END-OF-COURSE REPORT

Student: *Nguyen Tuan Hung*

Student ID: *K194141723*

1. Literature Review:

1.1. Cash holding:

Cash holding is the amount of cash and cash equivalents held by the company.

Author Komikos and colleagues (2018) with research paper “Corporate Cash Holdings in the Shipping Industry” believes that cash holding is essential to the company because the firm can eliminate the cost of liquidating assets by using the cash held in investment, etc. A. Herlambang (2019) stated that cash holding demonstrates the company’s ability to fulfill its obligations in a timely manner, so cash held is significant for the company.

1.2. Factors affecting the cash holding of the companies:

According to Kim et al. (1998), high operating cash flow companies regularly tended to use internal cash, which means that they use cash to cover an external debt, this leads to low cash holdings.. Research paper “The Impact Of Cash Holding On Market Performance Of Listed Firms In The Vietnamese Stock Market” by Do Thi Thanh Nhan, Pham Ha, Ngo Kim Thanh, Pavelkova Drahomira (2017) argues that firm size is inversely proportional to the firm's cash holdings. Attaullah Shah (2011) demonstrates that growing firms, large firms, companies with high cash inflows will have larger cash holdings and companies with faster conversion cycles will have smaller cash holdings. Research “Corporate cash holdings: An empirical investigation of UK companies” (2004) shows that firm leverage is inversely proportional to cash holding, firms with higher leverage will have lower cash holdings. Companies with high leverage will have higher interest costs leading to reduced cash holding capacity. Phung Anh Thu and Nguyen Vinh Khuong (2018) show a negative association between return on assets (ROA), operating cash flow and corporate cash holdings

1.3. Choose variables:

The target variable is Cash Holding. This variable can be calculated by dividing cash and equivalents by total assets.

Based on the above researches and the dataset, I will choose three variables which can affect cash holding variable as follows:

- *SIZE*: Firm size. I will calculate the size of the given firms based on their total assets. According to Attaullah Shah (2011), size is the natural log to total assets.
- *ROA*: Return on assets. ROA will be calculated by dividing net income by total assets.
- *OCF*: Operating Cash Flow. OCF will be calculated by dividing cash from operating activities by total assets.

2. **Data collection and input:**

The excel file contains data of Toyota Motor Corporation from the beginning of 2012 to the end of 2021.

First I will import the dataset and rename the columns.

```r
#Import the dataset.
library(tidyverse)
library(readxl)
data = read_excel("K194141751.xlsx", skip = 1)
#Name the columns.
names(data)[1] = 'Company'
names(data)[2] = 'Date'
names(data)[3] = 'Total Assets'
names(data)[10] = 'Cash from Operating Activities'
names(data)[11] = 'Net Income'
names(data)[12] = 'Tangible Assets'
names(data)[13] = 'Intangible Assets'
data = data[2:nrow(data),]
View(data)
```

The raw data has the following form:

Date	Total Assets	Total Debt	Operating Income	Market Value for Company	Cash and Equivalents	Total Equity	Revenue	Cash from Operating Activities	Net Income	Tangible Assets	Intangible Assets	Total Current Assets	Total Current Liabilities
2021-12-31	6.393808e+13	2.512233e+13	7.84370e+11	287094900000	5.770225e+12	2.493955e+13	7.785742e+12	2.641506e+12	7.91739e+11	1.224831e+13	1157947000000	2.166989e+13	2.018250e+13
2021-09-30	6.175881e+13	2.451980e+13	7.49978e+11	291514000000	4.954016e+12	2.451047e+13	7.545741e+12	1.872280e+12	6.28851e+11	1.215210e+13	1118611000000	2.068791e+13	1.977192e+13
2021-06-30	6.165115e+13	2.446490e+13	9.97489e+11	242600000000	5.013426e+12	2.412304e+13	7.985558e+12	9.588580e+11	8.97635e+11	1.194224e+13	1102561000000	2.142449e+13	1.962902e+13
2021-03-31	6.226714e+13	2.505964e+13	6.89826e+11	227516700000	5.100837e+12	2.340455e+13	7.089339e+12	2.727162e+12	7.77190e+11	1.180130e+13	1158634000000	2.277680e+13	2.140047e+13
2020-12-31	5.807032e+13	2.397485e+13	9.87941e+11	212341100000	4.679992e+12	2.180417e+13	8.150037e+12	1.834086e+12	8.18846e+11	1.115549e+13	1057987000000	2.192679e+13	2.061004e+13

Then I will select and calculate necessary variables and assign all those information to a new variable data_final.

```
data_final = data %>%
  arrange(Date) %>%
  transmute(Date,
            `Cash Holding` = `Cash and Equivalents`/`Total Assets`,
            ROA = `Net Income`/`Total Assets`,
            `Operating Cash Flow` = `Cash from Operating Activities`/`Total Assets`,
            Size = log(`Total Assets`))
```

After that I will define the Covid variable. We all know that the Covid - 19 pandemic appeared in early 2020 so the Covid variable will equal to 1 if it belongs to the period after 2019. Then I will format the time and the data frame for easier reading.

```
# Define the Covid variable.
data_final$Covid = ifelse(data_final$Date >= '2020-03-31', 1, 0)
#Format the time for easier reading.
data_final$Time = "0"
count = 0
for (i in 2012:2021){
  for (j in 1:4){
    count = count + 1
    data_final$Time[count] = print(paste('Q',j,'/',i),sep = '')
  }
}
data_final = data_final %>% select(-Date)
data_final = data_final %>% select(Time,everything())
```

After formatting, the data frame will look like this:

	Time	Cash Holding	ROA	Operating Cash Flow	Size	Covid
1	Q 1 / 2012	0.05478457	0.003948783	0.04738627	31.05369	0
2	Q 2 / 2012	0.05755248	0.009668637	0.02339981	31.03321	0
3	Q 3 / 2012	0.05532838	0.008568643	0.04118003	31.03557	0
4	Q 4 / 2012	0.03860738	0.003107065	0.05430301	31.10165	0
5	Q 1 / 2013	0.04842549	0.008848665	0.06908362	31.20008	0
6	Q 2 / 2013	0.04106425	0.015125646	0.02769400	31.24648	0
7	Q 3 / 2013	0.04558047	0.011463182	0.05143156	31.27508	0
8	Q 4 / 2013	0.04549549	0.012968610	0.06730345	31.33277	0

Finally I will split the data frame into two parts (before 2020 and after 2020).

```r
#Split the data into two datasets (before 2020 and after 2020)
before_2020 = data_final[1:32,]
after_2020 = data_final[33:nrow(data_final),]
view(data_final)
view(before_2020)
view(after_2020)
```

3. **Provide descriptive statistics of all the variables for BEFORE and AFTER periods:**

Before 2020:

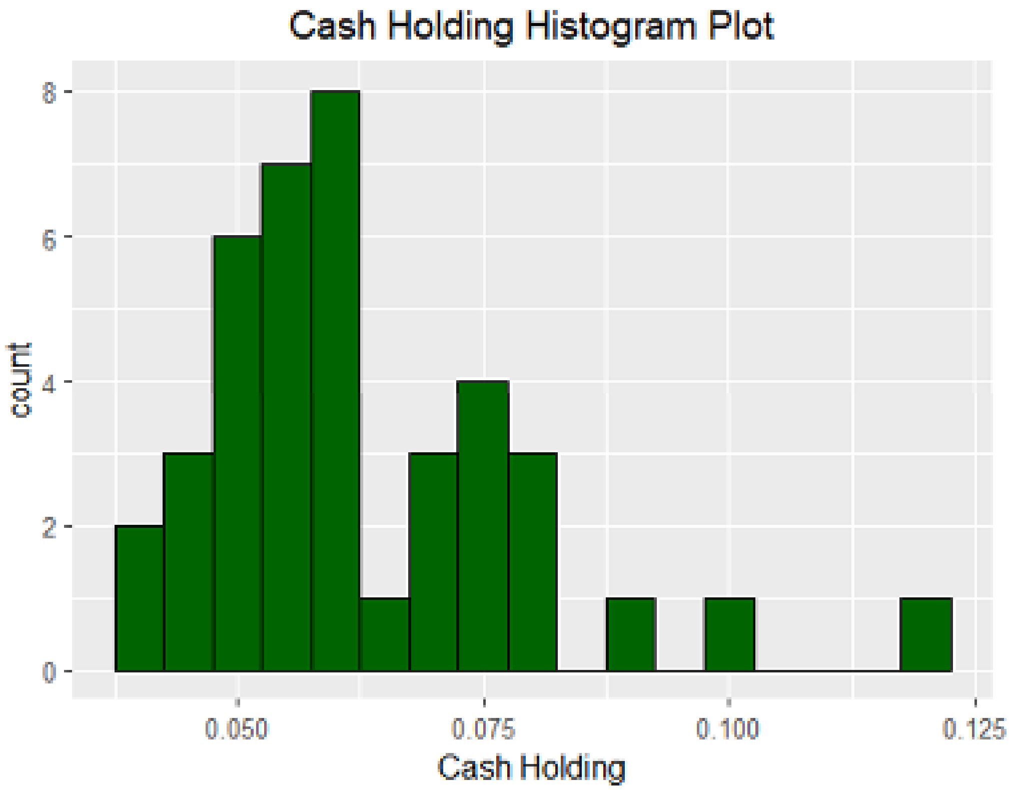
	Cash Holding	ROA	OCF	Size	Covid
min	0.038607378	0.003107065	0.01248904	31.0332105	0
max	0.076182199	0.018389576	0.09405615	31.6163157	0
median	0.055362327	0.010274485	0.04305126	31.5005599	0
mean	0.056456227	0.010475086	0.04728913	31.4237952	0
std.dev	0.009238255	0.003246824	0.02200100	0.1764818	0

After 2020:

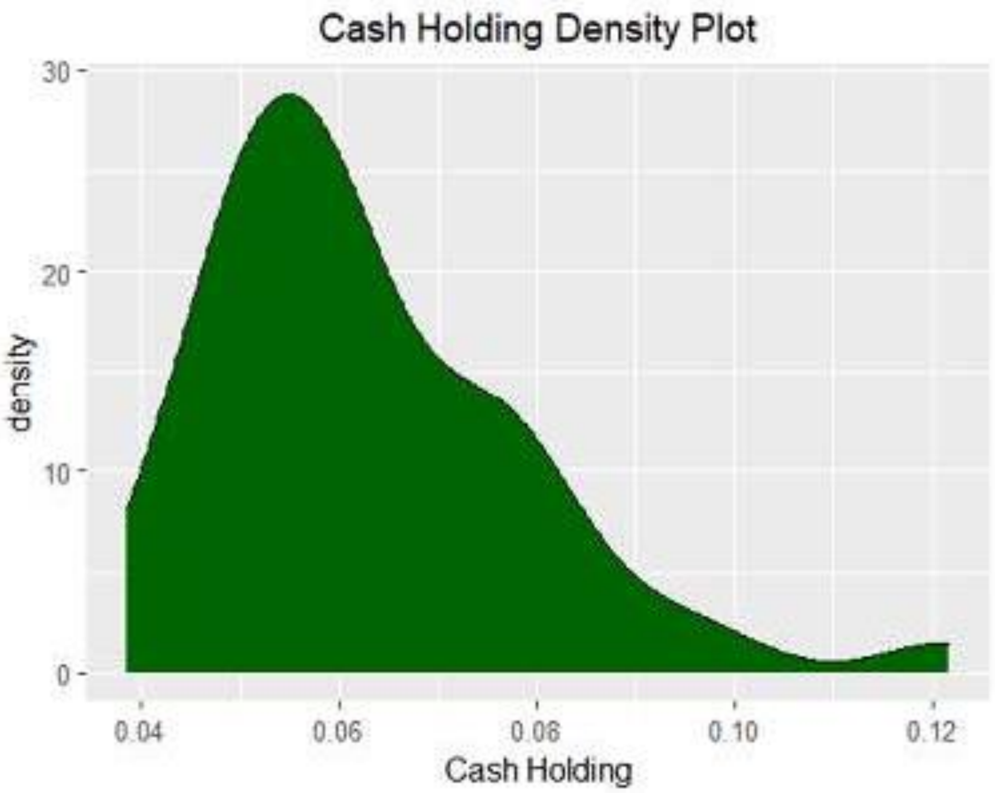
	Cash Holding	ROA	OCF	Size	Covid
min	0.07593609	0.002840716	0.01204030	31.61949324	1
max	0.12163774	0.014563119	0.04443936	31.78391856	1
median	0.08161909	0.011295967	0.03097034	31.72207747	1
mean	0.08832475	0.010166204	0.03026548	31.71051701	1
std.dev	0.01527953	0.004177571	0.01262878	0.06044276	0

ROA tends to decrease when the covid 19 pandemic occurs (the mean ROA after 2020 is smaller than before 2020). ROA has more volatility during pandemic (post-pandemic standard deviation is larger than pre-pandemic). Due to the Covid - 19 pandemic, business activities were quite stalled, so it is understandable that Toyota's return on assets will decline and fluctuate.
OCF also shows a similar trend with ROA when the mean of OCF before pandemic is greater than after pandemic. The only difference is that OCF became less volatile when the covid pandemic appear.
The impact of covid 19 on Cash Holding is the most obvious. Cash Holding tends to increase quite quickly when the pandemic occurs (most of the statistics are almost doubled). Business activities were halted during the pandemic, leading to frozen cash flow in the business, so the cash holding ratio increased.
The covid 19 pandemic has caused certain impacts on businesses and their financial ratios.

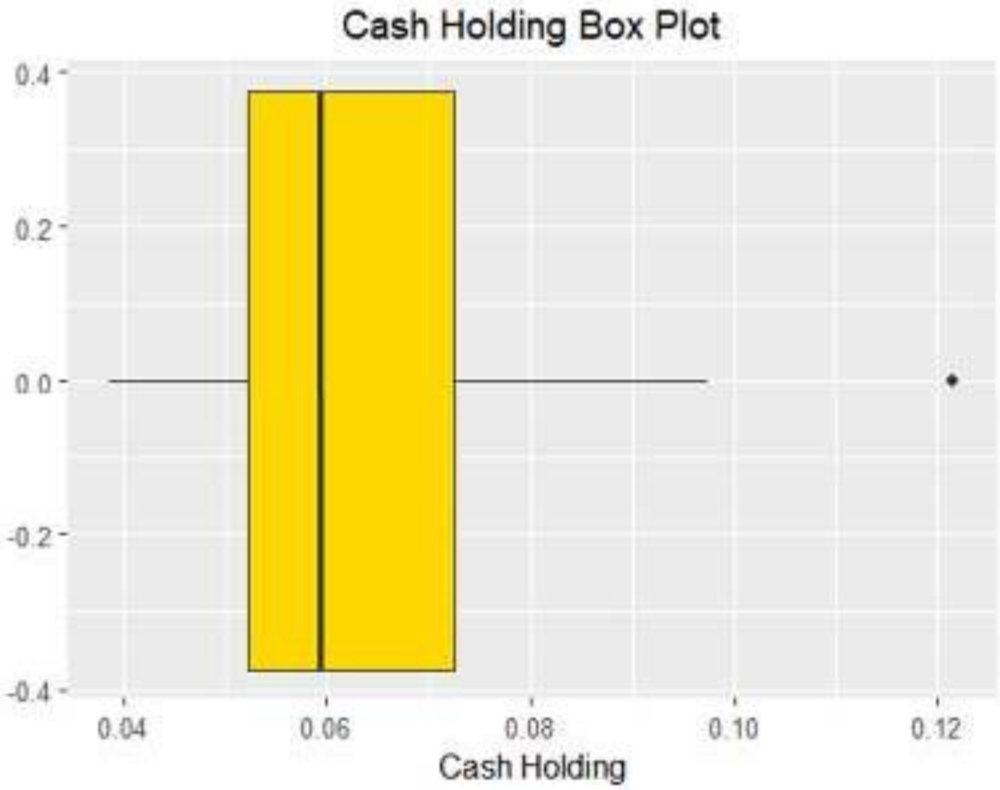
4. **Provide box & whisker plot and histogram of the variable of assigned topic (for the entire period):**



The histogram plot of cash holding is not really pretty and clear so I will draw the density plot for a better look.



The distribution of cash holding is slightly skewed to the left. Most of the time Toyota has a cash holding ratio between 5% and 7%.



The box plot also shows that the cash holding data is skewed to the left. Observations between the first and median quartiles (about 25% of the data) range in value from more than 5% to 6%. There was a quarter where Toyota had an unusually high cash holding ratio of 12%, which is shown by the outlier value on the box plot.

5. **Perform multiple regression to determine the significant determinants of the variable of assigned topic. The significance level is 10%. There are 2 models to run:**

5.1. With the usual individual variables (model 1):

```
Call:
lm(formula = `Cash Holding` ~ ROA + Size + OCF, data = data_final)

Residuals:
       Min         1Q     Median         3Q        Max 
-0.0183378 -0.0045480 -0.0009383  0.0062078  0.0225741 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.709381   0.260273   6.568 1.22e-07 ***
ROA         -2.185982   0.487757  -4.482 7.23e-05 ***
Size         0.057442   0.008286   6.932 4.03e-08 ***
OCF         -0.304611   0.075914  -4.013 0.000291 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.009809 on 36 degrees of freedom
Multiple R-squared:  0.6787,	Adjusted R-squared:  0.6519 
F-statistic: 25.35 on 3 and 36 DF,  p-value: 5.432e-09
```

R-squared:

The R^2 of a model is 0.6787, then approximately 67.87% of the variance for Cash Holding can be explained by ROA, Size and OCF.

Coefficients:

- With no other factors, the cash holding ratio is -1.709381.
- When the variable ROA increases by 1 unit, the Cash Holding variable will decrease by 2.185982 units.
- When the variable Size increases by 1 unit, the Cash Holding variable will increase by 0.057442 units.
- When the variable OCF increases by 1 unit, the Cash Holding variable will decrease by 0.304611 units.

Coefficient p-value:

p-value of all coefficients are less than significance level (10%), so all coefficients are statistically significant.

F-statistic p-value:

p-value of F-statistic is less than significance level (10%), so the model is significant.

5.2. With the usual individual variables and the interaction between Covid-19 dummy variable and the independent variables:

First I will define and create the interaction between Covid-19 variable and the independent variables in the data frame.

```
data_final$`ROA & Covid` = data_final$ROA*data_final$Covid
data_final$`Size & Covid` = data_final$`Size`*data_final$Covid
data_final$`OCF & Covid` = data_final$OCF*data_final$Covid
```

Then I will run the model 2.

```
Call:
lm(formula = `Cash Holding` ~ ROA + Size + OCF + `ROA & Covid` +
    `Size & Covid` + `OCF & Covid`, data = data_final)

Residuals:
      Min        1Q    Median        3Q       Max
-0.014030 -0.004501 -0.001548  0.005162  0.015921

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.9856376  0.2647124  -3.723 0.000732 ***
ROA            -1.1034390  0.4979829  -2.216 0.033713 *
Size            0.0337605  0.0084943   3.974 0.000362 ***
OCF            -0.1529226  0.0695122  -2.200 0.034922 *
`ROA & Covid`  -1.2840151  0.8752398  -1.467 0.151830
`Size & Covid`  0.0014289  0.0003923   3.643 0.000916 ***
`OCF & Covid`  -0.4298304  0.2519843  -1.706 0.097445 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.007864 on 33 degrees of freedom
Multiple R-squared:  0.8107,    Adjusted R-squared:  0.7762
F-statistic: 23.55 on 6 and 33 DF,  p-value: 1.297e-10
```

R-squared:

The R^2 of a model is 0.8107, then approximately 81.07% of the variance for Cash Holding can be explained by ROA, Size, OCF and the interaction between Covid-19 variable and the independent variables.

Coefficients:

- With no other factors, the cash holding ratio is -0.9856376.
- When the variable ROA increases by 1 unit, the Cash Holding variable will decrease by 1.1034390 units.
- When the variable Size increases by 1 unit, the Cash Holding variable will increase by 0.0337605 units.
- When the variable OCF increases by 1 unit, the Cash Holding variable will decrease by 0.1529226 units.
- When Covid 19 happens and ROA increases by 1 unit, the Cash Holding ratio will decrease by 1.2840151 units more.
- When Covid 19 happens and Size increases by 1 unit, the Cash Holding ratio will increase by 0.0014289 units more.

- When Covid 19 happens and OCF increases by 1 unit, the Cash Holding ratio will decrease by 0.4298304 units more.

Coefficient p-value:

p - value of variable ROA & Covid is greater than significance level (10%), so variable ROA & Covid is not statistically significant at 10% level. All other variables are significant at the 10% level (since their p-values are all less than 10%).

F-statistic p-value:

p-value of F-statistic is less than significance level (10%), so the model is significant.

5.3. Predict the value of the variable of assigned topic for all the quarters of the sample using Model:

```
predict_1 = data.frame("Period"=data_final$Time,"Prediction"=predict(model1, data_final))
predict_1
```

The prediction will be shown as follows:

	Period	Prediction
1	Q 1 / 2012	0.05134070
2	Q 2 / 2012	0.04496763
3	Q 3 / 2012	0.04209162
4	Q 4 / 2012	0.05382910
5	Q 1 / 2013	0.04242982
6	Q 2 / 2013	0.04398109
7	Q 3 / 2013	0.04639941
8	Q 4 / 2013	0.04158779
9	Q 1 / 2014	0.04925514
10	Q 2 / 2014	0.05432496
11	Q 3 / 2014	0.05648947
12	Q 4 / 2014	0.05530843
13	Q 1 / 2015	0.05587714
14	Q 2 / 2015	0.06519535
15	Q 3 / 2015	0.06105096
16	Q 4 / 2015	0.05411805
17	Q 1 / 2016	0.05116910
18	Q 2 / 2016	0.06066969
19	Q 3 / 2016	0.06426566
20	Q 4 / 2016	0.06451540
21	Q 1 / 2017	0.06186453
22	Q 2 / 2017	0.06818615
23	Q 3 / 2017	0.07051721
24	Q 4 / 2017	0.04678430

6. **Perform ARIMA model to predict the variable of interest for the 4 quarters in 2022 (if your data allows you to do so):**

Assume that the significance level is 5%.

First thing to do is check the stationary of Cash holding data by using ADF test:

```
        Augmented Dickey-Fuller Test

data:  data_final$`Cash Holding`
Dickey-Fuller = -2.6662, Lag order = 3, p-value = 0.312
alternative hypothesis: stationary
```

The p-value of the test is greater than 0.05 → the Cash holding data is non-stationary.

I will calculate the difference level 1 to remove the trend and make the data stationary and then conduct the ADF test again.

```
data_diff = diff(data_final$`Cash Holding`,differences = 1)
data_diff = na.omit(data_diff)

adf.test(data_diff)
```

```
        Augmented Dickey-Fuller Test

data:  data_diff
Dickey-Fuller = -3.9747, Lag order = 3, p-value = 0.02103
alternative hypothesis: stationary
```

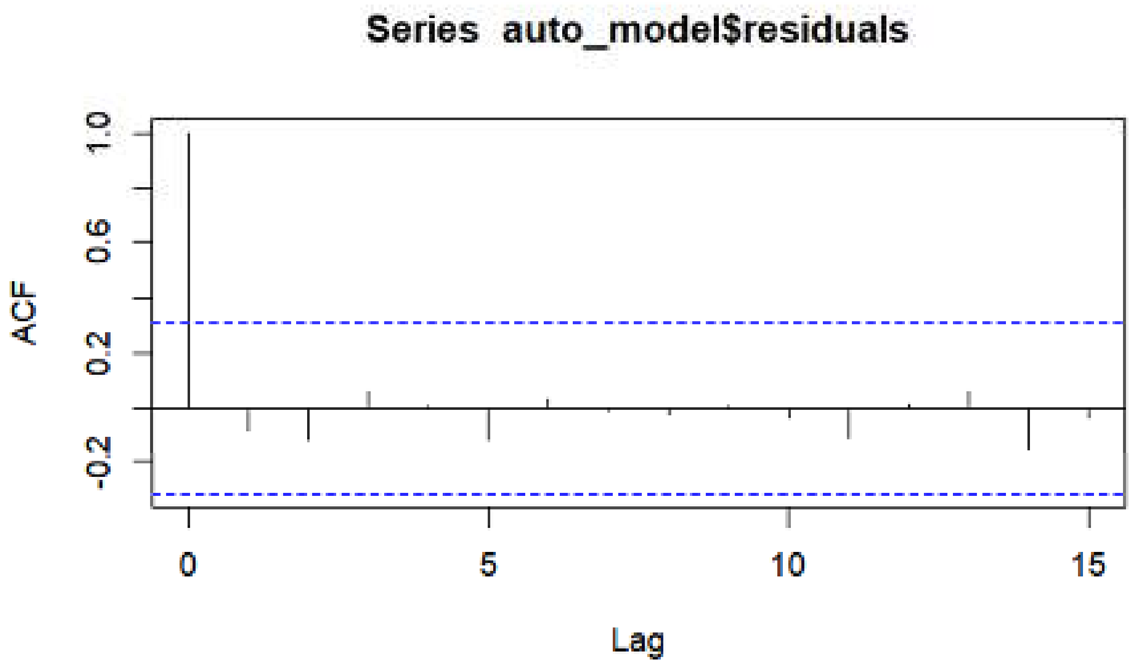
The p-value of the test is lower than 0.05 → the Difference level 1 of cash holding data is stationary.

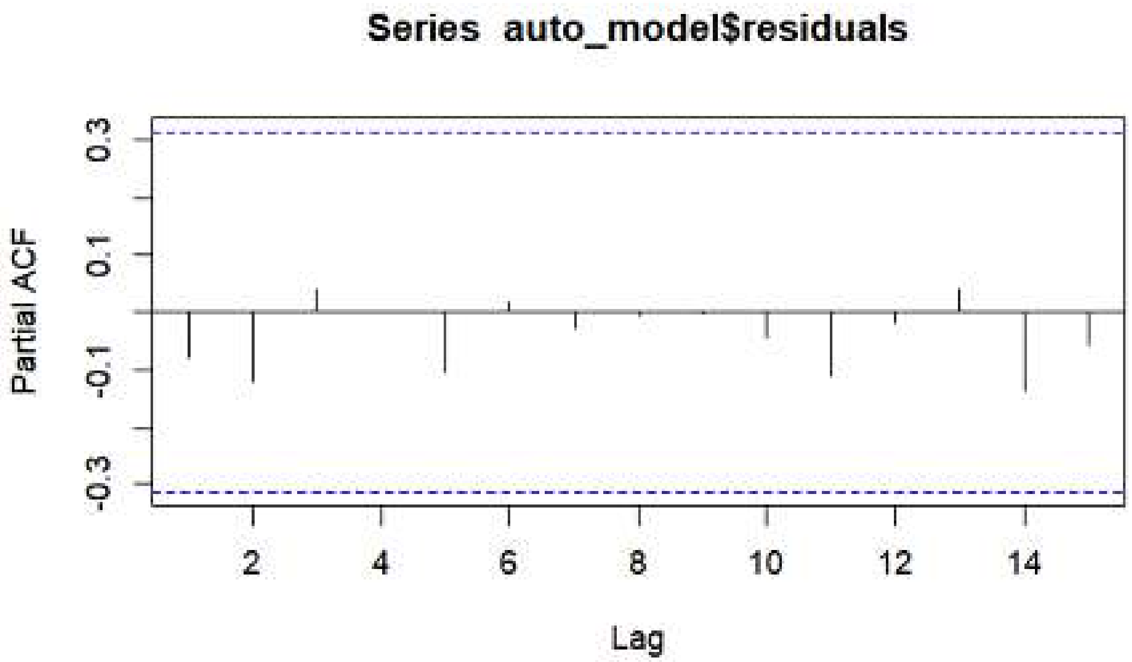
Then I will run the AUTO ARIMA function to find the best model:

```
ARIMA(2,0,2) with non-zero mean : Inf
ARIMA(0,0,0) with non-zero mean : -238.5239
ARIMA(1,0,0) with non-zero mean : -237.7667
ARIMA(0,0,1) with non-zero mean : -241.0656
ARIMA(0,0,0) with zero mean     : -240.2414
ARIMA(1,0,1) with non-zero mean : Inf
ARIMA(0,0,2) with non-zero mean : Inf
ARIMA(1,0,2) with non-zero mean : Inf
ARIMA(0,0,1) with zero mean     : -241.9692
ARIMA(1,0,1) with zero mean     : -241.7548
ARIMA(0,0,2) with zero mean     : -243.7792
ARIMA(1,0,2) with zero mean     : -242.1124
ARIMA(0,0,3) with zero mean     : -242.2617
ARIMA(1,0,3) with zero mean     : -240.3406

Best model: ARIMA(0,0,2) with zero mean
```

Best model for the data is ARIMA(0,0,2)
Now I will conduct the test for the residuals:



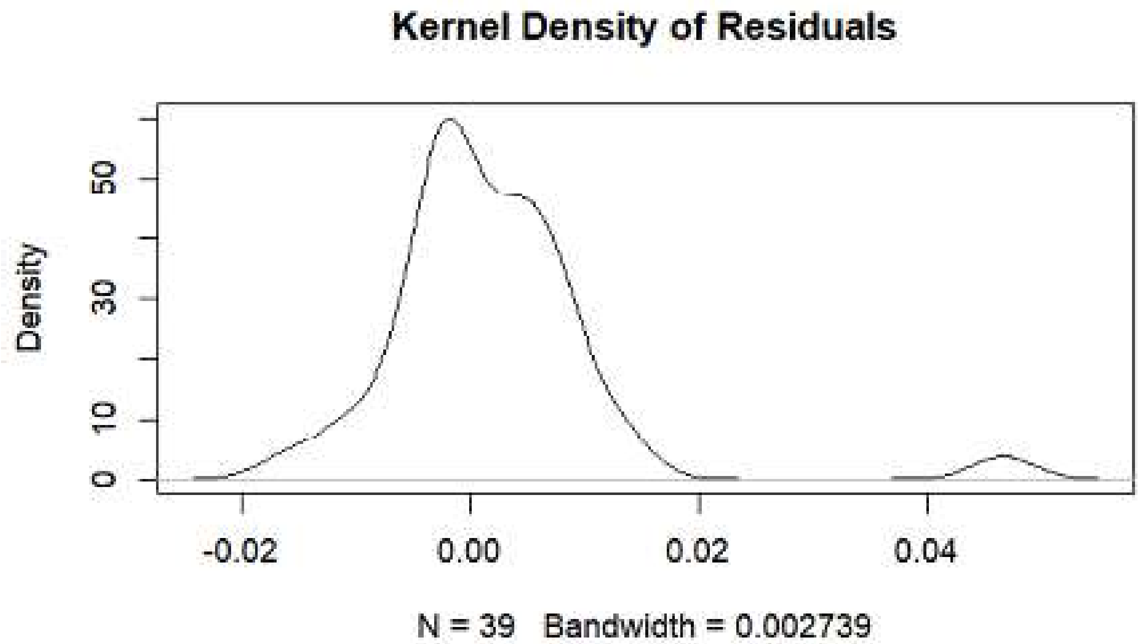


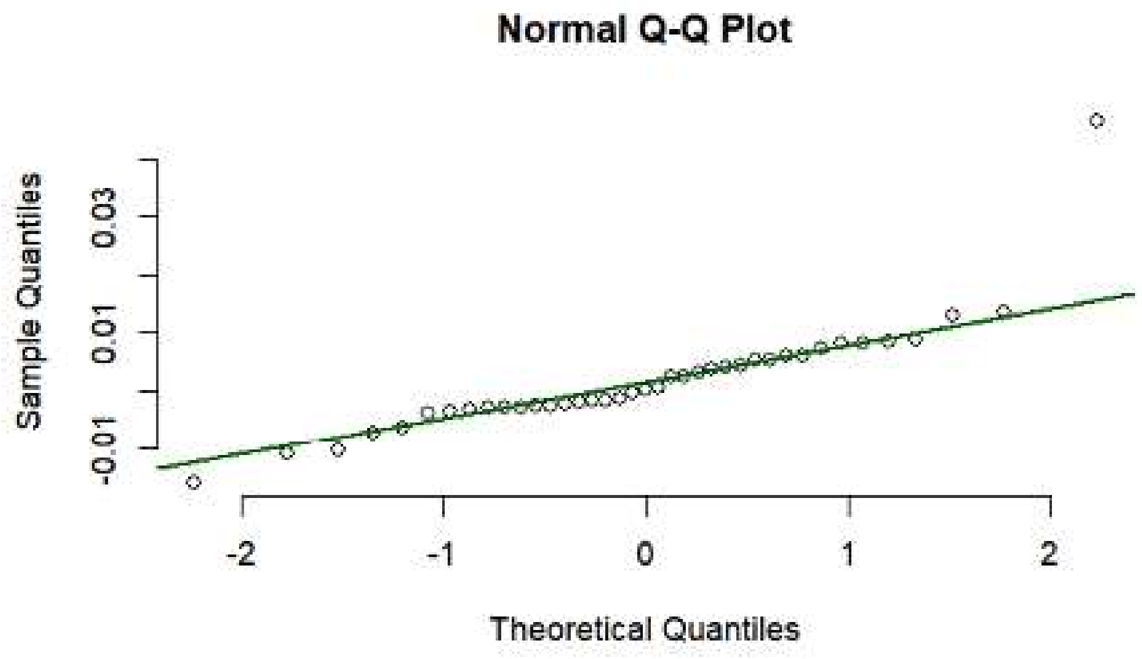
We can see that the residuals are independent because their correlation are 0 at 5% significance level.
Let’s conduct the Box-Ljung test

```
        Box-Ljung test

data:  auto_model$residuals
X-squared = 10.733, df = 20, p-value = 0.9529
```

The p-value is greater than 0.05 → The residuals are independent.
Then I will evaluate the distribution of residuals:

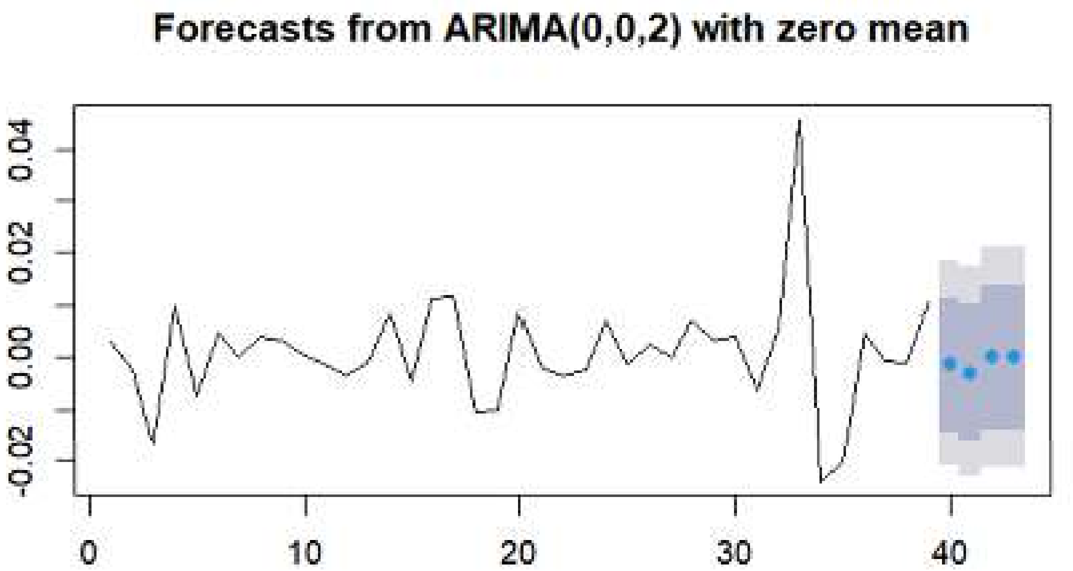




From the density plot and the normal QQ plot we can see that the residuals are normally distributed.
In conclusion, the residuals of the model is IID (independent and identically distributed).
→ The fitted model is good.

Finally, I will do the prediction for 4 quarters in 2022 using this model and plot the predictions:

Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
-0.001426829	-0.01431454	0.01146088	-0.02113688	0.01828323
-0.002974711	-0.01620563	0.01025620	-0.02320965	0.01726023
0.000000000	-0.01390418	0.01390418	-0.02126461	0.02126461
0.000000000	-0.01390418	0.01390418	-0.02126461	0.02126461



7. **Explain in fewer than 150 words how Random forest can be used in this case to predict the variable of interest for the 4 quarters in 2022.**
   When the Random Forest predicts the Cash Holding of period t. It will create decision trees by randomly picking the Cash Holding from the past (period t-n). Each tree will pick a different subset of the Cash Holding from the past (for example tree 1 picks Cash Holding in t-1 and t-2, tree 2 picks Cash Holding in t-2, t-4 and t-5). All the trees will analyze, learn and make their predictions. The final prediction is the average of predictions from all the trees.

REFERENCES

[1] Al-Najjar, B. (2013). The financial determinants of corporate cash holdings: Evidence from some emerging markets. *International Business Review*, *22*(1), 77–88. https://doi.org/10.1016/j.ibusrev.2012.02.004

[2] Attaullah Shah. (2011). The Corporate Cash Holdings: Determinants and implications. *AFRICAN JOURNAL OF BUSINESS MANAGEMENT*, *5*(34). https://doi.org/10.5897/ajbm11.2635

[3] Azmat, Q.-ul-ann. (2014). Firm value and optimal cash level: Evidence from Pakistan. *International Journal of Emerging Markets*, *9*(4), 488–504. https://doi.org/10.1108/ijoem-11-2011-0104

[4] Herlambang, A., R. Murhadi, W., & Cendrati, D. (2019). Factors affecting company's cash holding. *Proceedings of the 16th International Symposium on Management (INSYMA 2019)*. https://doi.org/10.2991/insyma-19.2019.7

[5] Nguyen Thanh. (2019). Optimal cash holding ratio for non-financial firms in Vietnam Stock Exchange Market. *Journal of Risk and Financial Management*, *12*(2), 104. https://doi.org/10.3390/jrfm12020104

[6] Nguyen, T. L., Nguyen, L. N., & Le, T. P. (2016). Firm value, corporate cash holdings and financial constraint: A study from a developing market. *Australian Economic Papers*, *55*(4), 368–385. https://doi.org/10.1111/1467-8454.12082

[7] Ozkan, A., & Ozkan, N. (2004). Corporate cash holdings: An empirical investigation of UK companies. *Journal of Banking & Finance*, *28*(9), 2103–2134. https://doi.org/10.1016/j.jbankfin.2003.08.003

[8] Racic, Z., & Stanisic, N. (2017). Analysis of the determinants of corporate cash holdings: Examples from companies in Serbia. *The European Journal of Applied Economics*, *14*(1), 13–23. https://doi.org/10.5937/ejae14-13574

[9] Yun, J., Ahmad, H., Jebran, K., & Muhammad, S. (2020). Cash holdings and firm performance relationship: Do firm-specific factors matter? *Economic Research-Ekonomska Istraživanja*, *34*(1), 1283–1305. https://doi.org/10.1080/1331677x.2020.1823241

[10] Thu, P. A. (2018). Factors effect on corporate cash holdings of the energy enterprises listed on Vietnam's stock market.

[11] Kim, C.S., Mauer, D.C., Sherman, A.E. (1998), The determinants of corporate liquidity: Theory and evidence. Journal of Financial and Quantitative Analysis, 33(3), 335-359.