# Chapter 2

## Research Methodology

### 2.1 Data and Sample Selection

This dissertation concentrates on examining interdependencies among ASEAN emerging stock markets (Indonesia, Malaysia, Philippines, Thailand and Vietnam), incorporating with the international gold market. Nowadays, the time series data in the financial markets are very plentiful. They are not only monthly or weekly, but also daily data and even the tick data. Commonly, many time series data on macro economics, financial markets, commodity markets, etc. can be conditionally accessed via some providing agencies *i.e.*, Datastream, Reuter, Bloomberg, etc. In addition, time series data are provided from many sources via internet. For instance, we can find a lot of time series data in stock markets over the world at www.finance.yahoo.com and daily international gold prices in London gold market at www.kitco.com. For some countries, the time series data on the stock markets may not be published there, but it can be accessed from some bank's websites *i.e.*, daily data of the Stock Exchange of Thailand can be downloaded from the website of the Siam Commercial Bank Asset Management (www.scbam.com) and the trading data in the Vietnam stock market are available on the website of the Vietnam Bank for Investment and Development's Securities Company (www.bsc.com.vn).

This study uses daily closing data in the 5 ASEAN emerging stock market indexes that can be obtained from Reuters or other aforementioned sources, and daily international gold market prices from the P.M London Gold Fix that can be accessed at www.kitco.com. Along with the daily time series data of the selected markets obtained, other sources of secondary data and information relating to the study are referred in the dissertation such as statistic data, articles, textbooks, theses, working papers and so on.

## 2.2 Data Analysis Methods

To obtain the objectives of the study, we work on time series data of the sample markets at both market price index series (the level series) and market returns series. In the dissertation, we use various methods for time series data analysis in terms of market linkages, cointegrations, volatility and volatility transmissions in the financial markets. Specifically, in order to see how the selected market price indexes are interlinked, methods are employed for the study including unit root test, Granger causality test and Johansen cointegration technique.

On the other hand, regarding daily return series of the sample markets, we examined volatility behavior in each market and volatility transmissions across the sample market returns, so both univariate and multivariate GACRH models are involved. For univariate volatility model, we select GARCH and GJR models, while multivariate GARCH models such as CCC, VARMA-GARCH, VARMA-AGARCH and DCC are applied for the study. Following the conventional approach, daily returns of the international gold and ASEAN emerging stock markets, $r_{i,t}$, are computed as the percentage of natural logarithmic difference in their daily prices *i.e.*, $r_{i,t} = 100 \times [\ln(p_{i,t}) - \ln(p_{i,t-1})]$, where $p_{i,t}$ and $p_{i,t-1}$ are the closing prices of market $i$ on the days $t$ and $t$-$1$, respectively.

### 2.2.1 Unit Root Test

When working on time series data, we need to check whether or not time series data used in the study are stationary, the Augmented Dickey Fuller (ADF) test is used for a set of the selected market price indexes. The ADF test for a unit root uses the formulation given in (2.1). The unit root test is carried out under the null hypothesis $\psi = 0$ against the alternative hypothesis of $\psi < 0$. If the null hypothesis in the ADF test, $\psi = 0$, can not be rejected, $x_t$ is non-stationary. However, if the null hypothesis of a unit root in the first differences of the level series can be rejected, these series are integrated of order one, denoted $I(1)$. Therefore, it is sufficient for performing cointegration tests for the level series.

$$\Delta x_t = \alpha_0 + \alpha_1 trend + \psi x_{t-1} + \sum_{j=1}^{p} \alpha_j \Delta x_{t-j} + u_t \qquad (2.1)$$

where $\Delta x$ is the first difference of $x_t$ and $p$ is the lag-length of the augmented terms for $x_t$.

### 2.2.2 Granger Causality Test

To see how the selected market price indexes can explain each other Granger causality test (Granger, 1969) is employed to determine directions of causality between the market pairs. Usually causal relations are tested both ways *i.e.,* from $x_1$ to $x_2$ and vice versa as specified below,

$$x_{1,t} = \sum_{i=1}^{n} \alpha_i x_{1,t-i} + \sum_{i=1}^{n} \beta_i x_{2,t-i} + \varepsilon_{1,t} \qquad (2.2)$$

$$x_{2,t} = \sum_{i=1}^{n} \lambda_i x_{2,t-i} + \sum_{i=1}^{n} \delta_i x_{1,t-i} + \varepsilon_{2,t} \tag{2.3}$$

where, $x_{1,t}$ and $x_{2,t}$ are the market price indexes; $\alpha_i$, $\beta_i$, $\lambda_i$ and $\delta_i$ are coefficients in the regressions; $n$ is the lag length used; and $\varepsilon_{1,t}$ and $\varepsilon_{1,t-i}$ are assumed to be uncorrelated innovations *i.e.*, $E(\varepsilon_{1,t}, \varepsilon_{1,t-i}) = 0$. A time series $x_1$ is said to Granger-cause $x_2$ if it can be shown, through the overall F-test on lagged values of $x_1$ and of itself, that those lagged values of $x_1$ will provide statistically significant explanations about future values of $x_2$. The Granger test can be applied to each pair of variables only. Then, causal relations are inferred through the overall statistical significance of coefficients of each equation. Therefore, the null hypotheses of the test are given as

$H_0^{(1)}$: $\beta_1 = \beta_2 = \ldots = \beta_n = 0 \rightarrow x_{2,t}$ does not Granger-cause $x_{1,t}$ for Eq.(2.2),

$H_0^{(2)}$: $\delta_1 = \delta_2 = \ldots = \delta_n = 0 \rightarrow x_{1,t}$ does not Granger-cause $x_{2,t}$ for Eq.(2.3).

The generalized $F$ statistic used in the Granger causality test is written as follows:

$$F = \frac{(RSS_R - RSS_{UR})/n}{RSS_{UR}/(T-2n)} \tag{2.4}$$

where $RSS_R$ is the residual sum of squares in a restricted regression that lagged terms of the exogenous variable are not included; $RSS_{UR}$ denotes the residual sum of squares in an unrestricted regression, in which all lagged terms are included; $n$ is the lag length used; and $T$ is the total number of observations in the time series data.

According to the two hypotheses defined above, four cases may occur: *(i)* No rejection of $H_0^{(1)}$ and rejection of $H_0^{(2)}$ imply causality from $x_{1,t}$ to $x_{2,t}$, *(ii)* No rejection of $H_0^{(2)}$ and rejection of $H_0^{(1)}$ mean causality in the reverse direction, *(iii)* If both hypotheses are rejected, meaning that there is a lead-lag relation between the two variables, and *(iv)* If both can not be rejected, no Granger causality exists between the variables.

Although the test is very straightforward, the lag length used in the regression is unknown. To identify the number of lags, both the minimum Akaike information criteria (AIC) and Schwarz criteria (SC) should be considered to identify an appropriate lag length for each pair. Generally, more lags should be included, if the number of observations is large enough.

### 2.2.3 Johansen Cointegration Technique

The Johansen (1988) and Johansen & Juselius (1990) test framework is used for testing the presence of the long-run relationship between two or more $I(1)$ variables. If the test shows the presence of cointegrating relationships given by the linear combinations between them, called the cointegrating vectors, it implies a long-run equilibrium relationship. As a result, there exists the vector error correction model (VECM) that measures speed of adjustment to the long-run equilibrium in the cointegrated variables. The procedure of Johansen test begins with a vector autoregressive (VAR) model of order p below



$$x_t = \omega + \sum_{i=1}^{p} A_i x_{t-i} + \varepsilon_t \tag{2.5}$$

where $x_t$ is an $(m \times 1)$ vector of variables $(x_{1t}, x_{2t}, \ldots, x_{mt})'$, which are $m$ level series, $\omega$ is a vector of constants, $A_i$ is a $(m \times m)$ matrix of coefficients and $\varepsilon_t$ is a vector of error terms, and $p$ is the lag length in the variables in the system. If the variables in the vector $x_t$ are integrated of order one, $I(1)$, it implies that the linear combination of one or more of these series may exhibit a long-run relationship among them. This leads to using the Johansen (1988) and Johansen & Juselius (1990) method for further explorations in the sample market price indexes in our study. The method can be briefly expressed as follows

$$\Delta x_t = \omega + \sum_{i=1}^{p} \Gamma_i \Delta x_{t-i} + \Pi x_{t-1} + \varepsilon_t \qquad (2.6)$$

where $x_t$ is a $(m \times 1)$ vector of the sample market price indexes, $\omega$ is the $(m \times 1)$ vector of constant terms and $\varepsilon_t$ is a vector of error terms. $\Gamma_i$ denotes the $(m \times m)$ matrix of coefficients, containing information regarding the short-run relationships among the sample market price indexes. $\Pi$ are $(m \times r)$ matrix, reflecting the possible long-run relationship between the sample market price indexes, where $r$ is the rank of $\Pi$ so that $r \leq m - 1$. The Johansen procedure is to decompose the matrix $\Pi$ into two $(m \times r)$ sub-matrices, $\alpha$ and $\beta$, such that $\Pi = \alpha\beta'$. The matrix $\beta$ is called the matrix of cointegrating vectors, representing the possible long-run relationship between the sample market price indexes, and $\alpha$ is defined as the matrix of error correction coefficients that measure speed of adjustment in the cointegrated variables to their long-run equilibrium. The Johansen technique is based on the maximum likelihood estimation of $\alpha$ and $\beta'$, and the two computed statistics such as the trace statistic and the maximum eigen-value statistic in order to test for the presence of $r$ cointegrating vectors in the systems. The trace statistic tests the null hypothesis of at most $q$ cointegrating vectors against the alternative hypothesis of $r$ = $n$ cointegrating vectors. Meanwhile, the maximum eigen-value statistic also tests for the presence of $r$ cointegrating vectors against the alternative hypothesis of $r+1$ cointegrating vectors.

### 2.2.4 Symmetric Univariate GARCH Model

The autoregressive conditional heteroscedasticity (ARCH) model was first invented by Engle (1982) and extended then by Bollerslev (1986) to become the generalized ARCH (GARCH) model. The GARCH model is assumed that a positive shock ($\varepsilon_t > 0$) and a negative shock ($\varepsilon_t < 0$) with an equal magnitude has the same impact on the conditional variance, $h_t$. The mean and variance specifications in GARCH($p, q$) model is written as follow

$$r_t = E(r_t \,|\Psi_{t-1}) + \varepsilon_t, \qquad with\ \varepsilon_t | \Psi_{t-1} \sim N(\mu_t, h_t) \tag{2.7}$$

$$h_t = \omega + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{2.8}$$

where $r_t$ is $a$ market return series and $\varepsilon_t = r_t - \mu_t$ is the shocks to the market returns; $\Psi_{t-1}$ denotes the past information available at time $t$; $z_t = \varepsilon_t / \sqrt{h_t}$ are the standardized shocks to the market returns; $h_t$ is the univariate conditional variance of the market returns; $\alpha_j$ denote the ARCH effects, implying the short-run effects of shocks and $\beta_j$ denote the GARCH effects, indicating the contribution of such shocks to long-run persistence ($\sum\alpha_j + \sum\beta_j$). Bollerslev (1986) indicated that $\omega > 0$, $\alpha_j \geq 0$ for $j = 1, \ldots, p$ and $\beta_j \geq 0$ for $j = 1, \ldots, q$ are sufficient conditions for a positive conditional variance $h_{it} > 0$, and $\sum_{j=1}^{p}\alpha_j + \sum_{j=1}^{q}\beta < 1$ is the necessary and sufficient condition for the existence of the second moment. The simplest case is GARCH(1,1) model, *i.e.*, $h_t = \varpi + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$, but has been most widely used in practice so in our study we the GARCH(1,1) model.

To check the structural properties of the univariate volatility model, the second moment and log-moment conditions. Jeantheau (1998) built the log-moment condition for the GARCH(1,1) given by $E(\log(\alpha_1\eta_t^2 + \beta_1)) < 0$, which is sufficient for the QMLE to be consistent. Since the log-moment condition is a weaker regularity condition than the second moment condition, namely $\alpha_1 + \beta_1 \prec 1$. The log-moment condition can be satisfied even when $\alpha_1 + \beta_1 > 1$.

### 2.2.5 Asymmetric Univariate GARCH Model

Behavior of volatility in finance tends to be larger as the stock market index was decreasing than as it was increasing by the same magnitude. To capture asymmetric effects, Glosten, Jagannathan and Runkle (1993) proposed an asymmetric GARCH model, namely GJR. Variance equation in the GJR(*p, q*) model is formulated as follows



$$h_t = \varpi + \sum_{j=1}^{p}(\alpha_j + \gamma_j I(\varepsilon_{t-j} < 0))\varepsilon_{t-j}^2 + \sum_{j=1}^{q}\beta_j h_{t-j} \tag{2.9}$$



where $\omega$, $\alpha$, $\beta$ and $\gamma$ are estimated parameters and $I_t$ is an indicator function taking the values of 1 if $\varepsilon_{t-j} < 0$ (bad news) and zero, otherwise. As specified in (2.9), impact of positive and negative shocks that have the same magnitude on the conditional variance $h_t$ in the GJR model are different. We expect the estimated sign of $\gamma$ to be positive, so that the negative shocks "bad news" have stronger impact than positive ones.

Similar to the selection for the symmetric univariate GARCH(1,1) model in our study, the GJR(1,1) specification is applied. Therefore, when $\varepsilon_{t-1}$ is positive, its effect to the volatility is $\alpha\varepsilon_{t-1}^2$, however when $\varepsilon_{t-1}$ is negative, the total effect of shocks to the volatility is $(\alpha+\gamma)\varepsilon_{t-1}^2$. The GJR(1,1) model is asymmetric if $\gamma$ is significantly different from zero. Ling and McAleer (2002) established the regularity condition for the existence of the second moment of the GJR(1,1) model *i.e.*, $\alpha+\beta+\gamma/2<1$, and the log-moment condition for the GJR(1,1) model *i.e.,* $E(\log((\alpha_1+\gamma_1 I(\eta_t))\eta_t^2+\beta_1))<0$, which is sufficient for consistency and asymptotic normality of the QMLE for GJR(1,1). The log-moment condition for the GJR(1,1) model can be satisfied even when $\alpha_1+\gamma/2+\beta_1 \geq 1$.

### 2.2.6 Multivariate CCC-GARCH and VARMA-GARCH

In recent years, the univariate GARCH model have been extended to the multivariate GARCH (MGARCH) cases to examine the volatility spillovers as well as the conditional correlations between the markets. As implied in Bauwens *et al*.(2006), the spillover effects across markets are measured by lags in shocks and the conditional variances of a market or the covariance of two markets, which appear significantly in the conditional variance equation of other markets.

Generally, the default equation for the means in the MGARCH models could be constant, or AR($p$), or ARMA($p,q$). In our study, the conditional mean equations of daily returns in the selected markets, specified in MGARCH models, can be written as follows,

$$r_{it} = E(r_{it} \mid \Psi_{t-1}) + \varepsilon_{it}, \quad with\ \varepsilon_{it} \mid \Psi_{t-1} \sim N(\mu_{it}, h_{it}) \quad (2.10)$$

$$\varepsilon_{it} = \sqrt{h_{it}}\, z_{it}, \quad with\ z_{it} \sim iid(0,1).$$

where $i = 1...s$ is the number of the sample markets and $t= 1...n$ is the number of observations; $\varepsilon_{it} = r_{it} - \mu_{it}$ are shocks to the market returns; $r_{it}$ are return series of the sample markets that are conditional on the past information ($\Psi_{t-1}$) available at time $t$; $h_{it}$ are the univariate conditional variances of the market returns; $z_{it} = \varepsilon_{it} / \sqrt{h_{it}}$ the standardized innovations to the market returns.

In our study, two constant conditional correlation MGARCH models, namely VARMA-GARCH (Ling and McAleer, 2003) and VARMA-AGARCH (McAleer *et al*., 2009) are employed in order to measure spillovers across the selected markets and to capture the possible asymmetric effects of shocks to the conditional variances. Since the constant conditional correlation (CCC) is maintained in both VARMA-GARCH and VARMA-AGARCH models, we now take a view on how to construct the CCC multivariate GARCH model of Bollerslev (1990). As defined in (2.10), the conditional covariance matrix, $H_t$, in the CCC model is written as follows,

$$H_t = E(\varepsilon_t \varepsilon_t' \mid \Psi_{t-1}) = E(D_t z_t z_t' D_t) = D_t E(z_t z_t') D_t = D_t R D_t \quad (2.11)$$

where $D_{it} = diag(\sqrt{h_{it}})$ is a diagonal matrix of the univariate conditional variances of the sample markets, $R = E(z_t z_t') = D^{-1} H_t D^{-1} = (\rho_{ik})$ is a symmetric positive definite matrix that $(\rho_{ik}) = (\rho_{ki})$ with $\rho_{ik} = 1 \; \forall \; i = k$ (for $i, k = 1, \ldots, s$). Consequently, $R$ is the constant conditional correlation matrix, $\rho_{ik}$, between different pairs of the sample market returns. In the CCC model, Bollerslev (1986) assumed that the univariate conditional variances for the return series, $h_{it}$, follows a univariate GARCH process as

$$h_{it} = \omega_i + \sum_{j=1}^{p} \alpha_{ij} \varepsilon_{i,t-j}^2 + \sum_{j=1}^{q} \beta_{ij} h_{i,t-j} \qquad (2.12)$$

where $i = 1 \ldots s$ is the number of the selected markets, so $\alpha_{ij}$ denote the ARCH effects and $\beta_{ij}$ the GARCH effects on the conditional volatility equations. In the simplest case, the GARCH(1,1) model can be writeen as $h_t = \varpi + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$.

As specified in (2.12), the CCC model assumes that return volatility in each market is independent from others, so there are no shock and volatility spillovers across the markets. However, this assumption may not be realistic, particularly in the context of international integration and market liberation. To capture possibilities of the spillovers across markets, Ling and McAleer (2003) built the VARMA-GARCH model that the lagged terms of shocks and volatilities of other markets are added in the conditional volatilities of a market. As explained in (2.10) and (2.12) for the parameters and notations, they are now continuously used, the multivariate conditional variances of the VARMA-GARCH model can be expressed as,

$$h_{it} = \omega_i + \sum_{i=1}^{s} \sum_{j=1}^{k} \alpha_{ij} \varepsilon_{it-j}^2 + \sum_{i=1}^{s} \sum_{j=1}^{k} \beta_{ij} h_{it-j} \qquad (2.13)$$

### 2.2.7 Multivariate VARMA-AGARCH

As discussed in the GJR model, the existence of asymmetry in the volatility is a highlighted property of the financial time series data that reflects behavior of investors in responding to shocks. The asymmetry exists if the positive and negative shocks with an equal magnitude have different effects on the conditional volatility of a market. It is interesting to realize that both CCC and VARMA-GARCH models assumes that effect of shocks on the conditional volatility is symmetric, possible asymmetric effect is not taken into account. Consequently, McAleer *et al.* (2009) introduced VARMA-AGARCH model, in which the CCC and VARMA-GARCH models are nested within the VARMA-AGARCH. The multivariate conditional variances of the VARMA-AGARCH model can be expressed as,

$$h_{it} = \omega_i + \sum_{i=1}^{s}\left[\sum_{j=1}^{k}\alpha_{ij} + \sum_{j=1}^{k}\gamma_{ij} I(\varepsilon_{i,t-j} \le 0)\right]\varepsilon_{i,t-j}^{2} + \sum_{i=1}^{s}\sum_{j=1}^{k}\beta_{ij} h_{i,t-j} \qquad (2.14)$$

where $I(\square_{i,t-j} \le 0)$ is the indicator function that takes the values of 1 if $\varepsilon_{i,t-j} \le 0$ (*i.e.,* bad news) and zero, otherwise. It is obviously that the multivariate equation in (2.14) is simplified to the univariate asymmetric case of Glosten, Jagannathan and Runkle (1992) *i.e.,* GJR model, if $s=1$ (a single market only). If $\gamma_i = 0$ for all the cases, VARMA-AGARCH becomes VARMA-GARCH. The parameters in (2.10), (2.13) and (2.14) can be obtained from the quasi maximum likelihood estimator (QMLE), see Ling and McAleer (2003) and McAleer *et al.* (2009) for the details.

### 2.2.8 Multivariate DCC-GARCH

The assumption of the constant conditional correlations may often be reasonable over shorter time periods. To relax the assumption of the constant conditional correlations, Engle (2002) proposed the dynamic conditional correlation (DCC) model, which is a generalization of the CCC model. The DCC model can be expressed as follows

$$\varepsilon_t \mid \Psi_{t-1} \sim N(0, H_t)\,,$$
$$H_t = D_t R_t D_t \tag{2.15}$$

Let $H_t$ be the conditional covariance matrix, $D_t = diag(\sqrt{h_{it}})$ a diagonal matrix of the univariate conditional variance equations, $R_t$ the a conditional correlation matrix. The conditional variance ($h_{it}$) in the $D_t$ is assumed to follow a univariate GARCH model as given in (2.12). The difference between DCC and CCC models is that DCC model allows the conditional correlation matrix, $R_t$ , to be time varying *i.e.,* $R_t=\{\rho_{ik,t}\}$. The DCC estimation of conditional variances and correlations is conducted through two stages, so that the estimation of time-varying correlation matrix is easier. For instance, in the first stage univariate volatility parameters in (2.12) are estimated for each return series, using GARCH model and so the standardized shocks, $z_{it} = \varepsilon_{it} / \sqrt{h_{it}}$ , are obtained. In the second stage, the standardized residuals, $z_{it}$, obtained from the first stage are used to estimate the parameters of the dynamic conditional correlations, $q_{ik,t}$,. In our study, we employ the DCC(1,1) version of Engle (2002), so the model can be written as follows

$$R_t = \{diag(q_{ii,t})^{-1/2}\}\ q_{ik,t}\ \{diag(q_{kk,t})^{-1/2}\} = \{\rho_{ik,t}\}, \quad (2.16)$$

$$q_{ik,t} = (1-\theta_1-\theta_2)\overline{\rho}_{ik} + \theta_1 z_{i,t-1}z_{k,t-1} + \theta_2 q_{ik,t-1}, \quad (2.17)$$

(for $i, k = 1,..., s$ and $t=1,...,n$)

Engle (2002) defined that $\overline{\rho}_{ik}$ in (2.17) is the unconditional correlation between $z_{i,t}$ and $z_{k,t}$ that has unit variance, obtained from the first stage, and (2.16) is used to standardize the matrix estimated in (2.17). $\theta_1$ and $\theta_2$ are parameters to be estimated, if the estimates of $\theta_1$ and $\theta_2$ are significantly different from zero, then conditional correlation in the whole is not constant. On the contrary, if the estimates of $\theta_1$ and $\theta_2$ are not significant, then $q_{ik,t}$ in (2.17) can be interpreted as the CCC model. The DCC model is estimated using the maximum likelihood estimator (MLE). Engle (2002) showed the log-likelihood function as

$$L = -\frac{1}{2}\sum_{t=1}^{n} s\log(2\pi) + \log|R_t| + z_t^{'}R_t^{-1}z_t \quad (2.18)$$

It is assumed that $z_t$ in (2.18), the standardized residual series obtained from the first stage, $z_t = \varepsilon_t / \sqrt{h_t}$, is normally distriduted with zero mean and variance, $R_t$ (i.e., $z_t \sim N(0, R_t)$). Interstingly, $R_t$ plays the roles as a variance matrix of the standardized residuals and also as a correlation matrix of the residual series $\varepsilon_{it} = r_{it} - \mu_{it}$.