NATIONAL ECONOMICS UNIVERSITY

--------*--------**

<u>TOPIC:</u> FORECASTING PRICE AND GROWTH RATE VOLATILITY OF MSN STOCK USING TIME SERIES MODEL AND APPLICATION OF CAPM IN MSN, HVN AND STK STOCKS

Performed by: Nguyen Huynh Duc

Student ID: 11200853

Class: Actuary 62

Ha noi, 2022

Marking sheet

		A	B	C	D	E
1	Understanding of the key issues					
2	Ability to answer the question					
3	The logical structure of your arguments					
4	Evidence of reading					
5	Appropriate referencing					
6	Use of diagrams and theory					

Markers' comments

Advantages:

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………..……………………………

…………………………………………………………………………………

………..…………………………………………………………………………

Disadvantages:

…………………………………………………………………………………

…………………………………………………………………………………

…………………….…………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

……………………..…………………………………………………………

CONTENT

1. Introduction.

The strong re-emergence of the Covid-19 epidemic in 2021 had brought unprecedented challenges and significant impacts on the development of Vietnam's economy in general and businesses in particular, especially three sectors: Consumer, Textile and Aviation. In the report "Covid-19 impacts on Vietnam", Asean Security pointed out that these three industry groups are negatively impacted during the pandemic due to the influence of isolation and blockade policies. However, at the same time, this report also indicates that these industry groups will have positive prospects after the Covid-19 pandemic.

These are probably positive signals for the stock prices of companies in these industries, which clearly reflect the company's business results and development potential. However, two factors that also strongly influence stock prices are market confidence and macroeconomic developments. Therefore, in a market full of uncertainty and fear with concerns about rising inflation and how the State Bank will respond to inflationary pressures, by ordinary inferences and predictions, we cannot say anything about the future price and yield of the shares of these companies, even if it is potential.

To take a closer look and make forecasts related to these stock prices, I select three big companies in each industry group, namely Masan Group (Consumer), Century Synthetic Fiber Corporation (Textiles) and Vietnam Airlines (Aviation) and applied the ARIMA, ARCH, GARCH and CAPM models to provide forecasts regarding these three stocks.

This study is divided into three main parts. Section 2 is the research methodology which provides research methods, theories, and general descriptions of the data used in the paper. Section 3 and 4 cover the application of the ARIMA and ARCH, GARCH models for forecasting the price and volatility of log returns for stocks of Masan Consumer Corporation. Section 5 provides the results and discussion on the dependence of the returns of the three stocks above on the market returns through the application of the CAPM model.

2. Research methodology.

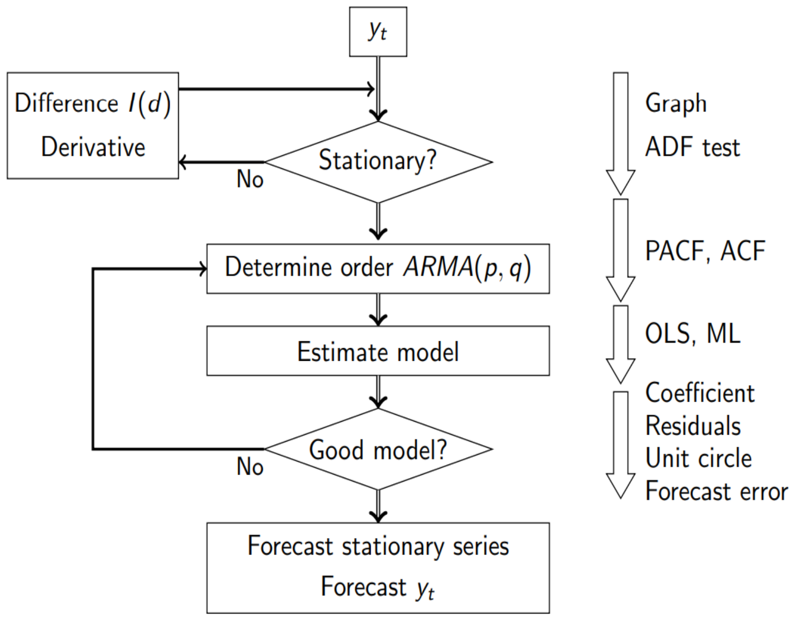
2.1. Methodology and Theoretical background.

2.1.1. ARIMA Model.

To predict MSN stock price, this study applies the ARIMA (Autoregressive Integrated Moving Average) model which was developed by Box and Jenkins. It is also known as the Box Jenkins methodology which consists of some major steps as identifying, estimating, and diagnosing.

Figure 1 depicts the ARIMA modeling and forecasting procedure flow chart.

Figure 1. Box – Jenkins Method's flow chart.



The ARIMA model is based on AR and MA models. While the AR model is used to show that the current observation is dependent on previous observations, the MA model is used to show that the current and previous residuals compose a linear function. The ARIMA (p, d, q) model takes the following form:

$$\Delta d_{y_t} = \mu + \left(\phi_1 \Delta d_{y_{t-1}} + \cdots + \phi_p \Delta d_{y_{t-p}}\right) + (\varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q})$$

ARIMA models have demonstrated their efficient ability to produce short-term predictions. ARIMA models frequently outperform more sophisticated structural models in terms of short-run forecasting ability (Meyler et al., 1998). It is expected that by applying the Box Jenkins methodology and the ARIMA model, the forecast results of the MSN stock price for the next 10 days of August will be close to actual data.

2.1.2. ARCH, GARCH Model.

The ARCH (Autoregressive conditional heteroskedasticity) model is a statistical model used to analyze volatility in time series to forecast future volatility. ARCH modeling shows that periods of high volatility are followed by more high volatility and periods of low volatility are followed by more low volatility. The ARCH(p) model has the following form:

$$\sigma_t^2 = w + \gamma_1 \varepsilon_{t-1}^2 + \cdots + \gamma_p \varepsilon_{t-p}^2 + v_t$$

ARCH has spawned many related models that are also widely used in research and in finance, including GARCH which is based on the supposition that forecasts of time changeable variance depend on the lagged variance of the asset. An unforeseen augment or diminish in returns at time t will generate an increase in the predictable variability in the subsequent period. The basic model GARCH(p,q) can be expressed as

$$\sigma_t^2 = w + (\delta_1\sigma_{t-1}^2 + \cdots + \delta_p\sigma_{t-q}^2) + \gamma_1\varepsilon_{t-1}^2 + \cdots + \gamma_p\varepsilon_{t-p}^2 + \ v_t$$

ARCH, GARCH models are chosen in this study to forecast the unconditional variance and volatility of growth rate series in the first ten days of August.

2.1.3. CAPM - Capital Assets Pricing Model

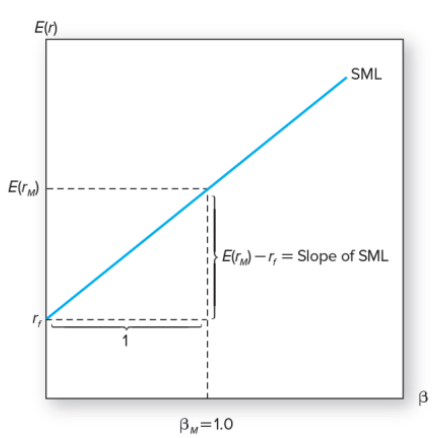
The Capital Assets Pricing Model (CAPM) describes the relationship between systematic risk, expected return and valuation of securities. The main features of this model show that the expected return on the portfolio and standard deviation depend on the structure of the portfolio and that the increase of the risk-free asset in the portfolio affects its expected return.

According to CAPM, the formula for calculating the expected return of an asset, given its risk, is as follows:

$$\mathrm{r}_y = r_f + \beta_y.(\mathrm{r}_M - r_f) + \varepsilon_y \ \ or\ E(\mathrm{r}_y) = r_f + \beta_y.(\mathrm{E}(\mathrm{r}_M) - r_f)$$

where $E(\mathrm{r}_y)\ is$ expected return of asset y, r_f is risk free rate, β_y is beta of asset y, $(\mathrm{E}(\mathrm{r}_M) - r_f)$ is market risk premium.

Figure 2. Security Market Line.



The model shows that the expected return on any portfolio associated with a risk-free asset and the portfolio beta results in the market line (SML-Securities Market Line (Figure 2)) describing a linear relationship between the expected rate of return and the systematic risk measured by beta.

With the aim to see the dependence of the returns of stocks in different industry sectors on the return of the market, the CAPM model will be used with log return series of three stocks: Masan Group (MSN), Century Synthetic Fiber (STK) and Vietnam Airlines (HVN). This study is applied to the Vietnam market, so the VN Index is used for the market index and the 10-year government bond is used as a risk-free asset. Then the daily risk-free rate is calculated by bond yield divided by 365.

CAPM model can be written as

$$\mathrm{ry}_t = r_{ft} + \beta_y.(\mathrm{rVNI}_t - r_{ft}) + \varepsilon y_t \quad or \qquad \mathrm{ery}_t = \beta_y.\mathrm{erVNI}_t + \varepsilon y_t$$

with $\mathrm{ery}_t = \mathrm{ry}_t - r_{ft}$ *and* $\mathrm{erVNI}_t = \mathrm{rVNI}_t - r_{ft}$ are respectively excess return of asset y and VNI at time t.

In turn, the excess return of the three stocks will be run using a regression model according to the excess return of the market portfolio or here the VN index. From that, the differences between the dependences of return of stocks on market portfolio are recognizable.

2.2. Data description.

This study collected stock price data of Masan (MSN), Century Synthetic Fiber (STK) and Vietnam Airlines (HVN) between the 4th of January to the 29th of July which includes 391 observations for estimation and forecast. Besides, to run CAPM model, the VN-index and the 10-year Vietnam Government bond yield over the same period are also taken from website investing.com. With MSN stock, the data is divided into two parts: first part is the in-sample data from 04/01/2021 to 15/07/2022 which includes 381 observations and are used as training set, second part is out of sample data which includes 10 observations covering the period from the 18/07/2022 to the 29/07/2022 and are used as validation set. The data are collected from 2 sources: website of Tan Viet Securities Joint Stock [2] and website investing.com [3].

2.3. Variables and formulas.

Table 1. Compilation of variables.

Variable	Symbol	Unit
Masan Group Stock Price	MSN	VND
Century Synthetic Fiber Stock Price	KDC	VND
Vietnam Airlines Stock Price	HVN	VND
VN Index	VNI	point
Vietnam Government bond yield	r_f	%

Table 2. Compilation of formulas.

Variable	Symbol	Formulas
1^{st} different of y	Δ_y	$= y_t - y_{t-1}$
Growth rate of y	gy_t	$= \frac{\Delta y_t}{y_{t-1}} . 100\%$
Log-return of y	ry_t	$= \ln \left(\frac{y_t}{y_{t-1}}\right). 100\%$
Excess return of y	ery_t	$= ry_t - r_f$

2.4. Descriptive statistics.

Table 3. Descriptive statistics.

Variables	QNS	HVN	STK	VNI	GVB
Observations	391	391	391	391	391
Mean	123,199	24,580	46,519	1,330	2.4425
Standard Deviation	24,202	4,299	12,002	121	0.4567
Kurtosis	-1.2719	-0.1638	-0.9671	-1.1227	0.2621
Skewness	0.1123	-0.2976	-0.4717	-0.1032	1.3423
Minimum	82,000	14,600	20,850	1,023	2.0380
Maximum	172,000	33,800	68,000	1,528	3.8630

Table 3 shows the descriptive measures for the study variables. The results indicate the mean value calculated for Masan stock price is 123,199. This can be inferred that the probability that this series, if stationary, will be stationary around a long run mean or trend stationary but cannot be stationary around zero. However, with the high standard deviation of 24,202, this series might not be stationary. A Unit Root test is required, though, to check the stationary of this series. It seems that from the beginning of 2021 to the middle of 2022, the stock prices are strongly affected by the Covid-19 epidemic and economic factors, so the variances of other stocks and VN-Index are also large. As a risk-free security, unlike stocks, the yield on government bonds is smaller but less volatile with a standard deviation of only about 0.46%.

3. Application of ARIMA Model for Forecasting MSN stock price.

3.1. Unit Root Test.

The first step in the Box Jenkins methodology is checking stationary and converting the non-stationary time series to stationary one (if necessary), as the estimation procedures are available only for stationary series. To assess whether the data come from a stationary process, Dickey–Fuller test (DF test) can be performed for as the Unit root test. There are three types of Unit root test: DF test without constant, DF test with constant and DF test with trend.

The approach to unit root testing implicitly assumes that the time series to be tested can be written as,

$$y_t - y_{t-1} = \beta_0 + (\phi - 1)\, y_{t-1} + \beta_1 t + \varepsilon_t$$

When $\phi = 1$, this reduces to a random walk model, as a result, the series is nonstationary. On the other hand, when $\phi < 1$, the series is stationary. We have the hypothesis pair:

$$\begin{cases} H_0\colon \phi = 1\text{: Unit root, nonstationary.} \\ H_1\colon \phi < 1\text{: Not Unit root, stationary.} \end{cases}$$

Let τ be the t-statistic associated with the y_{t-1} variable. According to Dickey–Fuller Test, if the absolute value of τ is greater than absolute value of 5% critical value then rejects Ho, which mean the series is stationary.

Table 4. *DF Test's Statistics with Critical values for MSN stock price series.*

Without constant		With constant		With trend	
τ	5% Critical value	τ	5% Critical value	τ	5% Critical value
-0.1181	-1.95	-1.5681	-2.87	-0.9101	-3.42

The results of DF Test for the series of MSN stock price have been illustrated in Table 4. From this table, it is noticeable that in all three cases the absolute value of τ is less than the absolute value of the critical value at 5%. It means the Null Hypothesis H_0 cannot be rejected and the MSN variable has a unit root. In other words, the variable is nonstationary, and it is necessary to transform this series to stationary one by applying first difference, growth rate and log return. Table 4 below shows the stationarity of all three series, as the absolute value of test statistics in all cases is greater than the absolute value of 5% critical value.

Table 5. *DF Test Statistics with Critical values for first difference, growth rate and log return of MSN stock price series.*

Without constant		With constant		With trend	
τ	5% Critical value	τ	5% Critical value	τ	5% Critical value
Results for first difference of MSN stock price series					
-18.5038	-1.95	-18.4808	-2.87	-18.5631	-3.42

Results for growth rate of MSN stock price series					
-18.4749	-1.95	-18.4621	-2.87	-18.5475	-3.42
Results for log return of MSN stock price series					
-18.4778	-1.95	-18.4558	-2.87	-18.549	-3.42

To provide a closer view of these three timeseries whether they are stationary around zero, long run mean or trend, Table 6 compares the results of estimated coefficients in three cases (without constant, with constant, with trend). The results given for these three variables are similar. When testing the autoregressive model with trend, the coefficient of the trend is not significant at 10%. This means that all three series are not trending stationary. Furthermore, they are also not stationary around the long run mean greater than zero when the intercept/drift coefficient is also not significant in the autoregressive model with constant.

Table 6. *DF Test estimated coefficients' results.*

		Without constant	With constant	With trend
First difference	Intercept	-	29.6343	463.6518
	Lagged value	- 0. 9507 [***]	- 0.9508 [***]	- 0.9565 [***]
	Trend	-	-	- 2.2832
Growth rate	Intercept	-	0.0667	0.4236
	Lagged value	- 0.9493 **[***]**	- 0.9499 **[***]**	- 0.9557 **[***]**
	Trend	-	-	- 0.00168
Log return	Intercept	-	0.03111	0.4024
	Lagged value	- 0.9494 **[***]**	- 0.9496 **[***]**	- 0.9558 **[***]**
	Trend	-	-	- 0.0019

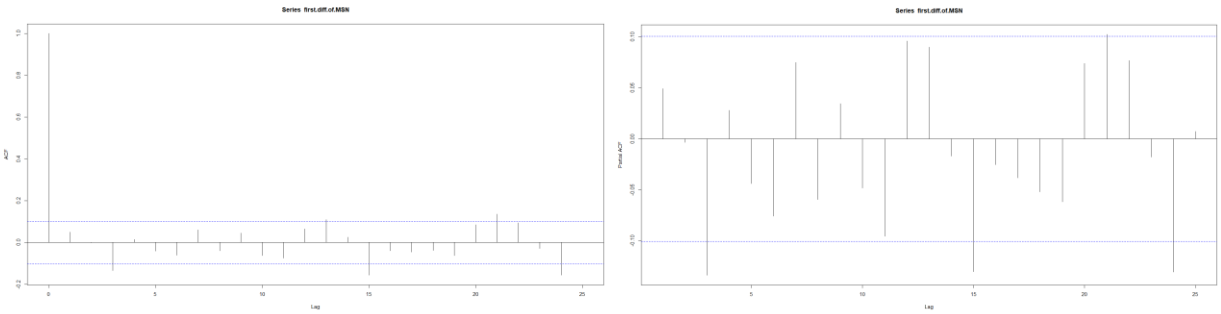
[*], [**], [***]: significant at 10%, 5%, 1%

In short, all three transformations of the MSN stock price series are stationary and ready to run the time series models below.

3.2. Model Identification.

The next step is model identification. The correlogram is checked to determine p (autoregressive parameter) for AR component and q (moving average parameters) for MA component of AR(I)MA Model. The numerical values for these could be obtained by looking for significant autocorrelation and partial autocorrelation coefficients. ACF and PACF may suggest diverse "possible models".

Figure 3. The Autocorrelation and Partial Autocorrelation Function Graphs in First Difference.



The autocorrelation and partial autocorrelation function graphs of the differenced series Δ_y have been illustrated in Figure 3. At the ACF bar indicates significant at lag 3 and nonsignificant after that till lag 15, as a result, it's safe to believe the data came from MA (3). According to the PACF graph, the bar at lag 3 is significant and PACF cut off after lag 3 until lag 15, as a result, AR (3) can be identified, and we get the initial ARIMA (3,1,3) model for MSN series. The next step is comparing among ARIMA (3, 1, 3) model and other ARIMA models such as ARIMA (3,1,2), ARIMA (2,1,3), ARIMA (2,1,1). and ARIMA (1,1,2).

AIC, Log likelihood and the parameter significant are all crucial criteria for selecting model. AIC (Akaike information criterion) is a metric used to compare the fit of different regression models. There is no value for AIC that can be considered "good" or "bad" because we simply use AIC to compare regression models. The model with the lowest AIC offers the best fit. Besides, the log-likelihood value which is, as the term suggests, a way to measure the goodness of fit for models; the higher, the better. However, log-likelihood values should only be used to compare models with the same predictor variables' number because adding more predictor variables to a model will almost always increase the log-likelihood value. Moreover, it is also necessary that all variables are statistically significant.

Table 7. Model results.

	Arima (3,1,3)	Arima (3,1,2)	Arima (2,1,3)	Arima (2,1,1)	Arima (1,1,2)
Number of variables significant at 10%/ Number of variables	5/6	1/5	4/5	2/3	3/3
AIC	7250.79	7255.38	7251.14	7257.09	7256.1
Log likelihood	-3617.39	-3620.69	-3618.57	-3623.54	-3623.05

Table 7 shows the tentative ARIMA (p,1,q) test results for various parameters. In five models above, only model ARIMA (1,1,2) has all variables significant at 10% and its AIC is also the smallest. Moreover, comparing to the model with the same number of variables ARIMA (2,1,1), this model’s log likelihood is larger. Based on these criteria, ARIMA (1,1,2) named model (1) is the optimal model.

Figure 4. The Autocorrelation and Partial Autocorrelation Function Graphs in growth rate.

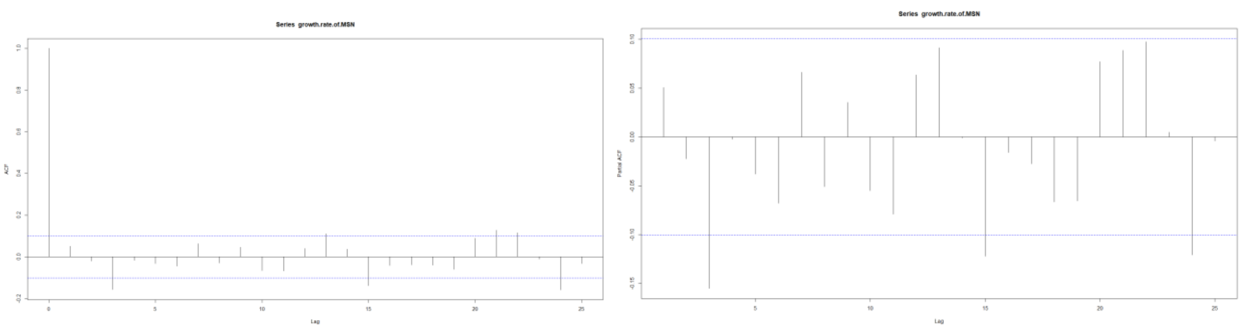
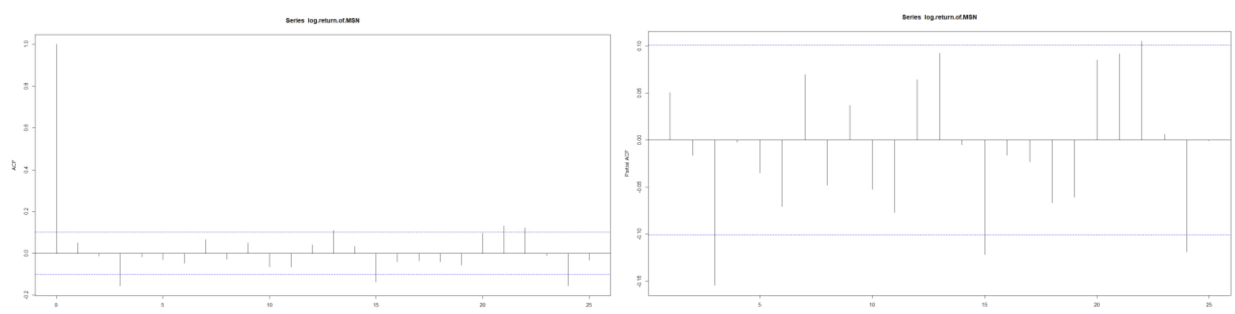


Figure 5. The Autocorrelation and Partial Autocorrelation Function Graphs in log return.



From figure 4 and 5, the Correlation Function (ACF) and Partial Auto Correlation Function (PACF) of the growth rate and log return series are the same and identical to those of the difference series. As a result, it is also reasonable to identify p = 3 and q = 3 for ARMA model of these two series and comparing among the five models like the table 8 and 9 below.

Table 8. Model results.

	Arma (3,3)	Arma (3,2)	Arma (2,3)	Arma (2,1)	Arma (1,2)
Number of variables significant at 10%/ Number of variables	**1/6**	**1/5**	**1/5**	**3/3**	**3/3**
AIC	1842.12	1843.04	1841.69	1843.91	1844.6
Log likelihood	-913.06	-914.52	-913.85	-916.96	-917.3

With the growth rate series, it is noticeable that the two model ARMA (2,1) and ARMA (1,2) are the only two models that have all variables significant at 10%. Comparing between these two models, the model ARMA (2,1) has the lower AIC and the higher log likelihood. Therefore, based on these criteria, ARMA (2,1), called model (2), is the optimal model for growth rate series.

Table 9. Model results.

	Arma (3,3)	Arma (3,2)	Arma (2,3)	Arma (2,1)	Arma (1,2)
Number of variables significant at 10%/ Number of variables	**1/6**	**1/5**	**1/5**	**3/3**	**3/3**
AIC	1846.82	1847.66	1846.35	1848.77	1849.45
Log likelihood	-915.41	-916.83	-916.17	-919.38	-919.72

Meanwhile, running ARMA models for log return series also gives the same result as for growth rate series in which ARMA (2,1) model for log return, called model (3) is chosen to go through the next steps of Diagnostic Tests and Model Selection.

3.3. Diagnostic Tests.

3.3.1. Residual Diagnostics

After having identified and fit a model, checking residual - left over value after fitting a model is still an important part to determine a model's validity. A good forecasting method will yield uncorrelated and zero residuals. It is also useful (but not necessary) for the residuals to be normally distributed. These patterns can be checked through a time plot, ACF plot and of the residuals and Ljung-Box test for the residual series.

Figure 6. The residual checking results for the model (1).

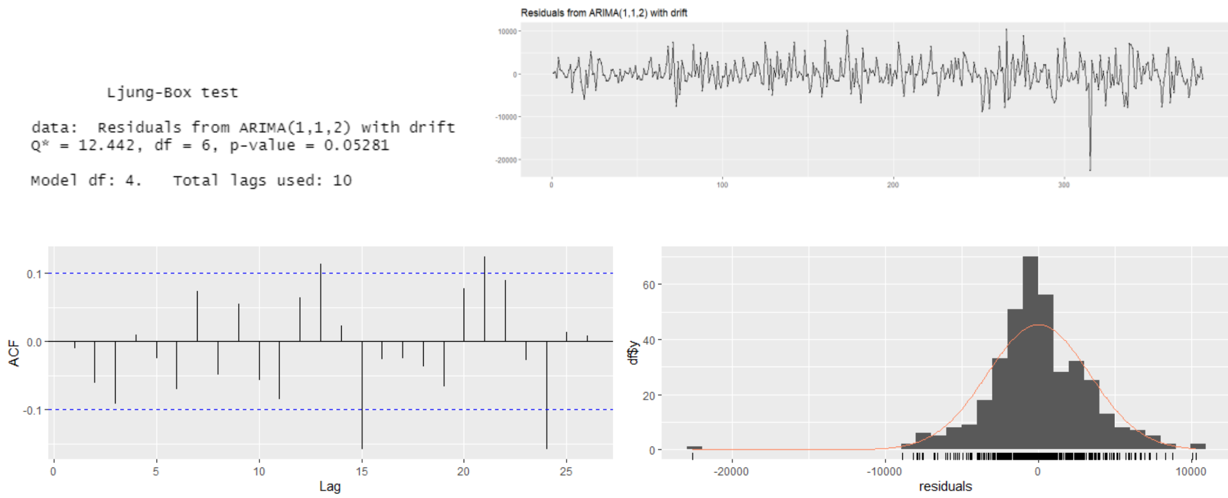


Figure 7. The residual checking results for the model (2).

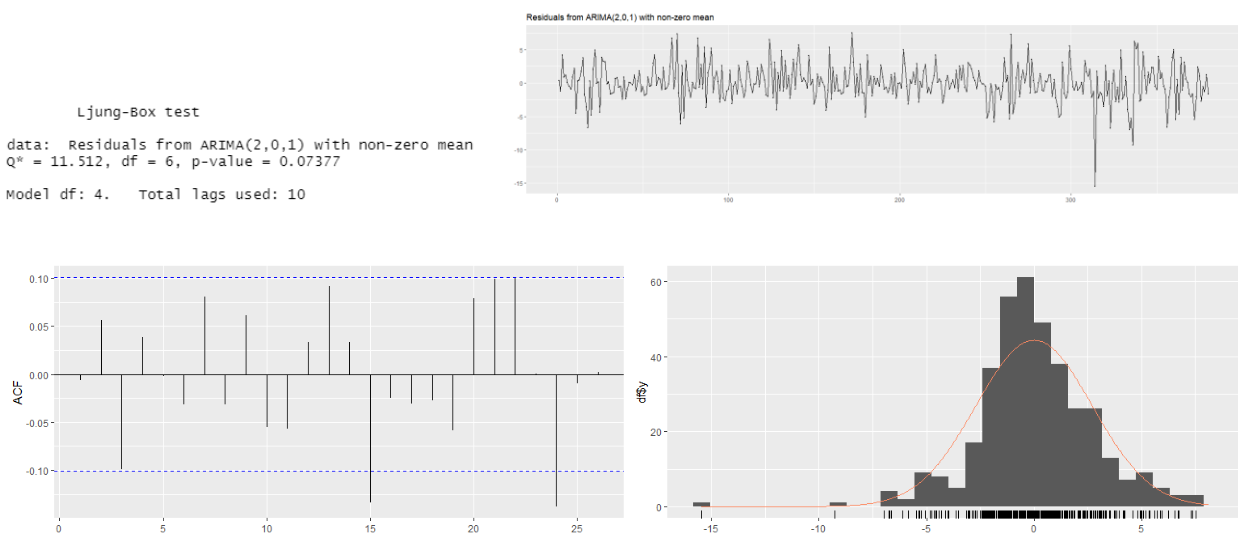
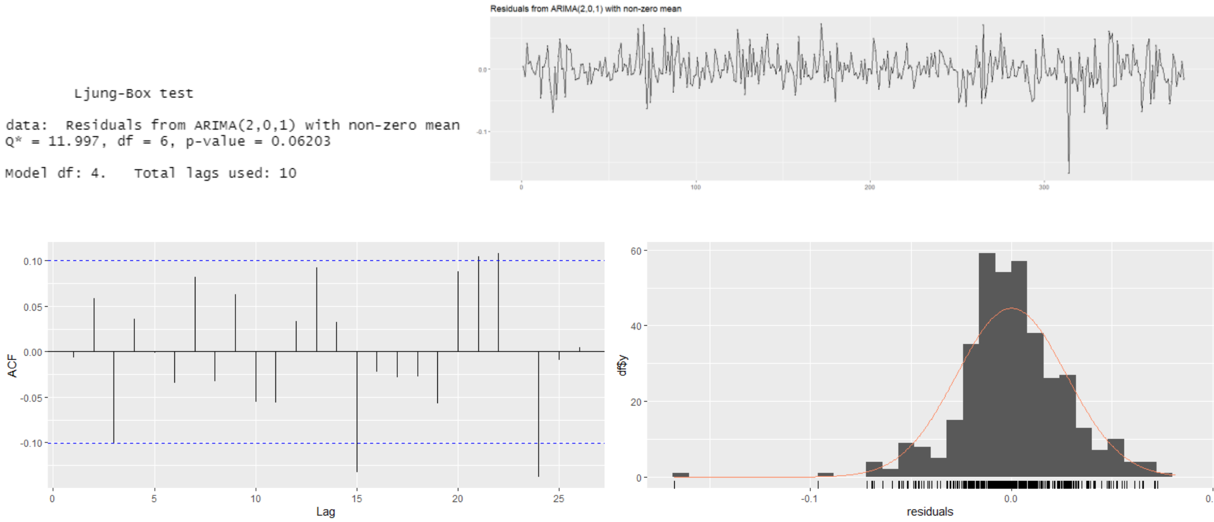


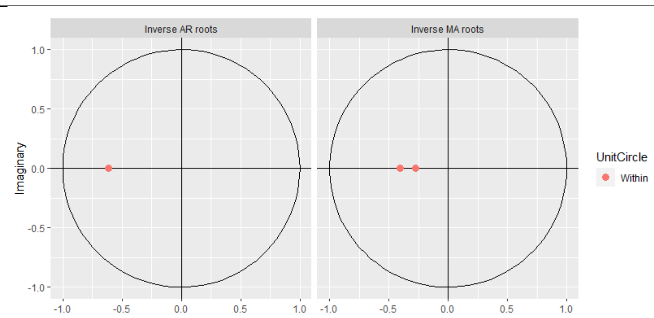
Figure 8. The residual checking results for the model (3).

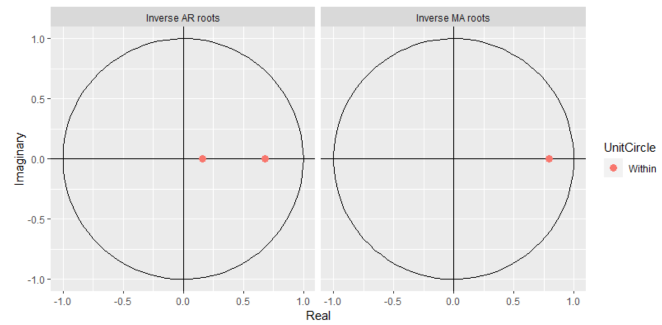
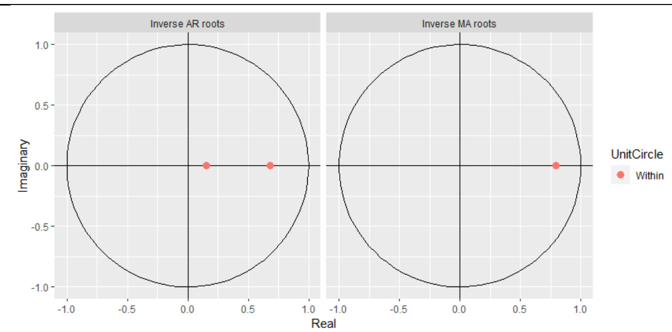


The figures 6, 7, 8 above respectively illustrate diagnostics for residuals from model (1), model (2) and model (3). The histograms for residuals from models show that the residuals seem to be slightly skewed. Although this is not essential for forecasting, it may also affect the coverage probability of the prediction intervals. In addition, the autocorrelation plots in all three cases show significant spikes at high lags (all above 10), but it is not enough for the test statistic of Ljung – Box test to be significant at the 5%. In these cases, the autocorrelations are not particularly large, and at these lags it is unlikely to have any noticeable impact on the forecasts or the prediction intervals. In short, the estimated models (1), (2) and (3) do not violate the assumption of no autocorrelation in the errors, and the forecasts by these three models are still efficient.

3.3.2. ARIMA Stationary Diagnostics.

Table 10. Inverse characteristic roots for 3 models.

Model	Unit Circle
Model (1) - ARIMA (1,1,2)	

<table>
<tr><td>Model (2) - ARMA (2,1)</td><td>
</td></tr>
<tr><td>Model (3) - ARMA (2,1)</td><td>
</td></tr>
</table>

The next step is to check if estimated ARIMA process is stationary or not and check ARIMA process is invertible or not. It is easy to see whether the model is close to invertibility or stationarity by a plot of the roots in relation to the complex unit circle and they should all lie within the unit circle. The results for the three models (1), (2) and (3) have been illustrated in Table 10. As seen from the table , AR and MA roots of these models are located inside the unit circle. So, it means that all three models (1), (2) and (3) processes are stationary and invertible.

3.4. Model Selection.

After ensuring that the residuals are white noise and ARIMA process is stationary and invertible, so we can forecast with three models: ARIMA (1,1,2) model for MSN stock price series, ARMA (2,1) model for growth rate and log return series.

The function forms of these three models are detailly written down in table 11.

Table 11. Equations of Model (1), (2) and (3).

Model (1) - ARIMA (1,1,2):

$$\Delta_{y_t} = 31.9782 + \left(-0.6132\Delta_{y_{t-1}}\right) + (\varepsilon_t + 0.6820\varepsilon_{t-1} + 0.1121\varepsilon_{t-2})$$

Model (2) - ARMA (2,1):

$$\text{gy}_t = 0.0713 + (0.84\text{gy}_{t-1} - 0.1090\text{gy}_{t-2}) + (\varepsilon_t + -0.7972\varepsilon_{t-1})$$

Model (3) - ARMA (2,1):

$$\text{ry}_t = 0.0338 + (0.8381\text{ry}_{t-1} - 0.1065\text{ry}_{t-2}) + (\varepsilon_t + -0.7940\varepsilon_{t-1})$$

As mentioned in data description, the data is broken in to two set: training test including 381 observations from 4/1/2022 to 15/07/2022 and validation set include 10 observations from 10 days left in July. The three selected models are used to forecast the stock price for these ten days and then the static forecast statistical performance measures will be used to evaluate and compare among the models. The table 12 below shows the actual and forecasting data for MSN stock prices from 18/07/2022 to 29/07/2022.

Table 12. Actual and Forecasting Data for MSN Stock prices (18/07/2022 - 29/07/2022)

Date	Actual Data	Model (1)	Model (2)	Model (3)
18/07/2022	102,600	101,311.1	101,330.3	101,277.5
19/07/2022	102,000	101,147.0	101,599.3	101,489.5
20/07/2022	103,000	101,299.2	101,831.2	101,668.5
21/07/2022	105,500	101,257.4	102,016.5	101,805.3
22/07/2022	108,500	101,334.6	102,166.6	101,910.3
25/07/2022	108,600	101,338.9	102,292.2	101,993
26/07/2022	110,000	101,387.9	102,401.1	102,060.4
27/07/2022	109,900	101,409.4	102,498.6	102,117.5
28/07/2022	109,900	101,447.8	102,588.3	102,167.4
29/07/2022	106,100	101,475.9	102,672.8	102,212.5

It is noticeable that three models give three different forecasts from each other and from the actual data. By ordinary speculation, nothing can be predicted about the accuracy of forecasts from models. Therefore, it is necessary to a closer look into the performance measuring criteria like MAE (Mean Absolute Error), MAPE (Mean Absolute Percent Error), RMSE (Root Mean Squared Error). These criteria are comparing in Table 13 below.

Table 13. Actual and Forecasting Data for MSN Stock prices (18/07/2022 - 29/07/2022)

	Model (1)	Model (2)	Model (3)
MAE	5269.071	4470.29	4739.832
RMSE	6053.324	5224.006	5504.742
MAPE	0.0487	0.04123832	0.04374102

It is easy to notice that model (2) - ARMA (2,1) for growth rate series has the smallest performance measures. In other words, model (2) is the most accuracy model, as a result, it is chosen to forecast for the next ten days in August.

3.5. Data Forecasting.

Now let's look at the selected model - model (2) ARMA for growth rate as below:

$$\mathrm{gy}_t = 0.0713 + (0.84\mathrm{gy}_{t-1} - 0.1090\mathrm{gy}_{t-2}) + (\varepsilon_t \pm 0.7972\varepsilon_{t-1})$$

From the model, at 10% level of significant, it is found that growth rate at one time depends on lags of the previous two days and the shock of the previous day by coefficients as shown in the models.

By applying static forecast method, forecast results for Masan Group's share price for the first 10 valuation days of August have been calculated and shown in Table 14. The price forecasts for the next days fluctuate around two values of 108000 and 109000. Hopefully these will be accurate forecasts because in the last days of July the price also fluctuated around these two-price value.

Table 14. Forecasting Data for MSN (01/08/2022 - 12/08/2022)

Date	Forecast Price
01/08/2022	108,193
02/08/2022	109,535
03/08/2022	109,085
04/08/2022	108,978
05/08/2022	109,095
08/08/2022	109,194
09/08/2022	109,268
10/08/2022	109,341
11/08/2022	109,417
12/08/2022	109,494

4. Application of ARCH Model to estimate unconditional variance of volatility.

In addition to the forecast for the stock price of the following days, the forecast of the volatility of the stock's return has received a great attention from investors, regulators, and academicians because it can be used as a measure of risk. In this section 4, ARCH (Autoregressive Conditional Heteroskedasticity) and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models will be used to forecast the growth rate volatility of Masan Group's stock.

4.1. Model Identification.

Before running the ARCH, GARCH models, it is required to check the stationary of the run series and find a best-fitting autoregressive model. These results for this log return time series are illustrated in Section 3. In summary, this log return series is stationary around zero and the ARMA (2,1) model is the most accurate and fit model for this series. Now, the log return series is ready to run the ARCH, GARCH models.

The next step is to identify parameters in p, q in model ARCH(p) and GARCH(p,q). These parameters can be determined by applying Lagrange Multiplier test for ARCH effect introduced by Engle (1982) at different orders.

Table 15. P-value of Arch Test at different orders

	At lag 1	At lag 1, 2	At lag 1, 2, 3
p-value	0.0949	0.0970	0.135

The table 15 above illustrates p-value of ARCH test respectively for the growth rate series in the number of lags 1, 2 and 3, respectively. It is recognizable that the p – value at lag 1 and at lag 1, 2 is smaller than 10%, as a result, at 10% level of significant, the null hypothesis, which is there are no ARCH effects at both lag 1 and 2, is rejected. In addition, testing for ARCH effect at the number of lags 3 is insignificant at 10% level of significant. From these results, it is determined that p, q value can be 1 or 2. ARCH(1), ARCH(2), GARCH(1,1), GARCH(1,2) and GARCH(2,1) are run and compared in the table 16 below.

Table 16. ARCH, GARCH models 'result.

	Arch (1)	Arch (2)	Garch (1,1)	Garch (1,2)	Garch (2,1)
w	5.97247 **[***]**	5.82877 **[***]**	6.5830 **[***]**	6.214 **[***]**	6.217
ε_{t-1}	0.2003 **[**]**	0.17125 **[**]**	0.1489 **[**]**	0.1736 **[**]**	0.1282 **[*]**
ε_{t-2}	-	0.0445	-	-	0.0493
σ_{t-1}	-	-	0.0000	0.0000	0.0000
σ_{t-2}	-	-	-	0.0000	-

[*], [**], [***]: significant at 10%, 5%, 1%

Of the five models, only model ARCH (1) has all coefficients significant at 5% significance level. For other models, the coefficients for ε_{t-2}, σ_{t-1}, σ_{t-2} are not significant. This means the volatility of growth rate series is only affected by shock of the previous day. In short, ARCH (1) model is selected to go through the next steps.

4.2. Diagnostic Tests

Next steps, it is necessary to do Diagnostic Test which is Jarque Bera Test and Box – Ljung Test for the model we have selected. The table 2 below provides the p-value in the Jarque Bera Test for the null hypothesis that the residuals are normally distributed and the Box – Ljung Test for checking autocorrelation within the ARCH model's residuals.

Table 17. Diagnostic Tests for ARCH (1) model.

Jarque Bera Test	Box – Ljung Test
p-value = 0.000	p-value = 0.563

P-value of the Jarque Bera Test for residuals is less than 5% which means the null hypothesis is rejected; in other words, the assumption to run ARCH model that the residuals are normally distributed is not satisfied. This may cause the error in forecasting. However, when applying Box – Ljung Test for squared residuals, the null hypothesis is not rejected meaning that this ARCH model has captured the autocorrelation condition.

4.3. Unconditional Variance Estimation and Volatility Forecasting.

Table 18. Forecasting Volatility for growth rate of MSN (01/08/2022 - 12/08/2022)

Date	Forecast Volatility
01/08/2022	$8.48261. 10^{-4}$
02/08/2022	$7.67154. 10^{-4}$
03/08/2022	$7.50908. 10^{-4}$
04/08/2022	$7.47654. 10^{-4}$
05/08/2022	$7.47002. 10^{-4}$
08/08/2022	$7.46872. 10^{-4}$
09/08/2022	$7.46845. 10^{-4}$
10/08/2022	$7.4684. 10^{-4}$
11/08/2022	$7.46839. 10^{-4}$
12/08/2022	$7.46839. 10^{-4}$

The selected model - ARCH (1) for the growth rate series are written as

$$\sigma_t^2 = 5.97247 + 0.2003\, \varepsilon_{t-1}^2 + \nu_t$$

The equation above implies that the volatility of growth rate depends on shock of the previous day by the coefficients of 0.2003 and w = 5.97247 is long- run volatility parameter. From this we can calculate the unconditional variance by the equations

$$\sigma^2 = \frac{w}{1 - \Sigma\gamma_j} = \frac{5.97247}{1 - 0.2003} = 7.4684$$

Noted that the growth rate variable has been multiplied by 100 and set to %; as a result, the result calculated for unconditional variance has been multiplied by 10^4 times. Therefore, the true value for unconditional variance must be $7.4684. 10^{-4}$.

From the selected model, volatility of the next 10 days in August also can be forecasted and shown in Table 18 above. It can be seen the forecast volatilities fluctuate around moving towards the unconditional variance σ^2.

5. Application of CAPM Model.

In the previous two sections, we went through the price and growth rate volatility forecast for Masan Group Stock. Now, the CAPM model is applied with panel data of log return of three stocks: MSN, STK and HVN to see the dependence of the returns of stocks in different industry sectors on the return of the market portfolio.

The table below shows the form of regression models for excess return of three stocks

Table 19. Regression models.

Model (4) – Regression Model of erMSN on erVNI:

$$\mathrm{erMSN}_t = \beta_0 + \beta_{MSN}.\mathrm{erVNI}_t + \varepsilon y_t$$

Model (5) – Regression Model of erHVN on erVNI:

$$\mathrm{erHVN}_t = \beta_0 + \beta_{HVN}.\mathrm{erVNI}_t + \varepsilon y_t$$

Model (6) – Regression Model of erSTK on erVNI:

$$\mathrm{erSTK}_t = \beta_0 + \beta_{STK}.\mathrm{erVNI}_t + \varepsilon y_t$$

Table 20. Regression models ‘result.

	Model (4)	**Model (5)**	**Model (6)**
Intercept	0.02498	-0.15808	0.21126 **[*]**
erVNI_t	1.11789 **[***]**	0.88199 **[***]**	0.63341 **[***]**
$R - squared$	0.3286	0.2263	0.1104
$Adjusted\ R - squared$	0.3269	0.2243	0.1081
P-value (F-test)	0.0000	0.0000	0.0000

[*], [**], [***]: significant at 10%, 5%, 1%

The results of running regression models (4), (5) and (6) are shown in the table above. It can be seen that at 5% significance level, all three coefficients $\beta_{MSN}, \beta_{HVN}\ and\ \beta_{STK}$ are significant, showing strong dependence on the market. However, all intercepts are statistically insignificant at 5%, this implies the intercept is not necessary in these regression models. These three models need to be rerun without intercept.

Table 21. Regression models ‘result.

	Model (4)	**Model (5)**	**Model (6)**
erVNI_t	1.11804 **[***]**	0.88099 **[***]**	0.6347 **[***]**
$R - squared$	0.3287	0.2251	0.1101
$Adjusted\ R - squared$	0.327	0.2231	0.1078
P-value (F-test)	0.0000	0.0000	0.0000

When running the models again, the coefficients are still significant, but only the Adjusted R - squared of model (4) is increased and the other two models are decreased. It is still reasonable to keep the model with intercept with excess return of STK and HVN stock and using model without intercept for that of MSN stock. As a result, we have the table 6 below show the regression models for excess return of these three stocks:

Table 22. Regression models.

Model (4) – Regression Model of erMSN on erVNI:

$$\mathrm{erMSN}_t = -1.11804\ .\mathrm{erVNI}_t + \varepsilon y_t$$

Model (5) – Regression Model of erHVN on erVNI:

$$\mathrm{erHVN}_t = -0.15808\ +\ 0.88199\,.\mathrm{erVNI}_t + \varepsilon y_t$$

Model (6) – Regression Model of erSTK on erVNI:

$$\mathrm{erSTK}_t = 0.21126\ + 0.63341.\mathrm{erVNI}_t + \varepsilon y_t$$

If $\varepsilon y_t\ is\ removed$, the above models are also the models that show the dependence of the expected return of stocks on the expected return of the market. In short, coefficient β in CAPM for stocks MSN, HVN and STK is$-1.11804, 0.88199, 0.63341$ respectively, in other words, when excess return of market portfolio increases 1 unit, the excess return of

HVN and STK will increase $0.88199, 0.63341$ respectively and the excess return of MSN will decrease -1.11804.

6. Conclusion and limitation.

6.1. Conclusion.

In conclusion, through three sections 4, 5 and 6, the research paper respectively solves the three set objectives:

First is applying ARIMA model to forecast stock price of Masan Group in the 10 first day of August 2022. After applying the Box - Jenkins Method, the research show that ARMA(2,1) for growth rate series is the best fit model for forecasting MSN stock price. This means that to forecast this stock price, the price time series must be transformed to the growth rate series to ensure stationary before running ARIMA and ARMA(2,1) is used to forecast for the growth rate series and then from these forecasts for growth rate, the stock price will be forecast as shown in table 14. Model ARMA(2,1) implies that growth rate at one time depends on lags of the previous two days and the shock of the previous day by coefficients as shown in the models.

Secondly, ARCH, GARCH models are used for estimating unconditional variance and forecast for the volatility of MSN stock's growth rate from 01/08/2022 to 12/08/2022. Through ARCH Test and comparing between suggested models, ARCH(1) is the most fitted models to explain the variation in the growth rate of MSN stock price in which the volatility of growth rate depends on shock of the previous day by the coefficients of 0.2003. From this model, unconditional variance is calculated to be $7.4684.\ 10^{-4}$. In addition, 10 volatilities for the next ten days of August are forecasted and illustrated in table 18.

Finally, by applying Capital Asset Pricing Model and Regression model, models for explaining the dependence of return from three stocks (MSN, HVN, STK) on the return of market portfolio are found out and shown in table 22. It is noticeable that coefficient β in CAPM for stocks MSN, HVN and STK who represent for Consumer, Aviation and Textile is$-1.11804, 0.88199, 0.63341$ respectively, in other words, when excess return of market portfolio increases 1 unit, the excess return of HVN and STK will increase $0.88199, 0.63341$ respectively and the excess return of MSN will decrease -1.11804.

6.2. Limitation.

Although the analysis nearly achieves the three purposes of the study, there are still some limitations:

First, although the ARMA(2,1) model for the growth rate series is the most accurate model, when forecasting for the validation set from 18/07/2022 to 29/07/2022, it can be seen that there is a certain difference with the actual data. This can create error in the forecasts for the upcoming periods. This can be explained by the fact that the selected model is still not the best model. Or this can be explained by the shock at each period is large due to the strong influence of the market after the covid-19 epidemic and the unstable macroeconomy.

Second, when running the ARCH(1) model for the growth rate series, the assumption for a normal distribution of residuals is not satisfied. This may affect the forecast for volatility of the growth rate series on the next ten days of August.

7. References.

1. (Book) Regression Modeling with Actuarial and Financial Applications – EDWARD W. Frees – Cambridge University press

2. (Book) Investments - Bodie−Kane−Marcus • Eleventh Edition.

3. Impact of the covid epidemic on industry groups - Aseansc Research [Internet]

4. Forecasting Irish inflation using ARIMA models by (Meyler et al., 1998)

5. Application of ARIMA Model for Forecasting Agricultural Prices of V. Jadhav, B. V. Chinnappa Reddy, and G. M. Gaddi (2017)

6. Time Analysis with ARIMA - ARCH/ GARCH model in R - Ly Pham

7. Forecasting volatility using GARCH models - Universidade do Minho

8. The application of the ARCH model for the assessment of transport routes in Northern Europe and Southeast Asia - K N Kikkas et al 2019 IOP Conf. S

9. The application of the CAPM model on selected shares on the Croatian capital market - Sandra Odobašić, Marija Tolušić, Zrinka Tolušić:

10. Proceedings of the 2nd International Conference On Systems Engineering and Modeling (ICSEM 2013) - Peng Zhang, Xianghuan Meng

11. Website of Tan Viet Securities Joint Stock [2] and website investing.com [3].