Chapter 4

Methodology

4.1 Introduction

For the purpose of empirical studies, this chapter will look into the methodology and statistic tests used to investigate the relationship between the stock markets. Section 4.2 outlines the scope of study, followed by the data of study in section 4.3. After that, we will use few econometric techniques to investigate long run as well as pre-crisis and post-crisis relationships between the variables in section 4.4. We will divide the study into three main parts. In the first part, we will employ the unit root test to examine whether the data used in the analysis is stationary or nonstationary. Then, in the second part, the Johansen and Juselius (JJ) (1990) Cointegration test is applied to test for the cointegrating relationships between the variables. Finally, we will impose the Granger causality test to verify the causality effect between the variables. Section 4.5 will conclude this chapter.

4.2 Scope of Study

The analysis is using weekly data from January 1994 to Jun 2001 and the stock indices invloved are Dow Jones Industrial Average (DJIA) – the US, Hang Seng Index (HSI) – Hong Kong, Taiwan Capitalization Weighted Price Index (TAIEX) – Taiwan, Korea Composite Stock Market Index (KOSPI) – South Korea, Kuala Lumpur Stock Exchange Composite Index (KLSECI) – Malaysia, Stock Exchange of Thai Index (SET) – Thailand, Straits Times Index (STI) – Singapore, Jakarta Composite Stock Price Index (JCSPI) – Indonesia and Philippines Stock Exchange Composite Index (PSECI) – Philippines. In this study, 391 weekly closing values of stock indices are used for the entire period of January 1994 to Jun 2001. 182 weekly observations for pre-crisis period (January 1994 - Jun 1997) and 209 weekly data for post-crisis period (July 1997 - Jun 2001).

4.3 Data

The data is provided by Morgan Stanley Capital International (MSCI) which is obtained from the Kuala Lumpur Stock Exchange (KLSE). The MSCI series is particularly been used for the study because all of the indices are constructed on a consistent common basis. The indices were converted and denominated into US dollar terms. All the series were transformed into natural logarithm. The purpose of utilizing weekly data is to avoid some problems associated with daily and monthly

data. Daily data are susceptible to the day-of-the-week effect, whereas monthly data are affected by the month-of-the-year effect.

4.4 Analytical Method

The econometric analysis is done by using a computer software named as *Econometric Views (EVIEWS) 4.0.*

Below an outline of unit root test, Johansen and Juselius (JJ) (1990) Cointegration test and Granger Causality test are presented.

4.4.1 Testing for Stationarity

We noted that the type of data used in this empirical analysis is time series data. In regressing a time series variable on another time series variable, a very high R^2 often been obtained although there is no meaningful relationship between the two (Gujarati, 1995). This situation exemplifies the problem of spurious regression. The problem arises because if both the time series involved exhibit strong trends (sustained upward or downward movements), the high R^2 observed is due to the presence of the trend. Nonstationarity in time series can be identified with the presence of trend, seasonality, and structural change, etc.. In order to study the true relationship

between variables which is based on time series, the empirical work is assumed to be stationary.

4.4.2 Stationary Stochastic Processes

Any time series data can be thought of as being generated by a stochastic or random process. *A stochastic process is said to be stationary if its mean and variance are constant over time and the value of covariance between two periods depends only on the distance of lag between the two time periods and not on the actual time at which the covariance is computed* (Gujarati, 1995).

To explain the above statement, let Y_t be a stochastic time series with the following properties:-

Mean : $E(Y_t) = \mu$ *(4.4.2.1)*

Variance : $var(Y_t) = E(Y_t - \mu)^2 = \sigma^2$ *(4.4.2.2)*

Covariance : $\gamma_k = E[(Y_t - \mu)(Y_{t-k} - \mu)]$ *(4.4.2.3)*

where γ_k, the covariance (or autocovariance) at lag k, is the covariance between the values of Y_t and Y_{t-k}, that is, between two Y values k periods apart. If $k = 0$, we obtain γ_0, which is simply the variance of $Y (= \sigma^2)$; if $k = 1$, γ_1 is the covariance

between two adjacent values of Y. Suppose we shift the origin of Y from Y_t *to* Y_{t+m}. Now if it is to be stationary, the mean, variance, and autocovariances of Y_{t+m} must be the same as those of Y_t. In short, if a time series is stationary, its mean, variance, and autocovariance (at various lags) remain the same no matter at what time we measure them.

4.4.3 The Unit Root Test of Stationarity

One simple test of stationarity that has recently become popular is known as the unit root test. The easiest way to introduce the test is to consider the following model:-

$$Y_t = Y_{t-1} + u_t \qquad (4.4.3.1)$$

where u_t is the stochastic error term that follows the classical assumptions, namely, it has zero mean, constant variance σ^2, and is nonautocorrelated. If the coefficient of Y_{t-1} is in fact equal to 1, we face what is known as the unit root problem, i.e., a nonstationary situation. Therefore, if we run the regression

$$Y_t = \rho Y_{t-1} + u_t \qquad (4.4.3.2)$$

and actually find that $\rho = 1$, then we say that the stochastic variable Y_t *has a unit root.* A time series that has a unit root is known as a random walk (time series). And a random walk is an example of a nonstationary time series (Gujarati, 1995).

For an example, assets prices, such as stock prices, follow a random walk, that is, they are nonstationary.

Equation *(4.4.3.2)* is often expressed in an alternative form as:-

$$\Delta Y_t = (\rho - 1)\, Y_{t-1} + u_t$$

$$= \delta\, Y_{t-1} + u_t \qquad (4.4.3.3)$$

where $\delta = (\rho - 1)$ and Δ is the first-difference operator, $\Delta Y_t = (Y_t - Y_{t-1})$. The null hypothesis is that $\delta = 0$.

Based on $\delta = 0$, the equation *(4.4.3.3)* can be written as:-

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \qquad (4.4.3.4)$$

It means that the first differences of a random walk time series ($= u_t$) are a stationary series because by assumption u_t is purely random.

A stationary process can be derived from a nonstationary process by differencing the series one or more times. If a time series is differenced once and the differenced series is stationary, the original (random walk) series is integrated of order 1, denoted by I (1). Similarly, if the original series has to be differenced twice before it becomes stationary, the original series is integrated of order 2, or I (2). Generally, if a

time series has to be differenced d times, it is integrated of order d or I (d). Thus, any time we have an integrated time series of order 1 or greater, we have a nonstationary time series.

4.4.4 Trend Stationary

In regressions involving time series data, to avoid the problem of spurious correlation, the common practice is to include variable t (trend variable) as one of the regressors. The introduction of the trend variable in the regression has the effect of detrending (i.e. removing the influence of trend from variables).

However, econometricians have challenged the common practice and argued that the standard practice may be acceptable only if the trend variable is deterministic and not stochastic.

Following that, two key concepts in time series analysis was introduced, namely, a trend-stationary process (TSP) and a difference-stationary process (DSP). A trend-stationary time series has a deterministic trend, stationarity is achieved by inclusion of a time trend variable. However, a difference-stationary time series has a variable, or stochastic, trend, requires differencing to achieve stationarity (Dickey and Fuller, 1979). It is important to check for the correct form of non-stationary behaviour of the time series because a difference-stationary process which is stochastic cannot be

cointegrated with a trend-stationary process, which, on the other hand, is deterministic. Nelson and Plosser (1982) have demonstrated that many economic time series appear to be difference-stationary processes.

4.4.5 Augmented Dickey Fuller (ADF) Test

Dickey and Fuller (1979, 1981) provide one of the most influential works in the field of unit root tests. The Dickey-Fuller unit root test is based on the following three regression forms:-

1. Without Constant and Trend $\Delta Y_t = \alpha Y_{t-1} + u_t$ *(4.4.5.1)*

2. With Constant $\Delta Y_t = \delta + \alpha Y_{t-1} + u_t$ *(4.4.5.2)*

3. With Constant and Trend $\Delta Y_t = \delta + \beta T + \alpha Y_{t-1} + u_t$ *(4.4.5.3)*

The hypothesis is:

$H_0 : \alpha = 0$ (Unit Root)

$H_1 : \alpha \neq 0$

Decision rule:

If $t^* >$ critical value - not reject null hypothesis, i.e., unit root exists

If $t^* <$ critical value - reject null hypothesis, i.e., unit root does not exist

Up to now we have used variations of AR (1) process. The following equation presents a higher AR process version of the Dickey-Fuller test:-

$$\Delta Y_t = \delta_0 + \alpha Y_{t-1} + \sum_{i=1}^{L} \Phi_i \Delta Y_{t-i} + \varepsilon_t \qquad (4.4.5.4)$$

where Y_t is a time series in levels. ε_t is the error term and L is the number of lagged changes in the necessary to make ε_t serially uncorrelated or white noise. This test is known as the Augmented Dickey-Fuller test (ADF). For the purpose of this study, ADF test is considered. The null hypothesis in ADF test, $H_0 : Y_t$ is $I(1)$, is rejected (in favour of $I(0)$) if α is found to be negative and significantly different from zero, Y_t is stationary.

If a time trend is included in equation *(4.4.5.4)*, we have the following equation *(4.4.5.5)*, which is used to determine whether the series is trend-stationary:

$$\Delta Y_t = \delta_0 + \theta t + \beta Y_{t-1} + \sum_{i=1}^{L} \Phi_i \Delta Y_{t-i} + \varepsilon_t \qquad (4.4.5.5)$$

where t is a time trend. If parameter β is negative and significantly different from zero then Y_t is said to be trend-stationary. The computed t-statistic on parameter $Z(t_\alpha)$ and $Z(t_\beta)$ are compared to the critical value tabulated in MacKinnon (1991).

The unit root tests were also carried out for first-difference of the variable, that is, we estimate the following regression equation;

$$\Delta^2 Y_t = \delta_0 + \alpha Y_{t-1} + \sum_{i=1}^{L} \Phi_i \Delta^2 Y_{t-i} + \varepsilon_t \qquad (4.4.5.6)$$

As mentioned above, the role of lagged dependent variables in the Augmented Dickey-Fuller (ADF) regression equation *(4.4.5.4)* and *(4.4.5.5)* is to ensure that ε_t is white noise. Thus, we have to choose the appropriate lag length L. In this study, we used Schwert's (1987, 1989) criteria which sets the lag length equal to the integer (*int*) portion of two values of L, that is,

$$L_4 = int\ \{4(T/100)^{\frac{1}{4}}\} \text{ and } L_{12} = int\ \{12(T/100)^{\frac{1}{4}}\} \qquad (4.4.5.7)$$

where T is the total number of observations. Following Lutkepohl (1984), we used Akaike's Information Criterion (AIC) to differentiate the optimal lag length between the two L_4 and L_{12}. The lag length that produced the lowest AIC value will be the optimal lag length. Finally, the critical values we used in the unit root tests are from MacKinnon (1991).

4.4.6 Testing for Cointegration

After determining that the series are of the same order of integration, next we have to determine whether linear combination of the series that are non-stationary in levels are cointegrated. The main idea behind cointegration is a specification of models that include beliefs about the movements of variables relative to each other in the long run, such as to test whether the Malaysia stock market is pairwise cointegrated with each of the stock markets in the US, Hong Kong, Taiwan, South Korea, Singapore, Thailand, Indonesia and Philippines.

In earlier studies, the Engle-Granger (1987) two-step stimulation procedure is frequently used to test for cointegration. However, this procedure has been criticized for being static and have several econometric problems. Banerjee et al. (1986) noted that although Engle-Granger procedure produces super consistent parameter estimates, for small sample the biases could be quite severe. Beside s that, when the

cointegration relationships are not unique as in the present case, then the Engle-Granger procedure performs less satisfactorily. The estimates are not invariant to the choice of normalization.

4.4.7 Johansen and Juselius (JJ) (1990) Cointegration Test

Johansen and Juselius (JJ) (1990) have suggested an alternative method, the maximum likelihood estimation procedure, to test for cointegrating relationship. The approach is free from the problems as mentioned above. The JJ procedure involves the identification of rank of the long-run impact matrix, Π , in the specification given by

$$\Delta X_t = \delta + \sum_{l=1}^{k-1} \Delta X_{t-l} + \Pi X_{t-k} + \varepsilon_t \qquad (4.4.7.1)$$

where X_t is a column vector of the two markets indices since our study is based on pairwise cointegration between two markets. All the long run information in the process is summarised by the marix Π, and it is the rank of this matrix that determines the number of cointegrating vectors.

There are three possibilities to consider. First, Π could be of full rank. In this case, the assumed stationarity of ε_t in equation *(4.4.7.1)* requires that the levels of the X_t process themselves be stationary, implying the absence of any stochastic trends

whatsoever in the data, contrary to $I(1)$ specification. Second, Π has zero rank, no stationary linear combination can be identified. In other words, the variables in X_t are non-cointegrated. The third case occurs when Π is of intermediate rank r $(0 < r < n)$, in which case Π can be factored as

$$\Pi = \alpha\beta' \qquad (4.4.7.2)$$

where α and β' are $n \times r$ matrices. In this case, there are $n-r$ unit roots in the system (i.e. there are $n-r$ common trends) and r cointegrating relationship. The i-th row tell us how important each of these r cointegrating relations are to the dynamics of the i-th series.

Two different likelihood ratio tests are provided to determine the value of rank r, of the matrix Π in equation *(4.4.7.2)*.

The first is known as the maximal eigenvalue test (λ -max)

$$L_{max} = -T\, ln(1-\lambda_{r+1}) \qquad (4.4.7.3)$$

where λ_{r+1} is the $(r+1)^{th}$ largest squared canonical correlation or eigenvalue; and T is the number of observations. The null hypothesis is H_0: $r = r_0$ against H_a: $r = r_0 + 1$, there are r cointegrating vectors, against the alternative $r+1$ cointegrating vectors.

The other test for cointegration is the trace test which is based on the following statistic

$$L_{trace} = -T \sum_{i=r+1}^{p} ln(1-\lambda_i) \qquad (4.4.7.4)$$

where $\lambda_{r+1} \ldots \lambda_p$ are the $p-r$ smallest squared canonical correlation or eigenvalue. The null hypothesis is $H_0: r \leq r_0$ against $H_a: r > r_0$ where r refers to the number of cointegrating vectors. Johansen and Juselius (JJ) (1990) indicated that the trace test might lack power relative to the maximal eigenvalue test. Based on the power of the test, the maximal eigenvalue test statistic is often prefered.

As mentioned in Section 4.4.5, we used Akaike's Information Criterion (AIC) from *Eviews* to choose the optimal lag length. The lag length that produced the lowest AIC value will be the optimal lag length. In the cointegration tests, critical values at 5 % level of significance are obtained from Osterwald Lenum (1992).

4.4.8 Granger Causality Test

In studying the relationship between the Malaysian stock price and other countries, it is also of interest, to examine the direction of Granger causality between the Malaysian stock market and other countries' stock markets, i.e., whether market trends tend to originate in the Malaysian stock market and spill overseas, or *vice versa*, or whether there is bi-directional feedback. Although there are many approaches to modelling causality in temporal systems, we applied the prototype model by Granger (1969) not only because it is the simplest and most straightforward, but also the existence of causal ordering in Granger's sense points to a low causation and implies predictability and exogeneity. We therefore consider Granger's four definitions of causality, based on the following model:

$$MY_t = \sum_{j=1}^{m} \alpha_j MY_{t-j} + \sum_{j=1}^{n} \beta_j SP_{t-j} + \varepsilon_t \qquad (4.4.8.1)$$

$$SP_t = \sum_{j=1}^{m} c_j MY_{t-j} + \sum_{j=1}^{n} d_j SP_{t-j} + \mu_t \qquad (4.4.8.2)$$

where MY = Malaysia stock index variable; SP = other country stock index variable; ε_t and μ_t are assumed to be serially uncorrelated with zero mean and finite covariance matrix.

Granger (1969) offers four definitions of causality, which is comprised:

1. unidirectional causality from *MY* to *SP*
2. unidirectional causality from *SP* to *MY*
3. feedback causality between *MY* and *SP*
4. independence between *MY* and *SP*

The four definitions, according to Granger, imply that

1. for *MY* to Granger-cause *SP*, $c_j \neq 0$ whereas $\beta_j = 0$
2. for *SP* to Granger-cause *MY*, $\beta_j \neq 0$ whereas $c_j = 0$

If we allow for the possibility of $j = 0$ in the summation symbol for Equations *(4.4.8.1)* and *(4.4.8.2)*, the relationship between the two time series is said to be instantaneous. Thus, instantaneous causality is established when the inclusion of the present values of the independent variable improves the prediction or goodness of fit (or R^2) of both equations.

4.5 Conclusion

At the beginning of this chapter, we discussed about the source of data and scope of study for the purpose of studying within the specified scope. Due to the data used in this study being time series data, therefore, the Unit Root test is applied to test whether the time series is stationary and to avoid the problem of spurious regression. After establishing the order of integration of the series, Johansen and Juselius Cointegration test is used to test for the cointegrating relationship between the variables. Finally, Granger Causality test is applied to test the causality effect of the variables.