CHAPTER 3

DATA AND METHODOLOGY

Since the data used would influence the results obtained, an informative description of the dataset to be utilized in the study is necessary. With five countries under study, the dataset must be prepared in a way that comparisons among them are appropriate. Moreover, a careful and justified selection of the durations of the three periods must be emphasized.

This chapter also elaborates the methodologies that will be employed to achieve the objectives of the study. The principles related to each methodology will be discussed, without going into too much of mathematical details. Where applicable, there will be a justification of the methodologies adopted here instead of the alternatives.

3.1 Data

To overcome the need to adjust for public holidays or non-trading days, many researchers used monthly or weekly data in their studies. However, any potential interactions that last for only a few days may be obscured if monthly or weekly data are used (Eun and Shim, 1989). For this reason, it is more appropriate to use daily closing prices to study the behaviour and relationships of stock indices.

The ASEAN-5 stock markets are represented by their widely followed leading stock market indices. They are the Kuala Lumpur Stock Exchange Composite Index, Singapore Stock Exchange All-Share Index, Stock Exchange of Thailand Index, Jakarta Composite Index and the Philippines Composite Index. The data are obtained from the data financial provider Bloomberg and cover the period from 2 January 1992 to 8 August 2002. Since all the markets examined are operating almost simultaneously during the day, the setting for the analysis of market linkages is thus appropriate as there is no problem of non-synchronous return observations.

All these five markets are trading on a five-day week basis. To obtain a common dataset of the stock market indices, we uniformly exclude those days that are national holidays in any one of the countries. A total of 2309 observations are then made available for this study.

The data are partitioned into three periods, each characterized according to the timing in relation to the Asian financial crisis. The pre-crisis period, which runs from January 1992 to January 1997, corresponds to the period of stable market environment before the crisis. The period of high volatility and accompanied by a sharp decline in the market indices coincides with the duration from February 1997 to September 1998. This period is identified as the crisis period. Thereafter, all the five ASEAN markets seem to be recovering and hence the period from October 1998 to the end of the dataset is named as the post-crisis period. Figure 3.1 shows the graphs of the market indices of Indonesia, Malaysia, Philippines, Singapore and Thailand, respectively. The three identified periods are marked by the letters A, B and C.

Figure 3.1

Graphs of ASEAN-5 market indices: January 1992 – August 2002

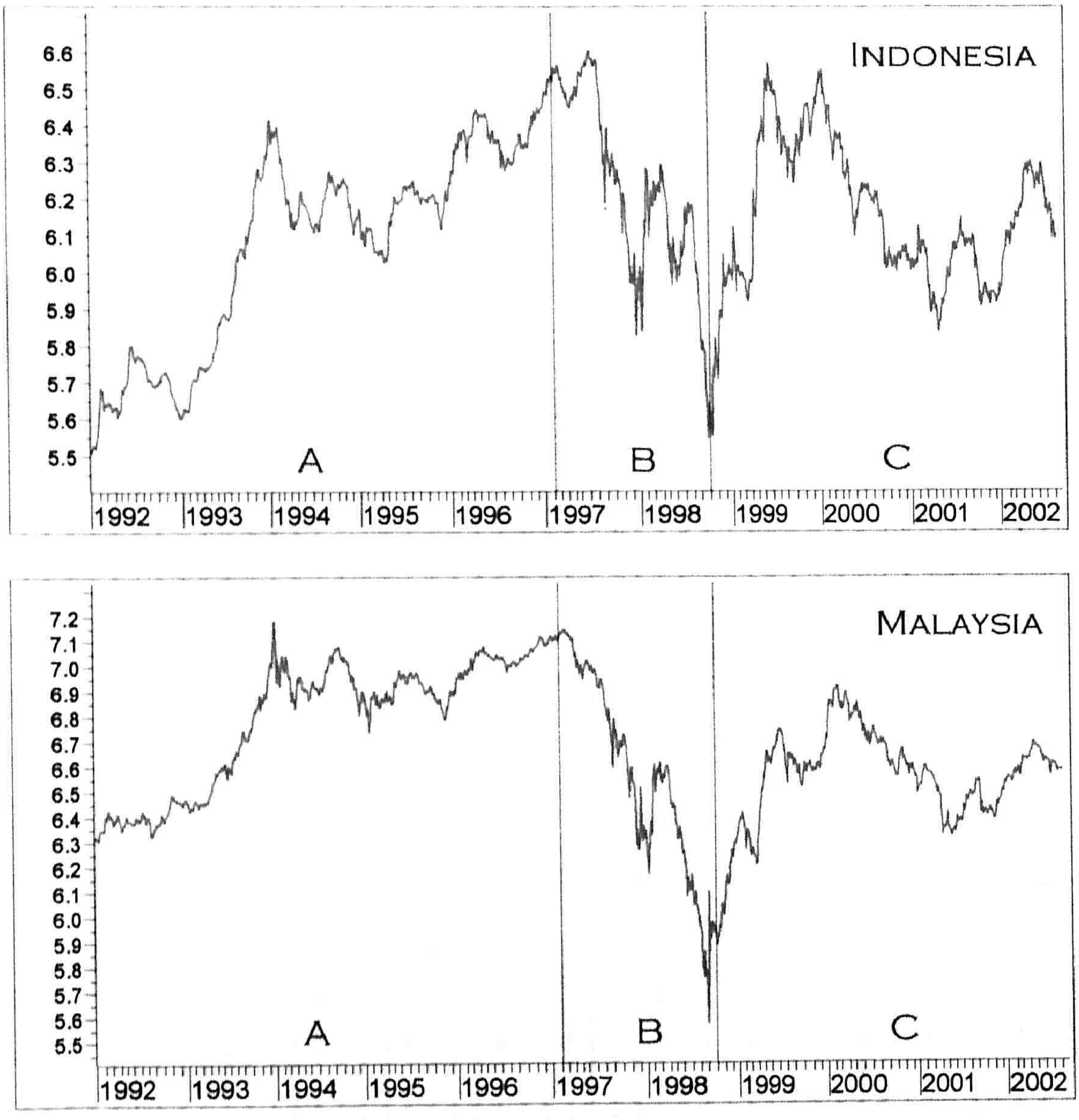
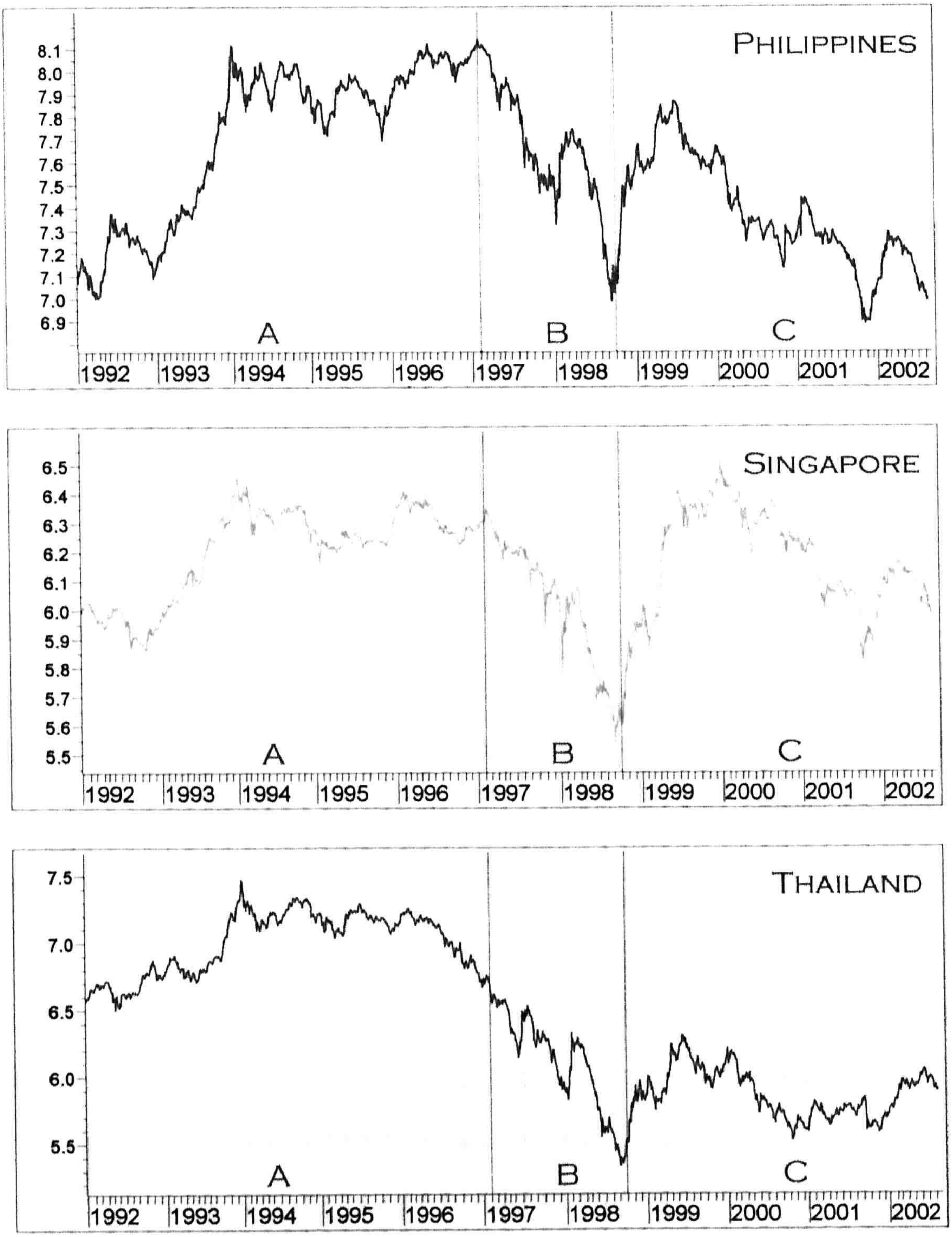


Figure 3.1 (continued)



A corresponds to the pre-crisis period, B, the crisis period and C, the post-crisis period

3.2 Methodology

3.2.1 Measurement of Market Return

The daily market returns are computed as:

$$R_{it} = P_{it} - P_{i,t-1} = \Delta P_{it} \tag{3.1}$$

where P_{it} represents the natural logarithm of the daily price index for market i (i = 1,2,....5) at time t, t = 1,2,.....,n, and n being the number of observations in the sample period. Δ represents the difference operator.

3.2.2 Model Selection Criteria

There are several criteria that can be used to choose among competing models. The more common ones are (1) R^2, (2) adjusted R^2, (3) Akaike Information Criterion (AIC), (4) Schwarz Information Criterion (SIC), (5) Mallow's C_p criterion, and (6) forecast χ^2. These criteria, each in its own way, aim at comparing the residual sum of squares (RSS) across models. However, only criteria (2), (3), (4) and (5) impose a penalty for including an increasingly large number of regressors. Therefore, there is a tradeoff between the goodness-of-fit of a model and its complexity (as judged by the number of regressors).

Each criterion has a different penalty function. The econometric software EViews (which is used in this study, see Section 3.3) displays only AIC and SIC in the estimation of a regression equation. Of these four criteria, SIC penalizes the inclusion of an additional regressor most heavily and thus, has the highest tendency to select models that are more parsimonious. In this study, SIC is used in the choice of models. The smaller the value of this criterion, the 'better' is the model. Therefore, in all autoregressive processes, the lag length chosen is the one associated with the smallest SIC.

For a system of q equations with a total of K parameters, the SIC is based on the following formula:

$$\text{SIC} = \frac{2\ell}{n} + \frac{K \log n}{n} \tag{3.2}$$

where ℓ and $|\hat{\Omega}|$ are given by

$$\ell = -\frac{nq}{2}(1 + \log 2\pi) - \frac{n}{2}\log|\hat{\Omega}| \tag{3.3}$$

and

$$|\hat{\Omega}| = \det\left(\sum \tilde{\varepsilon}_t \tilde{\varepsilon}_t' / n\right) \tag{3.4}$$

where $\tilde{\varepsilon}_t$ is the q x 1 vector of residuals. The log-likelihood value, ℓ, is computed under the assumption that the error term follows a normal (Gaussian) distribution.

3.2.3 Dummy Variable Approach – Testing Structural Changes

The dummy variable approach can be used to test for structural changes in the market returns that may be attributed to the Asian financial crisis. The following autoregressive process is fitted:

$$R_{it} = \beta_0 + \beta_1 D_{1t} + \beta_2 D_{2t} + \sum_{k=1}^{p} \alpha_k R_{i,t-k} + D_{1t} \sum_{k=1}^{p} \theta_k R_{i,t-k} + D_{2t} \sum_{k=1}^{p} \phi_k R_{i,t-k} + \varepsilon_{it} \qquad (3.5)$$

where ε_{it} is the error term, assumed to be identically independently distributed and p is the optimally chosen lag length based on minimizing SIC.

Since three periods are identified in this study, only two dummy variables are used. The dummy variables are defined as follows: $D_{1t} = 1$ for observations in the crisis period (February 1997 to September 1998) and 0, otherwise. $D_{2t} = 1$ for observations in the post-crisis period (October 1998 to August 2002) and 0, otherwise.

The mean returns for market *i* for each period are given below:

Pre-crisis period $$E(R_{it}) = \beta_0 + \sum_{k=1}^{p} \alpha_k R_{i,t-k} + \varepsilon_{it} \qquad (3.6)$$

Crisis period $$E(R_{it}) = \beta_0 + \beta_1 + \sum_{k=1}^{p} (\alpha_k + \theta_k) R_{i,t-k} + \varepsilon_{it} \qquad (3.7)$$

Post-crisis period $$E(R_{it}) = \beta_0 + \beta_2 + \sum_{k=1}^{p} (\alpha_k + \phi_k) R_{i,t-k} + \varepsilon_{it} \qquad (3.8)$$

The t-test is used to statistically ascertain whether β_1 and β_2 are significantly different from zero individually. β_1 and β_2 are the differential intercept coefficients which indicate how much the intercepts of the crisis and the post-crisis periods differ from the benchmark pre-crisis period. The F-test is used to test the joint significance of θ_k and the joint significance of ϕ_k, where k = 1,2,...,p. θ_k and ϕ_k represent the differential slope coefficients which indicate how much the slope coefficients of the crisis and the post-crisis periods differ from those of the pre-crisis period. Their significance would suggest structural changes in the behaviour of the market returns.

3.2.4 Unit Root Tests of Stationarity

A time series is said to be stationary if its mean, variance, and autocovariance at a given lag remain the same no matter at what point we measure them; that is, they are time invariant. Regressing non-stationary variables on each other can possibly result in misleading inferences about the estimated parameters and the degree of association. So to avoid such spurious results when models are fitted, it is imperative to firstly determine the stationarity of the market indices in the level form. A non-stationary series possesses one or more unit roots. To test for a unit root, we use both the augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979) and the Phillips-Perron (P-P) test (1988). The difference between these two tests lies in their treatment of any 'nuisance' serial correlation. The P-P test tends to be more robust to a wide range of serially correlated and heteroscedastic errors of the test equation.

A time series is said be stationary in level form (zero order of integration or I(0)) if it does not contain a unit root. A time series that contains a unit root and requires first-differencing in order to obtain stationarity is said to be first-order integrated, or I(1). As in all time series, autocorrelation of the residuals from different time periods is inevitably omnipresent. So to minimize this problem, the test equation for the ADF test includes lagged terms of the dependent variable. To accommodate the possible presence of deterministic time trend, a time trend term is included. A drift term is also being added. The test equation is:

$$\Delta P_{it} = \mu + \beta t + \delta P_{i,t-1} + \sum_{k=1}^{p} \theta_k \Delta P_{i,t-k} + \varepsilon_{it} \tag{3.9}$$

The null hypothesis of the ADF test states the presence of a unit root (H_0: $\delta = 0$) and the test statistic, t_τ, tests the rejection or acceptance of this null hypothesis. The computation of this test statistic is similar to a Student-t statistic. As mentioned earlier, the lag length p is chosen by SIC. The distribution of t_τ follows the empirical distribution as tabulated by Mackinnon (1991) and not the usual Student-t distribution.

A rejection of the null hypothesis means that P_{it} is integrated of order 0 and is said to be stationary. On the other hand, if the null hypothesis is not rejected, then P_{it} is said to be non-stationary and its order of integration is one or higher. If the latter happens, the next step is to test for the presence of unit root in the first difference of P_{it}, represented by ΔP_{it}. The null hypothesis (H_0: $\delta = 0$) is being tested in the following equation:

$$\Delta^2 P_{it} = \mu + \beta t + \delta \Delta P_{i,t-1} + \sum_{k=1}^{p} \theta_k \Delta^2 P_{i,t-k} + \varepsilon_{it} \tag{3.10}$$

where $\Delta^2 P_{it}$ denotes the second difference of P_{it}.

Similarly, a rejection of the null hypothesis would imply that the first difference of P_{it} is integrated of order zero, and henceforth P_{it} is integrated of order one. In other words, the natural log of the index series is non-stationary while its first difference is stationary.

The ADF test assumes that the errors are statistically independent and have a constant variance. To circumvent these limiting assumptions, Phillips and Perron (1988) developed a generalization of the ADF test. The regression equation of the P-P test for a unit root in P_{it} is given as follows:

$$\Delta P_{it} = \mu + \beta t + \delta P_{i,t-1} + \varepsilon_{it} \qquad (3.11)$$

This equation is estimated by the ordinary least squares method and the t-statistic of the δ coefficient is corrected for serial correlation in the error term by using a non-parametric approach for testing $H_o : \delta = 0$. The lag length employed for the correction of the P-P test statistic in this study follows that of the ADF test equation.

If the unit root tests show that all the five natural log price indices in this study are integrated of order one, we shall then proceed to test for the presence of co-integration among these non-stationary series.

3.2.5 Contemporaneous Relationship

On the study of co-movements among a group of markets, many researchers include a preliminary analysis on the contemporaneous correlations between markets. The correlation pattern may, to a certain extent, reflect the economic integration between countries. Generally, the more integrated the economies of two countries are, the stronger is the correlation between their stock movements (Eun and Shim, 1989).

The objective of correlation analysis is to measure the strength or degree of linear association between two market returns. Here, we use the correlation coefficient to measure the strength of contemporaneous relationship and its value ranges from –1 for perfect negative correlation up to +1 for perfect positive correlation.

The sample correlation coefficient, r, between two market returns, R_{it} and R_{jt}, can be computed directly using the following formula:

$$r = \frac{\sum_{t=1}^{n}(R_{it}-\bar{R}_i)(R_{jt}-\bar{R}_j)}{\sqrt{\sum_{t=1}^{n}(R_{it}-\bar{R}_i)^2}\sqrt{\sum_{t=1}^{n}(R_{jt}-\bar{R}_j)^2}} \qquad (3.12)$$

where $\bar{R}_i = \sum_{t=1}^{n} R_{it}/n$ and $\bar{R}_j = \sum_{t=1}^{n} R_{jt}/n$. Although r is a measure of linear association between two series, it does not imply any cause-and-effect relationship. If R_{it} and R_{jt} are normally distributed (or when n is large), the test statistic for the correlation coefficient for testing the null hypothesis H_0: $\rho = 0$ is given by

$$t = r\sqrt{\frac{n-2}{1-r^2}} \qquad (3.13)$$

where t follows a t-distribution with n-2 degrees of freedom under the null hypothesis and ρ is the true underlying (population) correlation coefficient.. The alternative hypothesis of interest in this study is H_1: $\rho > 0$, and thus the critical rejection region is $t > t_{\alpha,n-2}$. The null hypothesis is tested against a right side alternative hypothesis because the contemporaneous returns of the ASEAN-5 markets generally move in the same direction.

3.2.6 Error Correction Model and Cointegration

Conventionally, when two time series are non-stationary in the level form but some linear combination of them is stationary, they are said to be cointegrated and these series do not move far away from each other in the long run. The implication is that, in the short-run, any deviations from this long-run equilibrium will feed back on the changes in the variable in the system in order to force the movement towards the long-run equilibrium. When time series are cointegrated, an error correction model (ECM) exists and deviations from the long-run relationship are modeled through the inclusion of an error correction term (ECT). The long-run relationship is implied through the significance (or otherwise) of the t-test of the lagged ECT. This lagged ECT contains long-term information since it is derived from the long run cointegrating relationship.

Specifically, in this study, we shall employ the Johansen's (1991) methodology to investigate whether the ASEAN-5 equity markets are cointegrated in each of the three periods. If they are cointegrated in a particular period, then these markets are said to share one or more long-term equilibrium relationships in that period. The actual number of cointegration (or equilibrium) relationships found will result in a corresponding number of error-correction terms which are included in their lagged levels in the ECM. This will also rule out the use of modeling any dynamic relationships through first differenced vector autoregressions (which will be discussed in the next section) as these models will be misspecified.

The Johansen's (1991) cointegration test applies the maximum likelihood procedure to determine the presence of cointegrating vectors in the non-stationary series. It considers an error correction model of the following form:

$$\Delta \mathbf{y}_t = \gamma + \xi \mathbf{y}_{t-1} + \sum_{k=1}^{p} \Gamma_k \Delta \mathbf{y}_{t-k} + \varepsilon_t \qquad (3.14)$$

where $\mathbf{y}_t = [P_{1t}\ P_{2t}\ P_{3t}\ P_{4t}\ P_{5t}]'$, γ is the 5 x 1 column vector of constants, Γ_k and ξ are the 5 x 5 matrices of coefficients and $\varepsilon_t = [\varepsilon_{1t}\ \ \varepsilon_{2t}\ \ \varepsilon_{3t}\ \ \varepsilon_{4t}\ \ \varepsilon_{5t}]'$. The i,j th component of Γ_k measures the direct effect that a change in the return of the j th market would have on the i th market after a lag of *k* periods.

This way of specifying the system contains information on both short- and long-run adjustments to changes in $\mathbf{y}_t$, via the estimates of Γ_k and ξ respectively. If the coefficient matrix ξ has reduced rank *r* < 5, then there exist matrices α and β, both of order 5 x *r*, such that $\xi = \alpha\beta'$ and $\beta' y_t$ is stationary. Each column of the β matrix gives an estimate of a cointegrating vector. The cointegrating vector is not identified unless some arbitrary normalization is imposed. EViews adopts the normalization such that the *r* cointegrating equations are solved for the first *r* variables in the $\mathbf{y}_t$ vector as a function of the remaining k-*r* variables. However, one consequence of this normalization is that the normalized vectors that EViews provides will not, in general, be orthogonal, despite orthogonality of the unnormalized coefficients.

The elements of α are known as the adjustment parameters in the ECM and they indicate the speeds of adjustment. The rank r of the coefficient matrix ξ gives the number of co-integration relationships. This matrix is estimated in an unrestricted form, and then we test whether the restrictions implied by the reduced rank of ξ can be rejected. The null hypothesis for testing the presence of at least r cointegrating vectors against the alternative hypothesis of more than r cointegrating vectors) is based on the likelihood-ratio trace test-statistic given by:

$$Q_r = -n \sum_{k=r+1}^{5} \ln(1-\lambda_k) \tag{3.15}$$

where r = 0, 1, 2, 3 or 4 and λ_k is the defined k-th largest eigenvalue. The critical values for the test can be obtained from Osterwald-Lenum (1992). In this cointegration test, it is assumed that there is a linear deterministic trend in the data and that the cointegrating equations have only intercepts but no trend.

Thus, through the Johansen's (1991) cointegration test, we can determine statistically whether the stock markets move together in the long run in each of the three periods while allowing for the possibility of short-run divergence. In summary, if co-integration does exist among the non-stationary stock indices in the level form, an ECM would then be established prior to conducting the causality test for determining short-term causal effects. Otherwise, we would use a vector autoregression (VAR) model.

Compared to its predecessor, namely the Engle-Granger (1987) procedure, Cheung and Lai (1993) demonstrated that the Johansen's (1991) method is more robust and considerably more efficient. It is less sensitive to departure from the presumed Gaussian errors and serial independence, and has a higher empirical power. In fact, Masih and Masih (1999) highlighted several main advantages of the Johansen's procedure over the Engle-Granger two-step approach: (i) the Johansen's procedure does not, a priori, assume the existence of at most a single cointegrating vector. Instead, it explicitly tests for the number of cointegrating relationships; (ii) it is not sensitive to the choice of the dependent variable in the cointegrating equation as it assumes that all the variables are endogenous; (iii) this procedure is established on a unified framework for estimating and testing cointegrating relations within the ECM formulation.

3.2.7 Vector Autoregressive Model

As mentioned earlier, if a set of non-stationary series is not cointegrated, then there will not exist any long-term equilibrium among these series. An unrestricted VAR (without the error correction terms) will be applied to the stationary market returns to examine the interdependence structure of the group of ASEAN-5 stock markets.

The VAR analysis, developed by Sims (1980), estimates a set of unrestricted reduced form equations that have a uniform set of lagged dependent variables of every equation as regressors. This means that VAR estimates a dynamic simultaneous equation system that is not bounded by a priori restrictions on the structure of relationships. This form of analysis is equivalently a flexible approximation to the reduced form of the correctly specified, albeit unknown, model of the actual economic structure. To numerous researchers (for instance, Eun and Shim, 1989), it is appealing to use VAR analysis considering that large-scale structural models are very often misspecified. Besides studying market interdependence, the empirical findings of VAR analysis are also frequently employed to gain useful insights into the international transmission mechanism of equity market movements.

Mathematically, a vector autoregression equation of order p, VAR(p), will be given by:

$$\Delta \mathbf{y}_t = \gamma + \sum_{k=1}^{p} \Gamma_k \, \Delta \mathbf{y}_{t-k} + \varepsilon_t \qquad (3.16)$$

The notations are as defined for the ECM in Equation (3.14).

3.2.8 Granger-causality Test

The Granger-causality test in EViews runs the following bivariate regressions for all possible pairs of (R_i, R_j) series in the group:

$$R_{it} = \alpha_{10} + \sum_{k=1}^{p} \alpha_{1k} R_{i,t-k} + \sum_{k=1}^{p} \beta_{1k} R_{j,t-k} + \varepsilon_{it} \qquad (3.17)$$

$$R_{jt} = \alpha_{20} + \sum_{k=1}^{p} \alpha_{2k} R_{j,t-k} + \sum_{k=1}^{p} \beta_{2k} R_{i,t-k} + \varepsilon_{jt} \qquad (3.18)$$

In Equation (3.17), R_{jt} is said to Granger-cause R_{it} if the F-statistic leads to the rejection of the null hypothesis H_0: $\beta_{1k} = 0$ for all k. Similarly, in Equation (3.18), R_i is said to Granger-cause R_j if the F-statistic leads to the rejection of the null hypothesis H_0: $\beta_{2k} = 0$ for all k. Two-way causation or feedback occurs when both these null hypotheses are rejected. It is important to note that if R_{it} is found to Granger-cause R_{jt}, it does not imply that R_{jt} is the effect or result of R_{it}.

However, Masih and Masih (1999) pointed out that through the error correction terms, the error correction model opens up an additional channel for Granger-causality to emerge that is ignored by the simple bivariate lead-lag relationships. The standard F-test in the VAR framework is only useful in capturing short-run temporal causality. Moreover, Masih and Masih (1999) emphasized that while examinations of bivariate relationships may provide additional insight, they should serve at best as a pre-requisite for a more thorough analysis of relationships among stock market indices in a multivariate setting. The authors listed three approaches in which we can determine Granger-causal relationships through multivariate regression. The Granger-causality can be determined through the statistical significance of

(a) the coefficient of each of the lagged ECTs (ξ's in Equation (3.14)) by separate t-test;
(b) at least one coefficient of the lags of each explanatory variables (Γ's in Equation (3.14)), taken in turn, by using the Wald test;
(c) all the coefficients in (a) and (b) by using the Wald test.

In this study, we shall use a multivariate framework to test for Granger-causality. In the case of VAR, where there is no ECT, approach (b) will be more apt. Conversely, for those periods that the five equity markets are cointegrated (hence an ECM is used to analyze linkages), approach (c) will be adopted. As a simple example, suppose that the five markets are cointegrated with one cointegrating relation in a particular period and we aim to test whether the Malaysian equity market is Granger-caused by the other four ASEAN equity markets. The following single equation of the ECM system will be fitted:

$$R_{mal,t} = \delta_0 + \vartheta z_{t-1} + \sum_{k=1}^{p} \delta_k R_{phi,t-k} + \sum_{k=1}^{p} \varphi_k R_{tha,t-k} + \sum_{k=1}^{p} \psi_k R_{ind,t-k} + \sum_{k=1}^{p} \sigma_k R_{mal,t-k}$$

$$+ \sum_{k=1}^{p} \mu_k R_{\sin,t-k} \qquad (3.19)$$

where the ECT is given by

$$z_{t-1} = a_0 + R_{phi,t-1} + a_1 R_{tha,t-1} + a_2 R_{ind,t-1} + a_3 R_{mal,t-1} + a_4 R_{\sin,t-1} \qquad (3.20)$$

and R_{mal}, R_{phi}, R_{tha}, R_{ind} and $R_{\sin}$ denote the returns for the stock market of Malaysia, Philippines, Thailand, Indonesia and Singapore.

If the returns are entered in the order of the Philippines, Thailand, Indonesia, Malaysia and Singapore, the ECT will be normalized on the first variable, that is, the Philippines (see Section 3.2.6)

The hypotheses for testing whether the Singaporean equity market Granger-causes that of Malaysia are given below:

H_0: $\vartheta = 0$ and $\mu_k = 0$ for all k

H_1: At least one of the restrictions is not true

Rejection of the null hypothesis would suggest that the Singapore equity market does Granger-cause the Malaysian market.

The above regression is an unrestricted model and its corresponding restricted model used in the joint F-test is as follows:

$$R_{mal,t} = \delta_0 + \sum_{k=1}^{p} \delta_k R_{phi,t-k} + \sum_{k=1}^{p} \varphi_k R_{tha,t-k} + \sum_{k=1}^{p} \psi_k R_{ind,t-k} + \sum_{k=1}^{p} \sigma_k R_{mal,t-k} \tag{3.21}$$

The test statistic for causality is

$$F = \frac{(RSS_R - RSS_U)/(p+1)}{RSS_U /(n-5p-2)} \tag{3.22}$$

where RSS_R and RSS_U are the residual sums of squares of the restricted model (3.21) and unrestricted model (3.19), respectively. In computing this F-statistic, we assume that the errors are independent, as well as identically and normally distributed.

If the five markets are not cointegrated, then the unrestricted model will not contain the error correction term, z_{t-1}, in Equation (3.19). The null hypothesis is $H_o: \mu_k = 0$ for all k. EViews provides the computation of the F-statistic by imposing the coefficient restrictions under the null hypothesis.

3.2.9 Variance Decomposition and Impulse Response Analyses

By using VAR analysis, we can study the international transmission mechanism of equity market movements. Eun and Shim (1989) referred to such analysis as an innovative accounting technique. It basically comprises the forecast variance decomposition and impulse response analyses. These analyses allow us to measure the relative importance of random innovations of each market in generating variations of returns in itself as well as in the other four markets in response to the shock.

The autoregressive system given in Equation (3.16) contains complicated cross-equation feedbacks in which the practical implications of coefficients are difficult to describe. Sims (1980) showed that it would be better to analyze the system's reaction to typical random shocks by tracing out the system's moving average representation. Using Wold's theorem (Wold, 1983), Equation (3.16) can be transformed into a moving average representation via successive substitutions to yield:

$$\Delta \mathbf{y}_t = \sum_{k=0}^{\infty} \Pi_k \varepsilon_{t-k} \tag{3.23}$$

where the i,jth element of the Π_k matrix of dimension 5 x 5 is the impulse response of the return of market *i* after k periods, due to a one-unit increase in the innovation of market *j* in time t, holding all other innovations at all dates constant. This means the vector $\Delta \mathbf{y}_t$ can be read as a function of current and past one-step ahead forecast errors of its own, as well as from other variables.

The Π_k matrix can be used to trace out the time path of the effects of the shocks in any one market. Though the vector ε_t is by construction serially uncorrelated with $\Delta \mathbf{y}_t$, there is the possibility that the matrix portrays contemporaneous correlations. This means that the elements in ε_t can interact with one another instantaneously. The system for studying the impact of a unique shock of each market is therefore not identified. It is therefore necessary to impose additional restrictions on the VAR system in order to identify the impulse responses. For this purpose the Cholesky decomposition (see Appendix of Eun and Shim, 1989) is commonly used. By defining the obtained error vector as ε_t = $\boldsymbol{V}\boldsymbol{e}_t$, where $\boldsymbol{V}$ is a lower triangular matrix and $\mathbf{e}_t$ is a set of uncorrelated innovations given by $\mathbf{e}_t = [e_{1t} \quad e_{2t} \quad e_{3t} \quad e_{4t} \quad e_{5t}]'$, we would obtain:

$$\Delta \mathbf{y}_t = \sum_{k=0}^{\infty} \Pi_k \, \boldsymbol{V} \mathbf{e}_{t-k} \tag{3.24}$$

Alternatively,

$$\Delta \mathbf{y}_t = \sum_{k=0}^{\infty} \eta_k \, \mathbf{e}_{t-k} \tag{3.25}$$

where $\eta_k = \Pi_k \boldsymbol{V}$ is the matrix of decomposed impulse responses of dimension 5 x 5. Thus the *i,j*th element of the η_k matrix corresponds to the impulse response of market *i* to a unique one-unit increase in the return of market *j*, after k periods, through the dynamic structure of the VAR. Multiplying this value by the constant $\sqrt{Var(e_{jt})}$ would standardize that one-unit increase of the returns of market j to one standard error.

The forecast error variances of each stock market can be decomposed into innovations coming from the other four markets. The proportion of the m-step ahead forecast error (mean squared error) for every component of $\Delta \mathbf{y}_t$ explained by e_{1t}, e_{2t}, e_{3t}, e_{4t} and e_{5t} can be obtained for different values of m. This approach, which allows computation of the portion of the total variance of $\Delta \mathbf{y}_{it}$ that is due to the disturbance e_{jt}, is known as variance decomposition.

The orthogonalization of innovations based on the Cholesky decomposition imply the following: The first market has an immediate impact on the other four markets in the VAR system; a shock in the second market has an immediate impact on the other three markets, excluding the first market, and so on. The practical implication of this is that when performing impulse response and variance decomposition analyses, it is essential to pre-specify the ordering of the markets. The approach adopted in this study is to follow the rank of the total market capitalization, that is, from the largest to the smallest.

3.3 Statistical Software

Microsoft computer software, namely Access and Excel, are used in the preparation of the dataset. All computations and conducting of statistical tests are through EViews version 3.