CHAPTER III

Methodology

3.1 Scope of Research

This research focuses on factors that affect FDI inflows to Indonesia from provincial level, namely market size, infrastructure availability, labor availability and environmental quality. Moreover, in analyzing the determinants of FDI in Indonesia from regional perspective, this research used data over 33 provinces between 2010 and 2014.²

3.2 Data Type and Sources

Data utilized in this research is panel data, that has space as well as time dimension. The space dimensions are 33 provinces in Indonesia while the time dimension is that each unit (province) contains 5 years (2010-2014) of time series of annual data.

As has been elaborated in Chapter II (Conceptual Framework), annual secondary data of FDI inflows, GRDP, population, road length, labor force, and CO₂ emissions utilized in this study were collected from various government publications. Data on FDI inflows was collected from Badan Koordinasi Penanaman Modal (Indonesia Investment Coordinating Board) while data on

² This study only managed to utilize 165 datasets from 33 provinces between 2010 and 2014 due to data availability, particularly in CO₂ emissions.
GRDP, population, road length, and skilled labor force were taken from *Biro Pusat Statistik* (Indonesia Statistic). Moreover, data on CO₂ emissions was collected from the Ministry of Forestry and Environment Republic of Indonesia.

3.3 Data Analysis Method

3.3.1 Panel Data

This paper is a combination of qualitative and quantitative analysis. Quantitative analysis is used to increase the reliability of the data by reducing the possible bias from the researcher’s opinion as well as establishing a commonly accepted basis for drawing a conclusion (Taggart, 1997). Meanwhile, qualitative analysis is used to explain data interpretation and regression.

As mentioned in the preceding paragraph, panel data was used in this study and processed by quantitative method. Hsiao and Klevmarken in Baltagi (2005) stated that there are several benefits using panel data in regression model estimation as follows:

1) Controlling for individual heterogeneity.

2) Panel data give more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency.

3) Panel data are better able to study the dynamics of adjustment.

4) Panel data are better able to identify and measure effects that are simply not detectable in pure cross-section or pure time-series data.

5) Panel data models allows us to construct and test more complicated behavior models than purely cross-section or time-series data.
6) Micro panel data gathered on individuals, firms and households may be more accurately measured than similar variables measured at the macro level.

7) Macro panel data, on the other hand, have a longer time series and unlike the problem of nonstandard distributions typical of unit root tests in time-series analysis.

3.3.2 Econometric Specifications

As argued in Baltagi (2005), when the exogenous variables are highly correlated and/or the fundamental motivation of the study is to investigate the nature of the effect of each one of the specific endogenous variables on the exogenous variable, it may be preferable to run simple regressions separately for each one of the exogenous variables. Hence, the author adopted the similar approach in this study.

However, there is further complication regarding the econometric specification of the model, which is related to the correct specification of the nature of the cross-sectional unit (in this study, a province) specific effects captured by the intercept term ($\beta_0$) which is assumed to be constant over time. Given the assumption that $\beta_0$ is constant over time, there are three distinct possibilities for the values of $\beta_0$ across cross-sectional unit: (A) They are “fixed” and (in statistical sense) “different” from each other; (B) They are randomly drawn from a normal population distributed with 0 mean and constant variance; and (C) They are “fixed” and “common” across the provinces. The models described by cases (A), (B), and
(C) are referred to as Fixed Effects, Random Effects, and Pooled Classical Regression models which are specified below:

1) Fixed Effects Model

\[ y_{it} = \beta_0 + \beta_1 x_{it} + \epsilon_i, \]  

(1)

2) Random Effects Model

\[ y_{it} = \beta_0 + \beta_1 x_{it} + \epsilon_i + u_i, \]  

(2)

where \( \beta_0 \) is a constant term and \( u_i \) is the error (random) component of province specific effect for province \( i \) which is assumed to be distributed normally with 0 mean and constant variance;

\[ u_i \sim N(0, \sigma_u^2), \]

3) Pooled Classical Regression Model

\[ y_{it} = \beta_0 + \beta_1 x_{it} + \epsilon_t. \]  

(3)

The correct specification of the model for panel regressions (for each independent variable separately) involves applying statistical tests known as Hausman, Lagrange Multiplier (LM) and F tests and choosing the appropriate model out of the three models specified above. The empirical methodology in applying these specification tests is briefly as follows: After estimating all three models and the relevant test statistics, Hausman test is applied to make a choice between Fixed Effects and Random Effects model. If Hausman test preferred the Fixed Effects model, then F test is applied to determine whether or not the individual (province specific) effects given by \( \beta_0 \) are common across provinces. In case F test suggests the presence of “common intercepts” across provinces, the correct specification is given by the Pooled Classical Regression model. Otherwise,
the appropriate model is the Fixed Effects model. On the other hand, if the preferred model as a result of Hausman test is the Random Effects model, then LM test is applied to choose the Random Effects and Pooled Classical Regression models.

3.3.3 Hausman Test

The Hausman Test (also called the Hausman specification test) detects endogenous regressors (predictor variables) in a regression model. Endogenous variables have values that are determined by other variables in the system. Having endogenous regressors in a model will cause ordinary least squares estimators to fail, as one of the assumptions of OLS is that there is no correlation between a predictor variable and the error term. Therefore, to decide the best regression method, it is crucial to figure out the endogeneity of predictor variables in the first place. This is what the Hausman test will do.

In panel data analysis, Hausman Test is conducted to determine the appropriate model to be utilized between the Fixed Effects and Random Effects model. The hypothesis used in Hausman Test is:

H0 = The preferred model is Random Effects model,
H1 = The preferred model is Fixed Effects model,

with interpretation of the results as follows:

1) If (Prob > \( x^2 \)) < \( \alpha \), reject the null hypothesis (H0), means that the Fixed Effects model is more appropriate to be used than the Random Effects model.

2) If (Prob > \( x^2 \)) > \( \alpha \), then accept the null hypothesis (H0), means that the Random Effects model is more appropriate to be used than the Fixed Effects model.
3.3.4 Breusch Pagan Lagrange Multiplier (LM) Test

As has been elaborated in the preceding paragraphs, if the preferred model as a result of Hausman test is the Random Effects model, then LM test is applied to choose the Random Effects and Pooled Classical Regression models. The hypothesis used in this test is:

H0 = The preferred model is Pooled Classical Regression model,
H1 = The preferred model is Random Effects model,

with the following criteria:

1) If (Prob > \( x^2 \)) < \( \alpha \), reject the null hypothesis (H0), means that Random Effects model is more appropriate to be used than the Pooled Classical Regression model.

2) If (Prob > \( x^2 \)) > \( \alpha \), then accept the null hypothesis (H0), means that the Pooled Classical Regression model is more appropriate to be used than the Random Effects model.

3.4 Empirical Model

Within the empirical literature, researchers commonly assess the determinants of FDI by regressing that investments on set of potential determinant variables. As shown in the literature review, various scholars have utilized different determinants and, in this study, the author restricts these determinants in the models as follow:
Furthermore, the variables as shown in equation (4) above are elaborated explicitly as follows:

1) FDI is the annual foreign direct investment inflows in million US dollar. This variable is the dependent variable and transformed into natural logarithm in the model. The variable is taken from online database of Indonesia Investment Coordinating Board (BKPM).

2) GRDP is Gross Regional Domestic Product in billion rupiah on the constant price base year 2010. This variable is the independent variable in the model to
quantify the market size in province level. In this study, GRDP lagged by one year (GRDP_{t-1}) is utilized, to measure the feedback effect of market size on capital investment which is known as the acceleration principle. For estimation purposes, this variable is also transformed into natural logarithm. In addition, the variable was collected from online database of Indonesia Statistic (BPS).

3) POP is the number of populations at provincial level. Similar with GRDP, this variable is used as proxy of market size and transformed into natural logarithm form. The variable was obtained from online database of Indonesia Statistic (BPS).

4) ROAD is the total length of national, provincial, and city/regency road in kilometer at provincial level. This variable is employed to measure the effect of infrastructure availability on FDI inflows. This variable was transformed into natural logarithm for estimation purposes and collected from online database of Indonesia Statistic (BPS).

5) LABOR is the number of labor forces at provincial level who attained senior high school education and above. This variable is used as proxy of human capital and transformed into natural logarithm. This variable was also taken from online database of Indonesia Statistic (BPS).

6) CO2 is the number of CO2 emissions in tons at provincial level. This variable is employed to measure the effect of environmental regulation’s laxity on inward FDI. CO2 is chosen this this study since it is considered to be the primary greenhouse gas responsible for global warming. In addition, reliable time series
on CO$_2$ emission in provincial level is readily available at online database of the Ministry of Environment and Forestry Republic of Indonesia.