Untitled

Iffy

Friday, July 16, 2021

Table of Contents

## Install pacman  
  
if(!require(pacman)){  
   
 install.packages("pacman")  
   
}  
  
## Load required packages- ensure internet connection to load the packages  
pacman::p\_load(tidyverse, tidymodels, skimr, janitor, Amelia, knitr, kableExtra,   
   
 pglm, baguette, rpart.plot)

# Background

## Purpose of the study

The main objective of this study is to identify different the factors that influence misreporting of financial information amongst the top 40 Johannesburg Stock Exchange (JSE) listed firms. The presence of misreporting for firms listed on the JSE has been detected mostly by restatements firms have made through the Stock Exchange News Services (SENS), although this has not been the only way. Some firms may have wanted to perpetuate the misreporting for different reasons until the errors are exposed through media, rating agencies, whistle blowing, underground investigations, and so on. Rasyid and Ardana (2014) agreed with Ciesielski and Weirich (2006) that accounting complexities are the main drivers of most restatements. These restatements are the main triggers for firms presenting restated financial statements. The restated financial statements eventually lead to huge transaction costs emanating from the errors of the initial presentation. The errors mislead the interpretation of the state of the firm, which in turn influences the whole decision-making process in the stock market regarding appropriate capital and is thus misleading to investors at large.

## Research questions

Flowing from the stated research problems, the following are the research questions for this study:

* Do equity incentives to executives influence financial misreporting?
* Does financial distress influence financial misreporting?
* Does poor internal corporate governance influence financial misreporting?
* Does family ownership influence financial misreporting?
* Does state ownership influence misreporting?

#Hypotheses Based on the research questions, the following testable hypotheses are formulated in the positive form:

* Equity incentives to executives to encourage financial misreporting;
* Financial distress of a company influences financial misreporting;
* Poor internal corporate governance influences financial misreporting;
* Family ownership influences financial misreporting; and
* State ownership influences misreporting.

# Data and variables definitions

* RESTATEMENTS——-Restatement…dependent variable
* EB————————–Equity Incentives for executives
* FAMILY B—————-Family owned businesses
* STATE OWN————-State owned businesses
* FIN DS———————Financial distress
* POOR CORG————-Poor internal controls and corporate governance (Penalties, qualified audit and separation of control were used to determine poor internal controls and corporate governance) and the highest number here is 2.

All these are 1 or 0, except number 6 that has 0, 1, 2.

Poor internal controls and corporate governance is made up of Penalties, qualified audit and separation of control were used to determine poor internal controls and corporate governance. Where either variable was in any given year it was marked by 1 and 0 when not found and a total score was given for that year with a minimum of 0 and max of 3.

There is 1 independent variable: Restatement. In each year of the 20 years, I had to determine whether there was a restatement or not. Where a restatement was found in any year I allocated 1 and 0 for no restatement. And the same way I allocated a 1 or 0 for the variables 1 to 4.

## Read in the Data

iffy\_data <- readxl::read\_xlsx("data/iffy\_data.xlsx") %>%   
   
 janitor::clean\_names() %>%   
   
 select(name, company\_code, year, restatements, family\_b, eb,   
   
 state\_own, fin\_ds, poor\_corg) %>%   
   
 mutate(restatements = factor(restatements),  
   
 family\_b = factor(family\_b),  
   
 eb = factor(eb),  
   
 state\_own = factor(state\_own),  
   
 fin\_ds = factor(fin\_ds),  
   
 poor\_corg = factor(poor\_corg)  
 )  
  
head(iffy\_data)

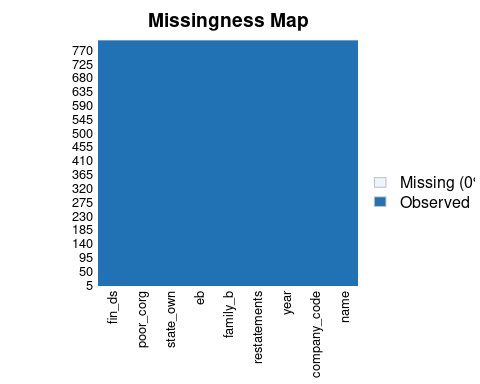
## # A tibble: 6 x 9  
## name company\_code year restatements family\_b eb state\_own fin\_ds  
## <chr> <chr> <dbl> <fct> <fct> <fct> <fct> <fct>   
## 1 PROSUS PRX 1 0 0 1 0 0   
## 2 PROSUS PRX 2 0 0 1 0 0   
## 3 PROSUS PRX 3 0 0 1 0 0   
## 4 PROSUS PRX 4 0 0 1 0 0   
## 5 PROSUS PRX 5 0 0 1 0 0   
## 6 PROSUS PRX 6 0 0 1 0 0   
## # … with 1 more variable: poor\_corg <fct>

# Exploring the Data

## Missing data

There are no missing data points for the key variables.

## Graphing missing data   
Amelia::missmap(iffy\_data)



## Count of missing data   
  
sapply(iffy\_data, is.na) %>%   
   
 colSums() %>%   
   
 tibble(VarNames = names(iffy\_data), missing = .) %>%   
   
 arrange(desc(missing)) %>%   
   
 kableExtra::kbl(booktabs = TRUE, caption = "Missing Data") %>%   
   
 kableExtra::kable\_classic(full\_width = FALSE)

Missing Data

VarNames

missing

fin\_ds

1

name

0

company\_code

0

year

0

restatements

0

family\_b

0

eb

0

state\_own

0

poor\_corg

0

## Data duplication

There are no duplicates

iffy\_data %>% filter(duplicated(.))

## # A tibble: 0 x 9  
## # … with 9 variables: name <chr>, company\_code <chr>, year <dbl>,  
## # restatements <fct>, family\_b <fct>, eb <fct>, state\_own <fct>,  
## # fin\_ds <fct>, poor\_corg <fct>

## First we check the Dependent Variable

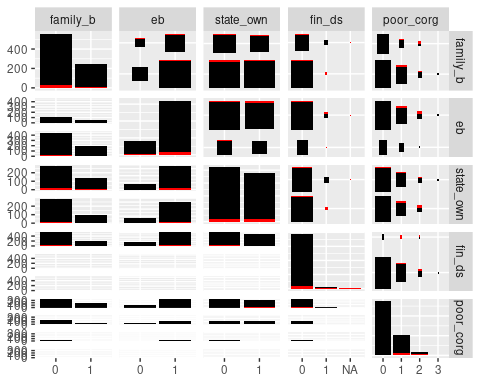
Here, we have an extreme case of class imbalance in the dependent variable, meaning we have very few restatements. See visualization on next page.

table(iffy\_data$restatements)

##   
## 0 1   
## 760 40

## Data Visualisation

iffy\_data %>%   
   
 GGally::ggpairs(columns = c("family\_b", "eb", "state\_own", "fin\_ds", "poor\_corg"),  
 mapping = aes(fill = restatements)  
 ) + scale\_fill\_manual(values = c("black", "red"))



# The Models

## First try panel logit model

* We fit a random effects panel logit model. It shows that the significant variables are financial distress and corporate governance.

iffy\_pglm <- pglm(restatements ~ family\_b + eb + state\_own + fin\_ds + poor\_corg,  
 data = iffy\_data,  
 family = binomial("logit"), R = 5, print.level = 3,  
 method = 'bfgs', index = c('company\_code', 'year'), model = "random")

## Initial function value: -112.1891   
## Initial gradient value:  
## (Intercept) family\_b1 eb1 state\_own1 fin\_ds1   
## 4.296996e+00 4.171236e+00 5.662097e+00 2.272641e+00 3.065357e+00   
## poor\_corg1 poor\_corg2 poor\_corg3 sigma   
## 3.827534e+00 3.699535e+00 -1.797644e-07 -3.681083e+00   
## initial value 112.189126   
## iter 2 value 109.472895  
## iter 3 value 104.665383  
## iter 4 value 100.060735  
## iter 5 value 98.282365  
## iter 6 value 97.031964  
## iter 7 value 95.240349  
## iter 8 value 95.001113  
## iter 9 value 94.830737  
## iter 10 value 94.739298  
## iter 11 value 94.666159  
## iter 12 value 94.648629  
## iter 13 value 94.638781  
## iter 14 value 94.635668  
## iter 15 value 94.634299  
## iter 16 value 94.634223  
## iter 17 value 94.634203  
## iter 18 value 94.634199  
## iter 19 value 94.634197  
## iter 19 value 94.634197  
## iter 19 value 94.634197  
## final value 94.634197   
## converged

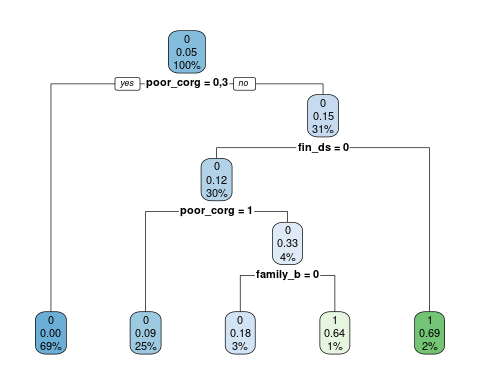
summary(iffy\_pglm)

## --------------------------------------------  
## Maximum Likelihood estimation  
## BFGS maximization, 32 iterations  
## Return code 0: successful convergence   
## Log-Likelihood: -94.6342   
## 9 free parameters  
## Estimates:  
## Estimate Std. error t value Pr(> t)   
## (Intercept) -7.68230 1.34541 -5.710 1.13e-08 \*\*\*  
## family\_b1 0.02711 0.82766 0.033 0.974   
## eb1 0.39778 0.78725 0.505 0.613   
## state\_own1 -0.14441 0.66897 -0.216 0.829   
## fin\_ds1 4.52469 1.13707 3.979 6.91e-05 \*\*\*  
## poor\_corg1 4.38823 0.90178 4.866 1.14e-06 \*\*\*  
## poor\_corg2 6.07086 1.07941 5.624 1.86e-08 \*\*\*  
## poor\_corg3 -8.73391 3435.17815 -0.003 0.998   
## sigma 2.54118 0.63259 4.017 5.89e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## --------------------------------------------

## The tree based model

* Even after adding name and year, they are not coming out as important factors.
* The tree shows that what is important is corporate governance, financial distress and whether or not a company is family owned.
* The variable importance is as follows: corporate governance comes first, followed by financial distress and then family ownership.

## The model   
iffy\_decision\_tree <- decision\_tree() %>%   
   
 set\_mode("classification") %>%   
   
 set\_engine("rpart")  
  
## Run the model  
  
iffy\_decision\_regression <- workflow() %>%   
   
 add\_model(iffy\_decision\_tree) %>%   
   
 add\_formula(restatements ~ family\_b + eb + state\_own + fin\_ds + poor\_corg) %>%   
   
 fit(data = iffy\_data)  
  
## Draw the tree  
  
iffy\_decision\_plot <- iffy\_decision\_regression %>% extract\_fit\_parsnip()  
  
rpart.plot(iffy\_decision\_plot$fit)



# Way forward

* Attempt up sampling (bootstrap the data to make the dependent variable more balanced between companies that have restated versus those that have not restated) of the data to deal with the low prevalence of the companies that have restated.
* re run the models using the up sampled data.