UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF FINANCE AND BANKING

❖❖❖

FINAL-TERM REPORT

PROGRAM PACKAGE IN FINANCE 2

Topic:

Cash Holding of the Masan business

Instructors: Nguyễn Thanh Liêm

Student: Nguyễn Đức Minh Tấn

Student ID: K194141745

Ho Chi Minh City, Jun 15, 2022.

Table of contents.

Abstract.

This study aims to study which factors affect the cash holding policy of Masan Group Corporation, a business in the manufacturing, food production, and other food production industries. This study collected data from the company's financial statements on finance.vietstock.vn website from the fourth quarter of 2009 to the first quarter of 2022. The results show that leverage, firm size, capital expenditures, and Covid-19 dummy variables are significant for the volatility of the cash holding of Masan company. In which, variables such as leverage, capital expenditures, and Covid-19 are the variables that have the opposite effect on cash holding. Meanwhile, the impact between firm size and cash holding at this company is positive.

1. Introduction.

Developing a reasonable cash holding policy to ensure the necessary cash source to maintain business activities, production, investment, debt liquidity, and provision purposes is a matter of great concern to business administrators. In addition, the research of Simionescu (2016) show that cash holdings is a significant indicator measuring corporate financial performance. However, holding a lot of cash higher or lower than necessary threshold will cause many risks for the business such as incurring the cost of holding cash, the profitability of the business and the ability to move capital for production business are reduced. Furthermore, the agency cost theory suggests that business managers can serve their self-interest by hoarding a lot of money, leading to conflicts between the owner and the agent (Harford, 1999). Therefore, the studies of factors affecting the tendency to hold cash play an important position in corporate governance, especially the enterprises in the manufacturing industry. There have been many studies that have shown a number of factors that affect the cash holding policy of each company. In this study, the author uses the previous results of papers to research the determinants of the Masan Joint Stock Company cash holding policy in the Vietnam market from the first quarter of 2009 to the first quarter of 2022.

Masan Group is one of the large business groups in the private economy of Vietnam, where domestic consumption is a major contributor to GDP and the engine of economic growth. The main business area that Masan Group focuses on is retail, consumer goods, including food and beverages, personal care products, etc., and financial services with many branches and affiliated companies across the country. All the more critically, every part organization is a significant piece of the customer ecosystem to meet the always-changing requirements of Vietnamese consumers.

2. Literature review.

Cash is the most liquid asset and it is determined by a company's capability to pay the debt. Corporate cash holdings always involve a trade-off between benefits and costs. The amount of cash being held by the company also reflects the asset structure and the liquidity of that company. Holding too much or not enough cash on total assets has different positive and negative effects at each stage of operation. If there is an excess of cash, the company incurs intangible fees such as inflation, investment opportunity costs, etc. If the business does not hold enough necessary cash threshold, it will cause certain difficulties in a turnover period. In addition, the amount of cash held affects the company's ability to handle unexpected difficulties. Therefore, determining the correct level of appropriate cash reserves is essential.

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

There have been many research articles in the world that have shown the factors affecting the cash holding policy of enterprises. In previous studies, there are many financial indicators that have positive and negative influences on the decision-making of cash holding strategies of an enterprise. Researches around the world show that continuous indicators such as financial strength, cash flows, leverage, dividends, intangibles, market to book ratio have effect on cash holdings (Aftab, Javid, & Akhter, 2018).

Leverage is a ratio that compares the total debts with the total assets of companies and this variable also indicates the company's ability to issue debt. The effect of leverage on cash holding can be a positive or negative relationship partly depends on country-specific characteristics (Guney, Ozkan, & Ozkan, 2007).

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

The pecking-order theory suggests that there is a negative relationship between the cash holding and leverage (Ferreira & Vilela, 2004). According to the free cash flow theory, it says that corporations with less leverage tend to hold more cash, meaning having a negative relationship between leverage and cash holding. Because investors may not supervise companies with low leverage so much, making more room for company managers to benefit themselves by holding a large amount of cash. In addition, the trade-off theory shows that there is a not clear relationship between leverage and cash holding, which means they can have a positive relationship or a negative relationship. In a research paper by Ferreira and Vilela (2004), it was demonstrated that there can be a negative relationship because high leverage can be a signal that the company can access the debt market resulting that they have no need to hold much cash. In addition, it can be that the costs of holding cash are greater for high leverage companies than for low leverage companies Ozkan and Ozkan (2004). In the other case, high leverage firms are considered to have higher financial risk. Thus, the firms need to hold more cash to decrease the risk, implying a positive relationship between leverage and cash holdings. Hence, on the basis of pecking order theory, free cash flow hypothesis, trade-off theory and previous empirical findings our next hypothesis are:

H_0: The leverage have a negative impact on the cash holding.

There are many researchers who have added the firm size variable to the research model, typically the study of Al-Najjar (2013), Gill and Shah (2012). Firm size is a commonly used determinant in empirical cash holding research (Aftab et al., 2018). The determinant is in general estimated by a firm’s total assets or the nature logarithm of the total asset.

$$Firm\ size = \ln(Total\ asset)$$

The trade off theory (Myers, 1977) and some previous research argue that large firms obtain edge of economies of scale over small firms so they do not require holding a large amount of cash reserves. However, there are still some theories that suggest a positive relationship between firm size and shareholdings, typically Pecking order theory (Myers & Majluf, 1984) and free cash flow theory (Jensen, 1986). The pecking-order theory investment implies that firm size and cash holding have a positive relationship because larger firms are considered to be more successful and therefore maintain more cash for the future. This theory has also been suggested by Opler, Pinkowitz, Stulz, and Williamson (1999) and Ferreira and Vilela (2004) finding evidence to support this point. Moreover, the free cash flow theory also suggests a positive impact of firm size on cash holdings. This is because bigger firms usually have more diversified shareholders, and managers can make decisions at their discretion. This leads to the increase of company cash holding because doing so can help managers increase the power of investing and financing decisions (Ferreira & Vilela, 2004). However, according to the Trade-off theory and due to the economy of scale, the demand for holding cash and the firm size has a negative relationship. Besides, it is arguable that small firms with higher developing potential, which can make them riskier, are likely to hold cash as it is more expensive for them to raise funds in the external market (Ferreira & Vilela, 2004), (Bover & Watson, 2005). Based on the above, it may be argued that firm size is significantly related to cash holdings, though the direction of their relationship cannot be estimated with certainty. Thus we hypothesize the following:

H_0: Firm size have the positive impact on the cash holding.

Capital expenditure is the amount of money a company spends to buy, upgrade and maintain tangible assets such as land, buildings, factories, technology or equipments. In the research of Gordon and Iyengar (1996), capital expenditure is estimated by a firm's total fixed assets divided by total assets.

$$Capital\ expenditure = \frac{Total\ fixed\ assets}{Total\ assets}$$

Empirical research by Mesfin (2016) and Arfan et al. (2017) have concluded that capital expenditures have a negative and statistically significant impact on cash holdings. There are also studies that have similar results to this study, Lee, Koh, and Kang (2011) also pointed out that firms with higher capital expenditures were shown to hold less cash. In addition, Mesfin (2016) evaluated the determinants of manufacturing cash holdings of share companies in Addis Ababa, Ethiopia also showed that capital expenditure is one of the significant variables of cash holding decisions of manufacturing companies in Addis Ababa. However, research by Mai (2010) in the UK private and public companies shows that there is no relationship between cash holding. Based on arguments and results from previous empirical studies, the author hypothesizes the following:

H_0: Capital expenditure has a negative impact on cash holding.

However, these studies only research on mature and developed economic markets, and the number of studies on this issue in developing countries is still limited. In Vietnam, Tran and Phan (2021) proved to us that leverage, profitability, cash conversion cycle, cash flow, and growth opportunities are variables that affect the cash holding of Vietnamese companies listed on the stock exchange. The research results of Tú (2017) also show that the factors affecting the cash holding ratio including company size, net working capital, cash flow, financial leverage, return on total assets, duration of operation and state ownership. In addition, the empirical study of Liêm, Thùy, and Thanh (2021) examined the effect of earnings management on cash holdings of non-financial firms listed on the Vietnamese stock market. The result of research shows that earnings management has a negative impact on cash holdings. The authors Liêm et al. (2021) used control variables in this study as tangible assets, firm size, operating cash flow, financial leverage, growth, and audit quality.

Based on an overview of previous studies, the author decided to use the leverage, firmsize and capital expenditure as control variables in studying cash holding policy at Masan Joint Stock Company.

3. Create dataframe

To perform steps such as reading data, processing, visualization and running model, the author import the libraries first. Next, the author check the data and handle the missing values. After making sure the dataset is clear, the author proceeds to calculate the necessary variables including cash holding, leverage, firm size, and capital expenditure. In addition, the author also included a dummy variable Covid-19 in the data frame.

- **<u>Code:</u>**

```r
# Ignore the warning messages
options(warn=-1)

# Import library
library(tidyverse)
library(readxl)
library(readr)
library(ggplot2)    # for plotting
library(forcats)    # for handling factors
library(scales)     # for axis scale formatting
library(xts)        # xts conclude index and core data (DataFrame have a timeseries columns)
library(tseries)    # adf.test
library(lmtest)     # coeftest
library(stats)      # Box.test
library(forecast)
library(PerformanceAnalytics)

# Import data from excel file.
df = read_excel("K194141745.xlsx")
View(df)

# Checking Na values.
sum(is.na(df))

####################################################    2. Calculated the control variable    ###########

covid_period = c(df$MSN[which(df$MSN == 'Q1/2020'):which(df$MSN == 'Q1/2022')])

# Calculate control varibles
data = df %>% transmute(Quarter = MSN,
                        CashHolding = round(Cash/TotalAssets,6),
                        Leverage = round(TotalDebt/TotalAssets,6),
                        CAE = round(FixAssets/TotalAssets,6),
                        Firmsizes = log(TotalAssets),
                        Covid19 = ifelse(MSN %in% covid_period,1,0)
)

# Checking Na values.
sum(is.na(data))
```

4. Descriptive statistics.

4.1. Descriptive statistics of Masan cash holding before Covid-19 (Q1/2009 – Q4/2019).

- **Code:**

```
#### Period Q4/2009 - Q4/2019
before= data %>% slice(1:which(Quarter == 'Q4/2019'))

# Descriptive statistics before Q1/2020
before_statis = before %>% summarise(variables = c('Cash Holding','Leverage', 'Capital expenditure','Firmsizes'),
                                     obs = nrow(before),
                                     min = c(min(CashHolding),min(Leverage),min(CAE),min(Firmsizes)),
                                     mean = c(mean(CashHolding),mean(Leverage),mean(CAE),mean(Firmsizes)),
                                     median = c(median(CashHolding),median(Leverage),median(CAE),median(Firmsizes)),
                                     std = c(sd(CashHolding),sd(Leverage),sd(CAE),sd(Firmsizes)),
                                     max = c(max(CashHolding),max(Leverage), max(CAE), max(Firmsizes))
)
before_statis = data.frame(before_statis)
before_statis
```

Table 1. Descriptive Statistics from Q1/2009 to Q4/2019

Descriptive Statistics	Cash holding	Leverage	Capital expenditure	Firmsizes
Observations	41	41	41	41
Min	0.047386	0.246514	0.544675	8.851807
Mean	0.1328106	0.5086328	0.7274480	10.6608426
Median	0.116750	0.494649	0.750640	10.803446
Standard deviation	0.06743804	0.13339969	0.06875038	0.67047146
Max	0.342349	0.721888	0.810980	11.485523

Through descriptive statistics, we can see that Masan Corporation has average cash holdings from Q1/2009 to Q4/2019 is 13,28%, while the average leverage is up to 50,86%, the average capital expenditure was 72,74% and the average firm size was 10,66. Thus, we can see that Masan tends to use more leverage than holding cash to serve operating activities such as buying tangible assets, investing in factories, buying production materials, etc. This can be explained by the fact that companies with a high leverage ratio have a higher return on equity. However, this also brings many financial risks and costs for the company. Masan's cash holdings during this period ranged from 4,7% to 34,23% while leverage ratios ranged from 24,65% to 72,19% and Capital expenditures ranged from 54,47% to 81,1%.

4.2. Descriptive statistics of Masan cash holding after Covid-19 (Q1/2020 – Q1/2022).

- **Code:**

```
#### Period Q1/2020 - now
after = data %>% slice(which(data$Quarter == 'Q1/2020'): nrow(data))

# Descriptive statistics after Q1/2020
after_statis = after %>% summarise(variables = c('Cash Holding','Leverage', 'Capital expenditure','Firmsizes'),
                                   obs = nrow(after),
                                   min = c(min(CashHolding),min(Leverage),min(CAE),min(Firmsizes)),
                                   mean = c(mean(CashHolding),mean(Leverage),mean(CAE),mean(Firmsizes)),
                                   median = c(median(CashHolding),median(Leverage),median(CAE),median(Firmsizes)),
                                   std = c(sd(CashHolding),sd(Leverage),sd(CAE),sd(Firmsizes)),
                                   max = c(max(CashHolding),max(Leverage), max(CAE), max(Firmsizes))
)
after_statis = data.frame(after_statis)
after_statis
```

Table 2. Descriptive Statistics from Q1/2020 to Q1/2022

	Cash holding	Leverage	Capital expenditure	Firmsizes
Observations	9	9	9	9
Min	0.035317	0.509241	0.653986	11.561754
Mean	0.08120822	0.70681467	0.71904622	11.66085962
Median	0.066712	0.737706	0.714997	11.663645
Standard deviation	0.04326452	0.08319910	0.03619362	0.07106462
Max	0.176893	0.783725	0.765936	11.744775

Through descriptive statistics, we can comment that Masan Joint Stock Company has average cash holdings in Q1/2020 to Q1/2022 period is 8,121%, down more than 5% compared to the previous period. At the same time, the company's average leverage also increased from 50,86% to 70,68% and the average firm size increased to 11,66. However, the average capital expenditure was 71,9%, unchanged so much from the previous period's 72.74%. The changes during this period may be due to the fact that the company is facing the global pandemic of Covid-19. Although this company in the food manufacturing industry is less affected by the pandemic than companies in other industries and has also taken measures to cope with the pandemic effectively Thành Trung (2022), the cash holding of The company has decreased significantly. Thus, we can concluded that the Covid-19 pandemic has had a negative effect on the company. This can be explained by the company having to cut raw materials, and limit production quantity, and at the same time, the company's cash flow is also blocked by the social distancing policies of the Government of Vietnam

5. Data visualization.

5.1. Histogram of cash holding

- **Code:**

```
# Histogram of cash holding
data %>% ggplot(aes(CashHolding)) +
  geom_histogram(binwidth = 0.03,fill = 'coral', color = 'white') +
  labs(title = "Histogram of Cash Holding",
       x = "Cash Holding",
       y = "Count")+
  scale_x_continuous(labels = comma)+
  scale_y_continuous()
```

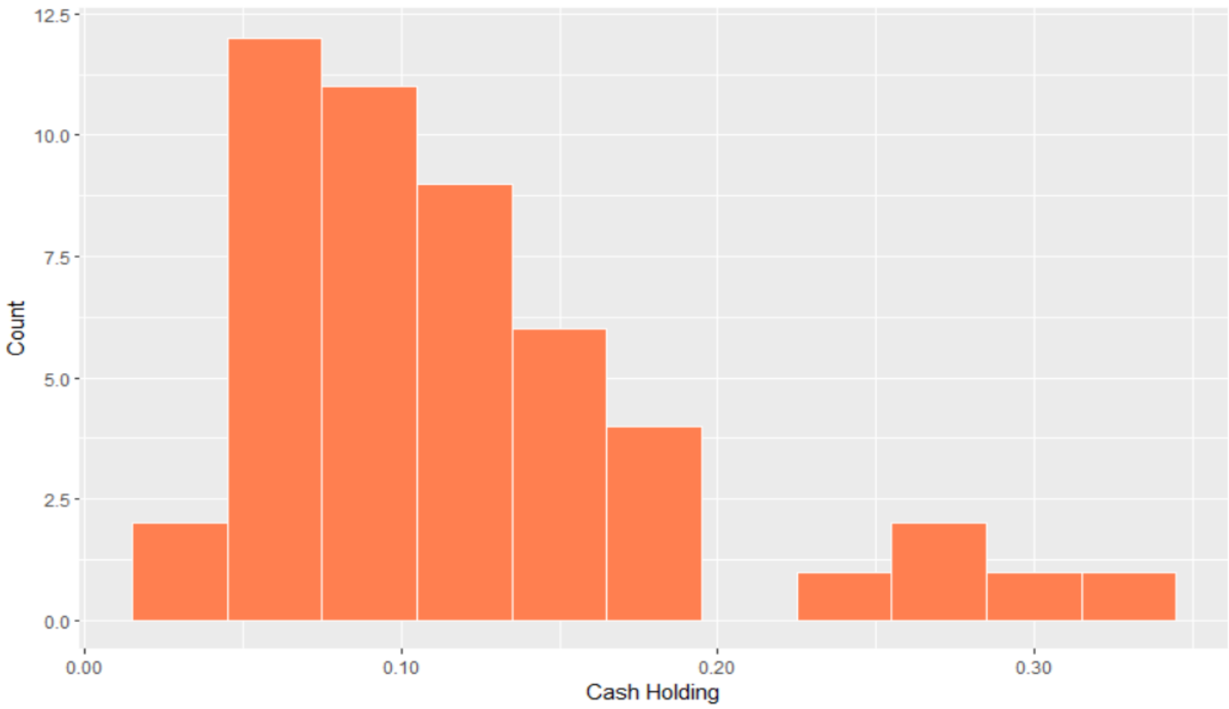

Figure 1. Histogram of cash holding

5.2. Box & whisker plot of cash holding.

- **<u>Code:</u>**

```
# Boxplot of cash holding
boxplot(data$CashHolding,
        main = "Box plot of Masan Cash holding",
        col = "orange",
        border = "brown",
        xlab= "Cash holding",
        horizontal = TRUE
)
```

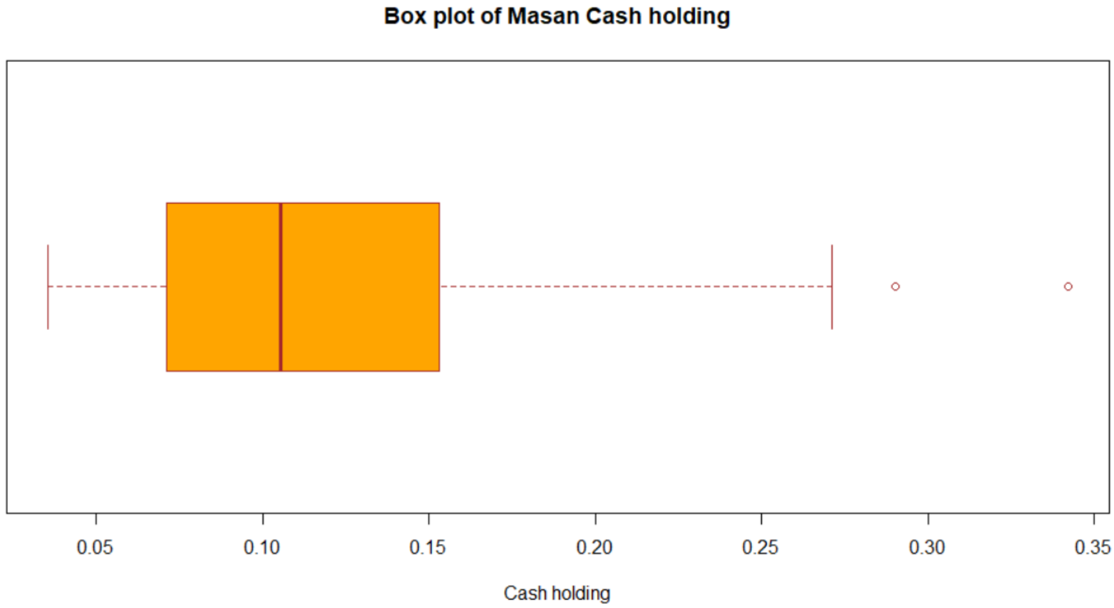

Figure 2. Boxplot of cash holding of Masan Group.

Both of the histogram plot and the box & whisker plot the cash holding of Masan company show that the cash holding distribution of Masan is left skewed and mostly range from as lowest is 5% to as highest is around 35%. However, the Masan Company regularly maintains a cash-to-total assets ratio from 5% to 15%, of which the the most frequently cash holding level that the company targets is 12%. Hence, we can see that Masan company tends to hold money at an average level, neither too high nor too low compared to other companies in the market. This result is similar to the study of Tú (2017), the average cash holding ratio of companies listed on the Vietnam stock exchange is 13%.

6. Running model.

Before the author performs the multiple linear regression model with the OLS method, the author visualize the relationship of the dependent variable is cash holdings with the independent variables including leverage, firmsize, and capital expenditures. The results show that these control variables tend to be linear characteristic for cash holding shown in figure 3. That is the reason that author decided to run the linear regression model with OLS method in this study.

- **Code:**

```
# Scatter for checking the linear characteristic in the data sets.
par(mfrow=c(1,3))
plot(CashHolding ~ Leverage, data=data, col = "blue", lwd = 2)
plot(CashHolding ~ CAE, data=data, col = "chartreuse4",lwd = 2)
plot(CashHolding ~ Firmsizes, data=data, col = "red",lwd = 2)
```

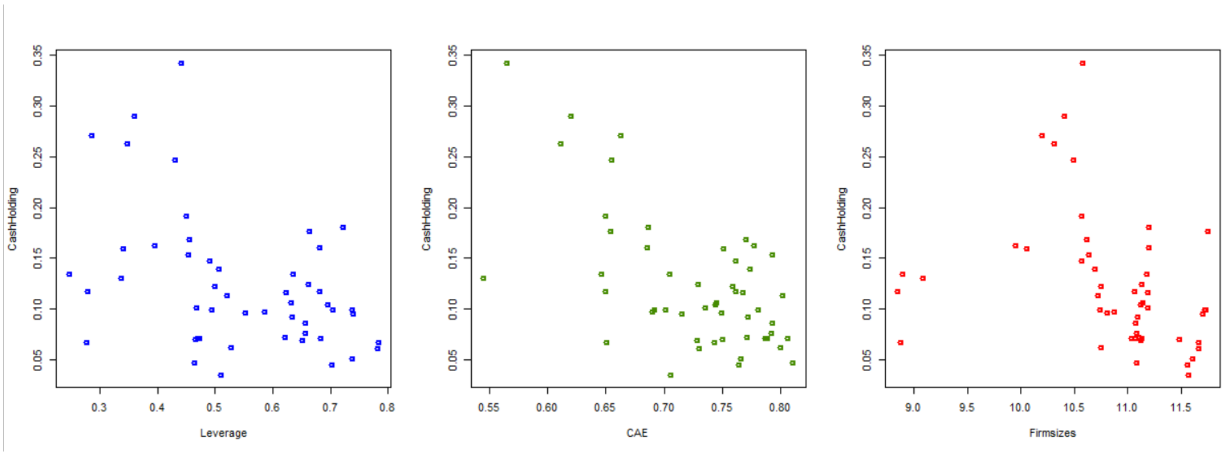

Figure 3. Relationship between cash holding and control variables.

However, during the period we studied the cash holding of Masan company, there was a huge economic shock not only to the Masan company but also to all the firms in the Vietnam market, which is the Covid-19 pandemic. According to the author's expectation, the variable Covid will be significant and have a negative effect on the cash holding of this business. In **model 2**, the author adds a dummy variable Covid-19 with 0 being the cash holding in the pre-Covid-19 period (Q4/2009 - Q4/2019) and 1 being the cash holding during the Covid-19 period (Q1/2020 - Q1/2022).

Model 1: $CASH_t = \beta_0 + \beta_1 LEVEAGE_t + \beta_2 FIRMSIZE_t + \beta_3 CAE_t + \varepsilon$

Model 2: $CASH_t = \beta_0 + \beta_1 LEVEAGE_t + \beta_2 FIRMSIZE_t + \beta_3 CAE_t + \beta_4 COVID19_t + \varepsilon$

In that:

β_0 : intercept of the model.

β_1, β_2, β_3, and β_4: is the slope of the model.

$CASH_t$: cash holding of Masan Group Joint Stock Company at the t time.

$LEVEAGE_t$: leverage of Masan company at the t time.

$FIRMSIZE_t$: firm size of Masan company at the t time.

CAE_t : capital expenditure of Masan company at the t time.

$COVID19_t$: dummy variable with 0 is the pre-Covid period and 1 is the Covid period.

We first perform regression with model 1 to see whether leverage, firm size, and capital expenditures really affect the cash holding of Masan firm. Besides, the author also wants to know whether the above variables have a positive or negative impact on the cash holding policy of this enterprise. The model results are shown in Table 3.

- **Code:**

```
##### 5.1. Multiple Linear Regression with the usual individual variables (model 1)
cash.holding.lm<-lm(CashHolding ~  Leverage + Firmsizes + CAE, data = data)
summary(cash.holding.lm)
```

Table 3. Regression result of model 1

```
Call:
lm(formula = CashHolding ~ Leverage + Firmsizes + CAE, data = data)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.131078 -0.034647  0.008655  0.033492  0.106227 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.34302    0.16468   2.083   0.0428 *  
Leverage    -0.23095    0.08943  -2.583   0.0131 *  
Firmsizes    0.03156    0.01876   1.682   0.0994 .  
CAE         -0.60047    0.12380  -4.850 1.45e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05023 on 46 degrees of freedom
Multiple R-squared:  0.4641,	Adjusted R-squared:  0.4292 
F-statistic: 13.28 on 3 and 46 DF,  p-value: 2.252e-06
```

Through model regression results, we can see that all three variables Leverage, Firmsize, and CAE are statistically significant. In which, leverage and capital expenditure have a negative effect on cash holding and firm size has a positive effect on Masan's cash holding. This is completely true with the expected hypothesis that the author set out at the beginning of the article. All other things being unchanged, a company's leverage increases by 1%, the company's cash holding decreases by 0,231%; the company's capital expenditure increases by 1%, Masan's cash holding decreased by 0,6%; Masan's firm size increases by 1%, the cash holding will increase by 0,031%. In this model, all control variables have explanatory significance at the significance level is 10% and these independent variables explained 42,92% of the variation of Masan's cash holdings through the adjusted R-squared index. The rest 57,08% is explained by the out-of-model variables and random error.

- **Code:**

```
# Check normality
shapiro.test(resid(cash.holding.lm)) # Null hypothesis is normality

# Check homoskedasticity
bptest(cash.holding.lm) #Null hypothesis is homoskedasticity
```

Table 4. Result testing of model 1

```
        Shapiro-Wilk normality test

data:  resid(cash.holding.lm)
W = 0.98317, p-value = 0.6909
```

```
        studentized Breusch-Pagan test

data:  cash.holding.lm
BP = 17.691, df = 3, p-value = 0.0005093
```

After running model 1, the author uses two tests: Shapiro-Wilk normality to check the normal distribution of the data with the null hypothesis is normality and the Breush-Pagan test to check whether there is a homoscedasticity phenomenon in this dataset or not with the null hypothesis is homoscedasticity. After performing the above two tests, this data set follows a normal distribution with p-value = 0.6909 > α = 0,05 (accepting H_0) at the Shapiro-Wilk test. At the same time, this data set does not have homoscedasticity with p-value = 0.0005093 < α = 0,05 (reject H_0).

Next, the author tests whether the impact of the Covid-19 pandemic really affects Masan's cash holding by model 2. The result regression is shown in Table 5.

- **Code:**

```
##### 5.2. Multiple Linear Regression with the usual individual variables and the interaction
# between Covid-19 dummy variable (model 2)
data$Covid19 = factor(data$Covid19)
cash.holding.lm.covid<-lm(CashHolding ~  Leverage + Firmsizes + CAE + Covid19 , data = data)
summary(cash.holding.lm.covid)
```

Table 5. Regression result of model 2

```
Call:
lm(formula = CashHolding ~ Leverage + Firmsizes + CAE + Covid19,
    data = data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.094248 -0.019667  0.000598  0.030951  0.081586

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.24052    0.15293   1.573  0.12279
Leverage    -0.18393    0.08254  -2.228  0.03090 *
Firmsizes    0.04966    0.01793   2.770  0.00812 **
CAE         -0.74725    0.12115  -6.168 1.76e-07 ***
Covid191    -0.07109    0.02175  -3.269  0.00207 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04565 on 45 degrees of freedom
Multiple R-squared:  0.5669,    Adjusted R-squared:  0.5284
F-statistic: 14.73 on 4 and 45 DF,  p-value: 9.146e-08
```

Through model 2 regression results, after adding the Covid-19 dummy variable to the model, all four variables Leverage, Firmsize, CAE, and Covid191 are statistically significant at the significance level is 10%. In which, leverage, capital expenditure, and Covid-19 have negative effects on cash holding, but firm size has a positive effect on Masan's cash holding. Other things being unchanged, the company's leverage increases by 1%, the company's cash holding decreases by 0,18393%; if the company's capital expenditure increases by 1%, Masan's cash holding decreased by 0,74725%. The impact of Covid-19 will cause cash holdings to decrease by 0,07109%. If Masan's firm size increases by 1%, the cash holding increase by 0,04966% when other factors remain unchanged.

The negative impact of Covid-19 on the volatility of cash holdings is explained because the Covid-19 pandemic has caused difficulties for the whole society, including Masan Company. After including the Covid-19 factor in the model, all the control variables are significant at the significance level is 10% and these independent variables explain 52,84% of the variation of Masan's cash holding. The remaining percentage of causality can be explained by out-of-model variables and random error.

- **Code:**

```
# Check normality
shapiro.test(resid(cash.holding.lm.covid))# Null hypothesis is normality

# Check homoskedasticity
bptest(cash.holding.lm.covid) #Null hypothesis is homoskedasticity
```

Table 6. Result testing of model 2

```
        Shapiro-Wilk normality test

data:  resid(cash.holding.lm.covid)
W = 0.97341, p-value = 0.3168
```

```
        studentized Breusch-Pagan test

data:  cash.holding.lm.covid
BP = 19.495, df = 4, p-value = 0.000628
```

After including the Covid-19 dummy variable in the model, the data still ensures the normal distribution and does not have the homoscedasticity phenomenon shown in the table 6. At Shapiro-Wilk test, p-value = 0,3168 > α = 0,05 (accept H_0) and at Breush-Pagan test, p-value = 0.00628 < α = 0,05 (reject H_0).

7. Prediction results.

After forecasting the cash holding for all the quarters based on the regression result of model 1, we obtain the forecast results shown in Table 7.

- **Code:**

```
##### 5.3. Prediction
pred.model = predict(cash.holding.lm, data[,3:5])

pred.table =data.frame(Quanter = data$Quarter,
                       CashHolding.Real = data$CashHolding,
                      CashHolding.Pred = round(pred.model,6))
pred.table
```

Table 7. Prediction result of Cash holding		
Quarter	**Cash holding**	**Cash holding prediction**
Q4/2009	0.117218	0.167932
Q1/2010	0.067455	0.168541
Q2/2010	0.13399	0.178709
Q3/2010	0.130077	0.224958
Q4/2010	0.162269	0.098663
…	…	…
Q4/2020	0.066712	0.08389
Q1/2021	0.061221	0.091941
Q2/2021	0.095644	0.112148
Q3/2021	0.099679	0.121638
Q4/2021	0.176893	0.167552
Q1/2022	0.099361	0.135317

- **Code:**

```
>
> accuracy(pred.table$CashHolding.Real,pred.table$CashHolding.Pred)
              ME       RMSE        MAE       MPE     MAPE
Test set -4e-08 0.04817554 0.03897424 -1.084706 30.84666
>
```

In this study, the author uses indexes such as RMSE, MAE, and MAPE to evaluate the model's predictive performance. The closer these indicators are to zero, the more accurate the forecast model is compared to reality. The results show that the difference between the predicted value and the actual value is at a relatively acceptable level. Specifically, the RMSE index is 0,04818; the MAE index is 0,03897; MAPE index is 30,85%.

8. Predict cash holding by ARIMA.

Auto Regressive Integrated Moving Average (ARIMA) is actually a class of models that explains a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values. The process of ARIMA model is understood as a model that combines three main factors: the autoregressive component AR (p), the stationarity of the time series I (d) and moving average model being the full name of MA (p) to create a ARIMA(p, d, q) model.

First of all, we have to make the first ban to data time by creating an xts data frame through the library "xts".

- **Code:**

```
# Check class
class(data$Quarter)

# Convert "character" to "Date" data
date.time = seq(as.Date("2009/12/01"), by = "quarter", length.out = nrow(data))
date.time

# Create a xts Dataframe
msn.cash = xts(data[,2],date.time)

View(msn.cash)

plot(msn.cash$CashHolding)
```

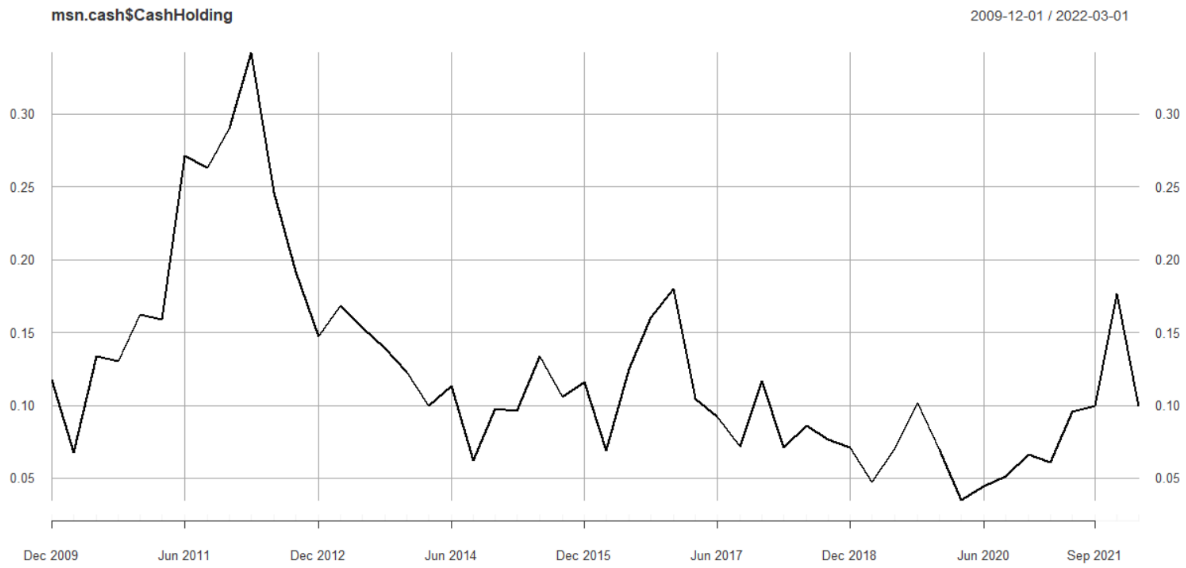

Figure 4. Cash holding of Masan.

To run the ARIMA model, we must ensure that the input data is a stationary series. This helps to ensure that spurious regressions of the dataset are avoided. However, in figure 4, the author initially stated that these time series are not stationary because they are trending up and down and are seasonal. To prove this statement, the author initially consider the stationarity of the data series through two auto-correlation function (ACF), and a partial auto-correlation function (PACF) visualized in Figure 5.

- **Code:**

```
# ACF/PACF
par(mfrow=c(1,2))
acf(msn.cash$CashHolding,main='ACF for CashHolding',lag.max = 24)
pacf(msn.cash$CashHolding,main='PACF for CashHolding',lag.max = 24)
```

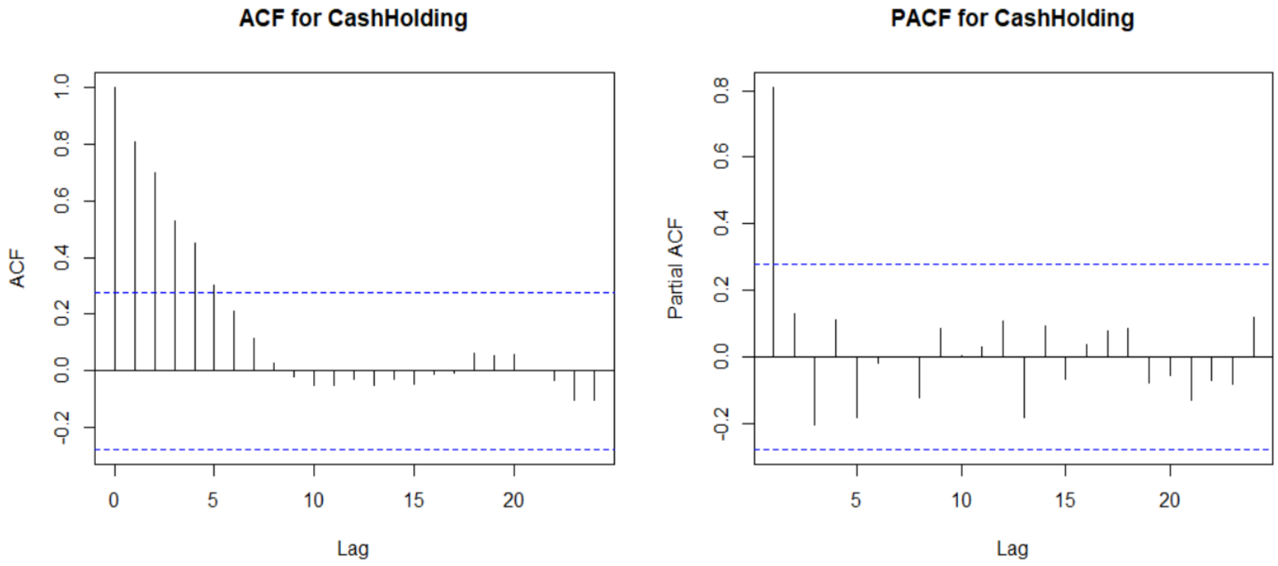

Figure 5. ACF/PACF plot of cash holding.

In Figure 5, we draw two characteristics: First, the ACF decreases very slowly, as in the ACF plot, the ACF up to 5 lags are all non-zero in statistical significance because they are outside the 95% confidence limit. Second, after the first delay, the PACF decreased sharply and all PACFs after the first delay were not statistically significant. Therefore, Masan's cash holding time series is not stationary.

After drawing initial conclusions from two ACF/PACF charts, we re-test the stationarity of Cash holding by two tests including the Augmented Dickey-Fuller Test (ADF test) and Kwiatkowski–Phillips–Schmidt –Shin test (KPSS test). As long as the data series satisfies one of the above two tests, we can conclude that it is a stationary series.

- For the ADF test, there is the following hypothesis:

$$\begin{cases} H_0: The\ time\ serie\ is\ non-stationary \\ H_1: The\ time\ series\ isstationary \end{cases}$$

- For the KPSS test, there is the following hypothesis:

$$\begin{cases} H_0: The\ time\ serie\ is\ stationary \\ H_1: The\ time\ series\ is\ non-stationary \end{cases}$$

- **Code:**

```
# ADF test and KPSS test
adf.test(msn.cash$CashHolding)
kpss.test(msn.cash$CashHolding)
```

Table 8. Cash holding stationary test results

```
        Augmented Dickey-Fuller Test

data:  msn.cash$CashHolding
Dickey-Fuller = -2.7456, Lag order = 3, p-value = 0.2752
alternative hypothesis: stationary
```

```
        KPSS Test for Level Stationarity

data:  msn.cash$CashHolding
KPSS Level = 0.62829, Truncation lag parameter = 3, p-value = 0.02006
```

The author tests whether the cash holding data set of Masan company is stationary or not. The results show that cash holding is non-stationary through two ADF tests with p-value = 0,2752 > 0,05 (accepting H_0) and KPSS test with p-value = 0,02006 < α = 0,05 (rejecting H_0) in the table 8. Realizing that the cash holding data series is not a stationary series, the author performs the percentage change of the firm's cash holding and tests for stationarity by the ACF/PACF plot, ADF test, and KPSS test.

- **Code:**

```
# Caculate the percent change of cash holding.
msn.cash$diff = diff(msn.cash$CashHolding)
which(msn.cash$CashHolding == 0)
sum(is.na(msn.cash))
lag.cash = msn.cash$CashHolding[-nrow(data)]
msn.cash = na.omit(msn.cash)
msn.cash$pct.change = msn.cash$diff/lag.cash
msn.cash = na.omit(msn.cash)

# The percent change of cash holding stationary check
# ACF/PACF
par(mfrow=c(1,2))
acf(msn.cash$pct.change,main='ACF for percent change of CashHolding',lag.max = 24)
pacf(msn.cash$pct.change,main='PACF for percent change of CashHolding',lag.max = 24)

# ADF test and KPSS test
adf.test(msn.cash$pct.change)
kpss.test(msn.cash$pct.change)
```

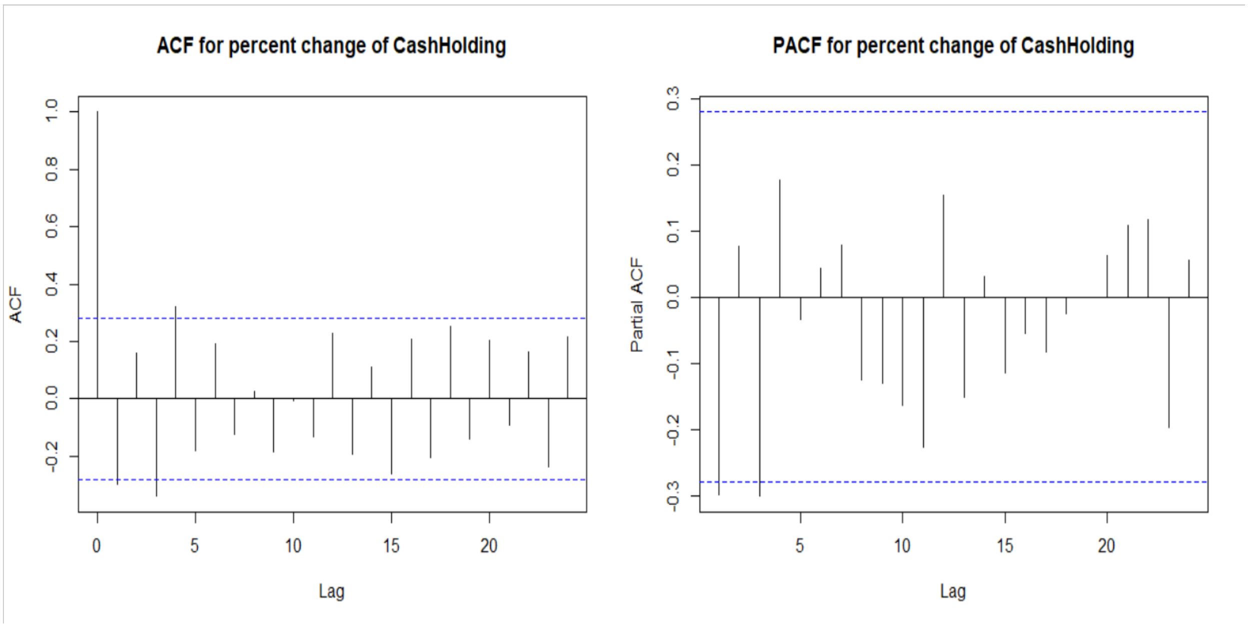

Figure 6. ACF/PACF plot of cash holding percentage change.

Table 9. The percentage change of cash holding stationary test results

```
        Augmented Dickey-Fuller Test

data:  msn.cash$pct.change
Dickey-Fuller = -3.0534, Lag order = 3, p-value = 0.1526
alternative hypothesis: stationary
```

```
        KPSS Test for Level Stationarity

data:  msn.cash$pct.change
KPSS Level = 0.092552, Truncation lag parameter = 3, p-value = 0.1
```

Based on figure 6, the author observes that this time series is a stationary series. At the ACF plot, the ACF dropped sharply to the threshold of statistical significance suddenly, all other lags were statistically significant at $\alpha = 0.05$. In addition, most of the lags at PACF are meant to be within the 95% confidence limit. However, in order to be more objective about the above statement, the author performs stationarity tests on the percentage change of cash holding time series to reinforce the above statement. The results from the table 9 show that the percentage change of cash holding is stationary. Although the stationary condition of the ADF test is not satisfied with p-value = 0,1526 > α = 0,05 (accepting H_0), this time series has p-value = 0,1 > α = 0,05 (accepting H_0). Thus, we can conclude that the percentage change of cash holding is a stationary series.

After determining the stationary degree of the series, the author performs Auto-ARIMA to find the remaining three parameters "p", "d", and "q". The author uses the entire data series to build an ARIMA model for the purpose of predicting the volatility of cash holdings in the 4 quarters of 2022. The results show that the optimal ARIMA model for the above data series is ARIMA(1, 0, 0).

- **Code:**

```
# Building Auto-ARIMA model
auto=auto.arima(msn.cash$pct.change,seasonal=F,trace = T, ic='aic')
auto
coeftest(auto)
```

Table 10. ARIMA model results

```
 ARIMA(2,0,2) with non-zero mean : Inf
 ARIMA(0,0,0) with non-zero mean : 44.3046
 ARIMA(1,0,0) with non-zero mean : 41.42832
 ARIMA(0,0,1) with non-zero mean : 42.10717
 ARIMA(0,0,0) with zero mean     : 43.51266
 ARIMA(2,0,0) with non-zero mean : 42.62395
 ARIMA(1,0,1) with non-zero mean : Inf
 ARIMA(2,0,1) with non-zero mean : Inf
 ARIMA(1,0,0) with zero mean     : 41.96929

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

> auto
Series: msn.cash$pct.change
ARIMA(1,0,0) with non-zero mean

Coefficients:
          ar1    mean
      -0.3166  0.0625
s.e.   0.1394  0.0379

sigma^2 = 0.1255:  log likelihood = -17.71
AIC=41.43   AICc=41.96   BIC=47.1
> coeftest(auto)

z test of coefficients:

           Estimate Std. Error z value Pr(>|z|)
ar1       -0.316568   0.139449 -2.2701  0.02320 *
intercept  0.062525   0.037871  1.6510  0.09874 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The results from the table 10 show that the ar1 and intercept are significant at the significance level is 5%. After finding the optimal ARIMA model, the author makes a cash holding forecast from the first quarter of 2022 to the fourth quarter of 2022 is shown in the Figure 7.

- **Code:**

```
# Prediction
term = 4
fcastauto=forecast(auto,h=term)
fcastauto # predicted values for 4 terms (Q1/2022 - Q4/2022)

# Visualization the prediction
par(mfrow=c(1,1))
plot(fcastauto)
```

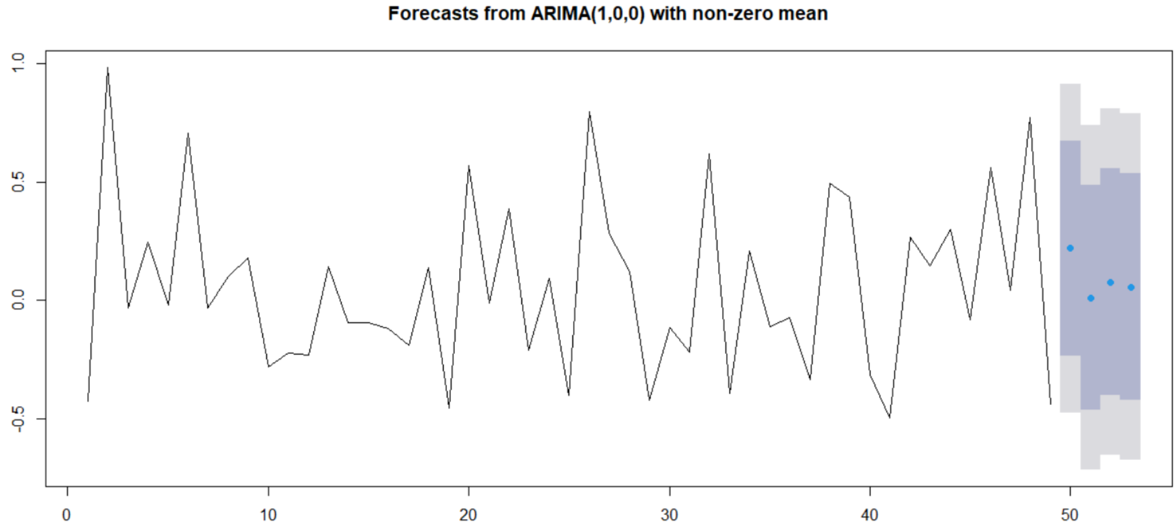

Figure 7. Forecasts result from ARIMA(1,0,0).

- **Code:**

```
> #train set
> accuracy(fcastauto)
                        ME     RMSE       MAE      MPE     MAPE      MASE       ACF1
Training set -0.002635366 0.346975 0.2856068 59.73147 142.1059 0.6115667 0.03824724
>
```

At the train set, RMSE index is 0,346975; MAE index is 0,2856068; and MAPE is 142,1059. That mean, the forecast result and the actual data errors from the ARIMA model are serious. Thus, the cash holding variation of Masan company can hardly be explained by the company's own cash holding in the past. Instead, Masan's cash holding is influenced by many other factors such as leverage, firm size, capital expenditures, etc as the author demonstrated in model 1 and model 2.

9. How Random forest can be used in this case to predict the variable of interest for the four quarters in 2022

The model is called random because the algorithm of this model choose predictors randomly at the time of training. The model is are called forests because they take the output of multiple trees to make a decision. Random forest outperforms decision trees as a large number of uncorrelated trees operating as a committee will always outperform the individual constituent models. To predict the variable of interest for the four quarters in 2022, the random forest model can be used as follows:

First, we need to divide the data into train set and test set, this helps to evaluate the model's predictive performance more objectively because the test set is a set of data not encountered by the model before, or it can be said that the test set also acts as a new input data sample. The train set is responsible for training the model to find out the predictive principles of the dependent variable while the test set is the set used to evaluate the predictive performance of this model.

The random forest does not use all user-supplied data, the model chooses random initial variables and low-correlation trees to try to build a model. These methods are considered advantageous because random forest does not depend too much on the model, does not need assumptions like traditional models, and can work with both linear and nonlinear datasets. Then, the model will randomly choose a number of samples from training data and grow a decision tree from the bootstrap sample. At each node of the tree, the model randomly selects a number of features. After that, the Random forest model split the node using features that provide the best split according to the objective function. Then, the model repeats the previous steps a number of times. Finally, the random forest model aggregates the prediction by each tree for a new data point to assign the class label by majority vote, which means they pick the group selected by the most number of trees and assign a new data point to that group.

After training and testing the predictive performance of the RF model on the test set, we feed the model a new dataset from first quarter of 2020 to first quarter of 2022. The model will return a series of future predictions for us. Then we can use indicators like RMSE, and MAPE to compare the difference between forecast and actual results.

Referrences

Aftab, U., Javid, A. Y., & Akhter, W. (2018). The determinants of cash holdings around different regions of the world. *Business and Economic Review, 10*(2), 151-181.

Al-Najjar, B. (2013). The financial determinants of corporate cash holdings: Evidence from some emerging markets. *International business review, 22*(1), 77-88.

Arfan, M., Basri, H., Handayani, R., Shabri, M., Fahlevi, H., & Dianah, A. (2017). Determinants of cash holding of listed manufacturing companies in the Indonesian stock exchange. *DLSU Business and Economics Review, 26*(2), 1-12.

Bover, O., & Watson, N. (2005). Are there economies of scale in the demand for money by firms? Some panel data estimates. *Journal of Monetary Economics, 52*(8), 1569-1589.

Ferreira, M. A., & Vilela, A. S. (2004). Why do firms hold cash? Evidence from EMU countries. *European financial management, 10*(2), 295-319.

Gill, A., & Shah, C. (2012). Determinants of corporate cash holdings: Evidence from Canada. *International journal of economics and finance, 4*(1), 70-79.

Gordon, L. A., & Iyengar, R. J. (1996). Return on investment and corporate capital expenditures: Empirical evidence. *Journal of Accounting and Public Policy, 15*(4), 305-325.

Guney, Y., Ozkan, A., & Ozkan, N. (2007). International evidence on the non-linear impact of leverage on corporate cash holdings. *Journal of Multinational financial management, 17*(1), 45-60.

Harford, J. (1999). Corporate cash reserves and acquisitions. *The Journal of Finance, 54*(6), 1969-1997.

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *The American economic review, 76*(2), 323-329.

Lee, S., Koh, Y., & Kang, K. H. (2011). Moderating effect of capital intensity on the relationship between leverage and financial distress in the US restaurant industry. *International Journal of Hospitality Management, 30*(2), 429-438.

Liêm, N. T., Thùy, C. T. M., & Thanh, N. P. (2021). Tác động của quản trị lợi nhuận đến mức nắm giữ tiền mặt của các doanh nghiệp niêm yết tại Việt Nam. *Economics-Law and Management, 5*(3).

Mai, D. (2010). The determinants of cash holdings in UK public and private firms. In: Lancaster University Management School Lancaster.

Mesfin, E. A. (2016). The factors affecting cash holding decisions of manufacturing share companies in Ethiopia. *International Journal of Advanced Research in Management and Social Sciences, 5*(3).

Myers, S. C. (1977). Determinants of corporate borrowing. *Journal of financial economics, 5*(2), 147-175.

Myers, S. C., & Majluf, N. S. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of financial economics, 13*(2), 187-221.

Opler, T., Pinkowitz, L., Stulz, R., & Williamson, R. (1999). The determinants and implications of corporate cash holdings. *Journal of financial economics, 52*(1), 3-46.

Ozkan, A., & Ozkan, N. (2004). Corporate cash holdings: An empirical investigation of UK companies. *Journal of banking & finance, 28*(9), 2103-2134.

Simionescu, M. (2016). The competition between London companies regarding their financial performance. *Journal of Competitiveness, 8*(2).

Thành Trung. (2022). Vượt qua đại dịch COVID-19, Masan đạt doanh thu hơn 77.000 tỷ đồng.

Tran, H. C., & Phan, T. T. D. (2021). Determinants of corporate cash holdings: evidence from Vietnamese materials firms. *Journal of International Economics and Management, 21*(3), 1-21.

Tú, P. T. (2017). CÁC YẾU TỐ TÁC ĐỘNG ĐẾN TỶ LỆ NẮM GIỮ TIỀN MẶT CỦA CÁC CÔNG TY NIÊM YẾT TRÊN THỊ TRƯỜNG CHỨNG KHOÁN THÀNH PHỐ HỒ CHÍ MINH.

APPENDIX

```
##############################    1. Create Datasets     ########################

# Ignore the warning messages
options(warn=-1)

# Import library
library(tidyverse)
library(readxl)
library(readr)
library(ggplot2)   # for plotting
library(forcats)   # for handling factors
library(scales)    # for axis scale formatting
library(xts)       # xts conclude index and core data (DataFrame have a timeseries
columns)
library(tseries)   # adf.test
library(lmtest)    # coeftest
library(stats)     # Box.test
library(forecast)
library(PerformanceAnalytics)

# Import data from excel file.
df = read_excel("K194141745.xlsx")
View(df)

# Checking Na values.
sum(is.na(df))

######################################################    2. Calculated the
control variable    ###################################################

covid_period = c(df$MSN[which(df$MSN == 'Q1/2020'):which(df$MSN ==
'Q1/2022')])

# Calculate control varibles
data = df %>% transmute(Quarter = MSN,
                    CashHolding = round(Cash/TotalAssets,6),
                    Leverage = round(TotalDebt/TotalAssets,6),
                    CAE = round(FixAssets/TotalAssets,6),
                    Firmsizes = log(TotalAssets),
                    Covid19 = ifelse(MSN %in% covid_period,1,0)
)

# Checking Na values.
sum(is.na(data))
```

```
# Descriptive statistics data
summary(data)
View(data)

# Export data to csv
# write.csv(data, file =
"C:/Users/ASUS/Desktop/Study/Year_3/HK_6/R_Program/Final/Code/MSN.csv",row.n
ames=FALSE, na="NA")

#########################################################    3. Descriptive
statistics
##################################################################

#### Period Q4/2009 - Q4/2019
before= data %>% slice(1:which(Quarter == 'Q4/2019'))

# Descriptive statistics before Q1/2020
before_statis = before %>% summarise(variables = c('Cash Holding','Leverage', 'Capital
expenditure','Firmsizes'),
                    obs = nrow(before),
                    min =
c(min(CashHolding),min(Leverage),min(CAE),min(Firmsizes)),
                    mean =
c(mean(CashHolding),mean(Leverage),mean(CAE),mean(Firmsizes)),
                    median =
c(median(CashHolding),median(Leverage),median(CAE),median(Firmsizes)),
                    std = c(sd(CashHolding),sd(Leverage),sd(CAE),sd(Firmsizes)),
                    max = c(max(CashHolding),max(Leverage), max(CAE),
max(Firmsizes))
)
before_statis = data.frame(before_statis)
before_statis

#### Period Q1/2020 - now
after = data %>% slice(which(data$Quarter == 'Q1/2020'): nrow(data))

# Descriptive statistics after Q1/2020
after_statis = after %>% summarise(variables = c('Cash Holding','Leverage', 'Capital
expenditure','Firmsizes'),
                   obs = nrow(after),
                   min =
c(min(CashHolding),min(Leverage),min(CAE),min(Firmsizes)),
                   mean =
c(mean(CashHolding),mean(Leverage),mean(CAE),mean(Firmsizes)),
                   median =
c(median(CashHolding),median(Leverage),median(CAE),median(Firmsizes)),
                   std = c(sd(CashHolding),sd(Leverage),sd(CAE),sd(Firmsizes)),
```

```
                            max = c(max(CashHolding),max(Leverage), max(CAE),
max(Firmsizes))
)
after_statis = data.frame(after_statis)
after_statis

################################################################     4.
Visualization
###########################################################################

# Histogram of cash holding
data %>% ggplot(aes(CashHolding)) +
  geom_histogram(binwidth = 0.03,fill = 'coral', color = 'white') +
  labs(title = "Histogram of Cash Holding",
       x = "Cash Holding",
       y = "Count")+
  scale_x_continuous(labels = comma)+
  scale_y_continuous()


# Box plot of cash holding
boxplot(data$CashHolding,
        main = "Box plot of Masan Cash holding",
        col = "orange",
        border = "brown",
        xlab= "Cash holding",
        horizontal = TRUE
)
############################    5. Perform multiple regression to determine the
significant determinants of the variable of assigned topic.   ##################

# Scatter for checking the linear characteristic in the data sets.
par(mfrow=c(1,3))
plot(CashHolding ~ Leverage, data=data, col = "blue", lwd = 2)
plot(CashHolding ~ CAE, data=data, col = "chartreuse4",lwd = 2)
plot(CashHolding ~ Firmsizes, data=data, col = "red",lwd = 2)


##### 5.1. Multiple Linear Regression with the usual individual variables (model 1)
cash.holding.lm<-lm(CashHolding ~  Leverage + Firmsizes + CAE, data = data)
summary(cash.holding.lm)

# Visualization
par(mfrow=c(2,2))
plot(cash.holding.lm)

# Check normality
```

```
shapiro.test(resid(cash.holding.lm)) # Null hypothesis is normality

# Check homoskedasticity
bptest(cash.holding.lm) #Null hypothesis is homoskedasticity


##### 5.2. Multiple Linear Regression with the usual individual variables and the
interaction
# between Covid-19 dummy variable (model 2)
data$Covid19 = factor(data$Covid19)
cash.holding.lm.covid<-lm(CashHolding ~  Leverage + Firmsizes + CAE + Covid19 ,
data = data)
summary(cash.holding.lm.covid)

# Visualization
par(mfrow=c(2,2))
plot(cash.holding.lm.covid)

# Check normality
shapiro.test(resid(cash.holding.lm.covid))# Null hypothesis is normality

# Check homoskedasticity
bptest(cash.holding.lm.covid) #Null hypothesis is homoskedasticity

##### 5.3. Prediction
pred.model = predict(cash.holding.lm, data[,3:5])

pred.table =data.frame(Quanter = data$Quarter,
                        CashHolding.Real = data$CashHolding,
                       CashHolding.Pred = round(pred.model,6))
pred.table

accuracy(pred.table$CashHolding.Real,pred.table$CashHolding.Pred)

##############################################     6. ARIMA
##############################################

# Check class
class(data$Quarter)

# Convert "character" to "Date" data
date.time = seq(as.Date("2009/12/01"), by = "quarter", length.out = nrow(data))
date.time

# Create a xts Dataframe
msn.cash = xts(data[,2],date.time)
```

```
View(msn.cash)

par(mfrow=c(1,1))
plot(msn.cash$CashHolding)

##### Cash Holding stationary check

# ACF/PACF
par(mfrow=c(1,2))
acf(msn.cash$CashHolding,main='ACF for CashHolding',lag.max = 24)
pacf(msn.cash$CashHolding,main='PACF for CashHolding',lag.max = 24)

# ADF test and KPSS test
adf.test(msn.cash$CashHolding)
kpss.test(msn.cash$CashHolding)


# Caculate the percent change of cash holding.
msn.cash$diff = diff(msn.cash$CashHolding)
lag.cash = msn.cash$CashHolding[-nrow(data)]

index(lag.cash) = seq(as.Date("2010/03/01"), by = "quarter", length.out = nrow(data)-1)
msn.cash$lag = lag.cash

msn.cash = na.omit(msn.cash)
msn.cash$pct.change = msn.cash$diff/lag.cash

##### The percent change of cash holding stationary check

# ACF/PACF
par(mfrow=c(1,2))
acf(msn.cash$pct.change,main='ACF for percent change of CashHolding',lag.max = 24)
pacf(msn.cash$pct.change,main='PACF for percent change of CashHolding',lag.max =
24)

# ADF test and KPSS test
adf.test(msn.cash$pct.change)
kpss.test(msn.cash$pct.change)

# Building Auto-ARIMA model
auto=auto.arima(msn.cash$pct.change,seasonal=F,trace = T, ic='aic')
auto
coeftest(auto)

# Prediction
term = 4
fcastauto=forecast(auto,h=term)
```

```
fcastauto # predicted values for 4 terms (Q1/2022 - Q4/2022)

# Visualization the prediction
par(mfrow=c(1,1))
plot(fcastauto)

#train set
accuracy(fcastauto)

#########################          The End          ########################
```