Chapter 4

Methodology

4.1.Overview

This chapter describes the method employed to accomplish the objectives of this study. Sources and the way of collection of data required for this study are described here. A description of the econometric analysis appropriated to the research is furnished.

A regression model was formulated, setting FDI inflows as dependant variable and the factors considered as possible key determinants of FDI as explanatory variables. The variables were first tested using the Dickey-Fuller for stationarity and all found to be integrated of order (1). It was determined that the cointegrating variables yield a linear combination, which is stationary. The model was regressed to find out the determinants of FDI inflows and used in the estimation of the parameters to quantify the effects of determinants. Based on the growth of values of explanatory variables under certain assumptions, FDI inflows are forecast.

In this study, data related to selected countries of Southeast Asia, namely Indonesia, Malaysia, Philippines, Singapore, and Thailand are used to analyze the model, assuming that the combination of these five countries are a good representation for the Southeast Asian region.

4.2.Model

The regression model, $FDI_t = \alpha_1 + \alpha_2 GNP_{t-1} + \alpha_3 \Delta GNP_t + \alpha_4 I/GNP_{t-1} + \alpha_5 XR_t + \alpha_6 V(XR)_t + \alpha_7 (M+X)/GNP_{t-1} + Ut$, postulated $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$, and $\alpha_7 > 0$, and $\alpha_6 < 0$ mentioned as the equation 1 in Chapter 3 was formulated based on the theoretical perspectives described in the same chapter.

In this model, GNP represents the size of the market or the economy. Change of GNP (ΔGNP) represents the growth or dynamics of the economy. Imports plus exports to GNP i.e. trade variable or openness ratio ((M+X)/GNP) implies the openness of the economy. Variation of the exchange rate (V(XR)) exhibits the fluctuation of the exchange rate. 'I' denotes domestic investment i.e. capital formation. I/GNP represents the ratio of domestic investment to GNP as a percentage. XR represents the foreign exchange rate. This shows the number of units of local currencies per U.S. dollar.

4.3.Research assumptions

In the process of this study, certain assumptions pertaining to collecting and analyzing data, formulation of the regression model, and forecasting of FDI inflows were taken into consideration. Those assumptions can be arranged under these three categories as follows.

(a) Assumptions related to data:

(1)There is no difference in definition between the countries concerned about the data collected.

(11) The combination of five countries concerned reflects a better representation of Southeast Asian region.

(111) Official exchange rates, applied by some countries are close to the market rates.

(b) Assumptions related to formation of the regression model:

(1) GNP is a better indication of the size of the economy or market.

(11) The model is correctly in relating the explanatory factors to FDI inflows of the Southeast Asian region.

(111) The dependant variable i.e. FDI inflows respond to the level of GNP, I/GNP and (M+X)/GNP of the previous year instead of the level of current year.

(c) Assumptions related to forecasting of FDI inflows:

(1) The period taken into consideration is sufficiently long to explain the long-run trend.

(11) Since the fluctuations of exchange rates during the period of Asian Financial Crisis were very high, for the purpose of forecasting FDI inflows, it was assumed that these fluctuations were not occurred.

(111) Only a baseline forecast is produced.

Subject to these assumptions, the methodology described above, were used to achieve the objectives of this study mentioned in chapter 1. Results and findings were presented in chapter 5.

4.4.Data collection and analysis

To analyze the model, which appeared in the previous section, annual data was used for the period from 1970 to 1999. Secondary data were collected for this purpose from several publications and from various institutions. The main source was the "International Financial Statistics Yearbook", published by International Monetary Fund.

Data relevant to the specifications of the variables of the model were collected for each countries concerned and aggregated for the region, using appropriate methods according to the nature and behavior of the variables. The way of aggregation **(UNCTAD, 1993)** of country data for the region related to each variable can be described as follows.

(1) GNP – Annual GNP values in current prices (in US dollars) for each of the five countries were aggregated in constant prices for the region using weighted average GDP deflators of the region. Weighted average GDP deflator was computed using each country's fraction of GNP out of the total. Based year for the GDP deflator is 1985.

(11) ΔGNP – Annual changes of aggregated GNP for the region computed according to the procedure mentioned in (1) above.

(111) I/GNP – Annual total domestic investment values of five countries were divided by the weighted average GDP deflator to compute the domestic investments in constant prices for the region. These values were used to compute the percentage of aggregated GNP values in constant prices for the region.

(1V) XR – Annual exchange rates (end periods) of five countries concerned (units of local currencies per unit of US dollar) were aggregated for the region using weights of sum of imports and exports of each country out of its total.

(V) V(XR) – Differences between each annual aggregated exchange rate of the region and the mean of those rates for the period from 1970 to 1999 were computed and then the square values of those figures were taken as V(XR).

(V!) (M+X)/GNP – Total annual imports and exports values in current prices for the five countries were converted into constant prices using weighted averaged GDP deflators of the region to get M+X values. Those values were used to compute percentage of aggregated GNP values in constant prices of the region.

4.5.Stationary tests

Any time series data, which is used to analyze a regression model, should have been generated by a stochastic or random process. It is also important that such a stochastic process is stationary for an estimation of a regression model to be meaningful. Therefore the data series, which used to analyze the model here, are tested for stationarity.

Stochastic process is a family of real valued random variables indexed by the time. **(Charemza and Deadman, 1997).** If the mean and the variance of a stochastic process are constant over time and the value of covariance between the two time periods and not on the actual time at which the covariance is computed, such a stochastic process is said to be a stationary stochastic process. **(Gujarati, 1995).** If a stochastic process which consists of the elements of $X_1, X_2, \ldots, X_k$ (denoted by X_t), X_t is said to be stationary

when $E(X_t) = \mu$ = constant, $var(X_t) = \sigma^2$ = constant and $cov(X_t, X_{t+j}) = \sigma_j$. If one or more of the conditions above are not satisfied, the process is non-stationary. For example, if the mean of a stochastic process is an increasing function of time, it is a non-stationary process.

Random walk stochastic process is a specific example of the non-stationary stochastic process. If a stochastic process is described as: $X_1 = Z_1$, $X_2 = X_1 + Z_2$, $X_3 = X_2 + Z_3$ etc.or in other words $X_t = X_{t-1} + Z_t$, where Z_t represents a series of identically and normally distributed random variables, such a process can be identified as a random walk. Variances of X_t of such a process are $var(X_1) = var(Z_1)$, $var(X_2) = var(X_1) + var(Z_2)$ etc.or in other words $var(X_t) = var(Z_1).t$. It implies that the variance of X_t is not constant. Therefore this type of a process is non-stationary.

Several methods are used to test for stationarity of a data series at present. The **Augmented Dickey-Fuller Unit Root Test[1] (ADF test)** is used here[1]. The concept of the unit root test can be explained in the following way. When we consider a model as: $Y_t = Y_{t-1} = U_t$, where Ut is the stochastic error term which has zero mean and constant variance σ^2 and is non-autocorrelated if the coefficient of Y_{t-1} is equal to 1, this has a unit root problem. In other words, in a similar model like $Y_t = pY_{t-1} = U_t$, if p is equal to 1 this has a unit root problem. Therefore such a process is not stationary. These types of time series are called random walk. The unit root tests are conducted to check whether such a unit root problem exists. The model of the form of $\Delta Y_t = \delta Y_{t-1} + U_t$, where δ is p-1 and U_t is the white noise error term is used for the Augmented Dickey-Fuller Unit Root Test here. According to the procedure of this test, the null hypothesis that $\delta = 0$ i.e. unit root

[1] Dickey, D.A. and W.A.Fuller, (1979), 'Distributions of the estimators for autoregressive time series with a unit root', Journal of the American statistical Association 74.

exists in Y is tested using the OLS. The computed t value of the variable Y_{t-1} is compared with the critical thau statistics computed by **MacKinnon**[2], at chosen level of confidence. If the computed thau value in absolute terms is smaller than the critical value, the null hypothesis is rejected. It means the data series exhibits a unit root i.e. non-stationary situation. **(Gujarati, 1995).**

If a data series is stationary, it means that the series is integrated of the order of zero I(0). Each data series used for estimation of a regression model is required to be stationary or in other words I(0). Some times data series may not be I(0), but each series of the model may be in a same order of integration. It was found that the all data series of the model used in this study are I(1). It is assumed that the long-run relationship holds and that this will be captured in the regression equation estimated. **(Gujarati, 1999).**

4.6.Test for multicollinearity

If a perfect or exact linear relationship is existed among some or all the explanatory variables of a regression model, such a situation is said to be a multicollinearity problem **(Frisch, 1936).** However presently the term multicollinearity is used in broader sense to include the case of perfect multicollinearity as well as the case where the explanatory variables are Interco related but not perfectly so. **(Gujarati, 1995).** When we consider a k-variable regression with explanatory variables with $X_1, X_2,\ldots,X_k$, exact linear relationship is said to exist in this regression, if the condition that $\lambda_1 X_1 + \lambda_2 X_2 + \ldots \lambda_k X_k = 0$, where $\lambda_1, \lambda_2,.., \lambda_k$ are constants but all of them are not zero simultaneously is satisfied. If the explanatory variables of a regression model are

[2] MacKinnon, J.G., (1991), 'Critical values of cointegration tests', in R.F.Engle and C.W.J.Granger (eds), Long-run economic relationships, Oxford University Press, Oxford.

intercorrelated but not perfectly so, the condition that has to be satisfied is $\lambda_1 X_1 + \lambda_2 X_2 + \ldots \lambda_k X_k V_i = 0$, where V_i is a stochastic error term.

The reason to care about the problem of multicollinearity is that it causes the regression coefficients be indeterminate and their standard errors be very large. Therefore it is necessary to check whether a multicollinearity problem exists in a regression model before using it for analysis. Presently several methods are used to detect the problem of multicollinearity. The method used here is detecting multicollinearity using variance inflation factor (VIF). **(Gujarati, 1995).** If we consider a k variable regression model we can formulate auxiliary regression models using each explanatory variables as dependent variable with remaining explanatory variables of the original regression model as explanatory variables of such auxiliary regression models. Now based on the estimated results of such auxiliary regressions, the variance of partial regression coefficient can be expressed as: $var(\hat{\beta}_j) = (\sigma^2/\Sigma X_j^2) \cdot (1/1- R_j^2)$, where β_j is the partial regression coefficient of the regressor X_j, R_j^2 is the R^2 in the auxiliary regression in which X_j is dependent variable. VIF is defined as$1/1- R_j^2$. When R_j^2 is increases toward unity, as the collinearity of X_j with the other regressors increase, the VIF increases toward infinity. Therefore larger VIF indicates higher collinearity among regressors. When R_j^2 exceeds 0.90, VIF exceeds 10. Such a situation can be considered, as the regressors are highly collinear. **(Kleinbaum, Kupper and Muller; 1988).**

The test conducted in this study, using the above procedure confirmed that certain explanatory variables had collinearity problem. The explanatory variable (M+X)/GNP exhibited high collinearity. Therefore as a remedial measure this variable had to be

dropped from the model. It was assumed that this does not create a specification problem, as the role of this variable is partly included in the variable XR too.

4.7.Test for heteroscedasticity

It is necessary to check whether the model used in this study exhibits heteroscedasticity problem. If there is no existence of such a problem, variance of each disturbance term conditional on the chosen values of the explanatory variables should be some constant number. In contrast, if the conditional variance of disturbance term increases or decreases as the values of explanatory variables are changed. Such a regression exhibits heteroscedasticity problem. This can be shown symbolically as $E(U_i^2) = \sigma^2$. **(Gujarati, 1995).** If there is a heteroscedasticity problem in a regression, the estimated parameters are no longer 'efficient' and 'best'. **(Gujarati, 1995).** Therefore if such a problem exists, it is necessary to use a remedial measure to overcome this problem.

Several methods are used to detect heteroscedasticity at present. In this analysis, the **White's test**[3] is used for this purpose. According to the procedure of White's test, first the estimated residuals should be obtained by regressing the model. Then an auxiliary regression model should be formulated using U_i^2 as dependent variable and using the explanatory variables of the original model, their squares, and their cross products as explanatory variables. This model has to be estimated using the OLS. Results of this estimation are used to test the null hypothesis that all the parameters of explanatory variables of this estimated model are simultaneously equal to zero. The rule

[3] White, H., 'A heteroscedasticity consistent covariance matrix estimator and a direct test of heteroscedasticity', Econometrica, vol.48, 1980.

of thumb is that the null hypothesis is rejected if $nR^2 > \chi^2_{\alpha jp}$, where n is the number of observations (sample size), $\chi^2_{\alpha jp}$ is the critical chi square value at α confidence level under p degree of freedom. P should be equal to the number of regressors excluding the intercept term of the auxiliary regression. If the null hypothesis is rejected it implies that there is heteroscedasticity. In other words the error variance of U_i that is σ^2_i of the auxiliary regression is functionally related to the regressors, their squares, and their cross products. **(Gujarati, 1995).**

The test conducted using White's heteroscedasticity test for the model after dropping the variable (M+X)/GNP, revealed that there is no a problem of heteroscedasticity in this model.

4.8.Statistical analyses

After dropping the trade variable, from the regression model described in section 4.2, the model was estimated using the OLS and the results were used to analyze the model statistically in order to achieve the objectives of this study. It is important to note the theoretical background of the statistical analyses used here.

Explanatory variables of the model used in this study represents the possible factors serve as determinants of FDI inflows. The parameters of these variables have to be estimated and tested to find whether these are significant. If those parameters are significant, it can be concluded that those are determinants of FDI inflows in the context of this study. It is useful to conduct t tests to ascertain whether the parameters of the regressors are significant. The null hypotheses that the each value of individual parameters is equal to zero can be tested against the alternative hypothesis that each value

of parameters are not equal to zero in the context of this model. The t statistic is computed dividing the difference between the estimated value of the coefficient and the true value of the coefficient in terms of the null hypothesis by its standard error. This computed t value is compared with the critical t value at a chosen level of confidence under the degree of freedom equal to n-k , where n is the sample size and k is the number of regressors. As a rule of thumb, if the computed t value exceeds the critical t value at a chosen of confidence level, the null hypothesis that the value of parameter is equal to zero is rejected. It means, the value of the parameter is significantly different from zero. **(Gujarati,1995).**

Another important indicator that used for statistical analysis of a regression is the R^2. R^2 can be defined as ESS/TSS, where ESS is the explained sum of squares, and TSS is the total sum of squares. If we consider a three variable multiple regression model, ESS can be explained as $\hat{\beta}_2\Sigma y_i x_{2i} + \hat{\beta}_3\Sigma y_i x_{3i}$ and TSS can be explained as Σy_i^2, where y_is are the deviations from the mean values of Y and x_{2i}s and x_{3i}s are the deviations from the mean of X_2 and X_3 respectively.

R^2 lies between 0 and 1. If R^2 is equal to 1, it means the fitted regression line explains 100 percent of the variation in the dependent variable. If R^2 is zero, the model does not explain any of the variation in the dependent variable. Therefore the fit of the model is said to be better if R^2 is close to 1. **(Gujarat, 1995).**

In addition to R^2, adjusted R^2, which is denoted by R^2 is also used to interpret the results of an estimated regression model. Adjusted R^2 can be defined as: $\bar{R}^2 = 1-(\hat{\sigma}^2/S_y^2)$, where $\hat{\sigma}^2$ is the residual variance and S_y^2 is the sample variance of Y. Describing the

relationship with R^2, adjusted R^2 can be defined as: $\bar{R}^2 = 1-(1-R_2).(n-1/n-k)$. $\bar{R}^2$ is also used to measure the goodness of fit of a model as R^2. **(Gujarati, 1995).**

In addition to significance tests of parameters, overall significance of the regression model can be tested by using F test. F statistics are computed as: F = (ESS/k-1)/(RSS/n-k)., where ESS is the explained sum of squares, and RSS is the residual sum of the squares. k is the number of regressors, and n is the sample size. **(Gujarati, 1995).**

According to the procedure of F test, a null hypothesis that the each value of parameters of the explanatory variables are simultaneously equal to zero can be tested against the alternative hypothesis that at least one explanatory variable is not equal to zero. For this purpose computed F statistic is compared with the critical F value at a chosen level of significance under the degree of freedom k-1 and n-k. If the computed F statistic exceeds the critical value, the null hypothesis is rejected. It means at least one parameter of an explanatory variable is different from zero. Therefore in such a situation, the overall regression model is said to be significant.

4.9.Forecasting

One of the objectives of this study is examining the future trend in FDI inflows of this region. The estimated model and the data applied for estimation of the model were used to forecast FDI inflows for each year from 2000 to 2005. The values of the explanatory variables for each year for the period from 2000 to 2005 were estimated according to the following procedure.

(1) GNP – Annual growth rate of aggregated GNP of the region was calculated using the same data, which were applied for the estimation of the model, for the period

from 1970 to 1999. Those growth rates were averaged and that averaged growth rate was used to project GNP values of the region for the period from 2000 to 2005, based on the actual value in 1999.

(2) ΔGNP – Absolute changes of the projected GNP values were Computed for the forecasting period.

(3) I/GNP – Annual growth rates of aggregated domestic investments of the region were computed for the period from 1970 to 1999. Those growth rates were averaged and used to project domestic investment values of the region for the forecasting period, based on the actual value in 1999. Those values of domestic investments and the GNP values projected were used to compute the I/GNP values of the region for the forecasting period.

(4) XR – The procedure applied to estimate the exchange rate of the region for the forecasting period was somewhat different from the above. Actual exchange rate of the region for the period from 1997 to 1999 exhibit high fluctuation from the normal trend. It can be suspected that it is a result of the Asian currency Crisis, which hit the economies of the five countries concerned severely in this period. If this figures are taken into consideration for forecasting the results may differ from the reality as the crisis began to recover from 1999. Therefore, when the exchange rate is projected for the forecasting period, the values of exchange rates in the period from 1997 to 1999 were omitted. According to the procedure applied to project the regional exchange rate here, the growth rates of regional exchange rate were computed for the period from 1970 to 1996 only. These growth rates were averaged and that average growth rate was used to estimate the aggregated

exchange rate of the region, based on the actual exchange rate in 1996 instead of 1999. Therefore it was assumed that there were no variations from the normal trend in 1997, 1998, and 1999.

(5) V(XR) – The estimated XR values computed according to the procedure described in (4) above were used to compute the squared deviation values of the exchange rates for the forecasting period. In this process, mean of exchange rate was computed for the period from 1970 to 2005.

The estimated values of explanatory variables for the period from 2000 to 2005 were applied to the estimated regression model to forecast FDI inflows of the region for each year from 2000 to 2005.

4.10.Conclusion

The method used in order to achieve objectives of this study was described in this chapter. A model of determinants of FDI inflows formulated based on the theoretical background on the subject is used to analyze by estimating using the OLS. The study is conducted subject to certain assumptions related to data applied to the model, formation of the regression model, and forecasting of FDI inflows.

Relevant data required to apply to the model were collected from secondary data sources. Data were analyzed to be appropriate to the model used here. Before estimating the regression, data series were tested for stationarity and cointegration. After being satisfied that the data series are stationary, the test for multicollinearity was done. Since there was a problem of multicollinearity, one explanatory variable had to be dropped.

After that, the test for heteroscedasticity was conducted. It was found that there was no such a problem in the model.

Some statistical analyses such as t tests and F tests and also some statistical indicators such as R^2, and adjusted R^2 are used to analyze the findings of the estimated regression model in order to fulfill the objectives of this study.

One objective of this study is to forecast the values of FDI inflows. For this purpose, the average growth rates of the actual values of the explanatory variables of the model are used. Based on such actual trend of the variables, FDI inflows are forecast using the estimated regression.