CHAPTER III

RESEARCH METHODOLOGY

3.1 The classification of the research

This research tends to highlight the relationship between the global markets from selected countries. The research based on secondary data and has recourse some theories and formula to sort out the basic queries leading the research. After using different methodology and tools analysis are expected to imply some findings. The finding then can contribute to the building of business and investment knowledge.

3.2 Location of the research

The location of the research takes place in Yogyakarta (Indonesia) where the researcher lives actually at the present. All the data type is secondary data and were extracted from Yahoo Finance (http://finance.yahoo.com), World Economic Forum (http://weforum.org), World Bank (http://worldbank.org). This research requires data across countries among the Asia-Pacific countries. The country that used in this research are Singapore (STI Index), Japan (N225 Index), Hongkong (HSI-Hang Seng Index), Taiwan (TWSE Index), New Zealand (NZ50 Index), Malaysia (KLSE Index), Australia (AXJO Index), South Korea (KOSPI Index), Republic of China (SSE Index), and Thailand (SET Index).

3.3 Data Procedures

The next step after choose the sample is to collecting data. The entire data variable that needed in this study gathered from the national bank of each country, another source is come from the World Bank (http://www.worldbank.org/), yahoo finance (http://finance.yahoo.com/) and also the capital market of each country, for example Indonesia Stock Exchange (http://www.idx.co.id). The other relevant data were collected from reputed books, journals, and articles.

The data used indicate the monthly closing price index. The closing price index used because it expresses the real ending market price at that time. As period of time, the data last for thirteen years from 2003 until 2015.

3.4 Data Analysis Technique

3.4.1 Descriptive Statistic

The descriptive statistic will be performed to analyze the distribution of the stock market, whether the data have normal distribution or not.

The compenent that the descriptive statistic that would be use is Mean, Maximum, Minimum, Range, and Standard Deviation.

$$Mean = \frac{Sum\ of\ all\ population}{number\ of\ values\ in\ the\ population}$$

$$s = \sqrt{\frac{\Sigma(\mathrm{X} - \mathrm{X})^2}{\mathrm{n} - 1}}$$

3.4.2 Run Test

Run test is used to detect statistical dependencies in return series. The null hypothesis of test is that the observed series is random. When $|Z|$ value is greater than 1.96, null hypothesis for Run test is rejected at 5 percent level of significance.

By examining how runs behave in a strictly random sequence of observations, one can derive a test of randomness of runs. Now let :

N = total number of observations = $N_1 + N_2$

N_1 = number of + symbols (i.e., + residuals)

N_2 = number of – symbols (i.e., - residuals)

R = number of runs

Then under the null hypothesis that the succesive outcomes are independent and assuming the number of runs is normally distributed with:

Mean : $E(R) = \frac{2N1N2}{N} + 1$ (Gujarati, 2003)

Variance : $\sigma_R^2 = \frac{2N_1N_2(2N_1N_2 - N)}{(N)^2\,(N-1)}$ (Gujarati, 2003)

If the null hypothesis of randomness is sustainable, following the properties of the normal distribution, we should expect that:

Prob [$E(R) - 1.96\sigma_R \leq R \leq + 1.96\sigma_R$] = 0.95 (Gujarati, 2003)

3.4.3 Unit Root Test

Unit root test is used for testing the existence of stationarity in time series, both Dickey-Fuller (ADF) and Phillips-Perron (PP) tests are used for the purpose of testing stationarity of time series.

The simple formula can be written as : (Gujarati, 2003)

$$\Delta Y_t = \delta Y_{t-1} + u_t$$

Where:

Δ : The first difference operator

ΔY_t : Variation of the series $Y_t - Y_{t-1}$

Y_t : Variable of interest

δ: ($\rho - 1$) : a coefficient

t: time trend

u_t : Random or white noise error term

The augmented Dickey-Fuller (ADF) and PP test (Gujarati, 2003)

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum \alpha_1 \Delta Y_{t-1} + \varepsilon_1$$

where ε_1 is a pure white noise error term and where $\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2})$, $\Delta Y_{t-2} = (Y_{t-2} - Y_{t-3})$, etc.

The difference with non-augmented DF lies on the addition of constraint δ and the extension of lag length α_1, to avoid correlation serial term. The process remains the same as the non-augmented DF. ADF allows for higher order autoregressive process.

3.4.4 Pearson Correlation

Pearson correlation is one measure of correlation that is used to measure the strength and direction of the linear relationship of two variables. Two variables are said to be correlated when changing one variable is accompanied by changes in other variables, both in the same direction or the opposite direction. It should be remembered that the value of the correlation coefficient is small (not significant) does not mean that the two variables are not interconnected. Perhaps the only two variables have the relationship is strong, but the value of the correlation coefficient is close to zero, for example in the case of non-linear relationship. Thus, the correlation coefficient only measures the strength of the linear relationship and not on a non-linear relationship. It should be remembered also that the strong linear relationship between variables does not always mean there is causality.

Pearson correlation is used to find short-run relation between the movements of stock markets. It is used to measure the extent of association between stock markets. (Gujarati, 2003)

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X^2)}{N}\right)\left(\sum Y^2 - \frac{\left(\sum Y^2\right)}{N}\right)}}$$

or

$$r = \frac{\sum(x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{\left(n \sum x_i^2 - \left(\sum x_i\right)^2\right)\left(n \sum y_i^2 - \left(\sum y_i\right)^2\right)}}$$

Where r is the coefficient correlation that valued between -1 to +1. N is sum of the data, x is indices of stock market *x*, and y is indices of stock market *y*.

3.4.5 Engle Granger Test

For causality testing, Engle Granger test is used, which identifies whether one series has significant explanatory power for another series. It is used to find out short-term relationship among sample markets.

The Engle Granger test is aiming to predict one series y_t from another series x_t. Therefore, the second series x_t is said to cause the first y_t. Granger, (1969) among the definitions given on causality, feedback, instantaneous causality and causality lag, Granger enchances from the first utterance the possibility of better prediction based on all avaiable information. The notation $x_t \rightarrow y_t$: to say x_t is causing y_t is clearly generalized specific over-time especially for non-stationary variables. In other words, x_t is said Granger-cause y_t that means x_t can be used to forecast y_t.

If two non-stationary variables are co-integrated, a VAR (Vector Auto Regression) in the first differences is unspecified.

If variables are co-integrated, an ECM (Error Correction Model or Error Correction Mechanism) must be constructed (Gujarati, 2003).

Equation for testing causal relations (Granger, 1969, 1988).

$$x_t = \alpha_0 + \sum_{j=1}^{k} \gamma_{\mathrm{j}} x_{t-j} + \sum_{j=1}^{k} \beta_j y_{t-j} + \varepsilon_{xt}$$

$$y_t = \alpha_0 + \sum_{j=1}^{k} \gamma_{\mathrm{j}} x_{t-j} + \sum_{j=1}^{k} \beta_j y_{t-j} + \varepsilon_{xt}$$

Where:

k: a suitably chosen positive integer

γ_j and β_j with j=0,1,2,...., k: parameters and constants

ε_{xt} : disturbance or error terms with zero means and finite variances.

3.4.6 Johansen's Co-integration Test

Differ from Granger bivariate model that analysis only the causality relationship between two series, Johansen's co-integration test is used for pinpointing long-run relationships among stock markets under study. Johansen (1988) co-integration analysis is said to be " a powerful way of analyzing complex interaction of causality and structure among variables in a system (Wong, Agarwal, & Du, 2005). Johansen uses the maximum likelihood forecast method to test the existence of all the co-integrated vectors.

The process of this method rely on the lag used, on one hand the lag used should be higher (default lag=4) in order to trace the number of co-integration equation present within the series. Furthermore, the lag used here serves to better estimate the white noise (error term) (Harris, 1995). The VECM requires that the series are co-integrated and the lag-length is welll determined in its optimum in order to shorten the period elapsed between the series for a better observation, as the lag length is important to avoid autocorrelation concern within the stationary data.

a. Maximum Eigenvalue Test

The maximum eigenvalue test examines whether the largest eigenvalue is zero relative to the alternative that the next largest eigenvalue is zero. The first test is a test whether the rank of the matrix Π is zero. The null hypothesis is that rank(Π) = 0 and the alternative hypothesis is that rank(Π) = 1. For further tests, the null hypothesis is that rank(Π) = 1,2... and the alternative hypothesis is that rank(Π) = 2,3,....

If the rank of the matrix is zero, the largest eigenvalue is zero, there is no cointegration and tests are done. If the largest eigenvalue is nonzero, the rank of the matrix is at least one and there might be more cointegrating vectors.

Equation of Eigenvalue: (Gujarati, 2003)

$$\lambda_{eig}(r, r+1) = -T \ \mathrm{In}(1 - \lambda_{r+1})$$

b. Trace Test

The trace test is a test whether the rank of the matrix Π is r0. The null hypothesis is that rank (Π) = r0. The alternative hypothesis is that r0 < rank(Π) ≤ n, where n is the maximum number of possible cointegrating vectors. For the succeeding test if this null hypothesis is rejected, the next null hypothesis is that rank(Π) = r0 +1 and the alternative hypothesis is that r0 + 1 < rank(Π) ≤ n.

Equation of trace statistic: (Gujarati, 2003)

$$\lambda_{trace} = -T \sum_{s=r+1}^{n} \mathrm{In}(1 - \lambda_s)$$

3.4.7 GARCH (1,1)

GARCH (1, 1) model is used to test the stock markets. This model captures the existence of volatility clustering in stock markets, which is a sign of market inefficiency. The GARCH (1, 1) model is also the simplest model of GARCH. The equation of GARCH (1, 1): (Gujarati, 2003)

$$\sigma_t^2 = \alpha_0 + \alpha_i u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2$$

This model stated that the conditional variance of *u* at time *t* depends not only on the squared error term in previous period, but also on its conditional variance in the previous time period.

This method requires some conditions to be met, among others, first, the data must be stationary, either stationary or stationary in the mean variance. second, the residuals should be white noise means that the residuals should be distributed uniformly. (Box and Jenkins, 1976). In its development, using data in fluctuation in the UK, found that the residuals are not distributed uniformly so that it appears the method Autoregressive Conditional Heteroscedasticity (ARCH) (Engle, 1982). This method is able to overcome heteroscedasticity in time series data. Then, ARCH models refined into Generalized Autoregressive Conditional Heteroscedasticity (GARCH) by adding the effect of the variant on the previous lag (Bollerslev, 1986).