CHAPTER 3

DATA AND METHODOLOGY

In this chapter, data that will be used in this research are mentioned. Then, also explained both variable, independent and dependent variable which will be used in this research. Lastly, the method, the system and steps in how to do the research will be explained as well.

3.1 Data

The analysis period for the achievement of the research purpose used daily market and closing price data for a total of 10 years from July 1, 2010 to June 30, 2020. Daily data sets were used to avoid weekly effects and asynchronous trading in Indonesia, the United States, China and Singapore stock markets. Analysis countries are the daily market and closing prices of the US (S&P500), China (SSEC), Singapore (STI Index), and Indonesia (COMPOSITE (IDX)) stock indices. As analysis data, all stock price data was used at www.capitaliq.com. Table 1 and Table 2 reviews the list of countries, stock markets and symbols and Number of data for each stock market analysis used in this study.

Table 1: Stock Market category

Country	Stock Market	Symbols
USA	Standard and Poor‟s 500	S&P 500
China	Shanghai Stock Exchange Composite Index	SSEC
Singapore	Straits Time Index	STI
Indonesia	COMPOSITE (IDX)	COMPOSITE

Source : Data Process

Table 2 : Amount of data for each stock market surveyed

	S&P 500	SSEC	STI	COMPOSITE(IDX)	Total
Daily data	2,517	2,430	2,518	2,609	10,074

Source : Total Data

In the empirical analysis, The stock price is the stock price when the volatility on the investor minus the growth rate is divided by the dividend.

The time difference problem plays an important role in analyzing the correlation between international stock markets. It was pointed out that the correlation between each country was less than real when the time difference was analyzed without considering it. The time difference between China, Singapore and Indonesia is not much different, so the same date can be compared, but in the case of the United States, coordination is necessary because there is a big time difference. The U.S. has set t-1 as the starting date for stock trading, while China, Singapore and Indonesia have set t as the starting date for stock trading. In addition, data were re-modified to take into account the fact that national holidays in each country were not the same.

3.2 Empirical Model and Research Method

In this paper, we use the VAR model to investigate the impact of information transfer on the stock markets of different countries. To understand the link between information transfer and stock markets in other countries, this study used several empirical models and methods.

3.2.1 VAR model

The Vector Autoregression (VAR) model is a model that can explain the relationship between variables only by model without economic theory. The Vector Autoregression (VAR) model is a multivariate time series model that combines the features of time series analysis and regression analysis to estimate correlations and causal relationships between variables. Therefore, since the past values (trailing values) of the endogenous variables appear in the right-hand term, there is no problem of simultaneity, and an appropriate estimation method is used.

The fact that the disturbances in the error term have no series correlation means that any series correlation can be absorbed by adding the trailing y values up to more lags.

The model Such as:

( ) (2)

Σ (3)

Stock Index (S&P, SSEC, STI, COMPOSITE)

Σ ( )

( ) (4)

Where is the k vector of the endogenous variable, and is the d vector of the exogenous variable. and β are coefficient matrices for model estimation, and may have a temporary correlation with each other, but there is no correlation between one's own past value and other exogenous variables.

Here, daily data on stock index and stock return data of the US, China, Singapore, and Indonesia from July 01, 2010 to June 01, 2020 used to analyze the vector autoregressive (VAR) model. At this time, a third-order trailing vector autoregression model with three past values of endogenous variables with one constant as the only exogenous variable can be expressed as follows.

( ) (5)

(( )

In order to carry out this analysis, the unit root test was first performed to confirm the time series stationary of the data.

3.2.3 unit root test

When a time series has a unit root, it means that the time series is a probabilistic trend. So, the stationary of the time series must be restored by the difference. However, since specific equations for time series cannot be determined experimentally, a certain test is required. Therefore, the unit root test method should be used. Selected when the null hypothesis that there is no serial correlation is rejected for the delay length. (PP) The test is considered

stronger than the ADF test due to its applicability. The PP test is based on the regression shown in Equation (6).

(6)

( ) (7)

The model of AR (1) in expression (6) can be used as the following expression (7). Thus, if the equation (7) does not have normality, it is possible to obtain stable time series data through differentials.

In this study, the ADF (Agulated Dicky–Fuller test) test is applied in consideration of the presence of dichotomy or time series correlation in the error term (Eagle and Granger, 1983)

This study examined the data by performing a stationary test first. Two widely used unit root tests examined the normality of the data. Augmented Dicky and Fuller (1979) tests and Phillips and Perron (1988) tests. The ADF test should be tested by estimating the regression equation (8), (9) and comparing it to the threshold value.(Dicky & Fuller, 1981)

Σ (8)

Σ (9)

Equation (9) is an equation that includes the trend variable T in Equation (8).

The hypothesis of the ADF test is as follows.

If the null hypothesis is rejected, it means stationary and if the null hypothesis is non-rejected and adopted, it means that the unit root exists and is abnormal.

3.2.4 Correlation

Correlation is a relationship that is estimated to be related to each other by calculating a certain number. In other words, it is used to determine which one is correct when there are two or more conflicting theories; it is the relationship between two variables when one of the two variables increases and the other increases or decreases. When it increases, it is called a positive correlation, and when it decreases, it is called a negative correlation. Here, it was used to determine the correlation between the four random variables. The formulas for A,B,C,D are as follows.

$$\frac{(\quad)}{\sqrt{(\quad)(\quad)(\quad)(\quad)}} \qquad (10)$$

( )

3.2.5 Cointegratoin

This paper will use Johansson's test among the tests of cointegration. It is a cointegration test method developed by Johansson in 1991 to test whether there is a stable long-term

equilibrium relationship between the integral series through a hypothesis test on the VAR model. The co-integral test using the VAR model calculates the statistic by analyzing the normal correlation coefficient ( ) between two vectors and the explanatory variable

.

Σ (11)

3.2.6 Lag Length Criteria Test

Information is based on initial measures that could be adopted to select the "lag length" in the proper time sequence. However, using these criteria, this model can get a conclusion that conflicts with regard to "lag length". This paper will explain using an example to estimate the VAR model. First, this paper need to use a large number of fully differential when estimating the VAR model based on the number of observations. Secondly, while reducing the time lag for the included variables, you must test that can estimate the same model. If some paper prefer a more determined LR test remediation models use less delay, some paper must estimate the VAR model.

3.2.7 Granger Causality Test

One of the characteristics of the Granger model is to verify that a variable has precedence and explanatory power over another variable. The VAR model estimated the Granger's causality test, the impulse response function and the Variance decomposition to analyze whether the stock indices in each country help predict each other's price discovery.

The Granger's causality test assumes that information suitable for predicting each variable X and Y is contained only in the time series data of the variables. This can be expressed in the following equation.

Σ Σ (12)

Σ Σ (13)

Σ

Σ

Σ (14)

Test whether the equation (12) has greater explanatory power than the model of the equation (14). For significance test, perform the regression equation (12) and (14) to obtain the sum of the Residual squared and then perform the F test.

( ) (15)
( )

If the estimated F value is greater than the threshold F, then the null hypothesis is rejected, and if it is smaller, the null hypothesis is adopted.

3.2.8 Impulse Response Function and Variance Decomposition

The impulse response function is a function that allows you to confirm the result that all variables in the model respond according to the time result after applying a certain impact to a specific variable in the model. By using this impulse response function, it is possible to analyze the correlation between variables or the ripple effect of changes in policy variables.

The impulse response function is transmitted to all endogenous variables through the dynamic structure of the VAR model. (Budhidharma, 2020)

Variance decomposition is the decomposition of changes in endogenous variables in the component impact of endogenous variables in the VAR. Variance decomposition gives the relative importance of each probability error for the variables within the VAR.

Therefore, through the analysis of variance decomposition of prediction error and the impulse response function, the ripple effect of policy change as well as the contribution of explanatory variables can be expressed as a percentage.