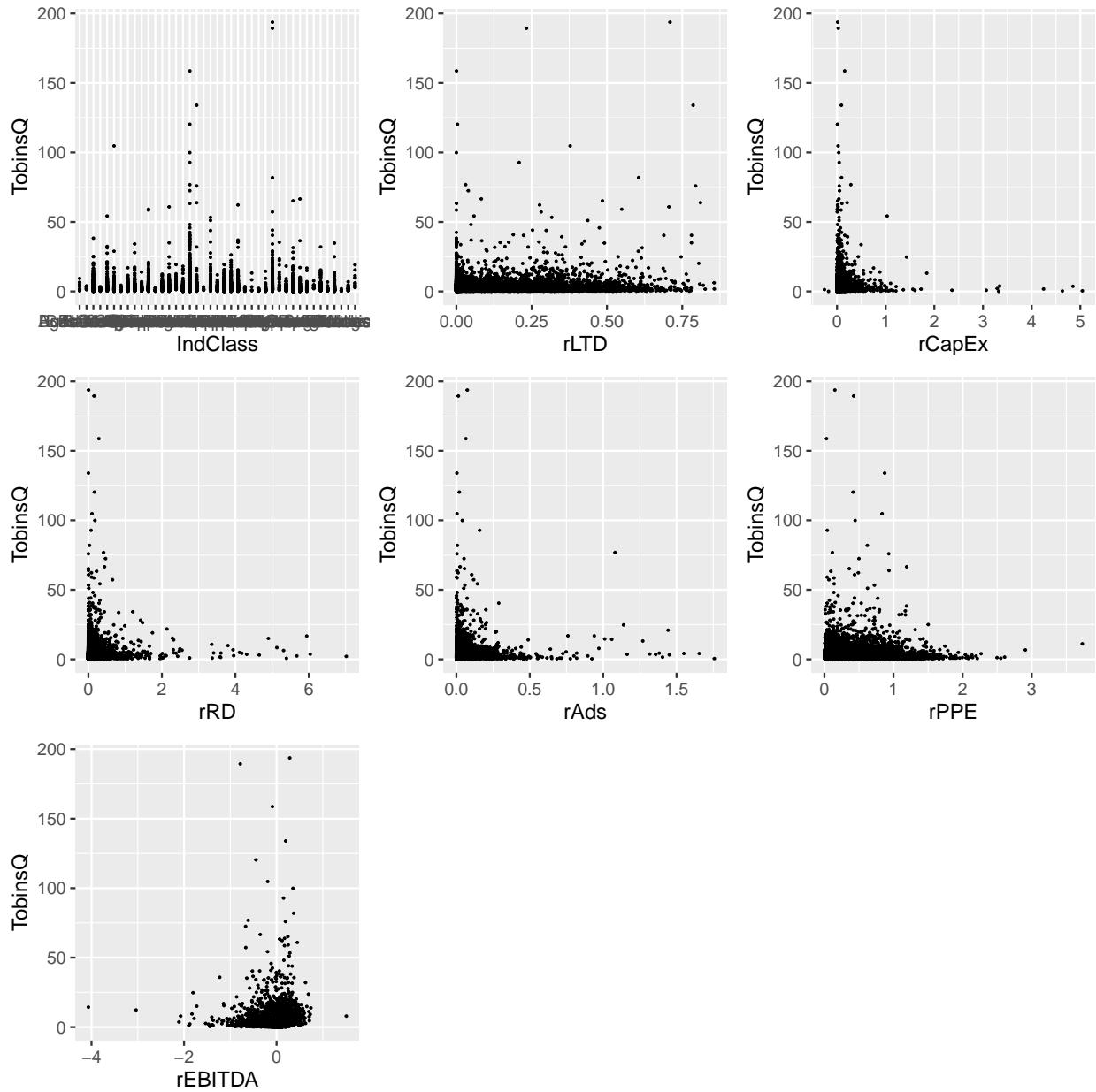
```
dim(newDf)
```

```
[1] 13525     20
```

Let's replot the new filtered data for both unlogged and logged values of `TobinsQ`.

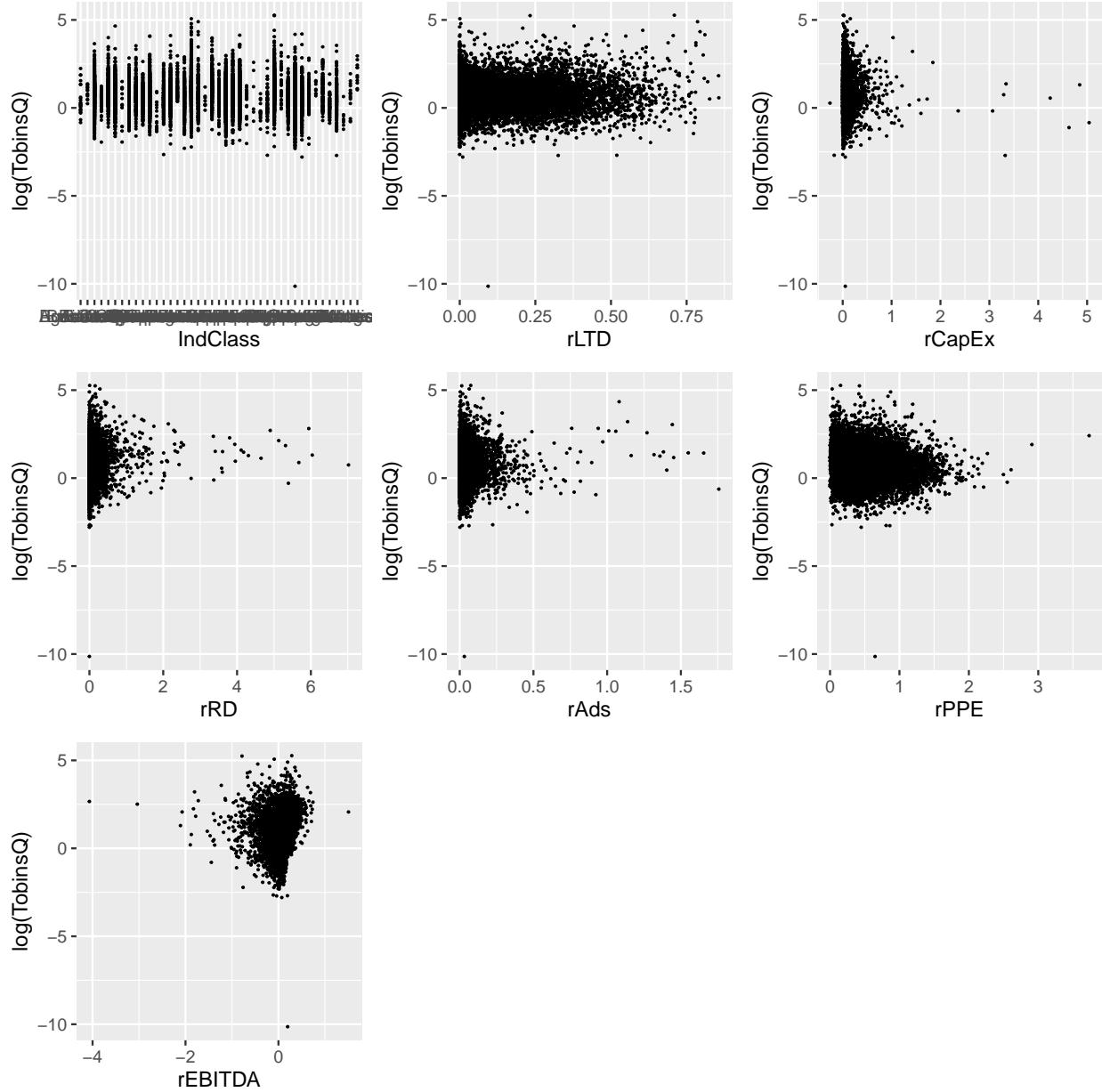
```
# Create list of plots of TobinsQ against other covariates
p <- list()
for (covariate in c("IndClass", "rLTD", "rCapEx", "rRD",
                     "rAds", "rPPE", "rEBITDA"))
{
  p[[covariate]] <- ggplot(newDf, aes_string(x=covariate,
                                              y="TobinsQ")) +
    geom_point(size=0.2)
}

# Arrange ggplots in a grid
grid.arrange(grobs=p)
```



```
# Create list of plots of log(TobinsQ) against other covariates
p <- list()
for (covariate in c("IndClass", "rLTD", "rCapEx", "rRD",
                    "rAds", "rPPE", "rEBITDA"))
{
  p[[covariate]] <- ggplot(newDf, aes_string(x=covariate,
                                              y="log(TobinsQ)")) +
    geom_point(size=0.2)
}

# Arrange ggplots in a grid
grid.arrange(grobs=p)
```



Model specification and fitting

Model 1

Assume TobinsQ is normally distributed. A model with all candidate variables included reveals not all covariates are significantly related to the response.

```
library(car) # for Anova and vif
mdl <- glm(TobinsQ ~ rLTD + rCapEx + rRD + rAds + rPPE + rEBITDA +
           as.factor(Year) + IndClass, data=newDf)
summary(mdl)
```

Call:

```
glm(formula = TobinsQ ~ rLTD + rCapEx + rRD + rAds + rPPE + rEBITDA +  
as.factor(Year) + IndClass, data = newDf)
```

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
rLTD	2.50805	0.32689	7.672	1.80e-14 ***
rCapEx	0.12032	0.37208	0.323	0.746425
rRD	1.07732	0.21614	4.984	6.29e-07 ***
rAds	2.64664	0.67291	3.933	8.43e-05 ***
rPPE	-0.10843	0.17013	-0.637	0.523906
rEBITDA	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 ***
IndClassAircraft	0.55182	1.73504	0.318	0.750457
IndClassBanking	0.21422	1.08986	0.197	0.844173
IndClassBeerLiquor	-0.94130	1.29173	-0.729	0.466191
IndClassBusinessServices	1.58045	1.11623	1.416	0.156834
IndClassBusinessSupplies	0.37124	1.08947	0.341	0.733293
IndClassCandySoda	-0.73503	1.77702	-0.414	0.679150
IndClassCarsTrucks	-0.07082	1.11455	-0.064	0.949336
IndClassChemicals	0.33799	1.12026	0.302	0.762879
IndClassClothing	-0.28114	1.13888	-0.247	0.805020
IndClassComputers	0.41276	1.08636	0.380	0.703991
IndClassConstruction	-2.00792	1.61473	-1.243	0.213705
IndClassConstructionMaterials	-0.42883	1.12354	-0.382	0.702709
IndClassConsumerGoods	0.25356	1.10428	0.230	0.818392

IndClassDefense	1.86536	1.28282	1.454	0.145939
IndClassElectricalEquipment	-0.29389	1.09709	-0.268	0.788798
IndClassElectronicEquipment	2.11349	1.07336	1.969	0.048970 *
IndClassEntertainment	1.17056	1.14407	1.023	0.306255
IndClassFabricatedProducts	-0.82631	2.10508	-0.393	0.694672
IndClassFoodProducts	2.09192	1.10514	1.893	0.058394 .
IndClassHealthcare	-0.13318	1.16676	-0.114	0.909123
IndClassLabEquipment	0.02149	1.07761	0.020	0.984086
IndClassMachinery	0.26428	1.08839	0.243	0.808150
IndClassMedicalEquipment	0.91036	1.08347	0.840	0.400799
IndClassMining	-0.54985	1.51186	-0.364	0.716096
IndClassOil	2.59045	2.35255	1.101	0.270862
IndClassOther	-1.04508	1.69889	-0.615	0.538463
IndClassPersonalServices	-0.52966	1.26963	-0.417	0.676555
IndClassPharmaceutical	2.34357	1.08253	2.165	0.030413 *
IndClassPublishing	4.43601	1.34011	3.310	0.000935 ***
IndClassRecreation	-0.49397	1.11166	-0.444	0.656795
IndClassRestaurants	-0.13223	1.07406	-0.123	0.902023
IndClassRetail	0.11785	1.10177	0.107	0.914822
IndClassRubberPlastic	0.08343	1.21011	0.069	0.945032
IndClassShipbuilding	1.63593	1.82283	0.897	0.369485
IndClassShippingContainers	1.08465	1.17290	0.925	0.355107
IndClassSteelWorks	-1.08432	1.27581	-0.850	0.395391
IndClassTelecom	0.24347	1.13418	0.215	0.830030
IndClassTextiles	-1.79135	2.01433	-0.889	0.373856
IndClassTobacco	-0.09236	1.69906	-0.054	0.956648
IndClassWholesale	5.90316	1.87708	3.145	0.001665 **

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

Anova(mdl)

Analysis of Deviance Table (Type II tests)

Response: TobinsQ

	LR	Chisq	Df	Pr(>Chisq)
rLTD	58.87	1	1.688e-14	***
rCapEx	0.10	1	0.7464	
rRD	24.84	1	6.215e-07	***
rAds	15.47	1	8.385e-05	***
rPPE	0.41	1	0.5239	
rEBITDA	23.54	1	1.223e-06	***
as.factor(Year)	231.40	24	< 2.2e-16	***
IndClass	450.32	40	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

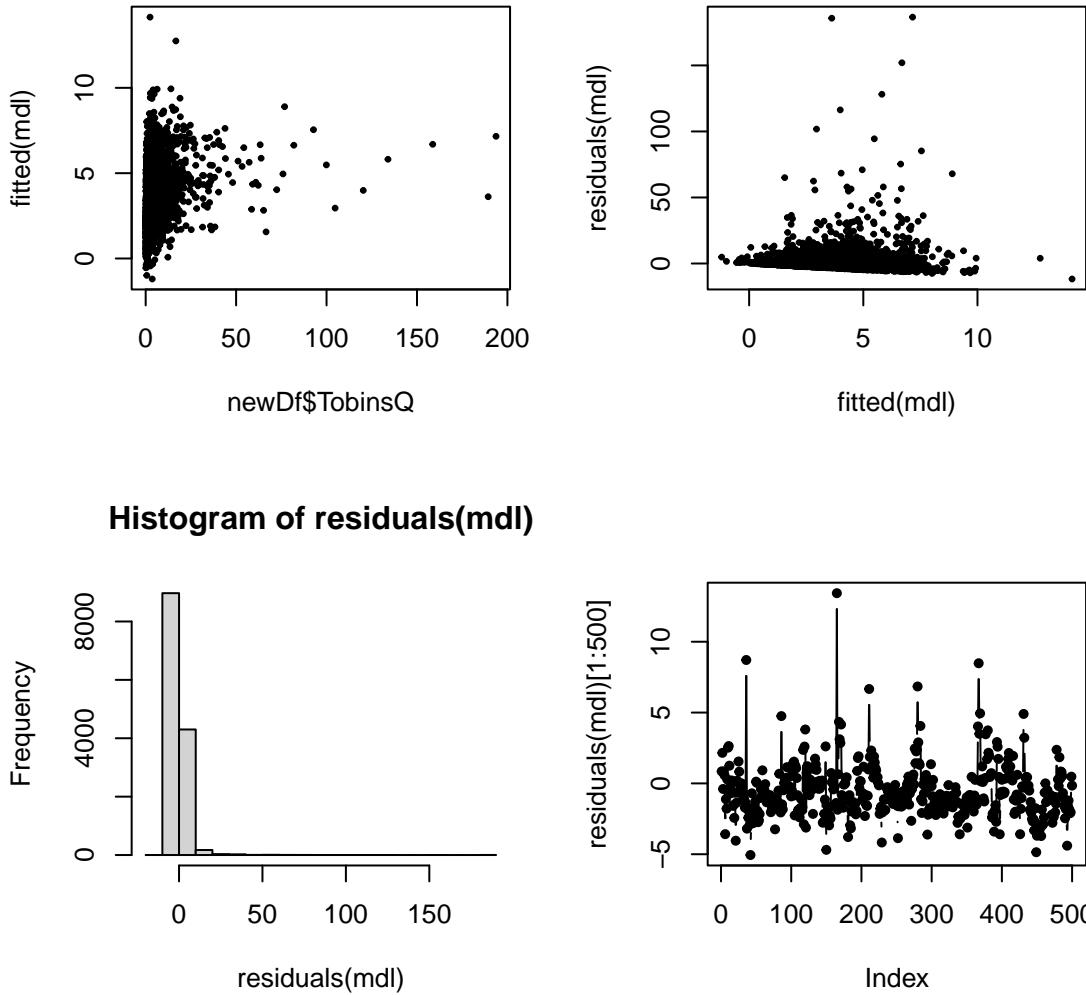
```
vif(mdl)
```

	GVIF	Df	GVIF^(1/(2*Df))
rLTD	1.232321	1	1.110099
rCapEx	1.213780	1	1.101717
rRD	1.395413	1	1.181276
rAds	1.182557	1	1.087454
rPPE	1.679157	1	1.295823
rEBITDA	1.287513	1	1.134686
as.factor(Year)	1.140738	24	1.002747
IndClass	2.534756	40	1.011694

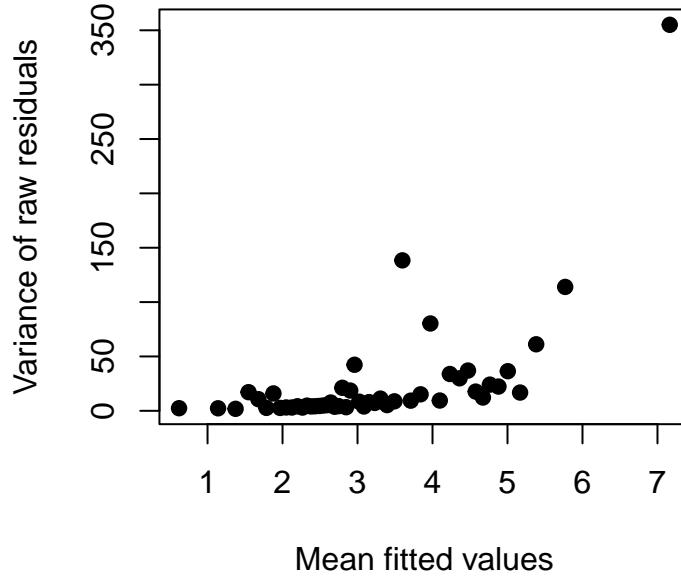
```
logLik(mdl)
```

'log Lik.' -41232.35 (df=72)

```
# Residual plots
par(mfrow=c(2, 2))
plot(newDf$TobinsQ, fitted(mdl), pch=19, cex=0.4)
plot(fitted(mdl), residuals(mdl), pch=19, cex=0.4)
hist(residuals(mdl))
plot(residuals(mdl)[1:500], type="b", pch=20)
```



```
# Mean-variance relationship plot
par(mfrow=c(1, 1))
# "Bin" the x-axis
xbin <- cut(fitted(mdl),
             breaks=quantile(fitted(mdl),
                             probs=c(seq(0, 1, length=50))))
# Compute mean of xbin
xmean <- tapply(fitted(mdl), xbin, mean)
# Compute var of residuals within xbin
resvar <- tapply(residuals(mdl), xbin, var)
# Plot
plot(xmean, resvar, xlab="Mean fitted values",
      ylab="Variance of raw residuals", pch=19)
```



Model 2

Assume MarketValue scaled to million is Poisson distributed.

```
# Scale market and book value to million
newDf <- newDf %>%
  mutate(MarketValueMil=MarketValue*1e6,
        BookValueMil=BookValue*1e6)

# Poisson GLM model with offset
mdlPois <- glm(MarketValueMil ~ rLTD + rCapEx + rRD + rAds + rPPE +
                 rEBITDA + as.factor(Year) + IndClass,
                 offset=log(BookValueMil), data=newDf, family=poisson)
summary(mdlPois)
```

Call:
`glm(formula = MarketValueMil ~ rLTD + rCapEx + rRD + rAds + rPPE +
 rEBITDA + as.factor(Year) + IndClass, family = poisson, data = newDf,
 offset = log(BookValueMil))`

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 ***

rLTD	2.594e-01	1.799e-06	144167.4	<2e-16 ***
rCapEx	1.917e-01	1.941e-06	98777.5	<2e-16 ***
rRD	7.717e-01	1.035e-06	745765.0	<2e-16 ***
rAds	5.612e-01	5.438e-06	103199.5	<2e-16 ***
rPPE	-2.787e-01	8.484e-07	-328520.1	<2e-16 ***
rEBITDA	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 ***
IndClassAircraft	3.110e-01	2.021e-05	15390.5	<2e-16 ***
IndClassBanking	2.210e-01	1.956e-05	11302.1	<2e-16 ***
IndClassBeerLiquor	-2.439e-01	1.960e-05	-12443.4	<2e-16 ***
IndClassBusinessServices	1.010e-01	1.966e-05	5138.6	<2e-16 ***
IndClassBusinessSupplies	-5.569e-02	1.955e-05	-2848.6	<2e-16 ***
IndClassCandySoda	-5.618e-02	1.994e-05	-2817.0	<2e-16 ***
IndClassCarsTrucks	-2.173e-01	1.956e-05	-11112.5	<2e-16 ***
IndClassChemicals	1.294e-02	1.956e-05	661.8	<2e-16 ***
IndClassClothing	-3.159e-01	1.968e-05	-16053.3	<2e-16 ***
IndClassComputers	-1.181e-01	1.955e-05	-6040.8	<2e-16 ***
IndClassConstruction	-4.689e-01	2.722e-05	-17227.5	<2e-16 ***
IndClassConstructionMaterials	-2.463e-01	1.975e-05	-12470.6	<2e-16 ***
IndClassConsumerGoods	1.382e-01	1.956e-05	7063.1	<2e-16 ***
IndClassDefense	-2.442e-01	2.069e-05	-11802.5	<2e-16 ***
IndClassElectricalEquipment	-8.837e-02	1.962e-05	-4504.3	<2e-16 ***
IndClassElectronicEquipment	1.978e-01	1.953e-05	10125.7	<2e-16 ***
IndClassEntertainment	1.390e-01	1.958e-05	7099.5	<2e-16 ***
IndClassFabricatedProducts	-3.865e-01	8.672e-05	-4457.2	<2e-16 ***
IndClassFoodProducts	1.988e-01	1.953e-05	10179.7	<2e-16 ***
IndClassHealthcare	-1.847e-01	2.004e-05	-9217.3	<2e-16 ***
IndClassLabEquipment	-1.611e-01	1.953e-05	-8246.8	<2e-16 ***
IndClassMachinery	-6.444e-02	1.954e-05	-3298.2	<2e-16 ***
IndClassMedicalEquipment	1.628e-01	1.955e-05	8325.6	<2e-16 ***
IndClassMining	-2.605e-01	2.017e-05	-12912.7	<2e-16 ***

```

IndClassOil           5.497e-01  7.309e-05   7520.4 <2e-16 ***
IndClassOther          -1.398e-01 3.742e-05  -3735.0 <2e-16 ***
IndClassPersonalServices -7.288e-01 2.074e-05 -35138.6 <2e-16 ***
IndClassPharmaceutical 7.766e-02 1.954e-05  3975.1 <2e-16 ***
IndClassPublishing     7.695e-01 1.965e-05  39165.7 <2e-16 ***
IndClassRecreation      -2.187e-01 1.962e-05 -11145.2 <2e-16 ***
IndClassRestaurants     -8.237e-02 1.953e-05 -4218.3 <2e-16 ***
IndClassRetail           -3.443e-02 1.958e-05 -1757.9 <2e-16 ***
IndClassRubberPlastic    3.947e-02 1.971e-05  2002.5 <2e-16 ***
IndClassShipbuilding     1.758e-01 2.534e-05  6936.2 <2e-16 ***
IndClassShippingContainers 3.524e-01 1.956e-05  18021.0 <2e-16 ***
IndClassSteelWorks       -8.485e-01 1.959e-05 -43314.3 <2e-16 ***
IndClassTelecom          -4.410e-01 1.956e-05 -22545.1 <2e-16 ***
IndClassTextiles          -1.094e+00 7.832e-05 -13965.0 <2e-16 ***
IndClassTobacco           -1.767e-01 1.963e-05 -9003.4 <2e-16 ***
IndClassWholesale         6.869e-01 2.017e-05  34051.0 <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

Anova(mdlPois)

Analysis of Deviance Table (Type II tests)

Response: MarketValueMil

	LR	Chisq	Df	Pr(>Chisq)
rLTD	2.0656e+10	1	< 2.2e-16	***
rCapEx	8.3793e+09	1	< 2.2e-16	***
rRD	3.0448e+11	1	< 2.2e-16	***
rAds	1.0340e+10	1	< 2.2e-16	***
rPPE	1.0951e+11	1	< 2.2e-16	***
rEBITDA	2.6136e+12	1	< 2.2e-16	***
as.factor(Year)	7.7533e+11	24	< 2.2e-16	***
IndClass	1.0698e+12	40	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

vif(mdlPois)

	GVIF	Df	GVIF^(1/(2*Df))
rLTD	1.419477	1	1.191418
rCapEx	1.378527	1	1.174107
rRD	1.547548	1	1.244005
rAds	1.432513	1	1.196876
rPPE	2.185007	1	1.478177
rEBITDA	1.288822	1	1.135263

as.factor(Year)	1.448504	24	1.007749
IndClass	5.116273	40	1.020615

Model 3

Fit a quasi-Poisson model to `MarketValue` scaled to million.

```
# quasi-Poisson GLM model with offset
mdlQPois <- glm(MarketValueMil ~ rLTD + rCapEx + rRD + rAds + rPPE +
                  rEBITDA + as.factor(Year) + IndClass,
                  offset=log(BookValueMil), data=newDf, family=quasipoisson)
summary(mdlQPois)
```

Call:

```
glm(formula = MarketValueMil ~ rLTD + rCapEx + rRD + rAds + rPPE +
     rEBITDA + as.factor(Year) + IndClass, family = quasipoisson,
     data = newDf, offset = log(BookValueMil))
```

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
rLTD	0.25938	0.46452	0.558	0.57660
rCapEx	0.19173	0.50114	0.383	0.70204
rRD	0.77168	0.26716	2.888	0.00388 **
rAds	0.56115	1.40393	0.400	0.68938
rPPE	-0.27873	0.21906	-1.272	0.20326
rEBITDA	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
IndClassAircraft	0.31102	5.21777	0.060	0.95247
IndClassBanking	0.22102	5.04917	0.044	0.96509
IndClassBeerLiquor	-0.24387	5.06016	-0.048	0.96156
IndClassBusinessServices	0.10102	5.07586	0.020	0.98412
IndClassBusinessSupplies	-0.05569	5.04784	-0.011	0.99120
IndClassCandySoda	-0.05618	5.14940	-0.011	0.99129
IndClassCarsTrucks	-0.21732	5.04930	-0.043	0.96567
IndClassChemicals	0.01294	5.05010	0.003	0.99795
IndClassClothing	-0.31590	5.08068	-0.062	0.95042
IndClassComputers	-0.11812	5.04871	-0.023	0.98133
IndClassConstruction	-0.46888	7.02718	-0.067	0.94680
IndClassConstructionMaterials	-0.24635	5.10042	-0.048	0.96148
IndClassConsumerGoods	0.13819	5.05142	0.027	0.97818
IndClassDefense	-0.24422	5.34257	-0.046	0.96354
IndClassElectricalEquipment	-0.08837	5.06563	-0.017	0.98608
IndClassElectronicEquipment	0.19776	5.04255	0.039	0.96872
IndClassEntertainment	0.13899	5.05464	0.027	0.97806
IndClassFabricatedProducts	-0.38654	22.39129	-0.017	0.98623
IndClassFoodProducts	0.19881	5.04249	0.039	0.96855
IndClassHealthcare	-0.18474	5.17501	-0.036	0.97152
IndClassLabEquipment	-0.16110	5.04365	-0.032	0.97452
IndClassMachinery	-0.06444	5.04430	-0.013	0.98981
IndClassMedicalEquipment	0.16277	5.04780	0.032	0.97428
IndClassMining	-0.26046	5.20786	-0.050	0.96011
IndClassOil	0.54969	18.87199	0.029	0.97676
IndClassOther	-0.13977	9.66196	-0.014	0.98846
IndClassPersonalServices	-0.72876	5.35481	-0.136	0.89175
IndClassPharmaceutical	0.07766	5.04405	0.015	0.98772
IndClassPublishing	0.76953	5.07297	0.152	0.87943
IndClassRecreation	-0.21872	5.06694	-0.043	0.96557
IndClassRestaurants	-0.08237	5.04170	-0.016	0.98696
IndClassRetail	-0.03443	5.05656	-0.007	0.99457
IndClassRubberPlastic	0.03947	5.08928	0.008	0.99381
IndClassShipbuilding	0.17576	6.54256	0.027	0.97857
IndClassShippingContainers	0.35240	5.04896	0.070	0.94436
IndClassSteelWorks	-0.84852	5.05796	-0.168	0.86677
IndClassTelecom	-0.44098	5.05024	-0.087	0.93042
IndClassTextiles	-1.09371	20.22112	-0.054	0.95687
IndClassTobacco	-0.17673	5.06807	-0.035	0.97218
IndClassWholesale	0.68686	5.20809	0.132	0.89508

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

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

Analysis of Deviance Table (Type II tests)

Response: MarketValueMil
Error estimate based on Pearson residuals

	Sum Sq	Df	F value	Pr(>F)		
rLTD	2.0656e+10	1	0.3099	0.5778		
rCapEx	8.3793e+09	1	0.1257	0.7229		
rRD	3.0448e+11	1	4.5674	0.0326 *		
rAds	1.0340e+10	1	0.1551	0.6937		
rPPE	1.0951e+11	1	1.6427	0.2000		
rEBITDA	2.6136e+12	1	39.2063	3.929e-10 ***		
as.factor(Year)	7.7533e+11	24	0.4846	0.9837		
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

```
vif(mdlQPois)
```

	GVIF	Df	GVIF^(1/(2*Df))
rLTD	1.419477	1	1.191418
rCapEx	1.378527	1	1.174107
rRD	1.547548	1	1.244005
rAds	1.432513	1	1.196876
rPPE	2.185007	1	1.478177
rEBITDA	1.288822	1	1.135263
as.factor(Year)	1.448504	24	1.007749
IndClass	5.116273	40	1.020615

Model 4

Fit a quasi-Poisson model to MarketValue scaled to million and include interaction terms with industry class.

```
# quasi-Poisson GLM model with offset
mdlQPoisInt <- glm(MarketValueMil ~ rLTD*IndClass + rCapEx*IndClass + rRD*IndClass +
                     rAds*IndClass + rPPE*IndClass +
                     rEBITDA*IndClass + as.factor(Year) + IndClass,
                     offset=log(BookValueMil), data=newDf, family=quasipoisson)
summary(mdlQPoisInt)
```

Call:
glm(formula = MarketValueMil ~ rLTD * IndClass + rCapEx * IndClass +
 rRD * IndClass + rAds * IndClass + rPPE * IndClass + rEBITDA *
 IndClass + as.factor(Year) + IndClass, family = quasipoisson,
 data = newDf, offset = log(BookValueMil))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-184811	-7241	-1071	6014	358792

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.554e-01	1.360e+01	0.048	0.9616
rLTD	2.327e+00	9.067e+01	0.026	0.9795
IndClassAircraft	2.631e+00	2.470e+01	0.107	0.9152
IndClassBanking	-3.106e-01	1.363e+01	-0.023	0.9818
IndClassBeerLiquor	-6.370e-01	1.397e+01	-0.046	0.9636
IndClassBusinessServices	-5.646e-01	1.364e+01	-0.041	0.9670
IndClassBusinessSupplies	-1.429e+00	1.361e+01	-0.105	0.9163
IndClassCandySoda	1.786e+00	3.978e+01	0.045	0.9642
IndClassCarsTrucks	-1.359e+00	1.366e+01	-0.099	0.9208
IndClassChemicals	-1.063e+00	1.364e+01	-0.078	0.9379
IndClassClothing	-1.080e+00	1.383e+01	-0.078	0.9377
IndClassComputers	-5.012e-01	1.360e+01	-0.037	0.9706
IndClassConstruction	-3.188e+00	3.391e+01	-0.094	0.9251
IndClassConstructionMaterials	-1.506e+00	1.382e+01	-0.109	0.9132
IndClassConsumerGoods	-9.874e-01	1.365e+01	-0.072	0.9423
IndClassDefense	9.459e-01	1.483e+01	0.064	0.9492
IndClassElectricalEquipment	-5.818e-01	1.367e+01	-0.043	0.9661
IndClassElectronicEquipment	-2.991e-01	1.359e+01	-0.022	0.9824
IndClassEntertainment	-1.686e+00	1.365e+01	-0.123	0.9017
IndClassFabricatedProducts	6.774e+00	8.574e+02	0.008	0.9937
IndClassFoodProducts	-2.057e+00	1.360e+01	-0.151	0.8798
IndClassHealthcare	-1.173e+00	1.399e+01	-0.084	0.9332
IndClassLabEquipment	-8.861e-01	1.360e+01	-0.065	0.9480
IndClassMachinery	-1.407e+00	1.361e+01	-0.103	0.9177
IndClassMedicalEquipment	-8.952e-01	1.361e+01	-0.066	0.9476
IndClassMining	-1.021e+00	1.711e+01	-0.060	0.9524
IndClassOil	3.666e+00	7.811e+02	0.005	0.9963
IndClassOther	-5.097e-01	6.440e+01	-0.008	0.9937
IndClassPersonalServices	-3.637e-01	1.444e+01	-0.025	0.9799
IndClassPharmaceutical	-5.343e-01	1.360e+01	-0.039	0.9687
IndClassPublishing	4.003e-01	1.391e+01	0.029	0.9770
IndClassRecreation	-2.098e+00	1.393e+01	-0.151	0.8803
IndClassRestaurants	-1.075e+00	1.359e+01	-0.079	0.9370
IndClassRetail	-7.731e-01	1.362e+01	-0.057	0.9547
IndClassRubberPlastic	-2.014e+00	1.405e+01	-0.143	0.8861
IndClassShipbuilding	1.311e+00	8.248e+01	0.016	0.9873
IndClassShippingContainers	-3.418e+00	1.381e+01	-0.247	0.8045
IndClassSteelWorks	-1.922e+00	1.378e+01	-0.139	0.8891
IndClassTelecom	-2.063e-01	1.360e+01	-0.015	0.9879
IndClassTextiles	3.680e+00	4.232e+02	0.009	0.9931
IndClassTobacco	-3.218e+00	1.400e+01	-0.230	0.8182
IndClassWholesale	3.994e+00	7.059e+01	0.057	0.9549
rCapEx	3.856e+00	8.789e+01	0.044	0.9650
rRD	2.205e+01	4.121e+02	0.054	0.9573
rAds	-2.634e+01	5.101e+02	-0.052	0.9588
rPPE	-6.717e+00	9.868e+01	-0.068	0.9457
rEBITDA	1.160e+01	1.339e+02	0.087	0.9309
as.factor(Year)1991	2.894e-01	5.599e-01	0.517	0.6053
as.factor(Year)1992	3.347e-01	5.483e-01	0.610	0.5417

as.factor(Year)1993	4.293e-01	5.406e-01	0.794	0.4272
as.factor(Year)1994	1.604e-01	5.728e-01	0.280	0.7794
as.factor(Year)1995	2.542e-01	5.472e-01	0.465	0.6422
as.factor(Year)1996	4.031e-01	5.368e-01	0.751	0.4527
as.factor(Year)1997	5.748e-01	5.190e-01	1.108	0.2680
as.factor(Year)1998	6.365e-01	5.164e-01	1.233	0.2178
as.factor(Year)1999	8.818e-01	4.940e-01	1.785	0.0743
as.factor(Year)2000	6.785e-01	4.931e-01	1.376	0.1689
as.factor(Year)2001	5.502e-01	5.005e-01	1.099	0.2717
as.factor(Year)2002	3.557e-01	5.033e-01	0.707	0.4798
as.factor(Year)2003	5.619e-01	4.881e-01	1.151	0.2496
as.factor(Year)2004	5.050e-01	4.842e-01	1.043	0.2970
as.factor(Year)2005	4.783e-01	4.831e-01	0.990	0.3222
as.factor(Year)2006	5.408e-01	4.796e-01	1.128	0.2595
as.factor(Year)2007	5.539e-01	4.798e-01	1.155	0.2483
as.factor(Year)2008	5.004e-02	4.967e-01	0.101	0.9198
as.factor(Year)2009	3.254e-01	4.858e-01	0.670	0.5030
as.factor(Year)2010	3.486e-01	4.803e-01	0.726	0.4680
as.factor(Year)2011	2.825e-01	4.820e-01	0.586	0.5578
as.factor(Year)2012	3.407e-01	4.806e-01	0.709	0.4784
as.factor(Year)2013	5.425e-01	4.748e-01	1.143	0.2532
as.factor(Year)2014	6.264e-01	4.744e-01	1.320	0.1867
rLTD:IndClassAircraft	-4.621e+00	1.028e+02	-0.045	0.9642
rLTD:IndClassBanking	-1.833e+00	9.069e+01	-0.020	0.9839
rLTD:IndClassBeerLiquor	-2.242e+00	9.083e+01	-0.025	0.9803
rLTD:IndClassBusinessServices	-2.942e+00	9.072e+01	-0.032	0.9741
rLTD:IndClassBusinessSupplies	-1.208e+00	9.068e+01	-0.013	0.9894
rLTD:IndClassCandySoda	-3.502e+00	9.499e+01	-0.037	0.9706
rLTD:IndClassCarsTrucks	-2.415e+00	9.071e+01	-0.027	0.9788
rLTD:IndClassChemicals	-8.159e-01	9.073e+01	-0.009	0.9928
rLTD:IndClassClothing	-2.103e+00	9.093e+01	-0.023	0.9815
rLTD:IndClassComputers	-2.344e+00	9.070e+01	-0.026	0.9794
rLTD:IndClassConstruction	-1.210e-01	1.375e+02	-0.001	0.9993
rLTD:IndClassConstructionMaterials	-2.156e+00	9.078e+01	-0.024	0.9811
rLTD:IndClassConsumerGoods	-8.535e-01	9.069e+01	-0.009	0.9925
rLTD:IndClassDefense	1.209e-01	9.437e+01	0.001	0.9990
rLTD:IndClassElectricalEquipment	-2.834e+00	9.072e+01	-0.031	0.9751
rLTD:IndClassElectronicEquipment	-2.244e+00	9.067e+01	-0.025	0.9803
rLTD:IndClassEntertainment	-4.488e-02	9.069e+01	0.000	0.9996
rLTD:IndClassFabricatedProducts	-1.449e+01	1.954e+03	-0.007	0.9941
rLTD:IndClassFoodProducts	-6.264e-01	9.068e+01	-0.007	0.9945
rLTD:IndClassHealthcare	-1.756e+00	9.112e+01	-0.019	0.9846
rLTD:IndClassLabEquipment	-1.932e+00	9.068e+01	-0.021	0.9830
rLTD:IndClassMachinery	-6.946e-01	9.068e+01	-0.008	0.9939
rLTD:IndClassMedicalEquipment	-3.128e+00	9.069e+01	-0.034	0.9725
rLTD:IndClassMining	-3.059e+00	9.327e+01	-0.033	0.9738
rLTD:IndClassOil	-2.257e+01	2.864e+03	-0.008	0.9937
rLTD:IndClassOther	1.775e+00	3.710e+02	0.005	0.9962
rLTD:IndClassPersonalServices	-5.208e+00	9.150e+01	-0.057	0.9546
rLTD:IndClassPharmaceutical	-1.522e+00	9.067e+01	-0.017	0.9866
rLTD:IndClassPublishing	1.335e+00	9.112e+01	0.015	0.9883
rLTD:IndClassRecreation	-7.458e-01	9.074e+01	-0.008	0.9934
rLTD:IndClassRestaurants	-2.103e+00	9.067e+01	-0.023	0.9815
rLTD:IndClassRetail	-2.711e+00	9.071e+01	-0.030	0.9762

rLTD:IndClassRubberPlastic	-7.090e-02	9.086e+01	-0.001	0.9994
rLTD:IndClassShipbuilding	-1.818e+02	5.788e+03	-0.031	0.9749
rLTD:IndClassShippingContainers	5.442e-01	9.072e+01	0.006	0.9952
rLTD:IndClassSteelWorks	1.634e-01	9.089e+01	0.002	0.9986
rLTD:IndClassTelecom	-2.785e+00	9.071e+01	-0.031	0.9755
rLTD:IndClassTextiles	8.225e-03	5.468e+02	0.000	1.0000
rLTD:IndClassTobacco	4.396e+00	9.136e+01	0.048	0.9616
rLTD:IndClassWholesale	8.019e+00	1.083e+02	0.074	0.9410
IndClassAircraft:rCapEx	-7.757e-01	1.606e+02	-0.005	0.9961
IndClassBanking:rCapEx	-4.401e+00	8.795e+01	-0.050	0.9601
IndClassBeerLiquor:rCapEx	1.560e+00	8.891e+01	0.018	0.9860
IndClassBusinessServices:rCapEx	-9.789e-01	8.794e+01	-0.011	0.9911
IndClassBusinessSupplies:rCapEx	-9.636e-01	8.827e+01	-0.011	0.9913
IndClassCandySoda:rCapEx	-3.114e+01	1.993e+02	-0.156	0.8759
IndClassCarsTrucks:rCapEx	-2.878e+00	8.829e+01	-0.033	0.9740
IndClassChemicals:rCapEx	-7.513e+00	8.892e+01	-0.084	0.9327
IndClassClothing:rCapEx	-1.416e+00	9.371e+01	-0.015	0.9879
IndClassComputers:rCapEx	-1.281e+00	8.799e+01	-0.015	0.9884
IndClassConstruction:rCapEx	-4.159e+00	9.159e+01	-0.045	0.9638
IndClassConstructionMaterials:rCapEx	-6.101e+00	8.942e+01	-0.068	0.9456
IndClassConsumerGoods:rCapEx	-2.437e+00	8.969e+01	-0.027	0.9783
IndClassDefense:rCapEx	4.911e-01	9.331e+01	0.005	0.9958
IndClassElectricalEquipment:rCapEx	-3.451e+00	8.821e+01	-0.039	0.9688
IndClassElectronicEquipment:rCapEx	-1.322e+00	8.790e+01	-0.015	0.9880
IndClassEntertainment:rCapEx	-2.642e+00	8.790e+01	-0.030	0.9760
IndClassFabricatedProducts:rCapEx	-1.261e+01	9.325e+02	-0.014	0.9892
IndClassFoodProducts:rCapEx	-3.141e+00	8.831e+01	-0.036	0.9716
IndClassHealthcare:rCapEx	-6.582e+00	9.292e+01	-0.071	0.9435
IndClassLabEquipment:rCapEx	-2.766e+00	8.790e+01	-0.031	0.9749
IndClassMachinery:rCapEx	-2.799e+00	8.829e+01	-0.032	0.9747
IndClassMedicalEquipment:rCapEx	-4.487e+00	8.817e+01	-0.051	0.9594
IndClassMining:rCapEx	-2.134e+00	1.006e+02	-0.021	0.9831
IndClassOil:rCapEx	4.406e+01	1.034e+04	0.004	0.9966
IndClassOther:rCapEx	-7.861e+00	2.304e+02	-0.034	0.9728
IndClassPersonalServices:rCapEx	-4.195e+00	8.817e+01	-0.048	0.9621
IndClassPharmaceutical:rCapEx	-4.576e+00	8.794e+01	-0.052	0.9585
IndClassPublishing:rCapEx	-1.093e+01	9.171e+01	-0.119	0.9051
IndClassRecreation:rCapEx	-3.459e+00	9.237e+01	-0.037	0.9701
IndClassRestaurants:rCapEx	-2.513e+00	8.799e+01	-0.029	0.9772
IndClassRetail:rCapEx	-3.707e+00	8.865e+01	-0.042	0.9666
IndClassRubberPlastic:rCapEx	-8.745e-01	9.132e+01	-0.010	0.9924
IndClassShipbuilding:rCapEx	1.937e+01	1.273e+03	0.015	0.9879
IndClassShippingContainers:rCapEx	-1.038e+01	8.910e+01	-0.117	0.9073
IndClassSteelWorks:rCapEx	2.868e-01	8.833e+01	0.003	0.9974
IndClassTelecom:rCapEx	-3.946e+00	8.789e+01	-0.045	0.9642
IndClassTextiles:rCapEx	9.640e+01	5.970e+03	0.016	0.9871
IndClassTobacco:rCapEx	8.635e+01	1.605e+02	0.538	0.5905
IndClassWholesale:rCapEx	2.179e+00	1.036e+02	0.021	0.9832
IndClassAircraft:rRD	-2.355e+01	4.228e+02	-0.056	0.9556
IndClassBanking:rRD	2.687e+02	5.879e+02	0.457	0.6477
IndClassBeerLiquor:rRD	-3.301e+01	4.246e+02	-0.078	0.9380
IndClassBusinessServices:rRD	-2.288e+01	4.121e+02	-0.056	0.9557
IndClassBusinessSupplies:rRD	-2.024e+01	4.121e+02	-0.049	0.9608
IndClassCandySoda:rRD	-2.114e+02	4.778e+03	-0.044	0.9647

IndClassCarsTrucks:rRD	-1.914e+01	4.123e+02	-0.046	0.9630
IndClassChemicals:rRD	-2.155e+01	4.121e+02	-0.052	0.9583
IndClassClothing:rRD	-2.248e+01	4.250e+02	-0.053	0.9578
IndClassComputers:rRD	-2.152e+01	4.121e+02	-0.052	0.9583
IndClassConstruction:rRD	3.947e+02	7.268e+03	0.054	0.9567
IndClassConstructionMaterials:rRD	-7.394e+00	4.213e+02	-0.018	0.9860
IndClassConsumerGoods:rRD	-2.281e+01	4.124e+02	-0.055	0.9559
IndClassDefense:rRD	-2.297e+01	4.141e+02	-0.055	0.9558
IndClassElectricalEquipment:rRD	-2.199e+01	4.121e+02	-0.053	0.9574
IndClassElectronicEquipment:rRD	-2.131e+01	4.121e+02	-0.052	0.9588
IndClassEntertainment:rRD	-2.597e+01	4.124e+02	-0.063	0.9498
IndClassFabricatedProducts:rRD	4.735e+01	1.091e+04	0.004	0.9965
IndClassFoodProducts:rRD	-1.488e+01	4.132e+02	-0.036	0.9713
IndClassHealthcare:rRD	-9.879e+00	4.144e+02	-0.024	0.9810
IndClassLabEquipment:rRD	-2.005e+01	4.121e+02	-0.049	0.9612
IndClassMachinery:rRD	-1.902e+01	4.121e+02	-0.046	0.9632
IndClassMedicalEquipment:rRD	-2.133e+01	4.121e+02	-0.052	0.9587
IndClassMining:rRD	-2.691e+01	7.217e+02	-0.037	0.9703
IndClassOil:rRD	NA	NA	NA	NA
IndClassOther:rRD	-6.676e+01	2.373e+03	-0.028	0.9776
IndClassPersonalServices:rRD	-3.096e+01	4.211e+02	-0.074	0.9414
IndClassPharmaceutical:rRD	-2.150e+01	4.121e+02	-0.052	0.9584
IndClassPublishing:rRD	-2.057e+01	4.122e+02	-0.050	0.9602
IndClassRecreation:rRD	-1.462e+01	4.133e+02	-0.035	0.9718
IndClassRestaurants:rRD	-2.106e+01	4.122e+02	-0.051	0.9593
IndClassRetail:rRD	-1.414e+01	4.124e+02	-0.034	0.9726
IndClassRubberPlastic:rRD	-5.793e+01	4.358e+02	-0.133	0.8942
IndClassShipbuilding:rRD	7.540e+01	3.140e+03	0.024	0.9808
IndClassShippingContainers:rRD	3.186e+01	4.176e+02	0.076	0.9392
IndClassSteelWorks:rRD	-3.403e+01	4.162e+02	-0.082	0.9348
IndClassTelecom:rRD	-2.759e+01	4.126e+02	-0.067	0.9467
IndClassTextiles:rRD	-4.360e+02	3.001e+04	-0.015	0.9884
IndClassTobacco:rRD	-3.564e+02	7.607e+02	-0.469	0.6394
IndClassWholesale:rRD	-3.821e+01	6.214e+02	-0.061	0.9510
IndClassAircraft:rAds	-1.299e+01	1.723e+03	-0.008	0.9940
IndClassBanking:rAds	1.860e+01	5.104e+02	0.036	0.9709
IndClassBeerLiquor:rAds	2.295e+01	5.105e+02	0.045	0.9641
IndClassBusinessServices:rAds	2.946e+01	5.102e+02	0.058	0.9540
IndClassBusinessSupplies:rAds	3.022e+01	5.111e+02	0.059	0.9529
IndClassCandySoda:rAds	2.156e+01	6.159e+02	0.035	0.9721
IndClassCarsTrucks:rAds	2.640e+01	5.107e+02	0.052	0.9588
IndClassChemicals:rAds	2.758e+01	5.105e+02	0.054	0.9569
IndClassClothing:rAds	2.239e+01	5.109e+02	0.044	0.9651
IndClassComputers:rAds	2.856e+01	5.103e+02	0.056	0.9554
IndClassConstruction:rAds	4.112e+01	1.131e+03	0.036	0.9710
IndClassConstructionMaterials:rAds	3.097e+01	5.105e+02	0.061	0.9516
IndClassConsumerGoods:rAds	2.843e+01	5.101e+02	0.056	0.9556
IndClassDefense:rAds	-1.429e+01	5.832e+02	-0.025	0.9805
IndClassElectricalEquipment:rAds	2.644e+01	5.103e+02	0.052	0.9587
IndClassElectronicEquipment:rAds	2.398e+01	5.101e+02	0.047	0.9625
IndClassEntertainment:rAds	1.381e+01	5.105e+02	0.027	0.9784
IndClassFabricatedProducts:rAds	-1.124e+02	2.394e+04	-0.005	0.9963
IndClassFoodProducts:rAds	2.735e+01	5.101e+02	0.054	0.9572
IndClassHealthcare:rAds	3.180e+01	5.268e+02	0.060	0.9519

IndClassLabEquipment:rAds	2.467e+01	5.102e+02	0.048	0.9614
IndClassMachinery:rAds	2.398e+01	5.104e+02	0.047	0.9625
IndClassMedicalEquipment:rAds	3.294e+01	5.102e+02	0.065	0.9485
IndClassMining:rAds	1.234e+01	5.421e+02	0.023	0.9818
IndClassOil:rAds	-4.292e+01	4.448e+03	-0.010	0.9923
IndClassOther:rAds	3.060e+01	6.198e+03	0.005	0.9961
IndClassPersonalServices:rAds	2.504e+01	5.125e+02	0.049	0.9610
IndClassPharmaceutical:rAds	2.642e+01	5.101e+02	0.052	0.9587
IndClassPublishing:rAds	1.444e+01	5.112e+02	0.028	0.9775
IndClassRecreation:rAds	2.991e+01	5.102e+02	0.059	0.9533
IndClassRestaurants:rAds	2.382e+01	5.101e+02	0.047	0.9628
IndClassRetail:rAds	2.250e+01	5.109e+02	0.044	0.9649
IndClassRubberPlastic:rAds	4.080e+01	5.122e+02	0.080	0.9365
IndClassShipbuilding:rAds	-3.759e+01	2.678e+03	-0.014	0.9888
IndClassShippingContainers:rAds	3.697e+01	5.105e+02	0.072	0.9423
IndClassSteelWorks:rAds	4.265e+01	7.095e+02	0.060	0.9521
IndClassTelecom:rAds	2.712e+01	5.101e+02	0.053	0.9576
IndClassTextiles:rAds	1.906e+01	8.951e+03	0.002	0.9983
IndClassTobacco:rAds	-1.099e+01	5.222e+02	-0.021	0.9832
IndClassWholesale:rAds	-1.183e+01	7.596e+02	-0.016	0.9876
IndClassAircraft:rPPE	-6.287e+00	1.190e+02	-0.053	0.9579
IndClassBanking:rPPE	6.141e+00	9.868e+01	0.062	0.9504
IndClassBeerLiquor:rPPE	6.068e+00	9.870e+01	0.061	0.9510
IndClassBusinessServices:rPPE	6.486e+00	9.870e+01	0.066	0.9476
IndClassBusinessSupplies:rPPE	6.962e+00	9.869e+01	0.071	0.9438
IndClassCandySoda:rPPE	1.546e+00	1.619e+02	0.010	0.9924
IndClassCarsTrucks:rPPE	6.763e+00	9.868e+01	0.069	0.9454
IndClassChemicals:rPPE	6.690e+00	9.869e+01	0.068	0.9460
IndClassClothing:rPPE	6.740e+00	9.875e+01	0.068	0.9456
IndClassComputers:rPPE	5.902e+00	9.868e+01	0.060	0.9523
IndClassConstruction:rPPE	7.348e+00	1.143e+02	0.064	0.9487
IndClassConstructionMaterials:rPPE	7.016e+00	9.870e+01	0.071	0.9433
IndClassConsumerGoods:rPPE	6.236e+00	9.869e+01	0.063	0.9496
IndClassDefense:rPPE	5.155e+00	9.887e+01	0.052	0.9584
IndClassElectricalEquipment:rPPE	6.816e+00	9.871e+01	0.069	0.9450
IndClassElectronicEquipment:rPPE	6.317e+00	9.868e+01	0.064	0.9490
IndClassEntertainment:rPPE	5.671e+00	9.869e+01	0.057	0.9542
IndClassFabricatedProducts:rPPE	4.664e+00	7.547e+02	0.006	0.9951
IndClassFoodProducts:rPPE	6.910e+00	9.868e+01	0.070	0.9442
IndClassHealthcare:rPPE	6.764e+00	9.872e+01	0.069	0.9454
IndClassLabEquipment:rPPE	6.009e+00	9.868e+01	0.061	0.9514
IndClassMachinery:rPPE	6.829e+00	9.868e+01	0.069	0.9448
IndClassMedicalEquipment:rPPE	6.416e+00	9.869e+01	0.065	0.9482
IndClassMining:rPPE	7.285e+00	9.883e+01	0.074	0.9412
IndClassOil:rPPE	5.889e+00	3.045e+02	0.019	0.9846
IndClassOther:rPPE	7.069e+00	2.258e+02	0.031	0.9750
IndClassPersonalServices:rPPE	6.434e+00	9.891e+01	0.065	0.9481
IndClassPharmaceutical:rPPE	7.771e+00	9.868e+01	0.079	0.9372
IndClassPublishing:rPPE	6.271e+00	9.913e+01	0.063	0.9496
IndClassRecreation:rPPE	7.621e+00	9.875e+01	0.077	0.9385
IndClassRestaurants:rPPE	6.456e+00	9.868e+01	0.065	0.9478
IndClassRetail:rPPE	6.463e+00	9.869e+01	0.065	0.9478
IndClassRubberPlastic:rPPE	7.416e+00	9.879e+01	0.075	0.9402
IndClassShipbuilding:rPPE	4.113e+00	2.240e+02	0.018	0.9853

IndClassShippingContainers:rPPE	7.992e+00	9.869e+01	0.081	0.9355
IndClassSteelWorks:rPPE	5.984e+00	9.870e+01	0.061	0.9517
IndClassTelecom:rPPE	6.140e+00	9.868e+01	0.062	0.9504
IndClassTextiles:rPPE	-1.972e+00	5.823e+02	-0.003	0.9973
IndClassTobacco:rPPE	2.746e+01	1.066e+02	0.258	0.7967
IndClassWholesale:rPPE	1.146e+01	2.218e+02	0.052	0.9588
IndClassAircraft:rEBITDA	-1.477e+01	1.408e+02	-0.105	0.9164
IndClassBanking:rEBITDA	-6.059e+00	1.339e+02	-0.045	0.9639
IndClassBeerLiquor:rEBITDA	-7.308e+00	1.347e+02	-0.054	0.9567
IndClassBusinessServices:rEBITDA	-8.609e+00	1.340e+02	-0.064	0.9488
IndClassBusinessSupplies:rEBITDA	-5.861e+00	1.339e+02	-0.044	0.9651
IndClassCandySoda:rEBITDA	2.040e-01	3.847e+02	0.001	0.9996
IndClassCarsTrucks:rEBITDA	-5.890e+00	1.339e+02	-0.044	0.9649
IndClassChemicals:rEBITDA	-6.429e+00	1.340e+02	-0.048	0.9617
IndClassClothing:rEBITDA	-7.083e+00	1.341e+02	-0.053	0.9579
IndClassComputers:rEBITDA	-9.084e+00	1.339e+02	-0.068	0.9459
IndClassConstruction:rEBITDA	-7.852e+00	1.456e+02	-0.054	0.9570
IndClassConstructionMaterials:rEBITDA	-6.357e+00	1.342e+02	-0.047	0.9622
IndClassConsumerGoods:rEBITDA	-6.887e+00	1.339e+02	-0.051	0.9590
IndClassDefense:rEBITDA	-9.867e+00	1.343e+02	-0.073	0.9414
IndClassElectricalEquipment:rEBITDA	-8.691e+00	1.340e+02	-0.065	0.9483
IndClassElectronicEquipment:rEBITDA	-8.672e+00	1.339e+02	-0.065	0.9483
IndClassEntertainment:rEBITDA	-8.277e-01	1.339e+02	-0.006	0.9951
IndClassFabricatedProducts:rEBITDA	-1.434e+01	1.040e+03	-0.014	0.9890
IndClassFoodProducts:rEBITDA	-2.881e+00	1.339e+02	-0.022	0.9828
IndClassHealthcare:rEBITDA	-6.863e+00	1.348e+02	-0.051	0.9594
IndClassLabEquipment:rEBITDA	-7.377e+00	1.339e+02	-0.055	0.9561
IndClassMachinery:rEBITDA	-6.268e+00	1.339e+02	-0.047	0.9627
IndClassMedicalEquipment:rEBITDA	-5.461e+00	1.339e+02	-0.041	0.9675
IndClassMining:rEBITDA	-1.167e+01	1.388e+02	-0.084	0.9330
IndClassOil:rEBITDA	-7.174e+00	4.468e+02	-0.016	0.9872
IndClassOther:rEBITDA	-1.356e+01	2.929e+02	-0.046	0.9631
IndClassPersonalServices:rEBITDA	-7.000e+00	1.358e+02	-0.052	0.9589
IndClassPharmaceutical:rEBITDA	-1.029e+01	1.339e+02	-0.077	0.9388
IndClassPublishing:rEBITDA	-6.276e+00	1.349e+02	-0.047	0.9629
IndClassRecreation:rEBITDA	-7.166e+00	1.340e+02	-0.053	0.9573
IndClassRestaurants:rEBITDA	-5.884e+00	1.339e+02	-0.044	0.9649
IndClassRetail:rEBITDA	-6.805e+00	1.339e+02	-0.051	0.9595
IndClassRubberPlastic:rEBITDA	-5.363e+00	1.348e+02	-0.040	0.9683
IndClassShipbuilding:rEBITDA	-1.172e+01	1.574e+02	-0.074	0.9407
IndClassShippingContainers:rEBITDA	-5.629e+00	1.339e+02	-0.042	0.9665
IndClassSteelWorks:rEBITDA	-6.791e+00	1.340e+02	-0.051	0.9596
IndClassTelecom:rEBITDA	-9.535e+00	1.339e+02	-0.071	0.9432
IndClassTextiles:rEBITDA	-7.379e+00	7.760e+02	-0.010	0.9924
IndClassTobacco:rEBITDA	-1.281e+01	1.363e+02	-0.094	0.9251
IndClassWholesale:rEBITDA	-1.304e+01	2.453e+02	-0.053	0.9576

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 33110186447)

Null deviance: 1.2915e+13 on 13524 degrees of freedom
 Residual deviance: 5.3928e+12 on 13215 degrees of freedom
 AIC: NA

```
Number of Fisher Scoring iterations: 10
```

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

Analysis of Deviance Table (Type II tests)

Response: MarketValueMil

Error estimate based on Pearson residuals

	Sum Sq	Df	F value	Pr(>F)
rLTD	1.0196e+11	1	3.0796	0.07930 .
IndClass	1.0698e+12	40	0.8078	0.80093
rCapEx	2.2106e+10	1	0.6677	0.41388
rRD	1.2995e+11	1	3.9251	0.04759 *
rAds	6.1836e+07	1	0.0019	0.96553
rPPE	1.2675e+11	1	3.8282	0.05042 .
rEBITDA	1.9278e+12	1	58.2261	2.496e-14 ***
as.factor(Year)	6.3096e+11	24	0.7940	0.74883
rLTD:IndClass	1.7843e+11	40	0.1347	1.00000
IndClass:rCapEx	1.7181e+11	40	0.1297	1.00000
IndClass:rRD	1.2028e+11	39	0.0931	1.00000
IndClass:rAds	1.4887e+11	40	0.1124	1.00000
IndClass:rPPE	1.9953e+11	40	0.1507	1.00000
IndClass:rEBITDA	4.0122e+11	40	0.3030	0.99999
Residuals	4.3753e+14	13215		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Questions

1. Which of the following is FALSE?
 - It is always wise to mention the removal of any values as part of any reproducible and transparent data analysis process.
 - 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.
 - Upon consulting field experts, it is often wise to remove extreme values for one or more covariates; there is little support otherwise for the assumed modelled relationship 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 overplotting.
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 **rCapEx** and/or **rRD** increases.
 - The variability in the Tobin's Q scores appears to decrease as **rCapEx** and/or **rRD** 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 wouldn't 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?
 - Compared to 1990 (all other things held constant), the average value of Tobin's Q in 1996 was significantly lower (at the 5% level).
 - All covariates show significantly non-zero (specifically linear) relationships at the 5% level with Tobin's Q except the **rCapEx** and **rPPE** variables.
 - We can confirm from the Anova table 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 about 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, 3 and 4 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; much fewer than the model assuming the dispersion parameter to be equal to 1.
 - Based on the quasi-Poisson model (with interactions), there is no evidence that any of the covariate relationships trialled vary with industry classes (`IndClass`).
 - Under the quasi-Poisson model (without interactions), 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). However, the estimated regression coefficients can vary between the quasi-Poisson and the Poisson model (i.e. assuming the dispersion parameter to be 1), and so the analyst should inspect both sets of output to be sure.
6. 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.
7. 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.
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 datasets generated from the same underlying function.
 - Models which are more simplistic than the underlying function tend to exhibit high bias since they give predictions which are systematically wrong, and different from the underlying function on average.
 - 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 nor 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.