library(ACCT337)

presentation.R

Statistical 'Programming' {

[Project Presentation]

< Examination of the Relationship between Executive Pay and Company Performance for US-listed companies using R Programming >

}

G1_Group5 ← c("Abigail", "Clarice", "Erinn", "Kaitlyn", "Spencer")

library(ACCT337)

presentation.R

#TABLE OF CONTENTS

01 {

[INTRODUCTION]

> BACKGROUND
> LITERATURE REVIEW
> PROJECT OBJECTIVES
> HYPOTHESES
> METHODOLOGY

}

BACKGROUND

Stock-related components of CEO Compensation

FOR:

- Solves agency problem
- Executives less inclined to act opportunistically against shareholders' interests

AGAINST:

- Lack of concrete evidence of a direct relationship between CEO compensation and corporate performance
- Symbol of income inequality: Median $14.2m

LITERATURE REVIEW

We incorporated the following techniques in our model according to similar previous research:

1. **Fixed Effect Model** to account for unobservable Firm & Year variances and increase robustness

2. **Lead dependent variables (n+1)** to prevent reverse causality
    - Payoff not fully observable until next period, but managers will still be compensated

3. Inclusion of **control variables** such as firm size, financial leverage, & CEO-Duality role

OBJECTIVES

compustat

execucomp

Each observation
=
Company's financial information
+
CEO's pay in that particular year (2010-19)

> Predict Pay-Performance linkage
> Provide insights to BOD and relevant stakeholders

Identify the best measure of corporate performance

Assess optimal proportion of incentive compensation

HYPOTHESES - dependent variables

Objective: Determine the best measure of firm performance.

VARIABLE	DEFINITION
oiadp/at	Return on Assets (ROA)
ni / (csho * prcc_f)	Return on Equity (ROE)
(prcc_f * csho + lt) / at	Tobin's Q

HYPOTHESES - independent variables

VARIABLE	DEFINITION	SIGN	EXPLANATION
(tdc1 - total_curr) / tdc_1	Ratio of Incentive to Total Compensation	+	When company performance is linked to how well-compensated CEOs are, Ratio ↑
execdir	Dummy variable for Dual-Role of CEO & Director	+	Dual roles = Compensation ↑ = Higher expectation to boost firm performance
age	CEO's age	+	Older = More experience = Compensation ↑ = Better firm performance
shrown_excl_opts_pct	% of company shares owned by CEO	+	CEO more incentivised for the firm to perform better
fyear - becameceo	CEO tenure	+	Specialised experience from CEO = Better firm performance
lt / at	Financial leverage	-	Decreased ability to meet financial obligations
capx / at	Ratio of Capex to Total Assets	+	Indicator of financial health and future performance
xrd / at	Ratio of R&D to Total Assets	+	Significantly boosts growth opportunities and productivity
log(at)	Firm size	+	Bigger firms = Higher profitability

Console Terminal Background Jobs
R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

METHODOLOGY

INITIAL SAMPLE SELECTION

1. Download WRDS execucomp & compustat, excluding financial services firms
2. Select US-listed companies
3. Remove non-CEO executives
4. Account for market volatility and economic uncertainty during the Global Financial Crisis and COVID-19
5. Remove CEOs who were replaced/appointed that year
6. Keep specific variables related to our research

SUB-TASKS

1. Setup workspace
2. Clean data
3. Understand data
4. Data Analysis - using Regression Models
5. Conclusion - Recommendations

DATA ANALYSIS TECHNIQUE

1. Identify best independent variable
2. Conduct Fixed-Effect linear regression
3. Conduct stepwise/backward/forward regression
4. Identify best regression model
   - i.e., incentive compensation ratio with the strongest relationship to firm performance

Console Terminal × Background Jobs ×

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

02 {

[EXPLORATORY DATA ANALYSIS]

> SETUP WORKSPACE
> UNDERSTANDING DATA
> CLEANING DATA
> EXPLORING DATA

}

SETUP WORKSPACE

INSTALL & LAUNCH PACKAGES

```
install.packages("readr")
install.packages("dplyr")
install.packages("ggplot2")
install.packages("corrplot")
install.packages("car")
install.packages("psych")
install.packages("lubridate")
install.packages("zoo")
install.packages("caret")
install.packages("lfe")
install.packages("broom")
install.packages("stargazer")
install.packages("fixest")
install.packages("hrbrthemes")
```

```
library(readr)
library(dplyr)
library(ggplot2)
library(corrplot)
library(car)
library(psych)
library(lubridate)
library(caret)
library(lfe)
library(broom)
library(stargazer)
library(fixest)
library(hrbrthemes)
```

IMPORT DATASET

```
execucomp <- read_csv("execucomp_19922022.csv")
compustat <- read_csv("compustat_19502023.csv")
```

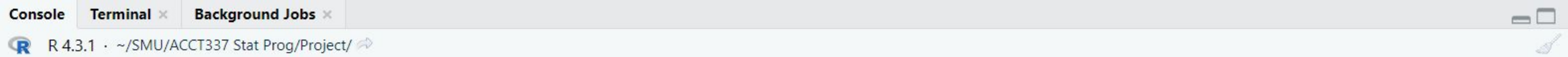
Console Terminal × Background Jobs ×

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

DATA CLEANING

INITIAL DATA CLEANING

1. Exclude missing observations in “tdc1”
2. Replace missing/negative values with 0 in “shrown_excl_opts_pct”
3. Replace all variable values with their industry averages
4. Exclude variables for industries with no financial data available
5. For companies where xrd data is unavailable, assume they do not have R&D expenses and replace their “xrd” with 0



UNDERSTAND THE DATA

EXPLORATORY DATA ANALYSIS: COMPANIES AND INDUSTRIES

```
length(unique(funda$SIC))
```

Synthesis: Identify the number of industries.
Output: 356

```
length(unique(funda$GVKEY))
```

Finding: Each company can have multiple records for various years.
Synthesis: Identify the number of unique companies.
Output: 2,399

```
compIndustries <- distinct(funda, GVKEY, SIC)
```

Synthesis: Identify companies which belong to more than one industry.
Output: Array with 2,399 rows (same as the number of companies)
Hence, each company belongs to a single industry.

	GVKEY	SIC
1	001004	5080
2	001045	4512
3	001072	3670
4	001075	4911
5	001076	6141

Showing 1 to 5 of 2,399 entries, 2 total columns

Console Terminal Background Jobs
R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

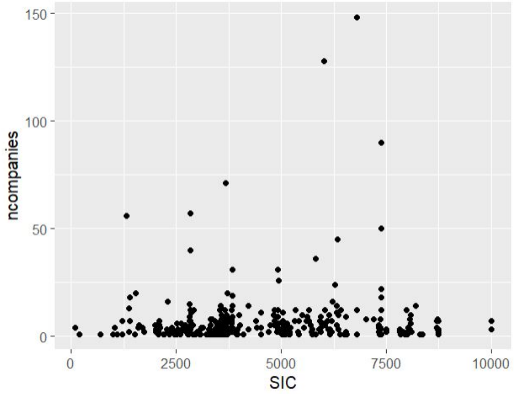
UNDERSTAND THE DATA

```
freqconm <- funda %>% group_by(SIC) %>% distinct(GVKEY)
                   %>% summarise(ncompanies=n())
                   %>% arrange(desc(ncompanies))
```

Synthesis: Overview of the number of companies in each industry
Outcome: Array with 356 rows

```
      SIC              ncompanies
 Min.   : 100    Min.   :  1.00
 1st Qu.:3043    1st Qu.:  2.00
 Median :3826    Median :  3.00
 Mean   :4466    Mean   :  6.74
 3rd Qu.:5919    3rd Qu.:  7.00
 Max.   :9997    Max.   :148.00
```

Key trends:
- 96% of industries: Less than 25 companies
- Only the first 14 industries have more than 25 companies



Industries with Most no. of Companies:
1. Real Estate Investment Trusts
2. Television Programming & Broadcasting
3. Services - Computer Programming, Data Processing, Etc.
4. Semiconductors & Related Devices
5. Pharmaceutical Preparations

Console | Terminal | Background Jobs
R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

UNDERSTAND THE DATA

EXPLORATORY DATA ANALYSIS: CONTINUOUS VARIABLES

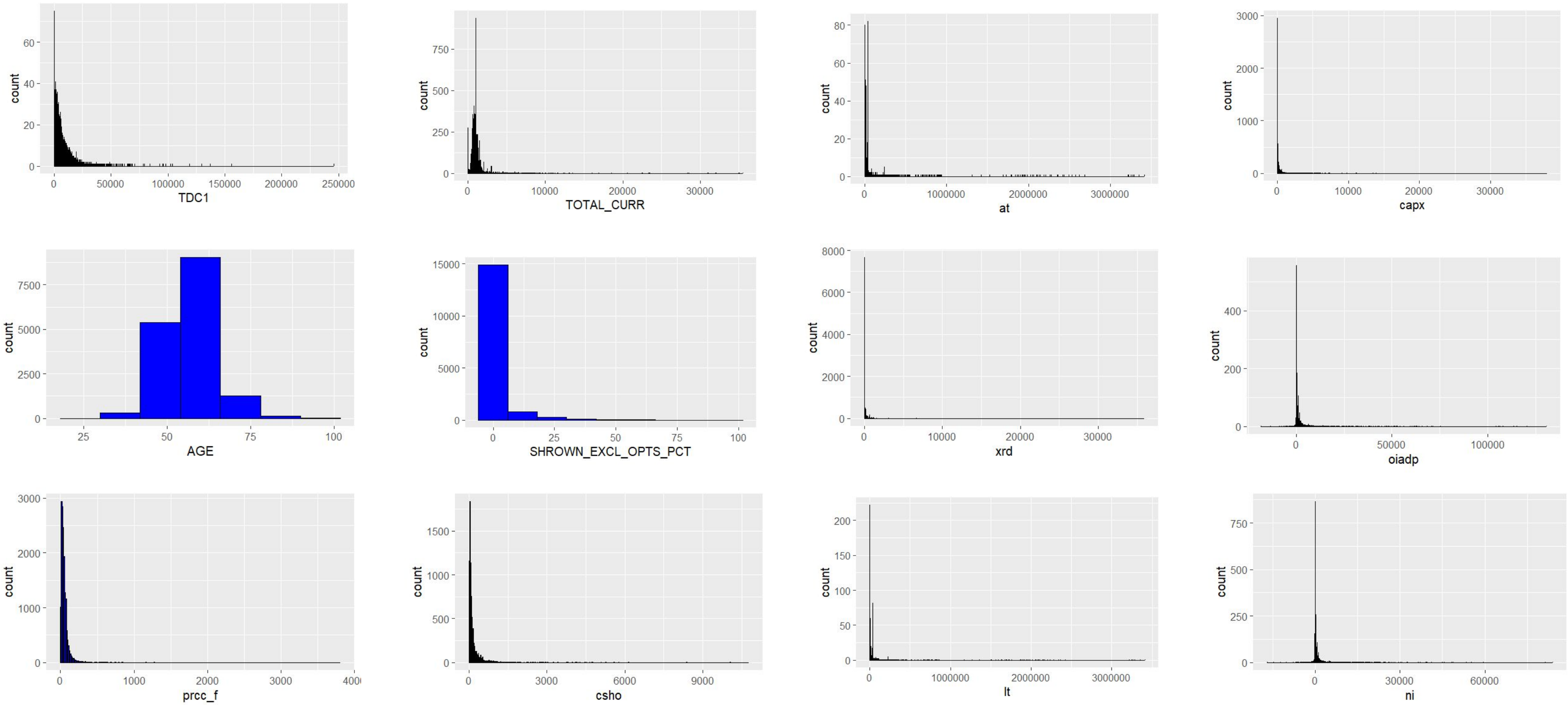
DISTRIBUTION OF SELECTED VARIABLES

VARIABLE	DISTRIBUTION	ACTION + EXPLANATION
TDC1, SHROWN_EXCL_OPTS_PCT, TOTAL_CURR, csho	Right-skewed	Able to see the overall distribution so no further action needed
AGE	Normal distribution	No further action needed; data near the mean more frequent in occurrence than the data far from the mean
at, capx, xrd, oiadp, prcc_f, lt, ni	N/A	Unable to view overall distribution due to outliers; will need to winsorize

Console Terminal × Background Jobs ×

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

UNDERSTAND THE DATA

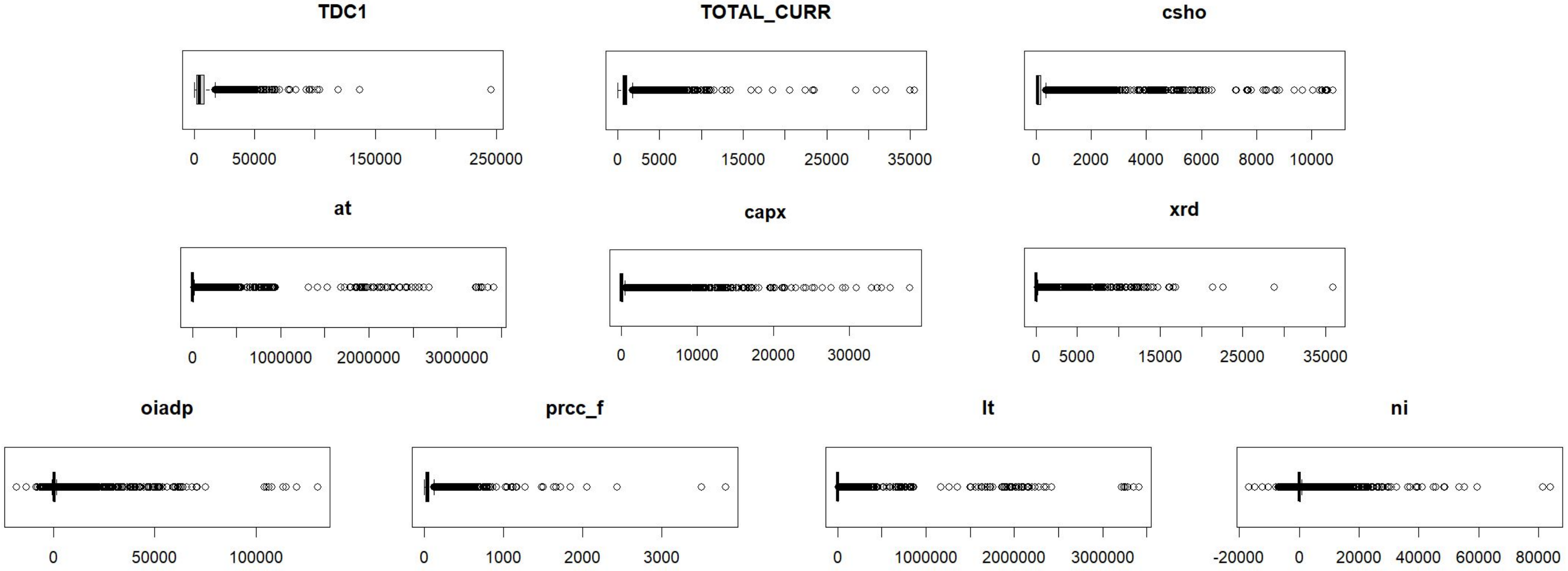


Console Terminal Background Jobs

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

UNDERSTAND THE DATA

BOXPLOT OF SELECTED VARIABLES - **BEFORE** ADJUSTMENT OF OUTLIERS



Console Terminal Background Jobs

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

SECONDARY DATA CLEANING

ADJUSTMENT OF OUTLIERS

Objective: Prevent significant impact on results

Method of Adjustment: *ifelse* and *winsorisation*
To maintain the number of observations that can be used for subsequent analyses

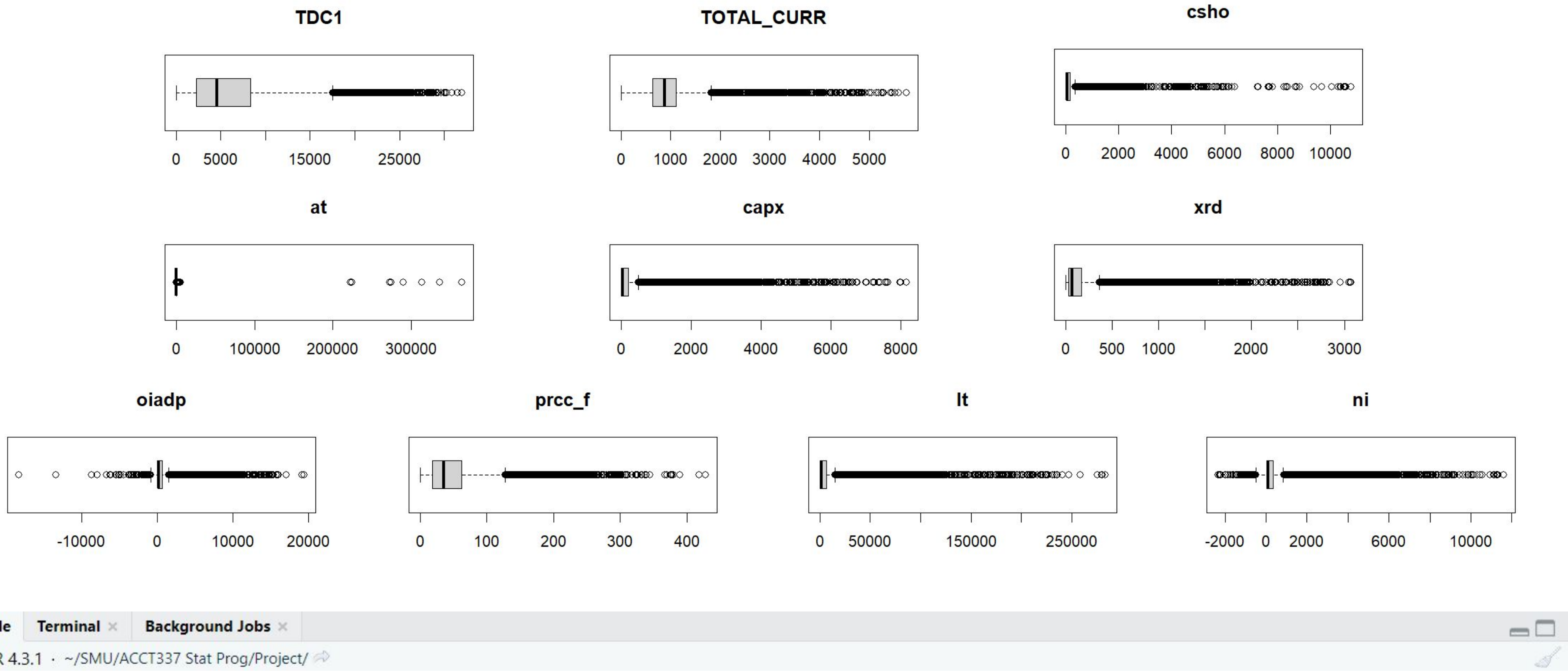
```
funda_winsor <- funda %>%
  mutate(at     = ifelse(at     > quantile(at,     0.99, na.rm = TRUE), quantile(at,     0.99, na.rm = TRUE), at)) %>%
  mutate(capx   = ifelse(capx   > quantile(capx,   0.99, na.rm = TRUE), quantile(capx,   0.99, na.rm = TRUE), capx)) %>%
  mutate(xrd    = ifelse(xrd    > quantile(xrd,    0.99, na.rm = TRUE), quantile(xrd,    0.99, na.rm = TRUE), xrd)) %>%
  mutate(oiadp  = ifelse(oiadp  > quantile(oiadp,  0.99, na.rm = TRUE), quantile(oiadp,  0.99, na.rm = TRUE), oiadp)) %>%
  mutate(prcc_f = ifelse(prcc_f > quantile(prcc_f, 0.99, na.rm = TRUE), quantile(prcc_f, 0.99, na.rm = TRUE), prcc_f)) %>%
  mutate(lt     = ifelse(lt     > quantile(lt,     0.99, na.rm = TRUE), quantile(lt,     0.99, na.rm = TRUE), lt)) %>%
  mutate(ni     = winsor(ni, trim=0.01))
```

Console Terminal Background Jobs

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

UNDERSTAND THE DATA

BOXPLOT OF SELECTED VARIABLES - **AFTER** ADJUSTMENT OF OUTLIERS

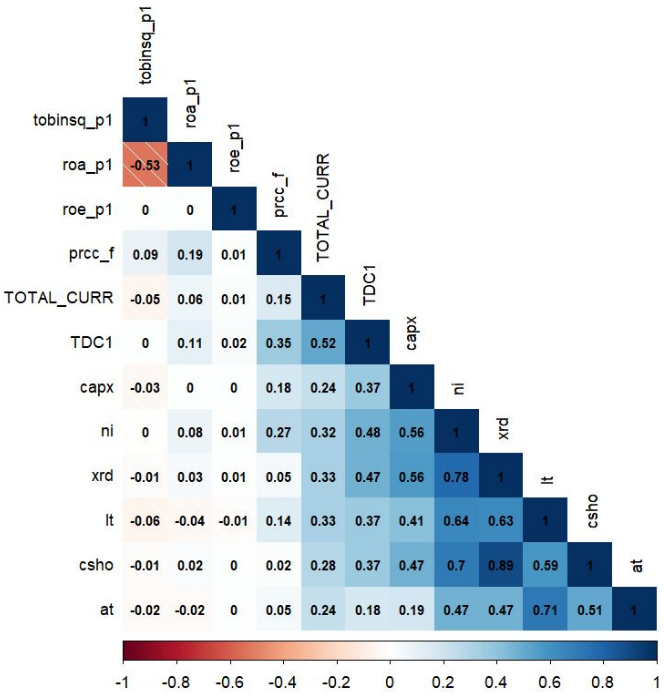


UNDERSTAND THE DATA

CORRELATION MATRIX

```r
funda_corr <- funda_winsor %>%
  select(TDC1, TOTAL_CURR, csho, at,
         capx, xrd, prcc_f, lt, ni,
         roa_p1, tobinsq_p1, roe_p1) %>%
  completify("TDC1", "TOTAL_CURR", "csho", "at",
             "capx", "xrd", "prcc_f", "lt", "ni",
             "roa_p1", "tobinsq_p1", "roe_p1")

corrplot(cor(funda_corr),
         method = "shade",
         order = "AOE",
         type = "lower",
         tl.pos = "ld",
         tl.col = "black",
         addCoef.col = "black",
         number.cex = .7,
         tl.cex = .8)
```



Console Terminal Background Jobs

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

03 {

[REGRESSION ANALYSIS]

> NORMAL REGRESSION
> FORWARD SELECTION MODEL
> BACKWARD ELIMINATION MODEL
> STEPWISE REGRESSION

}

PREPARE DATA FOR REGRESSION

REMOVE MISSING OBSERVATIONS

Objective: Normalise distribution of right-skewed variables

```r
funda_last <- funda_winsor[complete.cases(funda_winsor),]
```

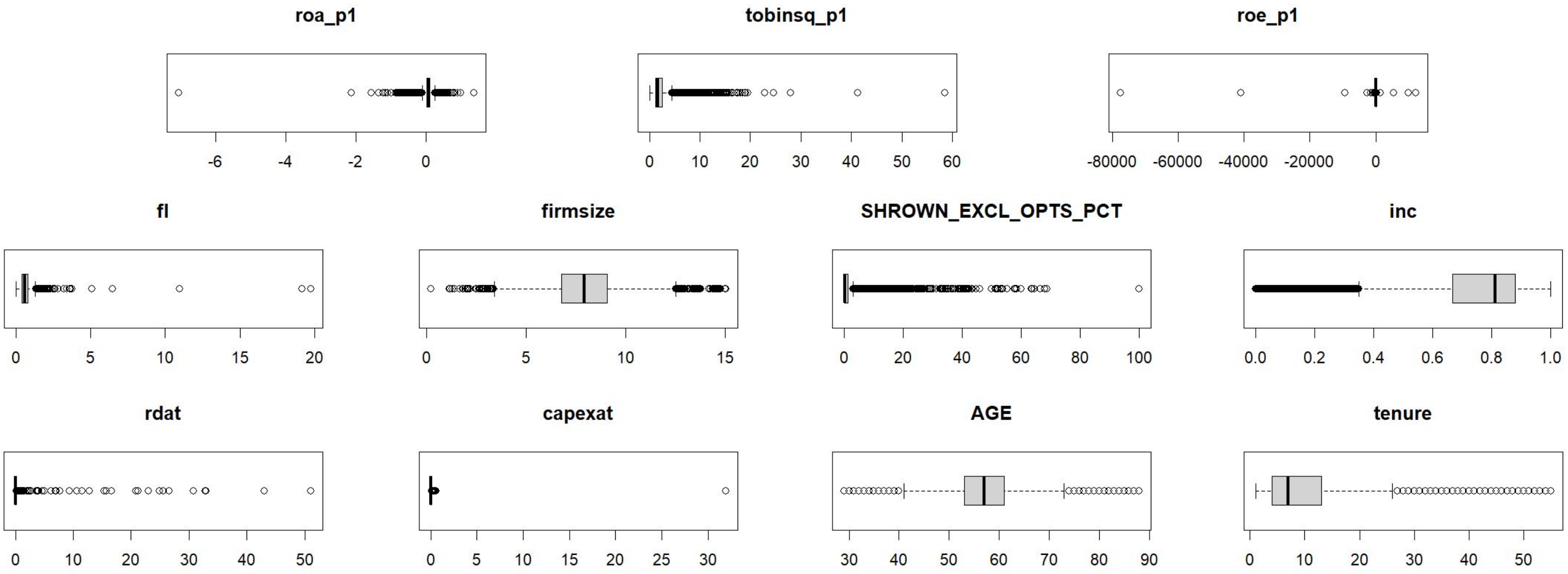
funda_last 8652 obs. of 38 variables

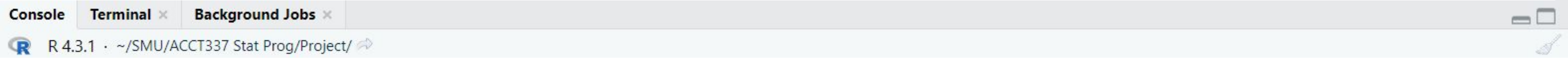
Console Terminal Background Jobs

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

OVERVIEW OF THE DATA

BOXPLOT OF CALCULATED VARIABLES





OVERVIEW OF THE DATA

DESCRIPTIVE STATISTICS

funda

Variable	Mean	Median	Min	Max	Stdev
TDC1	6391	4574	0	246027	7149
TOTAL_CURR	1038	875	0	35500	1152
csho	215.523	71.454	0.001	10778.264	604.036
at	21772	2912	0	3418318	135403
capx	413.354	52.600	-0.001	37985	1620.707
xrd	190.550	9.674	0	35931	953.207
prcc_f	51.364	35.755	0.007	3808.410	84.655
lt	17434	1767	0	3412078	125326
ni	566.58	95.65	-16855	83963	2503.88

funda_final

Variable	Mean	Median	Min	Max	Stdev
TDC1	6605.952	4676.242	0.001	246026.710	7544.209
TOTAL_CURR	1063.5	891.7	0	35500	1198.4
csho	218.161	70.258	0.001	10778.264	629.925
at	14756.861	3008.218	1.041	277797.670	39539.057
capx	345.5	50.2	0	6620	938.4
xrd	150.972	8.694	0	4389.610	527.308
prcc_f	50.57	37.79	0.05	294.07	47.97
lt	10874.790	1805.205	0.083	223523.166	31484.505
ni	497.9	101.4	-976.1	9845.1	1370.9

Console Terminal Background Jobs

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

PREPARE DATA FOR REGRESSION

CREATE TRAIN AND TEST DATASET

```
train <- sample_frac(funda_last, 0.6)
test <- anti_join(funda_last, train)
```

test	3461 obs. of 38 variables
train	5191 obs. of 38 variables

LOGARITHMIC TRANSFORMATION

Objective: Satisfy the linearity assumption & Normalise effects of distribution

Method:
Distributions with Positive values: log(x)
Distributions with Negative values: log(x + CEILING(MIN))

e.g., for roa, CEILING(-7.065) = 8

```
     roa
Min.   :-7.065
1st Qu.: 0.030
Median : 0.071
Mean   : 0.076
3rd Qu.: 0.123
Max.   : 1.374
```

```
reg_roa <- lm(log(roa_p1 + 8) ~ log(inc + 8)
           + SHROWN_EXCL_OPTS_PCT
           + firmsize
           + log(rdat + 8)
           + log(capexat + 8)
           + log(fl + 8)
           + EXECDIR
           + AGE
           + log(tenure + 8),
           data = train)
```

Console | Terminal | Background Jobs

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

NORMAL REGRESSION

SIMPLE REGRESSION

ROA

```
Coefficients:
                      Estimate Std. Error t value          Pr(>|t|)
(Intercept)          2.0732993  0.0225938   91.76 <0.0000000000000002 ***
inc                  0.0104791  0.0009549   10.97 <0.0000000000000002 ***
SHROWN_EXCL_OPTS_PCT 0.0000972  0.0000379    2.56              0.01 *
firmsize             0.0001025  0.0001285    0.80              0.43
log(rdat + 8)       -0.0000451  0.0030432   -0.01              0.99
log(capexat + 8)     0.0096954  0.0088959    1.09              0.28
log(fl + 8)         -0.0062374  0.0052068   -1.20              0.23
EXECDIR             -0.0005897  0.0010988   -0.54              0.59
AGE                  0.0000365  0.0000321    1.14              0.26
log(tenure + 8)     -0.0005738  0.0006434   -0.89              0.37
```

- inc
- SHROWN_EXCL_OPTS_PCT

TOBIN'S Q

```
Coefficients:
                     Estimate Std. Error t value             Pr(>|t|)
(Intercept)          1.410092   0.056930   24.77 < 0.0000000000000002 ***
inc                  0.398031   0.023842   16.69 < 0.0000000000000002 ***
SHROWN_EXCL_OPTS_PCT 0.003116   0.000936    3.33              0.00087 ***
firmsize            -0.079355   0.003392  -23.40 < 0.0000000000000002 ***
log(rdat + 1)        0.168772   0.028716    5.88         0.0000000044 ***
log(capexat + 1)     0.132961   0.077566    1.71              0.08656 .
log(fl + 1)          0.030784   0.031462    0.98              0.32790
EXECDIR              0.065980   0.027352    2.41              0.01589 *
AGE                 -0.002980   0.000790   -3.77              0.00016 ***
log(tenure + 1)      0.038385   0.008849    4.34         0.0000146732 ***
```

- inc
- SHROWN_EXCL_OPTS_PCT
- firmsize
- rdat
- EXECDIR
- AGE
- tenure

SELECTED MODEL

ROE

```
Coefficients:
                                Estimate   Std. Error t value Pr(>|t|)
(Intercept)                  31.28039328 128.75541145    0.24     0.81
inc                           0.00002038   0.00021944    0.09     0.93
log(SHROWN_EXCL_OPTS_PCT + 77552) -0.12806357   0.68254135   -0.19     0.85
firmsize                      0.00000346   0.00002869    0.12     0.90
log(rdat + 77552)             0.02588068   3.00872463    0.01     0.99
log(capexat + 77552)          0.16910692   8.22389868    0.02     0.98
log(fl + 77552)              -1.80025893   7.58147500   -0.24     0.81
EXECDIR                       0.00011262   0.00025323    0.44     0.66
AGE                           0.00000429   0.00000751    0.57     0.57
log(tenure + 77552)          -0.04502596   0.57482648   -0.08     0.94
```

No statistically significant variables.

Console   Terminal   Background Jobs

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

NORMAL REGRESSION

DROP INSIGNIFICANT VARIABLES

```
reg_tobinsq_normal <-
  lm(log(tobinsq_p1 + 1) ~ inc
     + SHROWN_EXCL_OPTS_PCT
     + firmsize
     + log(rdat +1)
     + log(capexat + 1)
     + EXECDIR
     + AGE
     + log(tenure + 1),
     data = train)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-1.1840 -0.2306 -0.0715  0.1668  2.0692

Coefficients:
                      Estimate Std. Error t value             Pr(>|t|)
(Intercept)           1.418463   0.056284   25.20 < 0.0000000000000002 ***
inc                   0.395588   0.023711   16.68 < 0.0000000000000002 ***
SHROWN_EXCL_OPTS_PCT  0.003084   0.000935    3.30              0.00098 ***
firmsize             -0.078104   0.003142  -24.86 < 0.0000000000000002 ***
log(rdat + 1)         0.167295   0.028676    5.83          0.000000057 ***
log(capexat + 1)      0.133739   0.077562    1.72              0.08471 .
EXECDIR               0.064817   0.027326    2.37              0.01773 *
AGE                  -0.002985   0.000790   -3.78              0.00016 ***
log(tenure + 1)       0.038032   0.008842    4.30          0.0000172830 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.365 on 5182 degrees of freedom
Multiple R-squared:  0.165,	Adjusted R-squared:  0.163
F-statistic:  128 on 8 and 5182 DF,  p-value: <0.0000000000000002
```

Console | Terminal | Background Jobs

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

NORMAL REGRESSION

VARIANCE INFLATION FACTOR (VIF)

Benchmark VIF:
VIF > 4: Multicollinearity might exist, Further investigation needed
VIF > 10: Serious indication of multicollinearity that requires correction

Objective: Detect multicollinearity
(more than one independent variable are correlated with each other)

Variable	inc	SHROWN_EXCL_OPTS_PCT	firmsize	rdat	capexat	EXECDIR	AGE	tenure
VIF	1.171	1.201	1.270	1.093	1.026	1.005	1.231	1.313

Conclusion:
Multicollinearity is insignificant. No further investigation required.

Console Terminal Background Jobs
R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

FORWARD SELECTION MODEL

1. Start with no independent variable
2. Add one variable each time
3. The variable added will increase R-squared the most

```
forward_tobinsq <- step(reg_tobinsq, direction = "forward")
tobinsq_forward_pred <- predict(forward_tobinsq, test)
```

Accuracy Results on Test Set	
Adjusted r-squared	0.163
MAE	1.177
RSME	2.388

```
Residuals:
    Min      1Q  Median      3Q     Max 
-1.188 -0.230 -0.071  0.168  1.989 

Coefficients:
                    Estimate Std. Error t value            Pr(>|t|)    
(Intercept)         1.410092   0.056930   24.77 < 0.0000000000000002 ***
inc                 0.398031   0.023842   16.69 < 0.0000000000000002 ***
SHROWN_EXCL_OPTS_PCT 0.003116  0.000936    3.33             0.00087 ***
firmsize           -0.079355   0.003392  -23.40 < 0.0000000000000002 ***
log(rdat + 1)       0.168772   0.028716    5.88        0.0000000044 ***
log(capexat + 1)    0.132961   0.077566    1.71             0.08656 .  
log(fl + 1)         0.030784   0.031462    0.98             0.32790    
EXECDIR             0.065980   0.027352    2.41             0.01589 *  
AGE                -0.002980   0.000790   -3.77             0.00016 ***
log(tenure + 1)     0.038385   0.008849    4.34        0.0000146732 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.365 on 5181 degrees of freedom
Multiple R-squared:  0.165,	Adjusted R-squared:  0.163 
F-statistic:  114 on 9 and 5181 DF,  p-value: <0.0000000000000002
```

Console | Terminal | Background Jobs

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

BACKWARD ELIMINATION MODEL

1. Start with all independent variables
2. Drops one variable each time
3. If the variable does not contribute to a higher accuracy, it is dropped.

```
backward_tobinsq <- step(reg_tobinsq, direction = "backward")
tobinsq_backward_pred <- predict(backward_tobinsq, test)
```

Accuracy Results on Test Set	
Adjusted r-squared	0.163
MAE	1.177
RSME	2.388

```
Residuals:
    Min      1Q  Median      3Q     Max
-1.1840 -0.2306 -0.0715  0.1668  2.0692

Coefficients:
                      Estimate Std. Error t value            Pr(>|t|)
(Intercept)           1.418463   0.056284   25.20 < 0.0000000000000002 ***
inc                   0.395588   0.023711   16.68 < 0.0000000000000002 ***
SHROWN_EXCL_OPTS_PCT  0.003084   0.000935    3.30              0.00098 ***
firmsize             -0.078104   0.003142  -24.86 < 0.0000000000000002 ***
log(rdat + 1)         0.167295   0.028676    5.83         0.0000000057 ***
log(capexat + 1)      0.133739   0.077562    1.72              0.08471 .
EXECDIR               0.064817   0.027326    2.37              0.01773 *
AGE                  -0.002985   0.000790   -3.78              0.00016 ***
log(tenure + 1)       0.038032   0.008842    4.30         0.0000172830 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.365 on 5182 degrees of freedom
Multiple R-squared:  0.165,	Adjusted R-squared:  0.163
F-statistic:  128 on 8 and 5182 DF,  p-value: <0.0000000000000002
```

Console Terminal × Background Jobs ×

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

STEPWISE REGRESSION

1. Start with no independent variables
2. Adds one variable each time
3. Existing variables that do not contribute to a higher accuracy will be dropped.

```
stepwise_tobinsq <- step(reg_tobinsq, direction = "both")
tobinsq_stepwise_pred <- predict(stepwise_tobinsq, test)
```

Accuracy Results on Test Set	
Adjusted r-squared	0.163
MAE	1.177
RSME	2.388

```
Residuals:
    Min      1Q  Median      3Q     Max
-1.1840 -0.2306 -0.0715  0.1668  2.0692

Coefficients:
                    Estimate Std. Error t value             Pr(>|t|)
(Intercept)         1.418463   0.056284   25.20 < 0.0000000000000002 ***
inc                 0.395588   0.023711   16.68 < 0.0000000000000002 ***
SHROWN_EXCL_OPTS_PCT 0.003084  0.000935    3.30              0.00098 ***
firmsize           -0.078104   0.003142  -24.86 < 0.0000000000000002 ***
log(rdat + 1)       0.167295   0.028676    5.83         0.0000000057 ***
log(capexat + 1)    0.133739   0.077562    1.72              0.08471 .
EXECDIR             0.064817   0.027326    2.37              0.01773 *
AGE                -0.002985   0.000790   -3.78              0.00016 ***
log(tenure + 1)     0.038032   0.008842    4.30         0.0000172830 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.365 on 5182 degrees of freedom
Multiple R-squared:  0.165,	Adjusted R-squared:  0.163
F-statistic:  128 on 8 and 5182 DF,  p-value: <0.0000000000000002
```

Console | Terminal × | Background Jobs ×

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

COMPARISON OF REGRESSION MODELS

ACCURACY RESULTS	FORWARD	BACKWARD	STEPWISE
ADJUSTED R-SQUARED the higher the better	0.163		
MAE the lower the better	1.177		
RMSE the lower the better	2.388		

Conclusion:
Keep the original simple regression model with no further changes.

Console Terminal Background Jobs
R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

04 {

[MODEL EVALUATION]

> SELECTION OF MODEL
> ADDITION OF FIXED EFFECTS & TWO-WAY CLUSTERING

}

SELECTED MODEL

SIMPLE LINEAR REGRESSION MODEL

```
linearreg <-
  lm(log(tobinsq_p1 + 1) ~ inc
     + SHROWN_EXCL_OPTS_PCT
     + firmsize
     + log(rdat +1)
     + log(capexat + 1)
     + EXECDIR
     + AGE
     + log(tenure + 1),
     data = train)
```

Console Terminal Background Jobs

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

SELECTED MODEL

LINEAR REGRESSION MODEL

SUMMARY STATISTICS

```
Residuals:
    Min      1Q  Median      3Q     Max 
-1.1834 -0.2330 -0.0717  0.1677  2.0602 

Coefficients:
                     Estimate Std. Error t value             Pr(>|t|)    
(Intercept)          1.410178   0.056965   24.75 < 0.0000000000000002 ***
log(inc + 1)         0.531154   0.035234   15.08 < 0.0000000000000002 ***
SHROWN_EXCL_OPTS_PCT 0.003040   0.000941    3.23               0.0012 ** 
firmsize            -0.075580   0.003136  -24.10 < 0.0000000000000002 ***
log(rdat + 1)        0.169821   0.028809    5.89          0.000000004 ***
log(capexat + 1)     0.142247   0.077918    1.83               0.0680 .  
EXECDIR              0.065243   0.027455    2.38               0.0175 *  
AGE                 -0.003134   0.000793   -3.95          0.000078909 ***
log(tenure + 1)      0.037620   0.008884    4.23          0.000023282 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.367 on 5182 degrees of freedom
Multiple R-squared:  0.157,	Adjusted R-squared:  0.155 
F-statistic:  120 on 8 and 5182 DF,  p-value: <0.0000000000000002
```

VARIANCE INFLATION FACTOR (VIF)

inc	1.171
SHROWN_EXCL_OPTS_PCT	1.201
firmsize	1.270
rdat	1.093
capexat	1.026
EXECDIR	1.005
AGE	1.231
tenure	1.313

Console Terminal × Background Jobs ×

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

SELECTED MODEL

ACCURACY OF MODEL

MEAN ERROR	ROOT MEAN SQUARED ERROR	MEAN ABSOLUTE ERROR	MEAN PERCENTAGE ERROR	MEAN ABSOLUTE PERCENTAGE ERROR
ME	**RMSE**	**MAE**	**MPE**	**MAPE**
1.133	2.388	1.177	33.15	39.17

Conclusion:
Since the relevant values (ME, RMSE, MAE) are low, the model's accuracy is adequate.

Console Terminal × Background Jobs ×
R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

ADDITIONS TO OUR MODEL

FIXED EFFECTS	TWO-WAY CLUSTERING
Objective: Control unobserved characteristics of individual entities in the dataset that might be systematically related to the dependent variable	**Objective**: Account for the presence of heteroscedasticity in the data, where the variability of error is not constant across all observations

```
foels_reg <- feols(data = train, log(tobinsq_p1 + 1) ~ inc
                   + SHROWN_EXCL_OPTS_PCT + firmsize
                   + log(rdat +1) + log(capexat + 1)
                   + EXECDIR + AGE
                   + log(tenure + 1) | GVKEY + YEAR)

final_pred <- predict(foels_reg, test)
final_error <- test$tobinsq_p1 - final_pred

final_final <- data.frame("Predicted" = final_pred,
                          "Actual" = test$tobinsq_p1,
                          "Error" = final_error)
```

```
cluster_reg <- feols(data = train, log(tobinsq_p1 + 1) ~ inc
                     + SHROWN_EXCL_OPTS_PCT + firmsize
                     + log(rdat +1) + log(capexat + 1)
                     + EXECDIR + AGE
                     + log(tenure + 1) | GVKEY + YEAR,
                     cluster = c("GVKEY", "YEAR"))
```

Overall Effect: Improve accuracy of our model & Provide better predictions

Console Terminal Background Jobs

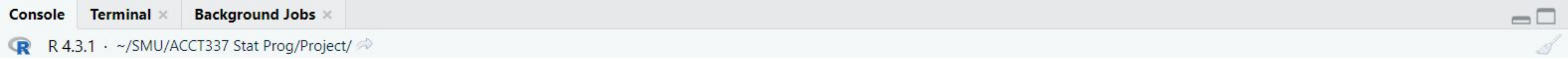
R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

SELECTED MODEL INCLUDING ADDITIONS

ACCURACY OF MODEL

ACCURACY RESULTS	PREVIOUS VALUE	NEW VALUE	EXPLANATION
ADJUSTED R-SQUARED the higher the better	0.163	0.839758	Even better fit for our model
MAE the lower the better	1.177	1.156	Remained relatively low
RMSE the lower the better	2.388	2.249	

Conclusion: Coefficients of the selected variables are significant - We will keep them for our final model.



05 {

[CONCLUSION]

> FINAL RESULTS
> ENDNOTES

}

FINAL MODEL

DROP INSIGNIFICANT VARIABLES

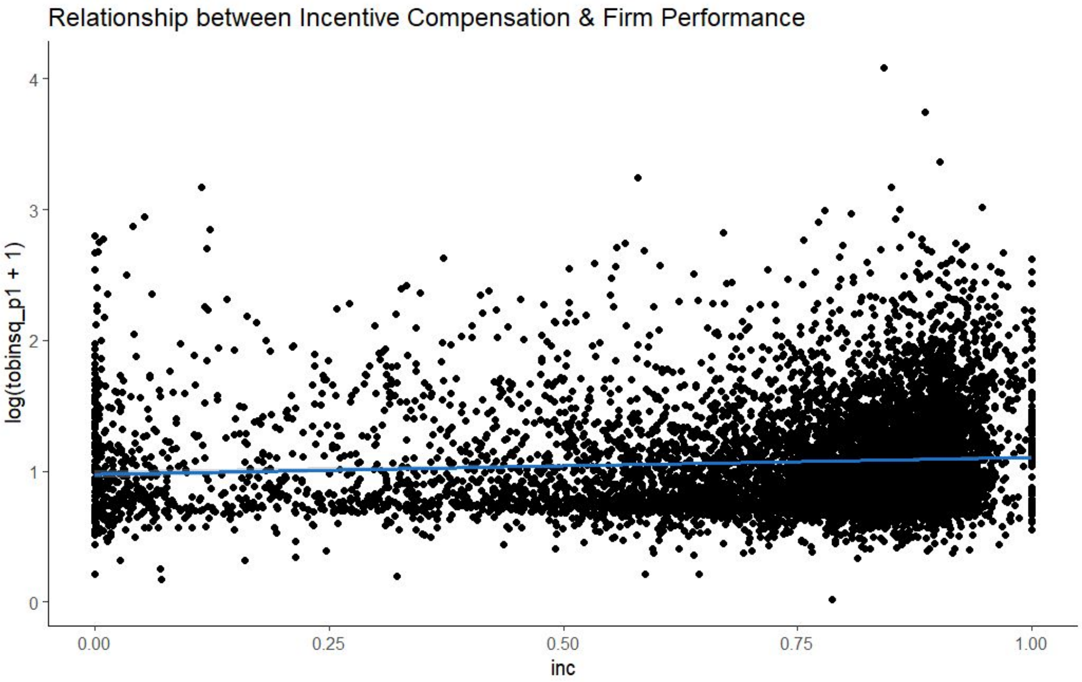
```
final_reg <- feols(data = train, log(tobinsq_p1 + 1) ~ inc
                   + SHROWN_EXCL_OPTS_PCT + firmsize | GVKEY + YEAR)
```

```
OLS estimation, Dep. Var.: log(tobinsq_p1 + 1)
Observations: 5,191
Fixed-effects: GVKEY: 1,845,  YEAR: 7
Standard-errors: Clustered (GVKEY)
                      Estimate Std. Error t value  Pr(>|t|)
inc                    0.07577   0.032364   2.341 0.0193330 *
SHROWN_EXCL_OPTS_PCT -0.01445   0.005555  -2.602 0.0093504 **
firmsize              -0.26027   0.021023 -12.380 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.127987     Adj. R2: 0.839937
                 Within R2: 0.165233
```

Console Terminal × Background Jobs ×

R 4.3.1 · ~/SMU/ACCT337 Stat Prog/Project/

OPTIMAL PROPORTION OF INCENTIVE COMPENSATION

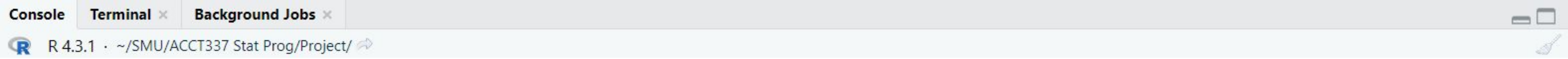


Weak linear relationship
between inc and Tobin's Q

Low Estimate for inc
0.07577 at 5% significance level

(after controlling for variables & finding the best fit through adjusted R-square)

Final Conclusion: No optimal proportion of incentive compensation.



ENDNOTES

OUR MODEL

To derive the best-fitted model for our use case, we included the following:

- Extensive data cleansing
- Replacement of NA values with the industry average for many variables
- Winsorization of outliers
- Elimination of reverse causality by using forward one-year dependent variable
- Log transformation to account for variables with right-skewed distributions
- Reduction of endogeneity by checking for multicollinearity and adding/dropping control variables
- Avoidance of omitted firm/year variable bias by performing fixed-effect linear regression

CONCLUSION

Positive (albeit weak) relationship between incentive compensation and firm performance
(consistent with the efficient market and agency theory hypothesis)

Possibility that managers accept large amounts of equity compensation in the form of option awards

Investors increase expectations on firm performance

Higher Tobin's Q values
as they overvalue the firm and its assets.

library(ACCT337)

presentation.R

```
Statistical 'Programming' {

    print("Thank You!")

}
```

G1_Group5 ← c("Abigail", "Clarice", "Erinn", "Kaitlyn", "Spencer")

Bivens, J., & Kandra, J. (2022, October 4). *CEO pay has skyrocketed 1,460% since 1978: CEOS were paid 399 times as much as a typical worker in 2021*. Economic Policy Institute. https://www.epi.org/publication/ceo-pay-in-2021/

CFI. (2022, December 5). *Variance inflation factor (VIF)*. Corporate Finance Institute. https://corporatefinanceinstitute.com/resources/data-science/variance-inflation-factor-vif/

Cheng, Q., Ranasinghe, T., & Zhao, S. (2017). Do high CEO pay ratios destroy firm value?. *Robert H. Smith School Research Paper No. RHS, 2861680.*

Conyon, M. J., & Leech, D. (1994). Top pay, company performance and corporate governance. *Oxford bulletin of Economics and Statistics*, *56*(3), 229-247.

Cooper, M. J., Gulen, H., & Rau, P. R. (2016). Performance for pay? The relation between CEO incentive compensation and future stock price performance. *The Relation Between CEO Incentive Compensation and Future Stock Price Performance (November 1, 2016).*

Economic Policy Institute. (October 4, 2022). Aggregated CEO-to-worker compensation ratio for the 350 largest publicly owned companies in the United States from 1965 to 2021 [Graph]. In *Statista*. Retrieved October 29, 2023, from https://www-statista-com.libproxy.smu.edu.sg/statistics/261463/ceo-to-worker-compensation-ratio-of-top-firms-in-the-us/

Grinstein, Y., Lauterbach, B., & Yosef, R. (2022). Benchmarking of pay components in CEO compensation design. *Journal of Corporate Finance*, *77*, 102308.

Kuo, C. S., Li, M. Y. L., & Yu, S. E. (2013). Non-uniform effects of CEO equity-based compensation on firm performance–An application of a panel threshold regression model. *The British Accounting Review*, *45*(3), 203-214.

Kweh, Q. L., Tebourbi, I., Lo, H. C., & Huang, C. T. (2022). CEO compensation and firm performance: Evidence from financially constrained firms. *Research in International Business and Finance*, *61*, 101671.

Li, M. Y. L., Yang, T. H., & Yu, S. E. (2015). CEO stock‐based incentive compensation and firm performance: a quantile regression approach. *Journal of International Financial Management & Accounting*, *26*(1), 39-71.

#REFERENCES