VIETNAM NATIONAL UNIVERSITY – HO CHI MINH CITY
UNIVERSITY OF ECONOMICS AND LAW
FACULTY OF FINANCE AND BANKING

FINAL REPORT

SUBJECT: PROGRAM PACKAGE IN FINANCE 2

TOPIC: CASH HOLDING

Lecturer: Nguyen Thanh Liem

Student: Do Thi Lan Phuong

MSSV: K194141741

Ho Chi Minh City, 2022, June

1. Literature review

Cash holding has always been an important issue in corporate finance, investments and financial decisions. The decision to hold cash is one of the important decisions made by managers to achieve the company's goals. Therefore, it is the responsibility of managers to choose the optimal cash holding ratio to maximize the value of the company. According to Basely and Brigham (2005), manufacturing firms hold a certain amount of cash to derive benefits from three motives. The three motives mentioned are transactional motive, precautionary motive and speculative motive. In addition, there are theories about holding cash including Trade-off theory, The pecking order theory and Free-cash flow model. According to Jensen and Meckling (1976), Trade-off theory suggests that firms can establish their optimal cash holdings by weighing marginal costs and marginal benefits. On the other hand, The pecking order theory of Myer's (1986) suggests that firms should finance investments first with retained earnings, then with safe-haven and risky debt, and finally with equity to reduce asymmetric information and financial distress costs. Finally, Jensen's (1986) Free cash flow theory shows that managers have an incentive to accumulate cash to increase the amount of assets under their control and to gain discretion over the firm's investment decisions. These theories have been extensively tested by various researchers and play an important role in determining optimal cash holdings.

The available literatures have 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... Enyew A. (2013) studied the determinants of cash holdings of manufacturing companies in Addis Ababa, Ethiopia. The study only focuses on firm-specific variables and shows that net working capital, cash outflow, capital expenditure, firm size and growth opportunities are important variables in a firm's decision to hold cash of 12 manufacturing companies selected in Addis Ababa. Research results of Ferreira and Vilela (2004) indicate that cash holding is positively influenced by investment opportunities and cash flow and negatively affected by asset liquidity, leverage, size and bank debt. According to Saddour K. (2006), cash holding increases with operational risks and growth opportunities, but inversely with financial leverage. For growing firms, there is a negative relationship between cash and firm size, current asset levels and current liabilities. While firms' cash levels increase with firm size, investment levels and shareholder dividend payouts, and decrease with trade credit and investment and development costs. The research of Gill A. and Shah C. (2012) on the determinants of corporate cash holdings in Canada shows that cash flow, network capital, leverage, firm size, association size, the board of directors and executives who have significant influence over the corporate cash holdings of Canadian companies. Hofmann C (2006) examined the determinants of cash holdings of non-financial firms and found that the main determinants of New Zealand firms' cash holding are the changes in cash flow, leverage, dividend payments, and availability of alternative liquid assets.

Summarizing the studies reviewed above, I choose 4 variables to consider and evaluate the level of influence on Cash holding including Capital Expenditure, Growth, Leverage and Firm size.

- Capital Expenditure

Bates et al. (2009) state that the cost of capital can increase the ability to borrow and thus reduce the need for cash because capital expenditures help acquire assets that can be used as collateral for debts. So there will be a negative correlation between capital expenditure and cash holding. Research of Lee and Song (2007) shows a negative correlation between the capital expenditure and cash holdings in firms after the Asian financial crisis. However, Opler et al. (1999) found that cash holdings increase with the capital expenditure.

Therefore, I would expect a negative correlation between Capital expenditure and Cash holding.

- Growth

According to Trade-off theory, there is a positive correlation between growth and cash holdings. Similarly, The pecking order theory holds the view that there is a positive association between growth and cash holdings. Businesses with higher growth opportunities need higher levels of cash to deal with any cash shortages.

Therefore, I would expect a positive correlation between Growth and Cash holding.

- 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.

Therefore, I would expect a negative correlation between Leverage and Cash holding.

- Firm size

The larger firm size allows companies to obtain external funding more easily, so the company's cash holding also increases. Ali and associates (2016) found that firm size has a positive effect on cash holdings, but Tayem (2017) and Suherman (2017) found that firm size has a negative effect on cash holdings. Ferreira and Vilela (2003) find that there is a negative correlation between cash holdings and firm size. However, according to the research results of Chireka and Fakoya (2017) and Basheer (2014), firm size does not affect cash holdings.

Therefore, I would expect negative/positive correlation between Firm size and Cash holding.

The following is a summary of variables, calculations, symbols and expectations:

Variable	**Formula**	**Sign**	**Expectations**
Cash holding	$\frac{Cash}{Total\ assets}$	Cash_holding	
Capital Expenditure	$\frac{Fixed\ assets}{Total\ assets}$	Capex	-
Growth	$\frac{Sales(t) - Sales(t-1)}{Sales(t-1)}$	Growth	+
Leverage	$\frac{Total\ Debt}{Total\ assets}$	Leverage	-
Firm size	log(Total assets)	Firm_size	-/+

2. Data collection and input

```
#Import library
library(tidyverse)
library(readxl)
library(zoo)
library(pastecs)
library(car)
library(Metrics)
library(forecast)
library(tseries)
library(lmtest)
library(stats)
```

There are all the imported libraries for using in the article. Those libraries include: tidyverse (a versatile library including many other libraries… used in data preparation), readxl (used to read excel files), zoo (supporting time-series format creation), car (used for analysis), pastecs (used in creating descriptive statistics), Metrics (used for the command to run error indexes) and the rest of the libraries used for ARIMA model.

```
13 ###### Task: Data collection and input ######
14 # Read file
15 filedata = read_excel('K194141741.xlsx')
16 view(filedata)
```

Use read_excel() to read data. The data used in the article is from Hoa Binh Rubber Joint Stock Company, in the manufacturing industry. The company is listed on HOSE and trading with the stock code – HRC. Data is collected from finance.vietstock.vn. In terms of time, data is collected from Q1/2010 to Q4/2021. This file data has 48 rows and 37 columns. The attributes are mainly taken from the Balance Sheet, Income Statement, Cash Flow Statement and Ratios.

[Out 16]:

	Time	CA	Cash	ST receivables	Inventories	FA	LT investment	TA	Liabilities	CL	Non-CL	TD	Owner's equity	Share premium	Retained earnings	Total E and L
1	2010 Q1	65	34	14	15	101	233	400	55	46	9	779.70	345	2	99	400
2	2010 Q2	50	13	5	30	103	244	398	37	28	8	850.70	362	2	116	398
3	2010 Q3	77	43	5	28	116	253	448	54	46	9	880.32	393	2	148	448
4	2010 Q4	115	25	16	70	125	262	502	84	75	9	861.08	418	-2	170	502
5	2011 Q1	143	74	20	46	128	265	537	78	70	8	888.52	458	-6	211	537
6	2011 Q2	115	28	19	64	134	252	503	70	62	8	766.41	433	-6	171	503
7	2011 Q3	155	60	16	73	158	254	569	112	85	27	854.59	457	-6	195	569
8	2011 Q4	177	37	36	101	173	268	620	164	128	35	3678.85	457	-6	195	620
9	2012 Q1	125	70	11	43	175	265	567	74	35	39	3801.03	493	-6	231	567
10	2012 Q2	123	61	12	47	183	272	584	111	65	46	3727.24	473	-6	189	584
11	2012 Q3	144	87	7	39	205	271	623	118	59	59	7120.50	505	-6	220	623
12	2012 Q4	161	80	4	67	225	269	657	172	95	78	7516.52	484	-6	200	657
13	2013 Q1	121	70	10	21	229	269	624	150	67	83	10560.72	474	-6	190	624
14	2013 Q2	121	44	15	55	242	272	636	166	73	93	10264.80	470	-6	23	636
15	2013 Q3	118	48	22	36	267	272	658	166	59	107	12836.28	492	-6	45	658
16	2013 Q4	119	70	14	33	302	273	700	194	62	131	13849.22	506	-6	59	700
17	2014 Q1	100	63	15	13	312	273	687	184	48	136	16599.00	503	-6	56	687

Showing 1 to 17 of 48 entries, 37 total columns

After reading the data file, I proceed to create a new dataframe to calculate dependent and independent variables as well as use only the necessary attributes for the project.

```
#Calculate independent and dependent variables
data <- filedata %>%
  mutate(Cash_holding = Cash/TA) %>%
  mutate(Capex = FA/TA) %>%
  mutate(Growth = (Sales-lag(Sales))/lag(Sales)) %>%
  mutate(Leverage = TD/TA) %>%
  mutate(Firm_size = log(TA))
view(data)
```

5 variables are calculated for using, including Cash holding, Capex, Growth, Leverage and Firm size. In which:

- Cash holding = $\frac{Cash}{Total\ assets}$
- Capex = $\frac{Fixed\ assets}{Total\ assets}$

- Growth = $\frac{Sales(t)-Sales(t-1)}{Sales(t-1)}$ => In the article, I used the lag() function to reduce one order of the current value, i.e. query to the value (t-1)
- Leverage = $\frac{Total\ Debt}{Total\ assets}$
- Firm size = log(Total assets) => Use the available log() function

```
27 # Select necessary columns
28 data <- data %>%
29   select(Time,Cash_holding,Capex,Growth,Leverage,Firm_size) %>%
30   na.omit(Growth)
```

After calculating the necessary variables, the dataset will retain 6 necessary columns including: Time, Cash_holding, Capex, Growth, Leverage and Firm_size. In which, in the above step, when calculating the Growth variable, the first row will have no data (NA), so delete the row with that NA.

```
32 # Check number of na in data
33 sum(is.na(data))
```

[Out 33]:

```
> sum(is.na(data))
[1] 0
```

Check the number of NA in the dataframe so that NA can be replaced by reasonable methods. But, because the number of NA is zero, no further processing is required.

```
35 # Check type of column Time, convert right format and create variable Covid
36 typeof(data$Time)
37 data$Covid <- ifelse(data$Time >= as.yearqtr('2010 Q2', format = '%Y Q%q') &
38                      data$Time < as.yearqtr('2020 Q2', format = '%Y Q%q'), 0, 1)
```

[Out 36]:

```
> typeof(data$Time)
[1] "character"
```

Checking the type of the Time column, it is in the ‘character’ form, so it needs to be converted to the correct type and format. I use the as.yearqtr() statement with format = ‘%Y Q%q’. Combining with the type conversion, I proceed to create a dummy Covid variable with the time from Q2/2010 to Q1/2020 which is the time before the Covid-19 pandemic appeared (not having Q1/2010 because it was deleted due to NA). The time after the Covid-19 pandemic starts from Q2/2020 to Q4/2021. The time before the Covid-19 pandemic will be assigned a value of 0 and after the Covid-19 pandemic will be assigned a value of 1.

[Out 37]:

	Time	Cash_holding	Capex	Growth	Leverage	Firm_size	Covid
1	2010 Q2	0.032663317	0.25879397	0.30434783	2.137437	5.986452	0
2	2010 Q3	0.095982143	0.25892857	1.50000000	1.965000	6.104793	0
3	2010 Q4	0.049800797	0.24900398	0.04000000	1.715299	6.218600	0
4	2011 Q1	0.137802607	0.23836127	0.22435897	1.654600	6.285998	0
5	2011 Q2	0.055666004	0.26640159	-0.45549738	1.523678	6.220590	0
6	2011 Q3	0.105448155	0.27768014	0.77884615	1.501916	6.343880	0
7	2011 Q4	0.059677419	0.27903226	0.12432432	5.933629	6.429719	0
8	2012 Q1	0.123456790	0.30864198	-0.47115385	6.703757	6.340359	0
9	2012 Q2	0.104452055	0.31335616	0.24545455	6.382260	6.369901	0
10	2012 Q3	0.139646870	0.32905297	-0.13138686	11.429374	6.434547	0
11	2012 Q4	0.121765601	0.34246575	0.07563025	11.440670	6.487684	0
12	2013 Q1	0.112179487	0.36698718	-0.33593750	16.924231	6.436150	0
13	2013 Q2	0.069182390	0.38050314	0.25882353	16.139623	6.455199	0
14	2013 Q3	0.072948328	0.40577508	0.20560748	19.508024	6.489205	0
15	2013 Q4	0.100000000	0.43142857	-0.33333333	19.784600	6.551080	0
16	2014 Q1	0.091703057	0.45414847	-0.46511628	24.161572	6.532334	0
17	2014 Q2	0.033536585	0.50304878	-0.06521739	24.559299	6.486161	0
18	2014 Q3	0.045261669	0.51768034	-0.23255814	21.223423	6.561031	0
19	2014 Q4	0.014577259	0.57725948	0.21212121	28.219592	6.530878	0
20	2015 Q1	0.010144928	0.11449275	-0.32500000	20.216304	6.536692	0
21	2015 Q2	0.005714286	0.11142857	-0.70370370	21.537857	6.551080	0

Showing 1 to 21 of 47 entries, 7 total columns

This is a dataframe containing the necessary variables to be used for the following tasks. In it, the columns will include: Time, Cash holding, Capex, Growth, Leverage, Firm size and Covid. The dataset has all 47 rows from Q2/2010 to Q4/2021.

3. Descriptive statistics of all variables for BEFORE and AFTER periods

- *Descriptive statistics of all variables for Before and After periods*

```
# 1. Descriptive statistics of period before Covid-19
Before <- data %>%
  filter(Covid == 0) %>%
  select(-Time,-Covid) %>%
  stat.desc() %>%
  t() %>%
  subset(select = c(min,max,median,mean,std.dev))
view(Before)
```

Firstly, using the filter() function to select the rows belonging to the group before the Covid-19 pandemic. Next, using the select() function to select all columns except for ‘Time’ and ‘Covid’ columns. The stat.desc() function is used to calculate the indexes in the descriptive statistics table. And using t() function to transpose the table and the subset() function in combination with select() to choose the 5 indexes: min, max, median, mean and standard deviation.

```r
50 # 2. Descriptive statistics of period after Covid-19
51 After <- data %>%
52   filter(Covid == 1) %>%
53   select(-Time,-Covid) %>%
54   stat.desc() %>%
55   t() %>%
56   subset(select = c(min,max,median,mean,std.dev))
57 view(After)
```

Similar to the pre-Covid-19 period, use functions and create descriptive statistics with 5 key indexes. The only difference is to use the filter() function to select the rows belonging to the group after the Covid-19 pandemic.

- *Impact of Covid-19 pandemic through descriptive statistics*

[Out 48]:

	min	max	median	mean	std.dev
Cash_holding	0.002743484	0.2341832	0.038832893	0.0623317	0.06435302
Capex	0.094488189	0.5772595	0.251661301	0.2642961	0.12380568
Growth	-0.928571429	30.0000000	-0.007777778	0.9501905	4.82650171
Leverage	1.501915641	34.2815403	24.360435416	19.9562136	10.48967368
Firm_size	5.986452005	6.9650803	6.556055500	6.5426964	0.20849327

[Out 57]:

	min	max	median	mean	std.dev
Cash_holding	0.01265823	0.2335907	0.02822201	0.08060955	0.10166993
Capex	0.25587959	0.3362069	0.26815102	0.28211997	0.03127243
Growth	-0.65306122	1.7647059	0.53191489	0.35817064	0.88999466
Leverage	25.03873185	32.6931034	26.51278689	27.70058549	2.74894601
Firm_size	6.69950034	6.9688504	6.94022247	6.88186581	0.11100246

As shown in the descriptive statistics, the average of cash holding of the company in the previous Covid-19 period was 0.062 and the following Covid-19 period was 0.08. It can be seen that cash holding and volatility of cash holding increase after the pandemic. Capital expenditure has an average of 0.264 and 0.282 in the two periods before and after Covid-19. However, in general, in the previous Covid-19 period, capital expenditure was in the range (0.09-0.577) and fluctuated quite a bit compared to after Covid-19. With the range (0.255-0.336), the company has a fairly balanced quarterly capital expenditure with a volatility of only 0.03. The heavy impact of Covid-19 made the company's revenue decrease, leading to a decrease in the company's growth rate. Before the Covid-19 outbroke, the average of growth was 0.95 with the highest rate of 30. Meanwhile, in the following Covid-19 period, the average of growth was at 0.35 and the highest rate was only 1.76. The financial leverage ratio of the two periods clearly shows that the impact of Covid-19 is huge. For the previous Covid-19 period, leverage fluctuated between (1.50-34.28) and had an average of 19.9. Meanwhile, leverage after Covid-19 pandemic is not too volatile but quite big, ranging from (25.03-32.69) with an average of 27.70. It can be seen that, after the pandemic, the company borrowed more than before the pandemic. Finally, with the firm size, both periods showed little change and had average at 6.54 and 6.88 respectively.

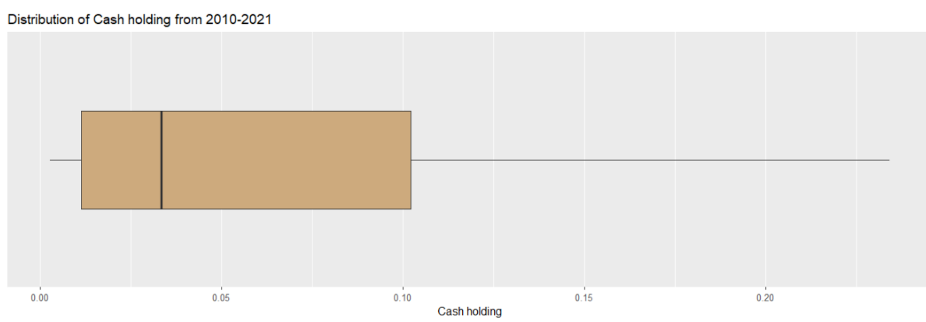
4. Box & whisker plot and histogram of the variable of cash holding

- *Box & whisker plot of the variable of cash holding*

```r
# 1. Box & whisker plot
ggplot(data = data, aes(y = Cash_holding)) +
  geom_boxplot(fill='burlywood3') +
  scale_x_discrete() +
  labs(title = 'Distribution of Cash holding from 2010-2021',
       y = 'Cash holding') +
  coord_flip()
```

Use ggplot() to plot. In which, use the geom_boxplot() function to draw a boxplot chart, attach a title, variable name and use coord_flip() to reverse the chart.

[Out 62]:



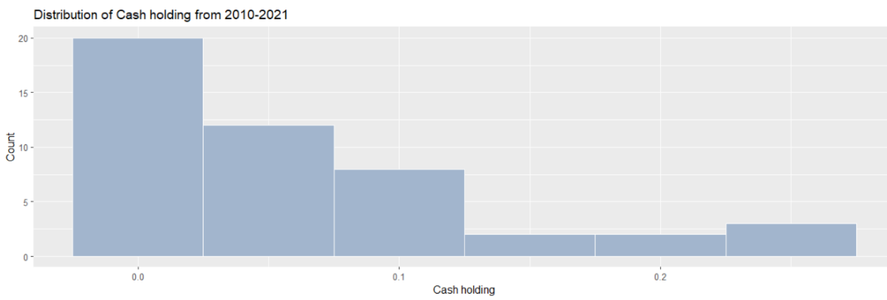
The chart shows the distribution of cash holdings from the beginning of 2010 to the end of 2021. Based on the chart, we see that cash holdings are mainly distributed in the range (0.011-0.102). In which, the average of cash holding for the past 11 years of the company is 0.033. Both min and max of cash holdings are 0.002 and 0.234. These are considered the outliers. It can be seen that the company’s cash holding ratio is at a low level, not exceeding 20%. Compared to total current assets, the company’s cash is small, it may be short of money or debt due to inefficient busi’ess operations as analyzed by the Growth variable in the descriptive statistics table.

- *Histogram of the variable of cash holding*

```r
# 2. Histogram
ggplot(data, aes(x=Cash_holding)) +
  geom_histogram(fill='lightsteelblue3', color='white', binwidth = 0.05)+
  labs(title='Distribution of Cash holding from 2010-2021',
       x='Cash holding',
       y='Count')
```

Use ggplot() to plot. In which, use the geom_histogram() function to plot histogram, attach title and variable name.

[Out 62]:



The chart shows the number of quarters for each cash holding level of the company from 2010 to 2021. Similar to the box-plot chart, the histogram shows that the company mainly has cash holdings ranging from 0-0.1. The number of quarters with cash holding at 0-0.3 accounts for a large number (20 quarters). The larger the cash holding, the lower the number of quarters. This shows that the company tends to hold fewer cash, no more than 30%. The amount of cash compared to total assets is quite low. Therefore, in my opinion, the company is having poor liquidity and may be having financial problems. The company has quite large total debt and increases over time, so this is also one reason why the company's cash holding is low. Another reason is that the company is facing difficulties in business operations, especially in the period after the Covid-19 epidemic.

5. Perform multiple regression to determine the significant determinants of the variable of cash holding

- *With the usual individual variables (model 1)*

```
78  # Build model
79  model_1 <- lm(Cash_holding ~ Leverage+Capex+Firm_size+Growth
80               , data = data)
81  summary(model_1)
82
83  # Correlation among independent variables
84  corr <- data[ , c('Leverage','Capex','Firm_size','Growth')]
85  cor(corr)
86
87  # Test multicollinearity with VIF method
88  vif(model_1)
89  mean(vif(model_1))
90
91  # Adjust model after considering
92  modeladjust_1 <- lm(Cash_holding ~ Leverage+Capex+Firm_size
93                 , data = data)
94  summary(modeladjust_1)
```

The steps to perform multiple regression for the cash holding variable include: building the model, running and reading the estimated results, looking at the correlation between the independent variables, checking for multicollinearity, and adjusting the model.

[Out 81]:

```
Call:
lm(formula = Cash_holding ~ Leverage + Capex + Firm_size + Growth,
    data = data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.123083 -0.020749 -0.003614  0.025816  0.092046

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.9908983  0.3334099  -5.971 4.37e-07 ***
Leverage    -0.0076310  0.0012127  -6.293 1.51e-07 ***
Capex        0.1507446  0.0664524   2.268   0.0285 *
Firm_size    0.3302386  0.0532085   6.206 2.00e-07 ***
Growth      -0.0006185  0.0016982  -0.364   0.7175
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04988 on 42 degrees of freedom
Multiple R-squared:  0.5366,    Adjusted R-squared:  0.4924
F-statistic: 12.16 on 4 and 42 DF,  p-value: 1.188e-06
```

This is the table of estimation results of the model between the dependent variable – Cash holding and 4 independent variables – Leverage, Capex, Firm size and Growth.

Firstly, in terms of R^2 and Adjusted R^2 are 0.5366 and 0.4924 respectively. This result shows that with the above model, the independent variables that explain 53.66% of the changes in the dependent variable or 53.66% of the change in Cash holding are due to the impact of 4 independent variables.

Next, analyzing the influence between independent and dependent variables based on coefficients. Looking at the estimation result, two variables Leverage and Growth are negatively correlated with the variable Cash holding. Meanwhile, the variable Capex and Firm size have a positive correlation with the variable Cash holding. From here, we also have the regression formula for the model like:

Cash holding = -1.990 – 0.008Leverage + 0.150Capex + 0.330Firm_size – 0.0006Growth

Thus, with 1 Leverage unit, the company reduces by 0.008 billion VND. Responding to a Capex unit, the company increased by 0.150 billion dong. Similar to Firm size and Growth, the company will increase by 0.330 billion VND and decrease by 0.0006 billion VND. Compared with the review at the beginning of the article, the results of the model are in line with the expectations of the article with Leverage and Firm size. Contrary to expectations are two variables Growth and Capex.

Finally, considering the significance of the model. All three variables Leverage, Capex and Firm size have p-value < 0.1, so all 3 are statistically significant. Meanwhile, the Growth variable has a p-value > 0.1 which is equivalent to no statistical significance, so it will be removed from the model.

[Out 84]:

```
          Leverage      Capex   Firm_size     Growth
Leverage   1.00000000 -0.1035264  0.79812058  0.06762231
Capex     -0.10352643  1.0000000 -0.13886960 -0.22544221
Firm_size  0.79812058 -0.1388696  1.00000000  0.02619098
Growth     0.06762231 -0.2254422  0.02619098  1.00000000
```

Running the correlation table between the independent variables, we see that Leverage and the Firm size have an index of 0.79 (quite large), so we conduct a multicollinearity test using the variable inflation factor (VIF).

[Out 88]:

```
 Leverage     Capex Firm_size    Growth
 2.774049  1.074555  2.792516  1.060780
```

[Out 89]:

```
> mean(vif(model_1))
[1] 1.925475
```

For the financial sector, VIF > 2 is considered multicollinear. However, if the mean of the VIF of the variables < 2, the model will be concluded that there is no multicollinearity. Thus, the model does not have multicollinearity

[Out 94]:

```
Call:
lm(formula = Cash_holding ~ Leverage + Capex + Firm_size, data = data)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.123310 -0.019607 -0.003328  0.026282  0.092491 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.000920   0.328905  -6.084 2.76e-07 ***
Leverage    -0.007667   0.001196  -6.409 9.28e-08 ***
Capex        0.156207   0.064081   2.438    0.019 *  
Firm_size    0.331573   0.052544   6.310 1.29e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04938 on 43 degrees of freedom
Multiple R-squared:  0.5351,	Adjusted R-squared:  0.5027 
F-statistic:  16.5 on 3 and 43 DF,  p-value: 2.799e-07
```

After considering the results and testing, Growth is excluded from the model. The estimation results of the new model with R^2 did not change much, but the Adjusted R^2 increased to 50.27%. Leverage continues to be negatively correlated with Cash holdings. The other two variables are positively correlated with the dependent variable. And all variables are statistically significant with p-value < 0.1.

- *With the usual individual variables and the interaction between Covid-19 dummy variable and the independent variables (model 2)*

```
# 2. With the usual individual variables and the interaction between Covid-19 dummy variable and
# The interaction between Covid-19 dummy variable and the independent variables
data <- data %>%
  mutate(Leverage_Covid = Leverage*Covid) %>%
  mutate(Capex_Covid = Capex*Covid) %>%
  mutate(Firm_size_Covid = Firm_size*Covid)
view(data)

# Build model
model_2 <- lm(Cash_holding ~ Leverage+Capex+Firm_size+Leverage_Covid+Capex_Covid+Firm_size_Covid
                   ,data = data)
summary(model_2)
```

Before running the model, create interaction variables between the independent variable and Covid dummy variable by multiplying the independent variable by the dummy variable to express the influence of Covid variable.

[Out 100]:

	Time	Cash_holding	Capex	Growth	Leverage	Firm_size	Covid	Leverage_Covid	Capex_Covid
1	2010 Q2	0.032663317	0.25879397	0.30434783	2.137437	5.986452	0	0.00000	0.0000000
2	2010 Q3	0.095982143	0.25892857	1.50000000	1.965000	6.104793	0	0.00000	0.0000000
3	2010 Q4	0.049800797	0.24900398	0.04000000	1.715299	6.218600	0	0.00000	0.0000000
4	2011 Q1	0.137802607	0.23836127	0.22435897	1.654600	6.285998	0	0.00000	0.0000000
5	2011 Q2	0.055666004	0.26640159	-0.45549738	1.523678	6.220590	0	0.00000	0.0000000
6	2011 Q3	0.105448155	0.27768014	0.77884615	1.501916	6.343880	0	0.00000	0.0000000
7	2011 Q4	0.059677419	0.27903226	0.12432432	5.933629	6.429719	0	0.00000	0.0000000
8	2012 Q1	0.123456790	0.30864198	-0.47115385	6.703757	6.340359	0	0.00000	0.0000000
9	2012 Q2	0.104452055	0.31335616	0.24545455	6.382260	6.369901	0	0.00000	0.0000000
10	2012 Q3	0.139646870	0.32905297	-0.13138686	11.429374	6.434547	0	0.00000	0.0000000
11	2012 Q4	0.121765601	0.34246575	0.07563025	11.440670	6.487684	0	0.00000	0.0000000
12	2013 Q1	0.112179487	0.36698718	-0.33593750	16.924231	6.436150	0	0.00000	0.0000000
13	2013 Q2	0.069182390	0.38050314	0.25882353	16.139623	6.455199	0	0.00000	0.0000000
14	2013 Q3	0.072948328	0.40577508	0.20560748	19.508024	6.489205	0	0.00000	0.0000000
15	2013 Q4	0.100000000	0.43142857	-0.33333333	19.784600	6.551080	0	0.00000	0.0000000
16	2014 Q1	0.091703057	0.45414847	-0.46511628	24.161572	6.532334	0	0.00000	0.0000000
17	2014 Q2	0.033536585	0.50304878	-0.06521739	24.559299	6.486161	0	0.00000	0.0000000
18	2014 Q3	0.045261669	0.51768034	-0.23255814	21.223423	6.561031	0	0.00000	0.0000000
19	2014 Q4	0.014577259	0.57725948	0.21212121	28.219592	6.530878	0	0.00000	0.0000000
20	2015 Q1	0.010144928	0.11449275	-0.32500000	20.216304	6.536692	0	0.00000	0.0000000

Because Growth variable is not statistically significant, it is not included in the model. I run the model with the remained three variables and the variables that interact with the Covid dummy variable. And model 2 is not used for the following tasks, so it will not adjust the model, only output the results for reviewing.

[Out 105]:

```
Call:
lm(formula = Cash_holding ~ Leverage + Capex + Firm_size + Leverage_Covid +
    Capex_Covid + Firm_size_Covid, data = data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.083037 -0.022134 -0.003194  0.023625  0.088431

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -2.730195   0.346632  -7.876 1.16e-09 ***
Leverage        -0.009155   0.001093  -8.377 2.46e-10 ***
Capex            0.175305   0.050936   3.442 0.001367 **
Firm_size        0.447658   0.055417   8.078 6.20e-10 ***
Leverage_Covid  -0.064702   0.015844  -4.084 0.000206 ***
Capex_Covid      6.425445   1.406580   4.568 4.64e-05 ***
Firm_size_Covid -0.012528   0.023560  -0.532 0.597845
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03869 on 40 degrees of freedom
Multiple R-squared:  0.7345,	Adjusted R-squared:  0.6947
F-statistic: 18.45 on 6 and 40 DF,  p-value: 3.9e-10
```

This is the table of estimating results of the model between the dependent variable – Cash holding and 6 independent variables.

Firstly, in terms of R^2 and Adjusted R^2 are 0.7345 and 0.6947 respectively. This result shows that with the above model, the independent variables explain 73.45% of the changes in the dependent variable or 73.45% of the change in Cash holding is due to the impact of 6 independent variables.

Next, analyzing the influence between independent and dependent variables based on coefficients. Looking at the results table, three variables Leverage, Leverage_Covid and Firm_size_Covid are negatively correlated with the Cash holding variable. Meanwhile, the variables Capex, Firm size and Capex_Covid have a positive correlation with Cash holding. From here, we also have the regression formula for the model as follows:

Cash holding = -2.730 – 0.009Leverage + 0.175Capex + 0.449Firm_size – 0.065Leverage_Covid + 6.425Capex_Covid – 0.013Firm_size_Covid

Finally, considering the significance of the model. All variables (except for Firm_size_Covid) have p-values < 0.1, so all 5 are statistically significant. Meanwhile, Firm _size_Covid with p-value > 0.1 is equivalent to not having statistical significance, so it will be excluded from the model.

- *Predict the value of the variable of assigned topic for all the quarters of the sample using Model 1*

```r
107 # 3. Predict the value of the variable of assigned topic for all the quarters
108 # Predict value
109 predictions <- predict(modeladjust_1, data)
110
111 # Create result dataframe of actual and predicted value
112 result <- data %>%
113   select(Time,Cash_holding)
114 names(result)[2] <- 'Actual'
115 result$Prediction <- predictions
116 view(result)
117
118 # RMSE,RSE
119 rmse(result$Actual,result$Prediction)
120 mae(result$Actual,result$Prediction)
```

Predicting Cash holding for all quarters are based on modified model 1. Then, creating a new dataframe consisting of 3 columns: Time, Actual (actual cash holding) and Prediction (predicted cash holding). Finally, using the RMSE and MAE indicators to check for errors.

[Out 116]:

	Time	Actual	Prediction
1	2010 Q2	0.032663317	0.008063080
2	2010 Q3	0.095982143	0.048645024
3	2010 Q4	0.049800797	0.086744599
4	2011 Q1	0.137802607	0.107894889
5	2011 Q2	0.055666004	0.091591325
6	2011 Q3	0.105448155	0.134399713
7	2011 Q4	0.059677419	0.129092926
8	2012 Q1	0.123456790	0.098183846
9	2012 Q2	0.104452055	0.111180519
10	2012 Q3	0.139646870	0.096368735
11	2012 Q4	0.121765601	0.115996267
12	2013 Q1	0.112179487	0.060694628
13	2013 Q2	0.069182390	0.075137731
14	2013 Q3	0.072948328	0.064533960
15	2013 Q4	0.100000000	0.086936818
16	2014 Q1	0.091703057	0.050709965
17	2014 Q2	0.033536585	0.039989120
18	2014 Q3	0.045261669	0.092677157
19	2014 Q4	0.014577259	0.038343194
20	2015 Q1	0.010144928	0.029348086
21	2015 Q2	0.005714286	0.023507424

[Out 119&120]:

```
> rmse(result$Actual,result$Prediction)
[1] 0.04723136
> mae(result$Actual,result$Prediction)
[1] 0.03391429
```

According to the results of two error indexes RMSE and MAE, we see that the difference between the actual value and the predicted value is quite little. RMSE = 0.047 and MAE = 0.034, both indicators show that the prediction results are quite effective and approximate to the actual value.

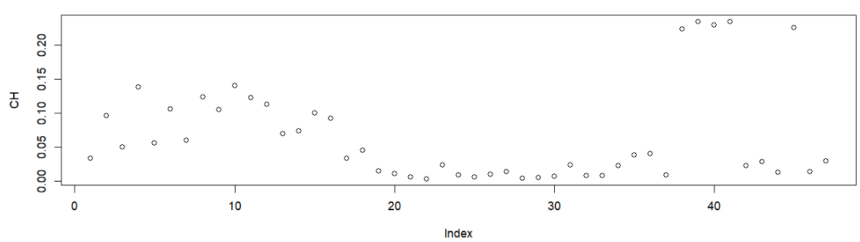
6. Perform ARIMA model to predict the variable of interest for the 4 quarters in 2022

The steps to perform the ARIMA model used in the lesson include:

- Test the stationarity of the time-series data. If the series is non-stationary, first order difference (Sai phân bậc 1) and continue to test
- Specify the model order by drawing acf, pacf or using auto.arima() to automatically select p,d,q
- Fit the model to the data and get the coefficients
- Apply test of the residuals to diagnostic the model
- Use the fitted model for forecasting

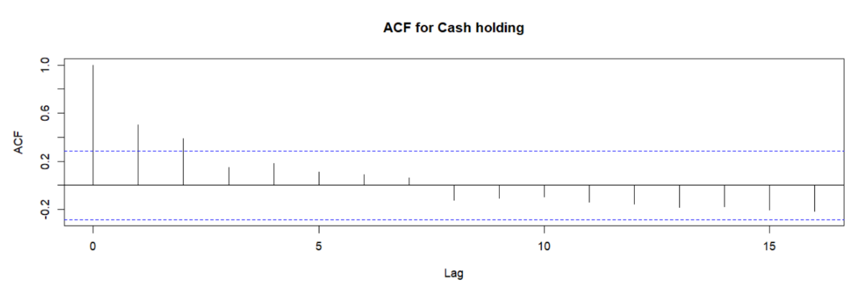
```
124 # Plot cash holding and ACF, PACF for cash holding
125 CH <- result$Actual
126 plot(CH)
127 acf(CH,main='ACF for Cash holding')
128 pacf(CH,main='PACF for Cash holding')
129
130 # Test stationary of cash holding
131 adf.test(CH)
```

[Out 126]:



Plot Cash holding

[Out 127]:



Plot ACF for Cash holding

[Out 131]:

```
        Augmented Dickey-Fuller Test

data:  CH
Dickey-Fuller = -1.991, Lag order = 3, p-value = 0.5776
alternative hypothesis: stationary
```

Based on the test results table, we see that p-value > 0.05, so this data series is not stationary. The first order difference is required to see if the data series is stationary.

```
133 # Diff cash holding with differences=1
134 d_CH <- diff(CH, differences =1)
135
136 # Test stationary of cash holding after diff
137 adf.test(d_CH,alternative='stationary')
```

[Out 137]:

```
        Augmented Dickey-Fuller Test

data:  d_CH
Dickey-Fuller = -4.3691, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

Based on the test results table, we see that p-value < 0.05, so the time-series data is stationary. Proceed to build the model by using the auto.arima() function with the 1st order difference data. The model will use the AIC to select the best model without considering the seasonality.

```
139 # Build model and select best model by auto.arima function
140 model=auto.arima(d_CH,seasonal=F,trace = T,lambda = 'auto',ic='aic')
```

[Out 140]:

```
 ARIMA(2,0,2) with non-zero mean : Inf
 ARIMA(0,0,0) with non-zero mean : -74.38097
 ARIMA(1,0,0) with non-zero mean : -81.26102
 ARIMA(0,0,1) with non-zero mean : -81.49165
 ARIMA(0,0,0) with zero mean     : 146.282
 ARIMA(1,0,1) with non-zero mean : -79.6613
 ARIMA(0,0,2) with non-zero mean : -79.54259
 ARIMA(1,0,2) with non-zero mean : -78.96639
 ARIMA(0,0,1) with zero mean     : 97.26731

 Best model: ARIMA(0,0,1) with non-zero mean
```

Model ARIMA(0,0,1) was selected as the best model based on the lowest AIC index. When having the model, apply diagnostic test of the residuals to validate the model.

```
142 # Obtain the coefficients
143 coeftest(auto.arima(d_CH,seasonal=F))
144
145 # Apply diagnostic test of the residuals to validate the model
146 acf(model$residuals)
147 pacf(model$residuals)
148 Box.test(model$residuals,lag=20,type='Ljung-Box')
```

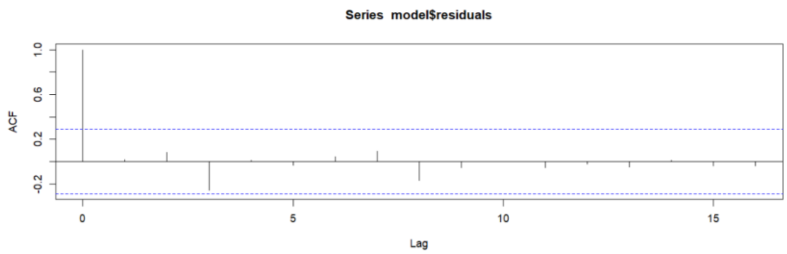
[Out 143]:

```
z test of coefficients:

    Estimate Std. Error z value Pr(>|z|)
ma1 -0.48532    0.16776  -2.893 0.003816 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

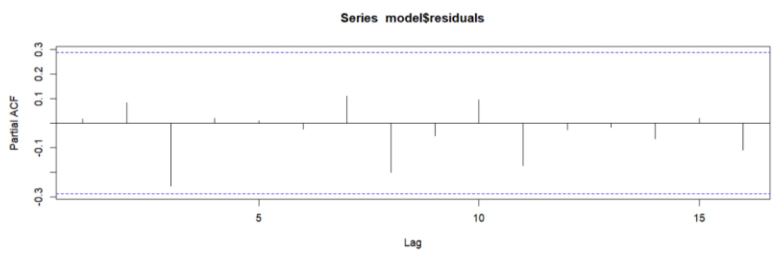
Use z-test of coefficients to test whether the coefficient is statistically significant or not. According to the test results, p-value < 0.05, so the coefficient is statistically significant.

[Out 146]:



Plot AFC for residuals of model ARIMA

[Out 147]:



Plot PACF for residuals of model ARIMA

[Out 148]:

```
        Box-Ljung test

data:  model$residuals
X-squared = 7.9623, df = 20, p-value = 0.9921
```

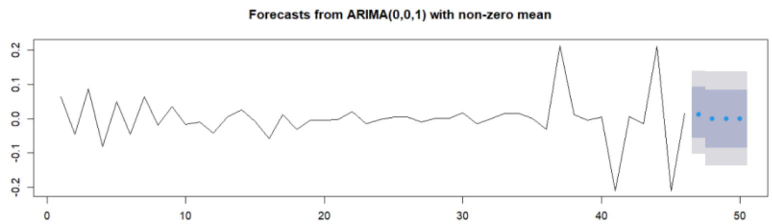
Use Ljung-Box test to test the degree of autocorrelation for the error-normalized squared series (Chuỗi bình phương giá trị chuẩn hóa sai số). For p-value > 0.05, there is no autocorrelation for the error-normalized squared series.

```
150  # Forecast cash holding for 4 quarters in 2022
151  forecast(model,h=4)
152
153  # Plot actual and forecasting cash holding
154  plot(forecast(model,h=4))
```

[Out 154]:

```
   Point Forecast       Lo 80      Hi 80      Lo 95     Hi 95
47    0.011768068 -0.05721827 0.09209002 -0.1031015 0.1403844
48   -0.000209002 -0.08478233 0.08369586 -0.1382481 0.1370871
49   -0.000209002 -0.08478233 0.08369586 -0.1382481 0.1370871
50   -0.000209002 -0.08478233 0.08369586 -0.1382481 0.1370871
```

Based on the table, it can be seen that the model predicts 4 quarters, in which, the first quarter has cash holding > 0, the next 3 quarters has cash holding < 0 but is equal and unchanged.



Plot for forecasting 4 quarters in 2022 from ARIMA(0,0,1)

7. Predict the variable of interest for the 4 quarters in 2022 with Random Forest

Because Cash holding variable is a continuous variable, when applying the Random Forest algorithm, the Random Forest regression model is used for prediction. The model's input includes the features: Leverage, Firm size, Capex and Growth. The output of the model is Cash holding. The sequence of steps is as follows: data processing (cleaning data, descriptive statistics, correlation of independent variables and visualizing), building the model (splitting training/test data, applying the model with training data, testing the model with test data and evaluating performance of the model) and predicting Cash holding for 4 quarters in 2022. Unlike Decision Tree, Random Forest will generate many decision trees randomly from 4 selected features. Each tree will be a subset of that feature set and the final result will be the sum of all decision trees that helps the predictive model to have higher accuracy than Decision Tree. After testing with the test dataset, evaluate the model's performance through error indexes such as RMSE, MAE, MSE and R2. Finally, apply the model Random Forest built above to predict 4 quarters in 2022.

REFERENCES

1. Mesfin, E. (2016) The Factors Affecting Cash Holding Decisions of Manufacturing Share Companies in Ethiopia, Papers.ssrn.com. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3386930 (Accessed: 15 June 2022).
2. Afza Talat and Adnan Sh.Muhammad (2007), Determinants of corporate cash holdings: A case of Pakistan, *Singapore Economic review and University of Manchester*, Singapore 164 165,Available:https://editorialexpress.com/cgibin/ conference/ download.cgi?db name=SERC2007 & paper id=166
3. Couderc, Nicolas. (2005) Corporate cash holdings: financial determinants and consequences. [Online] Available: http://www.univ-orleans.fr/deg/ GDR ecomofi/Activ/ couderc_strasbg05.pdf
4. Gill A. and Shah C. (2012), Determinants of Corporate Cash Holdings: Evidence from Canada, *International Journal of Economics and Finance*, Vol. 4, No. 1, January 2012