CHAPTER 4

METHODOLOGY

This chapter will cover several subsections associated with methodology used in this thesis. Overall, there shall be three subsections. First subsection covers shall explain the benchmark of the regression model. The second one covers panel data elucidation. The last one covers fixed effect model.

4.1 Benchmark of the Regression Model

Regression model used in this thesis uses benchmark model developed by HRV (2005)[13]. They just provide a tree for growth diagnostics framework of a nation. Thus, I use proxies for the branches of their diagnostics tree. However, explanatory variables used in this study have been used before in earlier studies[14]. As can be seen from table 4.1 above we cannot find investment variable. Investment variable is not included because I just want to see the impacts of determinants variable of investment to growth. In other words, it is assumed that investment is the main determinant for a nation's economic growth in this thesis, despite its controversies associated with it.

[13] There has been growing abundant research using their framework in analyzing countries' growth. A technical study is done by Jayaraman and Choong (2006) uses cointegration tests to identify growth constraints in Fiji. Among others Bocchi (2008) uses descriptive analysis to identify constraints of growth in the Philippines; Hausmann (2008) tries to identify growth constraints in Brazil.

[14] See appendix in Durlauf, Johnson and Temple (2005) to see a rather complete references of studies conducted before, along with the explanatory variables used in each study.

Table 4.1 Proxy Variables Used in This Thesis

Variables in HRV	Proxy Variables	Description
High Cost of Finance	Access to Sound Money	This variable reflects cost of borrowing because this variable shows inflation and money growth and freedom of owning foreign currency bank account. The higher the magnitude of this variable, the higher the cost of borrowing since inflation erodes real interest rate triggering banks to increases deposit rate and, somewhat necessarily, increases lending rate.
Low Social Returns	Educational Attainment for Secondary Level	This variable somehow reflects quality of human capital in a country since this variable shows how many people in the country who have completed secondary level education.
Government Failures -- Micro Risk	Regulations Quality and Legislation Structure and Security of Property Rights	This variable reflects governance quality. In this case it is assumed that the highest power of authority is held by the government. Thus, when law enforcement and regulations quality suck, this implicitly shows how suck the government is.
Government Failures -- Macro Risk	Government Expenditure	This variable somehow reflects government failures reflected in macroeconomic condition. It is assumed that government expenditure does not affect productivity but

		that distorts private decisions.
Market Failures – Coordination Failures	Degree of Openness	This variable tells about coordination failures between the government and private enterprises somehow. Since this variable shows volume of trade of a country, the higher the degree of openness, the more coordinated a country with its private sector because it implies the government knows what businesspersons need.

Source: Author's Creation

Before moving further to formation of regression equation, it is needed to inform that in addition to proxy variables based on variables in HRV (2005), I include political variable, democracy, used by Barro and Sala-i-Martin (2004). The reason to use this additional variable is to see the development of political economy in Indonesia as it seems that democracy has been so vital in explaining economic development in Indonesia and other Asian countries chosen here. Moreover, the quality of institutions somehow depends on the state of democracy level at the time being.

In accordance with proxy variables used the basic regression equation is as follows.

$$L\log(gdp_{it}) = \alpha_{it} + \beta_1 l\log(edu_{it}) + \beta_2 reg\_qual_{it} + \beta_3 l\log(gov_{it}) + \beta_4 acc\_som_{it} + \beta_5 \log(open_{it}) + \beta_6 leg\_prop_{it} + \beta_7 dem_{it} + \varepsilon_{it} \quad (4.1)$$

,where

$L\log(gdp_{it})$ = logarithm values of GDP per capita for country *i* and time *t*

$L\log(edu_{it})$ = logarithm values of educational attainment of secondary level in country *i* and time *t*;

reg_qual_{it} = regulations quality in country *i* and time *t*;

$Log(gov_{it})$ = logarithm values of government expenditure in country *i* and time *t*;

acc_som_{it} = quality of access to sound money in country *i* and time *t*;

$Log(open_{it})$ = degree of openness in country *i* and time *t*;

leg_prop_{it} = condition of legislations structure and security of property rights in country *i* and time *t*;

dem_{it} = democracy level in country *i* and time *t*.

In addition, in order to adjust to my purpose to diagnose of Indonesian economy, it is needed to analyze other economies as comparators, as suggested by Gooptu and Ianchovichina (2007). They argue that benchmarking the performance of an economy is an essential way to identify factors that cause low private investment level low in a country and, therefore, growth rate. The comparators should consist of countries that share similar characteristics with diagnosed country, but should also include good performers in the region. This is the reason why Singapore, Malaysia, Thailand, Philippines, South Korea, Japan, and China are used as comparators. To avoid redundancy, adjustment of the regression model is displayed in chapter 5.

4.2 Data Structure

4.2.1 Panel Data

In this thesis, the data take form of panel data. A sketched definition of panel data is data constructed out of time series and cross-sectional data. Wooldridge (2003) defines panel data as data that consist of a time series for each cross-sectional member in the data set.

According to Kennedy (2003), there are four main attractive features of panel data.

1. Panel data can be used to deal with heterogeneity in the micro units. Heterogeneity means that these micro units differ from one another in fundamental unmeasured ways and omit this variable causes biased estimation. The ability to deal with this omitted variable problem is sometimes claimed to be the main attribute of panel data.

2. Panel data create more variability, through combining variation across micro units with variation over time, alleviating multicollinearity problems. Thus, more efficient estimation is possible.
3. Panel data can be used to examine issues that cannot be studied using time series or cross-sectional data alone.
4. Panel data allow better analysis of dynamic adjustment, whereas cross-sectional data cannot tell anything about dynamics and time series data need to be very lengthy to provide good estimates of dynamic behavior.

Based on benefits of using panel data explained above the data structure employed in this thesis is panel data. Nevertheless, there are two specific reasons why employing panel data structure is preferable. First, the data are composed of several countries (cross-section) and several years (time series). It would be so tedious to run time-series regression for each country during the same given time period. It would be much simpler and less tedious to regress the data which take form in panel data. Second, this fits the purpose that to see the long-run trends and dynamic behavior of growth over country so that it is possible to make comparisons in diagnostics.

However, according to Baltagi (2005) there are several limitations of panel data that one needs to be aware of.

1. Design and data collection problem
2. Distortions of measurement errors
3. Selectivity problems, which include
    i. Self-selectivity
    ii. Non-response
    iii. Attrition
4. Short time-series dimensions
5. Cross-section dependence

He argues that the main message of limitations mentioned above is that panel data is not panacea and will not solve all the problems of time series and cross section data.

Form of the panel data can be given by equation below.

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it}\ , \quad i = 1,\ldots,N\ ;\ t = 1,\ldots,T \tag{4.2}$$

i denotes the cross-section dimension whereas *t* denotes time series dimension. In terms of completion of data used in panel data, there are two kinds of panel data. A panel is said to be balanced if each cross-sectional unit has the same number of time series observations and a panel is said to be unbalanced if there is difference in number of observations among panel members.

According to Gujarati (2003), depending on researchers' assumptions, there are four possibilities of estimation of panel data regression models. First, intercept and slope coefficients are constant across time and space and the error term captures differences over time and individuals. Second, the slope coefficients are constant but the intercept varies over individuals and time. Next, the slope coefficients are constant but the intercept varies over individuals and time. Last, intercept and slope coefficients vary over individuals.

4.3 Fixed-Effect Model

If one is interested in knowing unique characteristic of each cross-sectional unit of his observations, he should employ fixed effect model as his regression model. Fixed effect model is a regression model that has the ability to take into account uniqueness of each individual by letting its intercept differ one another, but invariant over time. In order to play its role, one should create one dummy variable for each individual's intercept. That's why this model is also called least-squares dummy variable (LSDV). In addition, dummy variable can be employed to see the time effect which means that one can see if regression's function shift over time or not.

In general, fixed effect regression model or LSDV model can be described as follows:

$$Y_{it} = \alpha_1 + \alpha D_i + \beta X_{it} + \varepsilon_{it}, \qquad i = 1,\ldots,N\ ;\ t = 1,\ldots,T \tag{4.3}$$

This regression model allows for differential intercept dummies to take place. However, one must notice that the number of dummy variables must not equal the number of individual, its number must be one less than the number of individuals.

This is to avoid dummy variable trap, the situation of perfect collinearity if there is more than one exact relationship among the variables.

Fixed effect model assumes only different intercept and constant slope. But, what if one believes that slope coefficient to vary over individual? This should be just easy. Gujarati (2003) suggests that one should extend his LSDV model to add interactive slope dummies. The extended model can be described as follows.

$$Y_{it} = \alpha_1 + \alpha D_i + \beta X_{it} + \gamma (D_i X_{it}) + \varepsilon_{it} \ , \qquad i = 1,\ldots,N \, ; \, t = 1,\ldots,T \qquad (4.4)$$

Nonetheless, one should be very cautious in employing this model, especially in introducing dummy variables. There are several major drawbacks associated with this model that need to be concerned with. First, too many dummy variables introduced would decrease degrees of freedom and thus lead to inefficient estimation. Next, too many variables in the regression equation increases probability of multicollinearity among independent variables. Third, fixed effect model does not have the ability to identify the impact of time invariant variables, such as individual's sex, ethnicity, etc. Last, there is possibility of error terms problem, or heteroskedasticity.

There are other few reasons why the method used in the study is fixed effects. According to Durlauf, Johnson, and Temple (2005), the main attractive feature of this method is its ability to eliminate biased estimate which is caused by correlation between omitted variables, which is common problem in research in growth, and disturbance term. Moreover, fixed effect is able to see the specific information of individual country, which is proclaimed as the future direction in panel data work.

In fact, another common method used in research in growth is 2SLS regression with instrumental variable. However, this method entails some difficulties with choosing the valid instrumental variable. The consequence that should be borne with is fatal. When the instrumental variables chosen are invalid, instrumental variables estimates will be inconsistent. Moreover, Durlauf, Johnson, and Temple argue most authors do not find it very useful. Further, they give example of Barro and Lee who find instrumental variable to be of little effect on coefficient estimates.

In addition, Wooldridge (2002) suggests potential problem associated with 2SLS method, which prevails when instrumental variables used are invalid or weak, is when standard errors of estimators are so large that nothing is insignificant. He argues that sometimes it is very difficult to choose between OLS estimator that is inconsistent but has relatively small standard errors than 2SLS that is imprecise that nothing could be concluded. This is because we are never able to know the size of inconsistencies in instrumental variables and OLS estimates.

Table 4. 2 Description and Operational Definition of Research Variables.

Variable	Description	Time Period	Unit
Dependent Variable			
Log(gdp)	Real GDP per Capita of countries that is measured with base price of year 2000 converted to USD.	1980-2005	Logarithm Values. The original values take form in its basic value (USD)
Independent Variables			
Log(edu)	Projection of amount of people who complete secondary level education. Backward projection was applied in time period 1980-2000, whereas forward projection is applied in time period 2000-2005.	1980-2005	Logarithm values. The original values take form in thousand.
leg_prop	This variable is composed of several other variables which closely relate to law enforcement and security of property rights situation in a country. Those variables are: • Judicial independence • Impartial courts • Protection of property rights • Military interference in rule of law and the political process • Integrity of the legal system • Legal enforcement of contracts • Regulatory restrictions on the sale of real property	1980-2005	Composite index which has range from 0-10. Value of 0 represents the worst, while value of 10 represents the best law enforcement and security of property rights in a country.
reg_qual	This variable is composed of several other variables which closely relate to regulations supporting economic activity in a country. Those variables are:	1980-2005	Composite index which has range from 0-10. Value of 0 represents the worst, while value of 10 represents

	A. Credit market regulations i. Ownership of banks ii. Foreign bank competition iii. Private sector credit iv. Interest rate controls B. Labor market regulations i. Minimum wage ii. Hiring and firing regulations iii. Centralized collective bargaining iv. Mandated cost of hiring v. Mandated cost of worker dismissal vi. Conscription C. Business regulations i. Price controls ii. Administrative requirements iii. Bureaucracy costs iv. Starting a business v. Extra payments/bribes vi. Licensing restrictions vii. Cost of tax compliance		the best supporting and competitive regulations in a country.
Log (gov)	General government final consumption expenditure	1980-2005	Million (constant 2000 USD).
acc_som	This variable is composed of several other variables which closely relate to access to sound money. Those other variables are: • Money growth • Standard deviation of inflation • Inflation: most recent year • Freedom to own foreign currency bank accounts	1980-2005	Composite index which has range from 0-10. Value of 0 represents the worst, while value of 10 represents the best supporting access to sound money in a country.
Log (open)	This variable is calculated by formula: $\frac{export - import}{GDP}$ x 100 % Thus, the values got here are said to be total trade as a percentage of GDP. This variable implies degree of openness of a country.	1980-2004	Percentage (%) The higher the of degree of openness, the more open a country is. The higher degree suggests minimum coordination problem, as well.

Dem	This variable reflects degree of democracy in a country. This variable is composed of two other variables as the proxies: • Political Right • Civil Liberty	1980-2005	Composite index which has range from 0-7. The value of 0 represents the most democratic country while the value of 7 represents the most authoritarian country.
Dummy	This variable is dummy variable for period after crisis, 1998-2005.	1998-2005	This is binary variable 0,1. The value of 1 represents years of period after crisis, 1998-2005; the value of 0, otherwise.
dummy_policy	This variable is dummy variable for period of policy reform up to a year before crisis, 1988-1997.	1988-1997	This is binary variable 0,1. The value of 1 represents years of period of policy reform; the value of 0, otherwise.

Table 4.3 Data Sources

Variable	Source
Log(gdp)	World Development Indicators (WDI) online issued by World Bank.
Log(edu)	IASA
leg_prop	www.freetheworld.com
reg_qual	www.freetheworld.com
Log (gov)	World Development Indicators (WDI) online issued by World Bank.
acc_som	www.freetheworld.com
Log (open)	Penn. World Table 6.2
Dem	Freedom House