CHAPTER III

DATA AND RESEARCH METHODOLOGY

A. Research Object

This research used quantitative method, so this research using deductive approach. Quantitative method is a method that stems from numerical data to be processed into information. So that the quantitative method is a method that is numeric and statistical analysis and then processed into information (Kuncoro, 2003).

In quantitative research there are two variables that serve as a model, the independent variables and the dependent variable. In this study, there are five variables that will be used i.e. one dependent variable, and four independent variables. The dependent variable used is economic growth denoted as gross domestic product, while the independent variable is foreign direct investment, exports, infrastructure, and inflation.

B. Type of Data

This research used secondary data annually in times series data. The observation period is from 1981 to 2014. The data that used in this study are as follows:

1. The data of Indonesia economic growth using gross domestic product (GDP in current US$ which is collected from World Bank Publications.

2. The data of Indonesia foreign direct investment in this research using the net inflows in current US$ which is collected from World Bank and The Investment Coordinating Board (BKPM).
3. The data of export in Indonesia using the data of export in current US$ which is collected from World Bank publications.
4. The data of Infrastructure is focused on the total length of roads in Indonesia (kilometers) which is collected from Central Bureau of Statistics (BPS) publications.
5. The data of Inflation Rate in Indonesia using the percentage data which is collected from World Bank publications.

C. Data Collection Technique

Data collection technique that is used in this study was a non-participant observer, where researchers only looked at data that is already available without become part of a data system.

D. Operational Data of Variables

The definition of research variables is used to prevent errors in analyzing the data. The definitions of each variable is described as follows:

1. Economic growth Gross Domestic Product (GDP) is the market value of all final goods and services, produced in the economy in a country in a period
2. Foreign direct investment is direct investment that has done by an individual or company of another country, that focus into production or business either

by buying a company or expanding operation of an existing business in that country.

3. Export is the activity of selling and sending goods from the origin country to other countries.
4. Infrastructure is a form of public capital, which formed from the investment made by the government. In this study, infrastructure including roads, bridges, and sewer system.
5. Inflation is an increase in the general price level of commodities and services during a specific time period. Inflation is regarded as a monetary phenomenon due to the impairment of the monetary calculation unit to a commodity.

E. Analysis Method

The analysis method in this research is Error Correction Model (ECM). By using descriptive quantitative approach, error correction model is used to determine the effect of independent variables on the dependent variable in the long term and short term. Short term is usually an economic behavior less than one year. It could be monthly, quarterly, or annually. Yet in this research, the short term could be define as the effect for one to two years. However, in the long term is usually an economic behavior for more than one year. The period is rarely determined by the researchers. This could be periods for more than two years. Especially in this research, the effect might occur during the periods of research (33 years).

The quantitative approach is used by using the econometric model calculations to test the stationary of each variable and see whether a variable is

stationary or not in a model. Then, to know the long-term equilibrium between the variables in the observation will be seen in the co-integration test.

Co-integration analysis was conducted to determine the equilibrium which is achieved in the long term, while the ECM (Error Correction Model) to correct imbalances in the short term (which is possible) toward long-term equilibrium. Therefore, this analysis uses time series data, then the stationary test data must be done first to ensure that the time series data is stationary. If the non-stationary data is still used, the result of regression data will be spurious.

Other than that, stationary test is absolutely necessary to meet the next assumption analyses are the co-integration and ECM. Once the data is stationary ensured, after it stationary at test level or degree of integration, the next test is a test co-integration. The co-integration tests will confirm whether the regression model co-integrated or not. Co-integrated models will show that the model are in equilibrium in the long run.

After co-integration test is done, the next analysis is developing ECM's regression model. This analysis is conducted to correct the imbalances in short term to long term. To simplify and to reduce manual errors, data processing in this analysis uses software tools E-views 4.0. For further explanation, these are the steps to be done in this research:

1. **Classical Assumption Test.**

   On estimate the linear equations by using OLS method (Ordinary Least Square), the assumptions of OLS must be proceeded. If the assumption are not

proceeded, then it will not be able to generate the parameter value becomes BLUE (Best Linear Unbiased Estimator).

Based on Gujarati (2003), BLUE assumptions can be achieved with the following criteria: The expected value of the average error is 0 (zero), fixed variance (homokedasticity), no autocorrelation in the disturbances, the variables that explain is non-stokastik, there is no multicollinearity among the variables that explain, and the assumption is normally distributed.

To determine whether it meets the assumptions BLUE or not, it needs to do some testing such as multicollinearity test, autocorrelation test, heteroscedasticity test and normality to ensure that the data are normally distributed.

After being tested DF is useful for testing residual generated, then if the critical value is less than the probability value it can thus be said that such data has co-integration and vice versa.

a. Autocorrelation Test.

Autocorrelation shows that there is a correlation between the variables observed. If the model has a correlation, the estimated parameters will be biased and its variations will no longer be a minimum and the model becomes inefficient. In this study, to detect autocorrelation in the model, then it uses the Lagrange Multiplier test (LM). If the value Obs* R-squared is less than the value of the table then the model can be said there is no autocorrelation. It also can be seen from the value of obs-

square probability, if the probability value is greater than the value of a selected it means there is no problem of autocorrelation.

b. Normality Test.

Normality test used to determine whether the residual is normal or not. To test whether the data were normally distributed or not, it can be done by using Jarque- Berra test (test J-B).

c. Heteroscedasticity Test.

Heteroscedasticity test aims to test whether the regression model occurred inequality residual variance from one observation to another observation. If the variance and residuals of the observations to other observations remain, then called as homoscedasticity and if the variance is different called as heterocedasticity.

In this study, the test used is a test of White Heteroskedasticity (no cross term) when the independent variable in small quantities, or White Heteroskedasticity (cross term) when the number of independent variables used in many models. If the probability of *obs * R-Squared* is smaller than the critical value; so that, there is heteroskedasticity occurred in the model.

d. Multicollinearity Test.

One of the assumptions of classical linear regression is there is no perfect multicollinearity or there is no linear relationship between the explanatory variables in the regression model. The consequence of multicolinearity, if the significance of variables, the amount of variable

coefficients and constants are not vaild multicolinearity occurs when the yield estimation R squared value is high (more than 0.8), high F value, and the value of t-statistics, almost all of the explanatory variables were not significant.

So, multicollinearity can be detected in various ways, such as the F-statistically significant but the majority of the value of the t-statistic is not significant, correlation between the coefficient value of independent variables is greater than 0.8 means there is multicollinearity problem in the data.

2. Dynamic Test.

a. Stationary Test.

In the data processing, the first thing to do is to test the stationary data. A stationary data is very important in data analyzes in the form of time series. A variable is said to be stationary if the average value and variance constant over time and the value of the covariance between the two time periods only depending on the difference or interval between the two time periods is not the actual time when the covariance is calculated (Gujarati, 2003).

The concept that used to test stationary time series data is a unit root test. If the data is not stationary, it can be said that data has the unit root problem.

In this study it will use Augmented Dickey-Fuller test to determine whether a time series data contains the root unit or non-stationary. To get a view of the unit root test, it can be explained by the following autoregressive models which is estimated by Ordinary Least Square:

$$DX_t = a_0 + a_1 BX_t + \sum_{i=1}^{k} b_i B^i DX_t$$

$$DX_t = a_0 + a_1 T + a_2 BX_t + \sum_{i=1}^{k} d^i B_i DX_i$$

Where, *T* = time trend, *Xt* = variable observed in period *t*. Furthermore, the next step is ADF calculation. The value of ADF is used to test the hypothesis that *a1 = 0* and *c2 = 0* is indicated by the value of t statistics calculated on *BXT* coefficients in the above equation. Number of inaction *k* is determined by *k = n1 / 5*, where *n* = total number of observations. Critical value (table) for the associated test can be seen in fuller, 1976; Guilky dan Schmidt, 1989 (Insukindro, 1999). The observed time series is stationary if the ADF has a value greater than the critical value.

The stationary decision on data is based on a comparison of the statistics obtained from the ADF t value coefficient with statistical critical values of Mackinnon. If the absolute value of the ADF statistic is greater than the critical value, the data Mackinnon stationary and otherwise the data is not stationary.

b. Integration degree test.

If the testing unit root in the time series is not stationary, then the next step is to do the degree of integration. It aims to determine an integration of data on the degree to how the data will be stationary (Basuki & Yuliadi, 2015).

Test the degree of integration is a test aims to see to what degree the observed data stationary. This test is similar to or an extension unit root test, carried out if the observed data was not stationary as recommended by the test unit roots. Common forms of regression are:

$$D2X_t = e_0 + e_t BDX_t + \sum_{i=1}^{k} f_i B^i D2X_t$$

$$D2X_t = g_0 + g_1 T + g_2 BDX_t + \sum_{i=1}^{k} h_i B^i D2X_t$$

Where, *D2Xt = DXT-DXT-1, BDXt = DXT-1,* further testing at the test unit roots. If the first degree of this data is still stationary, then the integration test needs to be continued at the next degree until obtaining a stationary condition.

c. Co-integration Test.

After knowing the data is not stationary, then the next step is to identify whether the data co-integrated. Co-integration test is required to provide an early indication that the model has a long-term relationship in this study.

The co-integration tests conducted to determine the long-term equilibrium between the variables observed. This test was developed based on the perception that the data model though individually not stationary but linear combination between two or more data time series and that will be stationary (Gujarati, 2003).

The co-integration tests are most often used is the Engle-granger (EG), augmented Engle-granger test (AEG) and Co-integration regression test Durbin- Watson (CRDW). To get the value of EG, AEG, and CRDW, the data used must be integrated at the same degree.

Co-integration test results obtained by forming a residual obtained by regressing the independent variable on the way leading independent variables OLS. Based on the regression result, residual has been obtained must be stationary at the level, so that can be said to have co-integration.

d. Error Correction Model.

Error Correction Model is a model that is used to correct an equation among variables that individually are not stationary. By using ECM, a variable that is not stationary will return to its equilibrium value in the long term with the main requirement is the existence of co-integration relationship between the variables of a constituent.

Co-integration is a long-term relationship among variables that are not stationary. With the non-stationary condition, it provides an opportunity for the data that individually are not stationary to generate a linear combination between those variables, so it will make stationary condition.

ECM is a dynamic model that is able to collect more variables in the analysis of the short and long-term economic phenomenon and analyze the empirical consistency with the econometric theory. Additionally, this model can provide solutions for time series variables which is not stationary and the regression is spurious in econometric analysis. In the long term, the Error Correction Model equation can describe as follows:

$$Y = a_0 + a_1\, X_1 + a_2 X_2 + a_3 X_3 + \cdots + a_n X_n + u$$

Short-term equation of ECM are as follows:

$$Y = a_0 + a_1\, \Delta X_1 + a_2 \Delta X_2 + a_3 \Delta X_3 + \cdots + a_n \Delta X_n + a_n + U(-1) + v$$

Based on the explanation above, co-integration concept is a way to see the long-term equilibrium of each variable. A system can be said to have long-run equilibrium, if the short-term movements responses to the long- run imbalance.

Basically, Error Correction Model has an Error Correction Term (ECT) which is ensure the fulfilment of the long run relationship in the

model (Endry, 2008). ECT is obtained from the estimated residual co-integration equation and the equation above with the symbol U (-1).

If the absolute value of the coefficient of Error Correction Term between zero and one and a p-value less than a = 1%, 5%, 10% then ECT may be said to be significant. If the ECT in Error Correction Model is significant then the model can be used to estimate the long-term relationship of a dependent variable. If the ECT is not significant, the possibility of specification error has occurred in the selection of relevant variables, so the balance relationships cannot be predicted on the early theories (Insukindro, 1999).