CHAPTER 4
RESEARCH METHODOLOGY

To test the existence of bubble in a stock market, the most important process is mentioned above: to define the excess between the stock market prices and the fundamental values, but the key is how to define the fundamental values? And that question was a barrier for every study in bubbles, then it could make a different conclusion in the same market. From that problem, each scholar employed different methodologies to detect the existence of bubbles:

The first methodology is applied to test bubble premium. Rappoport and White (1991) tried to prove that the bubble existed by used the broker's loan rate model for asset pricing instead of dividends which could not have any evidence of existence of bubble. And Hardouvelis (1988) implemented their test based on to find positive of bubble premium (the extra return rate investors expect to receive over their required rate of return) and the result was showed that there were exceeded between actual receive (discount factor) and the sum of risk premium with bubble premium in Japan and the United States or bubbles existed in those two countries, in contrast to Great Britain, evidence for the presence of bubble is weak.

The second one uses to detect directly the existence of bubbles in the stock market prices, that is excess volatility test, including variance bounds test; the test compares the variance of actual and fundamental values. This test is employed by Shiller (1981), Shiller's test proved that there is an excess in actual price volatility, Marsh and Merton (1986) failed to detect existence of bubble since the nonstationary of dividend and stock prices.

The third technique is to find the relationship between the stock market prices with their fundamental values via cointegration test, Fama and French (1988), and other scholars employed this to test the cointegration between dividend process and the stock market prices. If the dividend and stock prices are cointegrated, there is no exist bubble in the stock market.

All of above frameworks and methodologies contain some problem such as the misspecified components for the tests, and lead to different conclusions in the same markets. Especially, the dividend process plays an important role to do the test, since the dividend process cannot predict in the same technique from each study, the different conclusions are not avoidable. McQueen and Thorley (1994) constructed a new technique to detect the bubble in the stock market which supports to overcome the misspecification in kinds of tests above. Interestingly, the new test is suitable for discrete data or monthly data and does not concentrate into the fundamental values as well as the dividend process as traditional tests; the test was employed by: Chan, McQueen et al. (1998) to test speculative bubbles in six Asian stock markets (Hongkong, Japan, Korean Malaysia, Thailand and Taiwan) and the US stock market, after that Watanapalachaikul and Islam (2003) used in their study to test the speculative bubbles in Thai stock market, and Jirasakuldech, Emekter et al. (2007) applied the same technique to test again the rational bubbles in Thai stock market; Mokhtar, Ms.Nassir et al. (2006) detected the rational bubbles successful in Malaysian stock market; or recently, Haque, Shaoping et al. (2008) found the existence of bubbles in Chinese stock market with the same technique.

Besides that, the Vietnamese stock market has just established for 8 years, the data series is very small, before 2006, there were around 30 listed shares, and the rest had been listed during 2006-2008, the dividend series is not good enough for finding dividend process. Practically, the dividend is decided for a whole year, investors could use earning per share, which is provided in quarterly to the public, for their decision to hold a security in the stock market. From those barriers, to detect the existence of bubbles in the Vietnam stock market, the suitable framework is testing the bubbles in the stock market returns as mentioned above. To find the impact/responses of stock prices and other monetary variables, we develop an estimator that identifies the response of stock returns based on the heteroskedasticity of monetary policy shocks. The study applies VAR to capture the dynamics of stock returns and alternatively implements through instrumental variables. Similar finding is that an increase in short-term interest rates results in a decline in the stock market prices.

4.1 Econometric methodology

4.1.1 Unit root test

Mostly the data in a time series is nonstationary, a unit roots test must do initially for testing the stationary or nonstationary of economic data, since the OLS regression could make incorrect result if the data is nonstationary series, since the OLS could have a significant result in a high R2, t-statistics, but that result is without any theoretical meaning.

Theoretically, the properties of stationary process require that the data series must have a constant mean and variance period to period. It also needs a stationary for covariance in different periods:

$$E\left(X_t\right)=E\left(X_{t+m}\right) \quad \forall t,m$$

$$\operatorname{var}\left(X_t\right)=\operatorname{var}\left(X_{t+m}\right) \quad \forall t,m$$

$$\text{and } \operatorname{cov}\left(X_t,X_{t+k}\right)=\operatorname{cov}\left(X_{t+m},X_{t+k+m}\right) \quad \forall t,k,m$$

For example, $y_t=\phi y_{t-1}+u_t$, with $u_t \sim I(0)$: the null hypothesis test is $H_0:\phi=1$, if this hypothesis cannot be rejected, it means that the series has a unit root or the data is nonstationary. Hence, to test stationary of data, the test will be to check whether a unit root exists or not. And the most popular test which is employed in this study is Augmented Dickey Fuller (ADF) test.

Augmented Dickey Fuller test

Based on the Dickey Fuller test which allows testing one lag, the ADF test could expand to more than two lags and this test could be formed as:

$$\Delta y_t=\phi y_{t-1}+\sum_{i=1}^{k}\varphi_i\Delta y_{t-i}+u_t$$

$$\text{or } \Delta y_t=\beta+\phi y_{t-1}+\sum_{i=1}^{k}\varphi_i\Delta y_{t-i}+u_t$$

$$\text{or } \Delta y_t = \beta + \phi y_{t-1} + \gamma t + \sum_{i=1}^{k} \varphi_i \Delta y_{t-i} + u_t$$

where Δ is different operator; β, ϕ, γ and φ_i are coefficients to be estimated in the test; y_t is the variable to examine; t is time trend to test the trend-cycle of the variable; and the last term ut is white noise error term. The null hypothesis as mentioned above that $H_0 : \phi = 1$ (having a unit root or the series is nonstationary), and the alternative hypothesis is $|\phi| < 1$ (the root outside unit circle, the series is stationary). If the null hypothesis can be rejected, the series is stationary at level and $y_t \sim I(0)$. But, if the null hypothesis cannot reject, the other test could be conducted to test whether the series are cointergrated or not.

Since the model with lagged values are often used for forecasting, it needs to look for measures that have produced more better results for “assessing out of sample” prediction. The adjusted R2 is one possibility by increasing the value of adjusted R2. Others are Akaike (1973) information criterion (AIC(p)) and Schwatz’s Bayesian criterion (SC(p)):

$$AIC(p) = \ln \frac{e'e}{T} + \frac{2p}{T}$$

$$\text{and } SC(p) = AIC(p) + \left(\frac{p}{T}\right)(\ln T - 2)$$

where e’e is residual sum of square.

T is number of observation

p is appropriate lag lengths

As Greene (2003), if some maximum P is known, then p<P can be chosen to minimize AIC(p) then using SC(p) which has been seen to lead to underfitting in some finite sample cases.

4.1.2 Duration dependence test

As McQueen and Thorley (1994) provided a methodology for duration dependence testing; the main idea of this test is finding the positive and negative runs in the data series.

Let a data set S_T, with T observation from the random length, I. The definition of a run is a sequence of abnormal return in the same signs (if positive in period t, it will be positive in period t+1). So, I defined as a positive valuation of discrete random variable which is generated by some discrete function and corresponding cumulative density function as $f_i \equiv \Pr ob(I = i)$ and $F_i \equiv \Pr ob(I < i)$, respectively.

Let N_i and P_i are the count of completed and partial runs, respectively, of lengths i in the series. The density function of the log likelihood is defined as following:

$$L(\theta | S_T) = \sum_{i=1}^{\infty} N_i \ln f_i + P_i \ln(1 - F_i) \qquad (4.1.2.1)$$

where θ: a vector of parameters.

A partial run could be occurred at the beginning and at the end of the examined time period. The hazard function, $h_i \equiv prob(I = i | I \geq i)$, is probability that a run will end at i. The hazard functions specification show the data series in term of condition of probabilities which are opposite to the density function specifications that focus only on unconditional probabilities. The choice of hazard or density specifications is based on the economic question of interest.

4.1.3 Vector Autoregressive Analysis (VAR)

The VAR needs less structural specification of the model since we cannot sure the relationship between macroeconomic variables, it also is in the case of unknown which variables are causes and which are affects. So a VAR model could be a powerful tool to investigate the inter-connection among those variables.

In the VAR system, all variables are assumed to be endogenous. A vector of variables to forecast are selected, that allows all variables to interact linearly with their own and others' current and past values, and that uses the historical data to determine the quantitative impact that each variable has on its own and other variables' future values. Thus, each equation has the same set in each equation. Due to in VAR system, regressions are lagged variables, they can be assumed to be contemporaneously uncorrelated with the disturbance. So each equation could be estimated separately by ordinary least square (OLS) which will yield consistent and efficient estimators.

Consider the simple bivariate system:

$$\begin{aligned}
x_t &= a_{10} + a_{11}x_{t-1} + a_{12}y_{t-1} + a_{13}z_{t-1} + a_{14}w_{t-1} + \varepsilon_{1t} \\
y_t &= a_{20} + a_{21}x_{t-1} + a_{22}y_{t-1} + a_{23}z_{t-1} + a_{24}w_{t-1} + \varepsilon_{2t} \\
z_t &= a_{30} + a_{31}x_{t-1} + a_{32}y_{t-1} + a_{32}z_{t-1} + a_{34}w_{t-1} + \varepsilon_{3t} \\
w_t &= a_{40} + a_{41}x_{t-1} + a_{42}y_{t-1} + a_{42}z_{t-1} + a_{44}w_{t-1} + \varepsilon_{4t}
\end{aligned}$$

$$\Leftrightarrow \begin{pmatrix} x_t \\ y_t \\ z_t \\ w_t \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \\ a_{40} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \\ w_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{pmatrix}$$

where $\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{pmatrix} \sim IID \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22}^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33}^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44}^2 \end{pmatrix} \right)$

So it has a standard form as:

$$X_t = A_0 + A_1 X_{t-1} + \mu_t \qquad (4.1.3.1)$$

where X_t is an (nx1) vector of endogenous variables

A_0 is an (nx1) vector of intercept term in the original model.

A_1 denotes an (nxn) matrix of coefficients of lags endogenous.

μ_t states an (nx1) vector of error terms or shock in the structural model with the following properties

$$E(\mu_t) = 0$$

$$E(\mu_t \mu_{t-k}) = \begin{cases} \Omega(cons\tan t)\, if\ k = 0 \\ \quad 0 \qquad if\ k \# 0 \end{cases}$$

$$E(\mu_t X_{t-k}) = 0$$

The equation (4.1.3.1) expresses that each variable in an (nx1) of Xt is a linear function of:

(1). Its own lagged values.

(2). The lagged value of all other regressors in the system.

(3) A white noise error term.

This specification is an unrestricted reduced form relationship of a structural simultaneous equations which makes minimal theoretical demand on the structure of a model.

4.2 Analysis implementation

4.2.1 Testing the existence of bubble

As mentioned above, the testing will be based on the question of the study that will answer whether the probabilities that the stock returns run continues based on the length of the runs, that is caused from the appropriately hazard specification. One more reason to apply the hazard specification is from missing of closed form multi parameter discrete density function.

From, the equation (3.2.1.12):

$$\varepsilon_{t+1} = \begin{cases} \eta_{t+1} + \dfrac{1-q}{q}\left((1+r_{t+1})b_t - a_0\right) with \qquad probability\, q \\ \quad \eta_{t+1} - (1+r_{t+1})b_t + a_0 \qquad with \qquad probability 1-q \end{cases} \tag{3.2.1.12}$$

the probability of a negative observation of innovation with the condition of a sequence of i prior positive innovations:

$$h_i = prob(\varepsilon_t < 0 | \varepsilon_{t-1} > 0, \varepsilon_{t-2} > 0, ..., \varepsilon_{t-i} > 0, \varepsilon_{t-i-1} < 0)$$

decreases in i, and they proved that $h_{i+1} < h_i$ for all i if the bubble present. And since the bubble is not negative, a similar inequality does not hold for runs of negative abnormal returns

Hence, the hazard function which is related to the density function:

$$h_i = \frac{f_i}{(1-F_i)} \text{ and } f_i = h_i \prod_{j=1}^{i-1}(1-h_j) \quad (4.2.1.1)$$

With the relationship in (4.1.1) and from the equation (4.B.1), the hazard function of the log likelihood could be:

$$L(\theta|S_T) = \sum_{i=1}^{\infty} N_i \ln h_i + M_i \ln(1-h_i) + Q_i(1-h_i) \quad (4.2.1.2)$$

where M_i and Q_i: the count of completed and partial runs with a length greater than i, respectively.

The terms contain P_i in the equation (4.B.1) and Q_i in (4.1.2) are covered incorporate information contained in partial runs and could be skipped in large samples. All the real returns are converted into lengths of positive and negative returns. The numbers of positive or negative runs of particular length i are then counted. The sample hazard rate for each length i is computed as

$$\hat{h}_i = \frac{N_i}{M_i + N_i} \quad (4.2.1.2b)$$

which is derived from maximizing the hazard function log likelihood (4.2.1.2).

To detect the rational bubbles, log logistic function form is applied (this method was applied by McQueen and Thorley (1994); and Chan, McQueen et al. (1998)):

$$h_i = \frac{1}{1+e^{-(\alpha+\beta \ln i)}} \quad (4.2.1.3)$$

The log logistic function transforms the unbounded range: $(\alpha + \beta \ln i)$ into the bound (0,1) space of h_i, the conditional probabilities of ending a run.

The null hypothesis, no bubbles, illustrates that the probabilities of a run ending is dependent of the prior returns (or it could be said that positive and negative abnormal return are random). In numerically for the null hypothesis, that must be shown $\beta = 0$ (it implies the constant hazard rate). The bubble alternative says that the probabilities of a positive run ending should decline with the run length or β is negative (decreasing hazard rates). The equation (4.2.1.2) will be maximizing by substituting (4.2.1.3) and use likelihood ratio test for testing β.

Under the null hypothesis of no bubble ($\beta = 0$), the likelihood ratio test is asymptotically distributed chi-square with one degree of freedom

4.2.2 The relationship between monetary policy and stock returns

To implement a VAR analysis, it must be followed some steps as:

(1). Set of variables.

(2). Stability test, Unit roots test.

(3). Optimal number of lags.

(4). Exogeneity of the model.

(5). Evaluation of the model.

(6). Implications of the model: Impulse Response Function and Variance decomposition.

4.2.2.1 The set of variables

The set of endogenous are selected to implement VAR analysis including: stock returns, lending interest rate, consumer price index and exchange rate. The details of variables in VAR system are given as below:

- **Stock returns**

Stock returns are overall changing prices of stocks in the market. This is the proxy for the investors' profit/loss every period. The return rate is calculated as mentioned above (Section 3).

- **Interest rate**

Since the SBV would like employ its instrument tools to calm down the fever of investors, the SBV has its own instrument tools, basic interest rate, based on this rate, the commercial banks will adjust their lending interest rate to 150% for the ceiling,since the basic interest rate from SBV cannot show any effect to the stock market with unchangeable rate in a long time. The study will employ two kinds of rate: basic interest rate and lending interest rate (to identify the relationship between interest rate and lending interest rate, the study also employed some tests to confirm that the lending interest rate could be a proxy for basic interest rate)[1] which is chosen the minimum lending interest rate from commercial banks as a proxy of interest rate. The lending interest rate affects directly to investors' decision, since if they intend to speculate in the short term, they must need more financial aid from the banks or other sources. With high interest rate, they cannot reach the loan and they must consider finding other profitable investments. There is other result to employ the lending interest rate, because the lending interest rate is affected by the basic interest rate and other monetary instrument such as reserve rate or the broad money supply, for example, although the basic rate is not changed, but the State Bank of Vietnam increases broad money supply, the commercial bank could adjust their lending rate to lower in case of increasing the reserve rate or of decrease the broad money supply.

- **Inflation rate**

The SBV has a responsibility to control the inflation rate via its instrument policy. It always has a trade-off among the economic growth and the inflation. The inflation rate will indirectly impact to the fluctuation of VNindex (as mentioned in the introduction section).

- **Exchange rate**

Vietnam uses fixed exchange rate as an official rate for commercial banks and any remittance with a small adjustment for commercial banks (±0.25% for in 2006 and previous years, ±0.5% in 2007, ±0.75% for the first quarter of 2008, then

[1] The study will do a cointegration test and Granger causality to find the relationship between basic interest rate and lending interest rate. See in the appendix.

±1%->3% from second quarter to the end of 2008 and ±5% from March 2009). The exchange rate could affect directly to the capital source from foreign investors. Any adjustment in exchange rate could change the foreign investors' decision and belief in Vietnamese economy. Although the foreign investors generally could hold a third share in a listed corporation, but almost domestic investors look at them to make their decision as the followers. If the foreign investors pour more or withdrawn their money, it could be having a strong effect to the stock market and changes the stock returns.

4.2.2.2 The VAR model

As mentioned in section 3.2.B, the VAR system defined as:

$$RET_t = a_{11} + \sum_{i=1}^{k} a_{12,i} * INT_{t-i} + \sum_{i=1}^{k} a_{13,i} * INF_{t-i} + \sum_{i=1}^{k} a_{14,i} * EXC_{t-i} + \sum_{i=1}^{k} a_{15,i} * RET_{t-i} + e_{1t} \quad (3.2.2.5)$$

$$INT_t = a_{21} + \sum_{i=1}^{k} a_{22,i} * RET_{t-i} + \sum_{i=1}^{k} a_{23,i} * INF_{t-i} + \sum_{i=1}^{k} a_{24,i} * EXC_{t-i} + \sum_{i=1}^{k} a_{25,i} * INT_{t-i} + e_{2t} \quad (3.2.2.6)$$

$$INF_t = a_{31} + \sum_{i=1}^{k} a_{32,i} * RET_{t-i} + \sum_{i=1}^{k} a_{33,i} * INT_{t-i} + \sum_{i=1}^{k} a_{34,i} * EXC_{t-i} + \sum_{i=1}^{k} a_{35,i} * INF_{t-i} + e_{3t} \quad (3.2.2.7)$$

$$EXC_t = a_{41} + \sum_{i=1}^{k} a_{42,i} * RET_{t-i} + \sum_{i=1}^{k} a_{43,i} * INT_{t-i} + \sum_{i=1}^{k} a_{44,i} * INF_{t-i} + \sum_{i=1}^{k} a_{45,i} * EXC_{t-i} + e_{4t} \quad (3.2.2.8)$$

Each equation could be separately estimated by using OLS, then we obtain the optimal lags and select appropriate model (from step (2) to step (5)). From that result, we could express the relationship between those variables. In order to achieve this objective, we implement the step (6).

4.2.2.3 Impulse response function and variance decomposition

a. Impulse response function

Impulse response function is to trace the effect of a shock in one innovation of each endogenous variable on current and future values of other variables. In an attempt to make an interpretation of VAR model by impulse response function, we could express as:

From the general form, the equation (4.1.3.1), the stability property:

$$\begin{pmatrix} x_t \\ y_t \\ z_t \\ w_t \end{pmatrix} = \begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \\ \bar{w} \end{pmatrix} + \sum_{i=0}^{\infty} \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}^{i} \begin{pmatrix} \varepsilon_{1t-i} \\ \varepsilon_{2t-i} \\ \varepsilon_{3t-i} \\ \varepsilon_{4t-i} \end{pmatrix}$$

Derive to be impulse response function:

$$\begin{pmatrix} x_t \\ y_t \\ z_t \\ w_t \end{pmatrix} = \begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \\ \bar{w} \end{pmatrix} + \sum_{i=0}^{\infty} \begin{pmatrix} \phi_{11}(i) & \phi_{12}(i) & \phi_{13}(i) & \phi_{14}(i) \\ \phi_{21}(i) & \phi_{22}(i) & \phi_{23}(i) & \phi_{24}(i) \\ \phi_{31}(i) & \phi_{32}(i) & \phi_{33}(i) & \phi_{34}(i) \\ \phi_{41}(i) & \phi_{42}(i) & \phi_{43}(i) & \phi_{44}(i) \end{pmatrix}^{i} \begin{pmatrix} e_{1t-i} \\ e_{2t-i} \\ e_{3t-i} \\ e_{4t-i} \end{pmatrix}$$

Applies to the VAR system from (3.2.2.5) to (3.2.2.8), we have:

$$\begin{pmatrix} RET_t \\ INT_t \\ INF_t \\ EXC_t \end{pmatrix} = \begin{pmatrix} \overline{RET} \\ \overline{INT} \\ \overline{INF} \\ \overline{EXC} \end{pmatrix} + \sum_{i=0}^{\infty} \begin{pmatrix} \phi_{11}(i) & \phi_{12}(i) & \phi_{13}(i) & \phi_{14}(i) \\ \phi_{21}(i) & \phi_{22}(i) & \phi_{23}(i) & \phi_{24}(i) \\ \phi_{31}(i) & \phi_{32}(i) & \phi_{33}(i) & \phi_{34}(i) \\ \phi_{41}(i) & \phi_{42}(i) & \phi_{43}(i) & \phi_{44}(i) \end{pmatrix}^{i} \begin{pmatrix} e_{1t-i} \\ e_{2t-i} \\ e_{3t-i} \\ e_{4t-i} \end{pmatrix} \qquad (4.2.2.3.1)$$

The elements $\phi_{jk}(0)$ are impact multipliers, representing the immediate responses of the system to unit innovations, and the coefficients $\phi_{jk}(i)$ are called the impulse response functions. The interpretation of coefficient $\phi_{jk}(0)$ is as follows: The coefficient $\phi_{21}(0)$, for example, is the instantaneous impact of an autonomous change in stock returns at time t ($e_{1,t}$) on the lending interest rate at time t (INTt). In the same way, the coefficient $\phi_{32}(0)$ is the instantaneous impact of an autonomous change in lending interest rate at time t (e2,t) on inflation at time t (INFt). It is as same as for interpretation of $\phi_{jk}(i)$.

If the error terms are uncorrelated with other, the ordering of variables imposed in VAR model does matter. However, if the errors are correlated, so the ambiguity to identify shocks with specific variables will increase. When the errors are correlated, there are common components that affect more than one variable. A somewhat arbitrary method to deal with this problem is to attribute all effects of common components to the variable that appears first in the system. One problem with this procedure is that impulse response will depend on the particular ordering of the equations in the model.

Technically, the errors are orthogonalized by a Choleski decomposition so that the covariance matrix of the resulting innovations is diagonal. Changing the order of equations can drastically change the impulse responses and considerations should be given to interpreting the impulse response functions.

We, therefore, need to know correlation among the error terms. If the correlation coefficient between residuals is low, the ordering is not likely to be important. Moreover, in a VAR with several variables, it is improbable that all correlation will be small. After all, in selecting the variables to be included in a model, one is likely to choose variables that exhibit strong co-movements. Under the null hypothesis that the cross-correlation are all zero, the sample variance of cross-correlation coefficient i asymptotically converges to (T-i)-1, where T=number of observations. Let r_{yz}(i) denotes the sample cross-correlation coefficient between y and zt-i. So, $Var(r_{yz}) = (T-i)^{-1}$. Then, the standard deviation of the cross-correlation between y and z_{t-i} is (T-i)-1/2. If the calculated value r_{yz}(i) exceeds 2*(T-i)-1/2 (two standard deviations), the null hypothesis will be rejected.

b. Variance decomposition

The main idea of variance decomposition is to forecast error variance decomposition reveals the proportion of the movements in a sequence due to its own shocks versus shocks to the other variable.

From the general form (4.2.2.3.1) $X_t = A_0 + A_1 X_{t-1} + \mu_t$. Given all constant terms (A_0) and coefficients (A_1) are known, we want to forecast the various value of X_{t+h} conditional on observed value of Xt. The Variance decomposition will be based on the vector moving average. Consider, for example, the h-period ahead forecast error of X_1 which is one variable in the system [Xt]. The forecast error system can be represented as:

$$X_{1,t+h} - E_t X_{1,t+h} = \begin{array}{c} \phi_{11}(0)\mu_{X_{1,t+h}} + \phi_{11}(1)\mu_{X_{1,t+h-1}} + ...\phi_{11}(h-1)\mu_{X_{1,t+1}} + \\ \phi_{12}(0)\mu_{X_{2,t+h}} + \phi_{12}(1)\mu_{X_{2,t+h-1}} + ...\phi_{12}(h-1)\mu_{X_{2,t+1}} + \\ .... \\ \phi_{1n}(0)\mu_{X_{n,t+h}} + \phi_{1n}(1)\mu_{X_{n,t+h-1}} + ...\phi_{1n}(h-1)\mu_{X_{n,t+1}} + \end{array} \qquad (4.2.2.3.2)$$

or it could b reduced as a general form:

$$X_{1,t+h} - E_t X_{1,t+h} = \sum_{j=1}^{n}\sum_{s=0}^{h-1}\phi_{ij,s}\mu_{j,t+h-s} \qquad (4.2.2.3.3)$$

where $E_t X_{1,t+h}$ is the linear-least square forecast of $X_{1,t+h}$ given the information $X_{i,t}$, $X_{i,t-1}$, and so on. The h-period ahead forecast error variance of $X_{i,t+h}$ is given by:

$$E\left[\left(X_{1,t+h} - E_t X_{1,t+h}\right)\left(X_{1,t+h} - E_t X_{1,t+h}\right)\right] = \sum_{j=1}^{n}\sum_{s=0}^{h-1}\phi_{ij,s}^2 \qquad (4.2.2.3.4)$$

where $\phi_{ij,s}$ is the ijth component of the matrix ϕ_s. From (4.2.2.3.4), part of the expected h-period ahead squared prediction error of $X_{i,t+h}$ produced by the innovation in variable $X_{j,t+h}$ is $\sum_{s=0}^{h-1}\phi_{ij,s}^2$.

The percentage of the h-period forecast error variance of variable $X_{i,t+h}$ accounted by innovation in variable $X_{j,t+h}$ is:

$$\frac{\sum_{s=0}^{h-1}\phi_{ij,s}^2}{\sum_{j=1}^{n}\sum_{s=0}^{h-1}\phi_{ij,s}^2} * 100 \qquad (4.2.2.3.5)$$

The variance decomposition contains the same problem inherent in impulse response function analysis. Consequently, we need to analyze the correlation coefficient among the error terms as in the impulse response function analysis. The Choleski decomposition is also applied to restriction of identification.

Nevertheless, impulse response analysis and variance decompositions can be useful tools to examine the relationship among economic variables. If the correlations among the various innovations are small, the identification problem (the ordering of variables) is unlikely to be important. The alternative orderings should yield similar impulse response and variance decomposition.

4.2.3 Data

The data series is secondary data which is collected from official websites in Vietnam. The time to collect is 2003:1-2009:4- all series are monthly data.

a. Interest rate (INT_t)

As statement in the previous section, the monthly interest rate are basic interest rate which is collected from the State Bank of Vietnam and the lending interest rate for consumption in one year contract which collected from the lowest between two commercial banks: Sacombank and Asia Commercial bank, this rate is announced and changed based on the basic interest rate of State Bank of Vietnam which allowed the commercial bank to change their lending interest rate in a limited range.

b. Stock returns (RET_t)

The monthly stock returns are calculated from VNindex, the representative index of Hochiminh Stock Exchange. The start of stock returns will be picked up the index at the end of months, from January 2003 to April 2009. The stock returns are the rate of change of VNindex for a month in comparison to the VNindex from previous month (as the above formula).

c. Inflation (INF_t)

The study will employ the monthly inflation which will be collected from General Statistic Office (GSO- Vietnam).

d. Exchange rate (EXC_t)

The official exchange rate will be collected from Global Financial data, the official rate in the trading day at the end of each month, this rate is basic exchange rate which is announced by State Bank of Vietnam, the commercial banks will adjust based on the permitted rate from SBV.