ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MÌNH
TRƯỜNG ĐẠI HỌC KINH TẾ - LUẬT
KHOA TÀI CHÍNH-NGÂN HÀNG

REPORT

TOPIC: CASH HOLDING

Student: Do Thi Lan Phuong

Ho Chi Minh City, 2022, May

1. Perform literature review

Cash holding has always been an important issue in corporate finance, investments and financial decisions. The available literature has shown the existence of different factors affecting the cash holdings of enterprises, such as cash flow ratio, growth opportunities, cash dividend, firm size, leverage, volatility of cash flows, net working capital, capital expenditures... These factors are mainly viewed from the perspective of two widely theories named Pecking order and Trade-off. Empirical findings of Magerakis (2015) show that cash holding is positively related to investment opportunities and volatility of cash flows in the industry. Cash holding is also negatively affected by cash flow, net working capital, capital expenditure, leverage, and tax costs (Magerakis, 2015). In line with previous findings, cash holding decreases progressively with firm size and debt ratios, and increases with profitability, growth prospects, and dividend payout ratio (Nguyen, 2005). This report cannot test all variables to consider the degree of impact to the cash holding, but only takes two variables to test that are leverage and cash dividend. In which, the leverage is a continuous variable and the cash dividend is a discrete variable.

- Discrete variable: cash dividend

According to the Static Trade-off Theory, there is an inverse relationship of dividend payments by cash to the firm's cash holdings. Firms paying dividends can deduct the dividend payment if cash is short. In contrast to the Static Trade-off theory, the Pecking order theory predicts that dividend payments are positively related to the firm's cash holdings. Companies paying dividends to their shareholders will be reluctant to cut dividends and need more cash to pay dividends than firms that do not. Many prior researchers have found that paying dividends has a positive effect on a company's cash holdings. In this report, with the cash dividend variable, firms with a cash dividend = 0 are considered not paying a specified dividend of 0, firms with a cash dividend > 0 are considered to pay a specified dividend of 1 (Code is attached below).

- Continuous variable: leverage

According to Wenyao (2007), leverage is one of the factors that determine cash holding of a firm. Leverage is the ratio that compares the total debt to total assets of the company. The findings regarding the relationship between leverage and cash holdings show different empirical results. Studies have documented a positive (Guney, Ozkan and Ozkan, 2007; Schwetzler and Reimund, 2004) meaning that the higher leverage, the higher would be the firm's cash holding. Howerver, studies 's Wijaya, Bandi, and Hartoko (2010) in Indonesia and Couderc (2005), Sadoour (2006) revealed that financial leverage has a negative influence on cash holding.

In the data file, the variables kept for calculation and testing are firm name, total asset, total debt, cash, cash_dividend and industry. In which, leverage and cash holding variables are calculated according to the formula:

Leverage = $\frac{Total\ debt}{Total\ asset}$

Cash holding = $\frac{Cash}{Total\ asset}$

```
data <- data %>%
  select(firmname,totalasset,totaldebt,cash,cash_dividend,industry) %>%
  mutate(leverage = totaldebt/totalasset) %>%
  mutate(cashholding = cash/totalasset) %>%
  mutate(cash_dividend=ifelse(cash_dividend > 0,'1','0'))
view(data)
```

2. Create Dataset

- Code

```
#Fill NA with the median value of the corresponding variable
data <- filedata %>%
  mutate_if(is.numeric, function(x) ifelse(is.na(x), median(x, na.rm = T), x))

#Create dataset
set.seed(741)
data <-
  data[sample(1:nrow(data),100,replace=F), ]
```

In order to facilitate the selection of variables along with the calculation of variables, I filled NA first. NA values are assigned with the median value of the

corresponding variable. Then I create a random sample that is replicable. Use set.seed() with the last 3 numbers in ID student as 741. Next is to use sample() to extract data with sample size of 100 firms, from row 1 to row 753.

- Result

	firmname	totalasset	totaldebt	cash	cash_dividend	industry	leverage	cashholding
1	Vinacomin Northern Coal Trading JSC	1.010851e+12	4.240000e+11	1.380877e+10	1	Energy	0.419448534	0.0136605383
2	NinhBinh Thermal Power JSC	3.366492e+11	0.000000e+00	5.589054e+10	1	Utilities	0.000000000	0.1660201033
3	Quoc Cuong GiaLai JSC	9.817180e+12	4.548724e+11	3.100985e+10	1	Real Estate	0.046334326	0.0031587326
4	Din Capital Investment Group JSC	2.223028e+11	4.017018e+10	1.070604e+10	0	Basic Materials	0.180700247	0.0481596858
5	Lilama 10 JSC	1.262122e+12	1.042423e+11	4.014127e+11	1	Industrials	0.082592898	0.3180457981
6	Clever Group Corp	4.496008e+11	8.334411e+10	2.356471e+11	1	Consumer Cyclicals	0.185373581	0.5241251571
7	Mechanics Construction and Foodstuff JSC	2.287814e+11	8.284263e+10	5.216581e+09	1	Consumer Non-Cyclicals	0.362103842	0.0228015938
8	Idico Corporation JSC	1.625207e+13	3.532382e+12	2.663887e+12	1	Consumer Non-Cyclicals	0.217349697	0.1639106484
9	Transimex Corp	5.734084e+12	1.300027e+12	1.271480e+12	1	Industrials	0.226719148	0.2217407086
10	Dong Trieu Viglacera JSC	3.249808e+11	1.770817e+11	1.563721e+10	1	Basic Materials	0.544898951	0.0481173530
11	Vietnam Electric Cable Corp	6.648569e+12	2.677408e+12	6.560001e+10	1	Industrials	0.402704477	0.0098667865
12	Thu Duc Electro Mechanical JSC	7.419711e+11	2.214439e+11	8.996502e+09	1	Industrials	0.298453601	0.0121251393
13	Hoa Phat Group JSC	1.782364e+14	5.721258e+13	4.070753e+13	1	Basic Materials	0.320992614	0.2283906266
14	Mediplantex National Pharmaceutical JSC	4.936377e+11	6.859456e+10	9.450679e+10	1	Healthcare	0.138957306	0.1914497214
15	Ngan Son JSC	3.119955e+11	4.406656e+10	3.681143e+09	1	Consumer Non-Cyclicals	0.141241021	0.0117987064
16	IPA Investments Group JSC	8.463797e+12	4.301788e+12	1.176806e+11	1	Financials	0.508257504	0.0139039986
17	Central Power Real Estate JSC	9.373386e+11	2.979822e+11	1.139166e+10	1	Real Estate	0.317902387	0.0121531973
18	Viet Duc Welding Electrode JSC	1.378786e+11	1.834715e+10	2.348592e+10	1	Industrials	0.133067405	0.1703376990
19	Real Estate 11 JSC	3.657196e+11	5.974683e+10	4.267106e+10	0	Real Estate	0.163367881	0.1166769743
20	Educational Book JSC in Ha Noi City	1.517194e+11	0.000000e+00	7.929479e+09	1	Consumer Cyclicals	0.000000000	0.0522640992
21	Viet Nam Gas and Chemicals Transportation Corp	3.128533e+11	2.641071e+09	4.886459e+10	1	Industrials	0.008441884	0.1561900993

Showing 1 to 21 of 100 entries, 8 total columns

The result of that piece of code is a remaining dataframe with 100 rows and 8 columns in all.

3. Report

- 5 firms with highest cash holding

- Code

```
#1. 5 firms with highest cash holding
highest <- data %>%
  arrange(desc(cashholding)) %>%
  slice_head(n=5)
view(data1)
```

To find the 5 firms with the highest cash holding, I used the function arrange(desc()) to sort in descending order and the function slice_head()with n = 5 to cut out the first 5 firms.

- Result

	firmname	totalasset	totaldebt	cash	cash_dividend	industry	leverage	cashholding
1	Danang Housing Investment Development JSC	1.591019e+12	0	1.039938e+12	1	Real Estate	0.000000000	0.6536302
2	Clever Group Corp	4.496008e+11	83344108150	2.356471e+11	1	Consumer Cyclicals	0.185373581	0.5241252
3	Vneco1 Electricity Construction JSC	3.768047e+10	0	1.873099e+10	1	Industrials	0.000000000	0.4971007
4	Hoa Binh Securities JSC	4.020100e+11	0	1.980422e+11	1	Financials	0.000000000	0.4926299
5	APG Securities Joint Stock Co	1.172976e+12	9800000000	5.233157e+11	1	Financials	0.008354815	0.4461435

Running the code, I have the result as shown above 5 firms with the highest cash holding are Danang Housing Investment Development JSC (cashholding = 0.6536302). Next are Clever Group Corp, Vneco1 Electricity Construction JSC, Hoa Binh Securities JSC and APG Securities Joint Stock Co.

- 5 firms with lowest cash holding

- Code

```
#2. 5 firms with lowest cash holding
lowest <- data %>%
  arrange(cashholding) %>%
  slice_head(n=5)
view(lowest)
```

To find the 5 firms with the lowest cash holdings, I used the function arrange() to sort in ascending order and the function slice_head() with n = 5 to cut out the first 5 firms.

- Result

	firmname	totalasset	totaldebt	cash	cash_dividend	industry	leverage	cashholding
1	Sai Gon Ha Noi Commercial Joint Stock Bank	5.066043e+14	5.310707e+13	91792563080	1	Financials	0.1048295	0.0001811918
2	Tien Phong Commercial Joint Stock Bank	2.928271e+14	6.966972e+13	91792563080	1	Financials	0.2379210	0.0003134702
3	Lien Viet Post Joint Stock Commercial Bank	2.891939e+14	4.787305e+13	91792563080	1	Financials	0.1655396	0.0003174084
4	Vinacomin Coc Sau Coal JSC	1.790082e+12	6.296252e+11	1163719560	1	Energy	0.3517297	0.0006500928
5	Vinacomin NuiBeo Coal JSC	3.475176e+12	2.410961e+12	3096429020	0	Energy	0.6937668	0.0008910137

Running the code, I have the result as shown above 5 firms with the lowest cash holding are Sai Gon Ha Noi Commercial Joint Stock Bank (cashholding = 0.0001811918). Next are the firms like Tien Phong Commercial Joint Stock Bank, Lien Viet Post Joint Stock Commercial Bank, Vinacomin Coc Sau Coal JSC and Vinacomin NuiBeo Coal JSC.

- The name of industries which the firms belong to

- Code

```
#3. The name of industries which the firms belong to
unique(data$industry)
```

To find out the name of the industries firms belong to, I use the function unique(). The unique function is used to remove or delete duplicate values or rows contained in a vector, data frame, or matrix. So when using the function unique(), all sectors can be accessed.

- Result

```
> #3. The name of industries which the firms belong to
> unique(data$industry)
 [1] "Energy"                "Utilities"
 [3] "Real Estate"           "Basic Materials"
 [5] "Industrials"           "Consumer Cyclicals"
 [7] "Consumer Non-Cyclicals" "Healthcare"
 [9] "Financials"            "Technology"
```

The result shows that there are a total of 10 industries including: Energy, Utilities, Real Estate, Basic Materials, Industrials, Consumer Cyclicals, Consumer Non-Cyclicals, Healthcare, Financials and Technology.

- Provide descriptive statistics with median, mean, max, min, standard deviation of cash holding of:

➢ *Different categories of the discrete variable*

- Code

```
# 4.1 Different categories of the discrete variable
describe1 <- data %>%
  group_by(cash_dividend) %>%
  summarise(
    median = median(cashholding, na.rm = T),
    mean = mean(cashholding, na.rm = T),
    max = max(cashholding, na.rm = T),
    min = min(cashholding, na.rm = T),
    sd = sd(cashholding, na.rm = T)
  )
describe1
```

Before summarizing the descriptive statistics table, I use the function group_by() to group together two categories of discrete variables consisting of two

values 0 and 1. Then use the function summarise() to create a table consisting of the follwing values:

+ Median: use the median() to calculate the median for the cash holding variable

+ Mean: use the mean() to calculate the mean for the cash holding variable

+ Max: use the max() to calculate the max for the cash holding variable

+ Min: use the min() to calculate the min for the cash holding variable

+ SD: use the sd() to calculate the standard deviation for the cash holding variable

- Result

```
# A tibble: 2 x 6
  cash_dividend median  mean   max      min    sd
  <chr>          <dbl> <dbl> <dbl>    <dbl> <dbl>
1 0             0.0824 0.114 0.253 0.000891 0.112
2 1             0.0525 0.120 0.654 0.000181 0.143
```

The above table is a descriptive statistics table of cash holding according to 2 categories of cash_dividend. In terms of median, the value 0 (pay no dividend) has a higher median of cash holding than the value 1 (pay dividend). However, when comparing the mean, the value 0 is lower than the value 1. The max index of the value 0 is lower than the value of 1 while the min index is higher. Finally, the standard deviation of 0 is lower than 1, which means that the fluctuation around the mean of cash holdings of firms not paying dividend is smaller than that of firms paying dividend. Because of the imbalance in the data between the two values (value 0 has 6 elements while value 1 has 94 elements) there can be certain discrepancies and errors. Based on the results of the descriptive statistics table, the value 1 will have more outliers than the value 0 and there is much difference in the data. The value 1 has a wider distribution and the dispersion is not as concentrated as the value 0. Thus, firms paying cash dividend tend to hold less cash than firms not paying dividend. So it makes sense and is similar to Static Trade-off Theory.

➢ *Groups of above/below median of the continuous variable*

- Code

```r
# 4.2 Groups of above/below median of the continuous variable
dd <- data %>%
  mutate(leverage=ifelse(leverage>median(leverage),'Higher median','Lower median'))
view(dd)
describe2 <- dd %>%
  group_by(leverage) %>%
  summarise(
    median = median(cashholding, na.rm = T),
    mean = mean(cashholding, na.rm = T),
    max = max(cashholding, na.rm = T),
    min = min(cashholding, na.rm = T),
    sd = sd(cashholding, na.rm = T)
  )
describe2
```

Separate the leverage variable into two components: above median and below median named Higher median and Lower median before creating the descriptive statistics table.

Before summarizing the descriptive statistics table, I use the group_by() to regroup the two categories of the continuous variable after separation named the Higher median and the Lower median. Then use the summarise() function to create a table containing the values like the previous task.

- Result

```
# A tibble: 2 x 6
  leverage      median   mean   max      min     sd
  <chr>          <dbl>  <dbl> <dbl>    <dbl>  <dbl>
1 Higher median 0.0290 0.0696 0.398 0.000313 0.0915
2 Lower median  0.140  0.170  0.654 0.000181 0.163
```

The table is a descriptive statistics table of cash holding according to 2 categories of leverage variable. In terms of median, firms with lower median of leverage have higher median of cash holdings than firms with higher median of leverage. When comparing the mean, the Higher median continues to be lower than the Lower median. The max index of the Higher median is lower than the Lower median while the min index is higher. Finally, the standard deviation of the Lower median is higher than the Higher median, that is the volatility around the mean of cash holdings of firms whose leverage higher median is smaller than firms whose leverage lower median. Based on the results of the descriptive statistics table, the Lower median has a wider distribution and dispersion. Meanwhile, the Higher

median has a narrow distribution and more concentrated dispersion. Thus, firms with a higher leverage median will have cash holdings lower than firms a lower leverage median. So that the variable leverage and variable cash holding have an inverse relationship similar to the opinion of Wijaya, Bandi, and Hartoko (2010) Couderc (2005), Sadoour (2006).

4. Data visualization

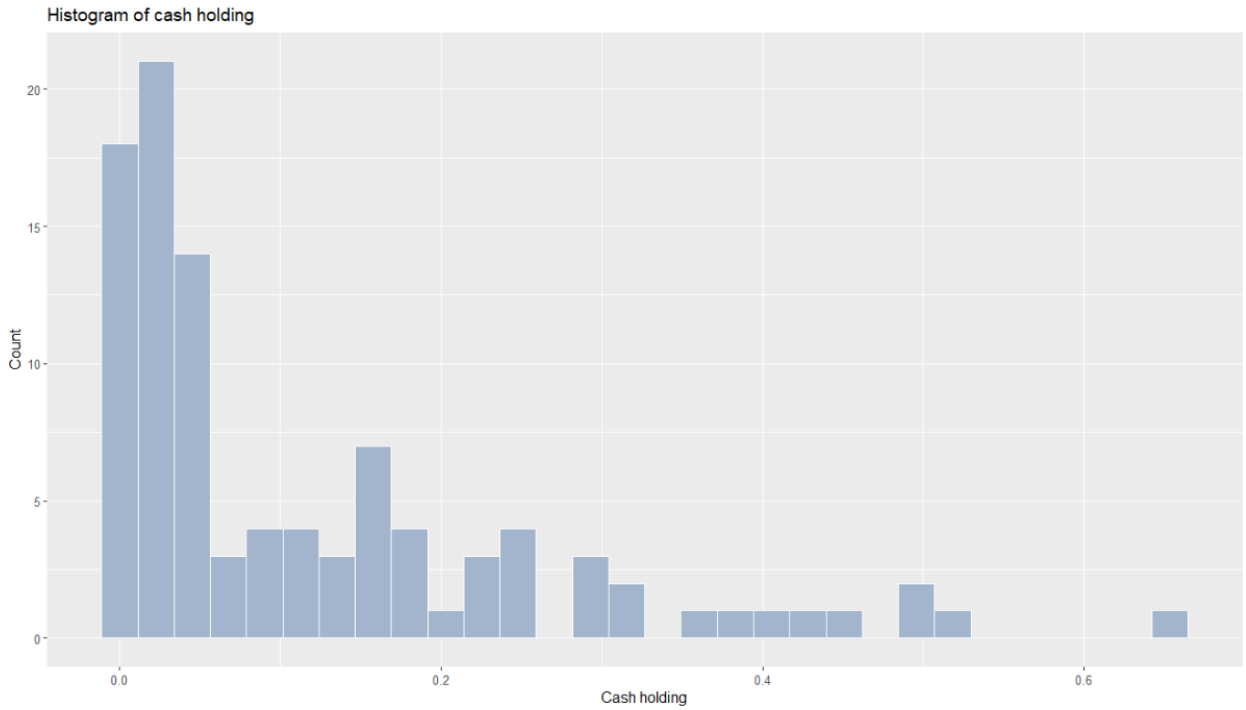
4.1. Provide histogram of cash holding

- Code

```
#1. Provide histogram of cash holding
data %>%
  filter(!is.na(cashholding)) %>%
  ggplot(aes(x=cashholding))+
  geom_histogram(fill='lightsteelblue3', color='white')+
  labs(title='Histogram of cash holding',
       x='Cash holding',
       y='Count')
```

Histogram is drawn with 4 elements. In which, data is a file data after calculating and selecting random 100 firms, filter to remove NA values, ggplot() with x = cash holding variable, use geom_histogram() to plot histogram and fill color of columns with lightseeblue palette. Finally, rename the chart and variables.

- Result



The graph shows the relationship between the cash holding variable and the number of firms with the cash holding ratio. According to the chart, the cash holding ratio of 100 firms in the range from 0 to 0.1 accounts for a large number with more than half of the number of firms. In which, the rate at about 0.02 has the most firms. Cash holding is evenly distributed from 0 to 0.2. From around 0.16, there is a downtrend. Looking at the chart, it can be seen that the cash holding is unevenly distributed in levels, valleys in many proportions and there are outliers at some points.

4.2. Provide scatter plot of cash holding with the continuous variable

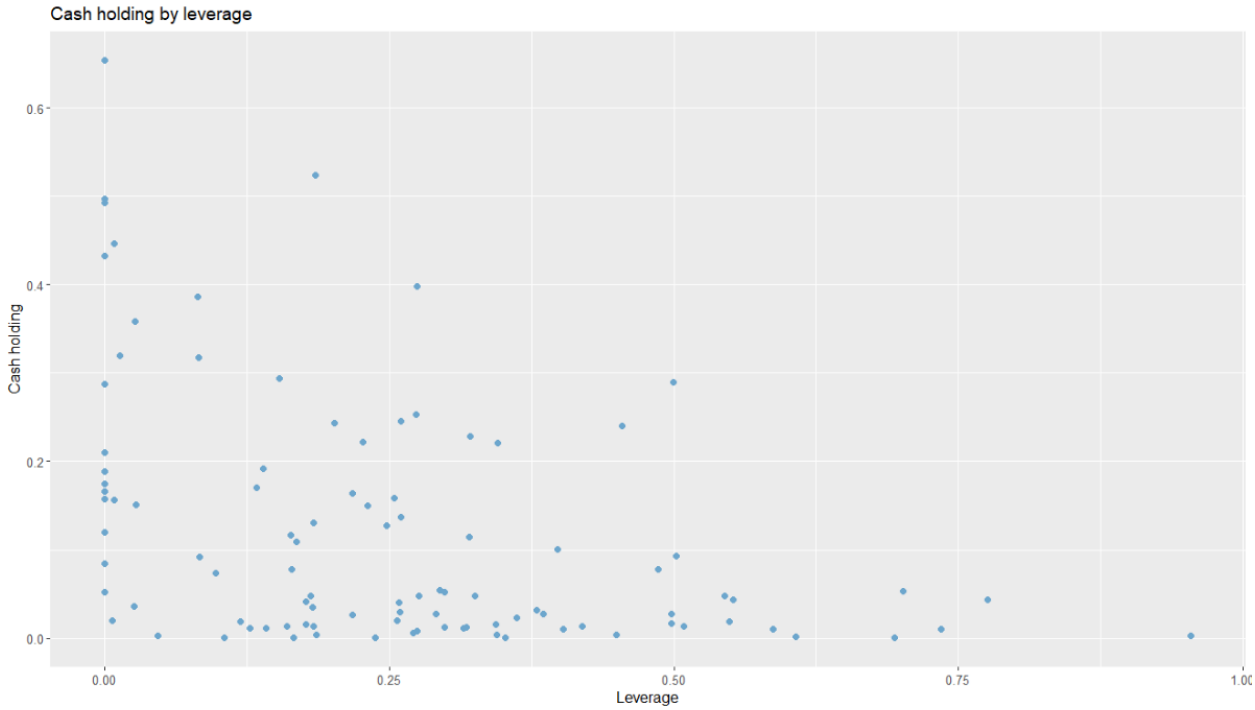
- Code

```
#2. Provide scatter plot of cash holding with the continuous variable
data %>%
  filter(!is.na(leverage), !is.na(cashholding)) %>%
  ggplot(aes(x=leverage,y=cashholding))+
  geom_point(size = 2, color='skyblue3')+
  labs(title='Cash holding by leverage',
       x='Leverage',
       y='Cash holding')
```

The scatter plot is drawn with 4 elements. In which, data is the file data after calculating and selecting random 100 firms, filter to remove NA values, ggplot()

with x = leverage variable and y = cash holding variable, use geom_point() to plot scatter plot and choose colors for the data points with the skyblue palette. Finally, rename the chart and variables.

- Result



The chart shows the relationship between the independent variable leverage and the dependent variable cash holding. In general, the data points tend to decrease, distributed mainly in leverage (from about 0.125 to 0.75) and cash holding (from about 0 to 0.2). At leverage = 0, the data points range from 0 to over 0.6 in the cash holding variable. It can be seen that there are still outliers and biases in the data points leading to the uneven distribution of the data. However, looking closely, we can still see the negative impact of leverage on cash holding. At the highest leverage, the cash holding is the lowest and vice versa. The descending trend is clearly seen in the chart. Thus, the results shown in the graph are similar to studies 's Wijaya, Bandi, and Hartoko (2010) in Indonesia and Couderc (2005), Sadoour (2006).

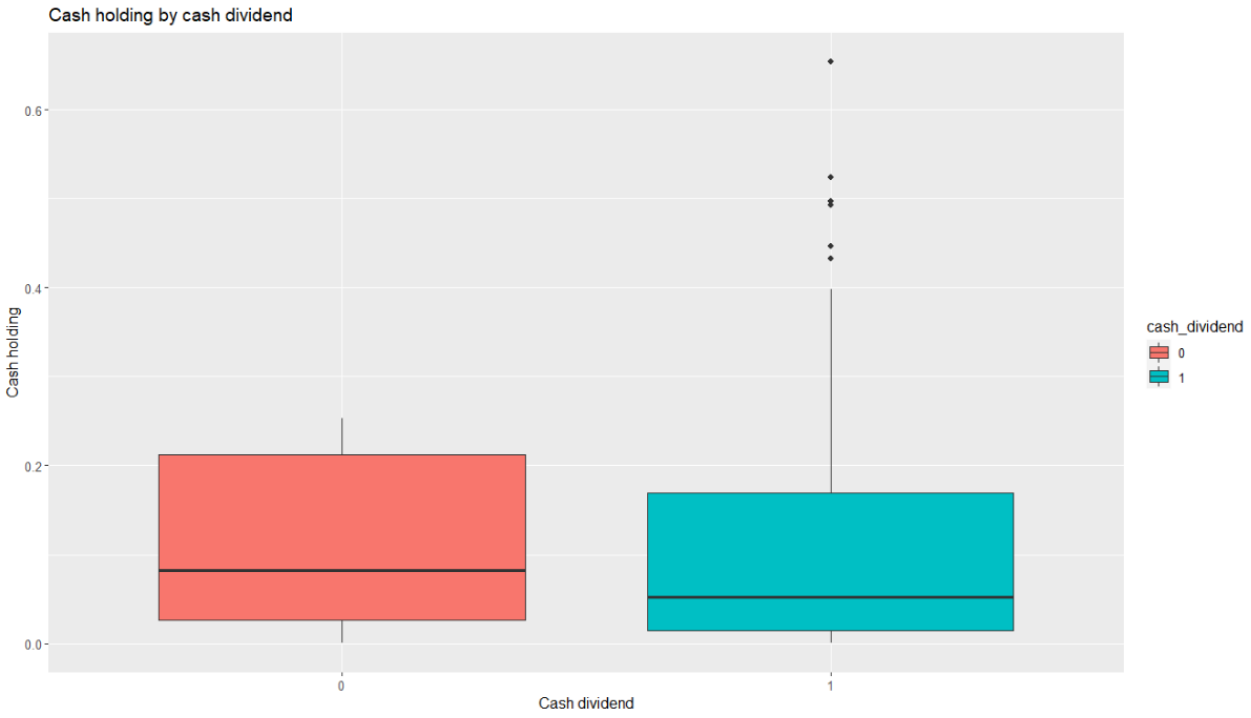
4.3. Provide boxplot of cash holding with the discrete variable (different colour for different categories of discrete variable)

- Code

```
#3. Provide boxplot of cash holding with the discrete variable
data %>%
  filter(!is.na(cash_dividend), !is.na(cashholding)) %>%
  ggplot(aes(x = cash_dividend, y = cashholding))+
  geom_boxplot(aes(fill=cash_dividend))+
  labs(title='Cash holding by cash dividend',
       x='Cash dividend',
       y='Cash holding')
```

The boxplot plot is drawn with 4 elements. In which, data is a file data after calculating and selecting random 100 firms, filter to remove NA values, ggplot() with x = cash_dividend variable and y = cash holding variable, use geom_boxplot() to plot boxplot and fill color for two categories of discrete variables. Finally, rename the chart and variables.

- Result



The graph shows the relationship between the cash_dividend variable and the cash holding variable. With two categories of 0 (no dividends) and 1 (cash dividends), the data shows that firms not paying dividends have a greater variability, range and dispersion than firms paying dividends. In particular, the median value of

0 is also higher than that of 1, indicating that cash holdings of firms not paying dividends are higher than those of firms paying dividends. Also, looking at the histogram, it can be seen that the value 1 has a lot more outliers than the value 0. Part of the reason is because the number of elements with the value 0 (6) is much more than the value 1 (94). Thus, it can be seen that cash holding and cash_dividend have an inverse relationship. Firms paying dividend tend to hold less cash than firms not paying dividends like Static Trade-off Theory.

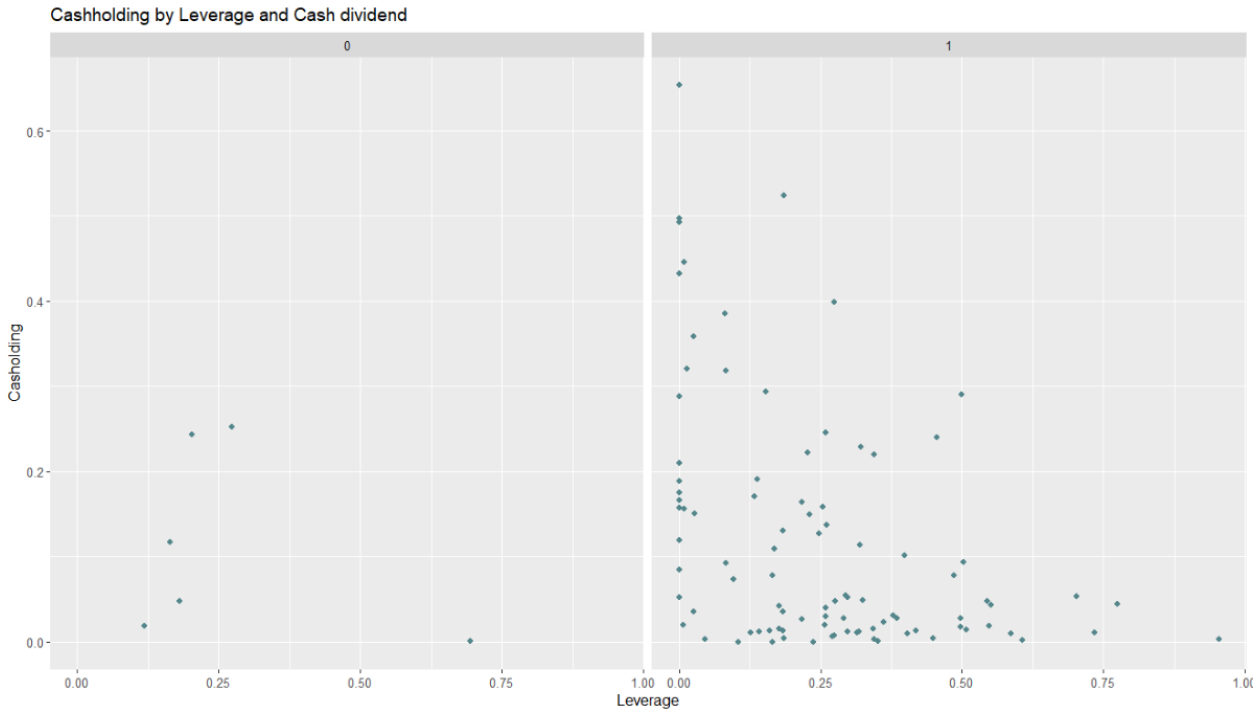
4.4. Provide a plot that allow the combination of continuous, discrete variables and cash holding

- Code

```
#4. Provide a plot that allow the combination of continuous, discrete variables and cash holding
data %>%
  filter(!is.na(cash_dividend), !is.na(leverage), !is.na(cashholding)) %>%
  ggplot(aes(x = leverage, y = cashholding)) +
  geom_point(color='cadetblue4')+
  labs(title='Cashholding by Leverage and Cash dividend',
       x='Leverage',
       y='Casholding') +
  facet_wrap(~ cash_dividend, nrow = 1)
```

The chart is drawn with 5 elements. In which, data is a file data after calculating and selecting random 100 firms, filter to remove NA values, ggplot() with x = leverage variable and y = cash holding variable, use geom_point() to plot plot and color the data points with the cadetblue palette. Next is to rename the chart and the variables. Finally, use facet_wrap() to classify the chart into 2 categories of the cash_dividend variable.

- Result



The chart shows the relationship between three variables, including two independent variables like leverage, cash_dividend and a dependent variable like cash holding. To firms not paying dividends, the distribution of data points tends to increase. The higher the leverage, the higher the cash holding. It is mainly distributed at leverage from 0.125-0.25 and cash holding from 0-0.26. To dividend-paying firms, the data points have a decreasing trend, showing a negative correlation between the leverage variable and cash holding. Allocation data is mainly at leverage from 0.125-0.5 and cash holding from 0-0.3. Thus, looking at the graph, we can conclude that firms that not paying dividends have a positive relationship between leverage and cash holding. Meanwhile, firms paying dividend have a negative relationship between leverage and cash holding similar to the studies of Wijaya, Bandi, and Hartoko (2010) in Indonesia and Couderc (2005), Sadoour (2006).

5. Using LOOP:

5.1 Count the number of firms in an industry (if you are given an industry name, you can count the number of firms in that industry)

```
#1. Count the number of firms in an industry
industry <- c(unique(data$industry))
count <- function(industry) {
  k = 0
  for (i in industry) {
    for (y in 1:nrow(data)){
      if (data$industry[y]==i)
        k = k+1
    }
  }
  return(k)
}
#Example count the number of firms in Financial industry
print(count('Financials'))
```

Firstly, I create a list where the elements are the industries which firms belong to. Then I write a function to count the number of firms in each industry. Initialize the counter variable k = 0. Run two nested loops. The first loop is to run 10 sectors in order of industries list. The second loop is to run from row 1 to row 100 of the dataframe. Then set the condition that if the value in the y^{th} row of the industry column is equal to the value i, the counter will increase by 1. Run all the loops and return the result k.

```
> #Example count the number of firms in Financial industry
> print(count('Financials'))
[1] 7
```

After writing the function, I test the number of firms in the Financials industry to verify whether the function was running correctly. As a result, there are 7 firms in the Financials industry. Then, I check again in the file data, the correct result is 7 firms.

5.2 Count the number of firms in an industry and with cash holding above a certain value (if you are given an industry name and a specific value of cash holding, you can count the number of firms in that industry and above that certain value)

```r
#2. Count the number of firms in an industry and with cash holding above a certain value
industry <- c(unique(data$industry))
cashholdings <- c(data$cashholding)
count1 <- function(industry,cashholdings) {
  k = 0
  for (i in industry) {
    for (y in cashholdings) {
      for (z in 1:nrow(data)){
        if (data$industry[z]==i & data$cashholding[z]>y)
          k = k+1
      }
    }
  }
  return(k)
}
#Example count the number of firms in Consumer Non-Cyclicals industry having cashholding above 0.0228015938
print(count1('Consumer Non-Cyclicals','0.0228015938'))
```

Just like the above task, first of all, I create a list where the elements are the industries to which firms belong and create a list of cashholdings. Then I write a function to count the number of firms in each industry and with cash holding above a certain value. Initialize counter variable k = 0. Run three nested loops. The first loop is to run 10 sectors in order of industry list. The second loop is to run the values in the cashholdings list in order of cash holding list. The third loop is to run from row 1 to row 100 of the dataframe. Then set the condition that if the value in the z^{th} row of the industry column is equal to the value i and the value in the z^{th} row of the cash holding column is greater than the value y, the counter will be increased by 1. Run all the rounds loop and return the result k.

```
> #Example count the number of firms in Consumer Non-Cyclicals industry having cashholding above
 0.0228015938
> print(count1('Consumer Non-Cyclicals','0.0228015938'))
[1] 4
```

After writing the function, I proceed to check the number of firms in the Customer Non-Cylicals industry with cash holding = 0.0228015938 to verify whether the function is running correctly. The output is that there are 4 firms in the Customer Non-Cylicals industry with cash holdings > 0.0228015938. Then, I checked again in the data file, the correct result was 4 firms.