CHAPTER 4

RESEARCH METHODOLOGY

4.1 Introduction to volatility

Volatility is the effect of random price movements which normally occur in the stock market. Since many factors can influence price movements, it is difficult to define the term volatility, however in finance it can be a measure of the dispersion of returns over time. Researchers have discovered that a relationship between volatility from one period to the next exists. The presence of this heteroskedastic relationship may be used when modelling and forecasting the future volatility of markets. Volatility is usually referred to as variance, and is symbolized as σ^2. Volatility is also known as standard deviation, which is symbolized as σ. The equation that represents variance is

$$\sigma_t^2 = \omega + \sum_{i-1}^{q} \alpha_i \, \varepsilon_{t=i}^2 + \sum_{j=1}^{p} \beta_j \, \sigma_{t-j}^2 \qquad \text{……………….(1)}$$

The assumption in the above formula is that all past prices have an equal relevance in influencing future volatility. This assumption is too simplistic since recent volatility should have more relevance than that from several years ago and should be given a relatively higher weight in the calculation. Furthermore, when volatility is calculated using the above formula, some contradictions may arise with regards to the estimate of the mean return because the accuracy of the volatility assessment depends directly on the accuracy of the mean estimation. In estimating the mean, the average return of the sample can be very noisy, and this makes the accuracy level go down. There are ways to improve the accuracy that have been suggested by many researchers but it's often felt the simplest approach to be by using simple average return.

4.2 Stylized facts about volatility

Studies have revealed that financial time series data exhibits certain characteristics and patterns. These characteristics and patterns have been crucial in analysis to identify, specify, estimate and forecast certain trends in the data being analysed and have been used in follow-up studies and research. A volatility model must be able to capture and reflect these stylized facts. The general stylized facts that are very common in the markets, financial instruments and bonds are:

1. Fat tail distribution – referred to as excess kurtosis at times, this characteristic in the unconditional distribution in return data and was true when return deviation was found to deviate consistently from normality. The deviation shows larger changes and is followed by small changes from normal distribution. According to Engle and Patton (2001), if the Gaussian distribution is used for conditional density, the unconditional density will have excess kurtosis because of different volatilities combined with the Gaussian mixture itself, generating even higher excess kurtosis.

They estimate that typical kurtosis ranges from 4 to 50 for very extreme deviation from normality.

2. Volatility clustering – this characteristic is referred to when large price variations are more likely to be followed by other large price variations or when small price variations are more likely to be followed by other small price variations. These types of movements indicate that the persistence in shocks exist, and volatility is predictable.

3. Mean-reverting – interpreted as that there is a normal level of volatility that volatility will return to eventually. A measure of persistence in a volatility model is determined by observing the 'half-life' of the volatility. 'Half-life' is the time taken for volatility to move halfway back from the unconditional mean after deviating from it. The following formula describes the 'half-life', τ :

$$\tau = \left\{ \frac{\log\left[\frac{(\alpha 1 + \beta 1)}{2}\right]}{\log(\alpha 1 + \beta 1)} \right\} \qquad \text{....................(2)}$$

4. Asymmetry – this characteristic is also known as the leverage effect or risk premium effect and is considered true when gains or losses are not symmetric. The leverage effect can be described as when the price of

stock falls, its debt-to-equity ratio rises, increasing its volatility. Risk premium is when the news of increasing volatility reduces the demand for a stock because of risk aversion. When a large price movement upward but not equal downward movement is observed, the distribution of returns becomes skewed.

Exponential GARCH (EGARCH) models have been designed to capture the leverage effect as noted by Black (1976) and French et al (1987).

Har, Lenan and Ong (2008) studied the leverage effect on KLCI data for the period between January 2004 and June 2007 by employing the EGARCH model. Their findings showed that the market did not exhibit the leverage effect.

Christos (2008) studied if the leverage effect existed on markets in Egypt and Israel using the EGARCH model. His data was made up of the closing prices of the Egyptian CMA generalized index from 1997 to 2007 and also the Israeli TASE-100 index for the same period. He observed negative gamma (γ) coefficient factor and confirmed the existence of leverage effect in both the Egyptian and Israeli markets (Gamma coefficient is explained in detail in section 4.4).

5. There are other stylized facts such as exogenous and decay of auto correlation. In the former, variables influence volatility – prices and returns evolve independently of the market around them and so it can be expected that other variables can influence the volatility of a series. Some exogenous factors include unrelated scheduled company announcements, a third government's policy announcements and even time-of-the-day effects. The latter is when linear autocorrelations are insignificant. The auto correlation often remains significant for even long time lags. This is also a manifestation of volatility clustering and also evidences the presence of conditional heteroskedasticity.

6. Apart from the above listed stylized facts, there are others such as co-movements of volatilities across markets and assets as well as long memory properties.

An optimal forecasting model should be able to capture these characteristics of the time series that is being analyzed.

4.3 Volatility modelling techniques

Volatility forecasting is an important area of research in financial markets as forecasting volatility can be used to hedge risks and manage asset portfolio. Kumar (2006) has pointed out that there is no conclusive evidence as to which volatility forecasting model has supremacy over others. Shamiri, Isa and Hassan

(2008) used the Kullback-leibler Information Criterion (KLIC) to evaluate KLCI returns and concluded that a good distribution model was more important than the specific GARCH volatility model. Zaharim, Zahid, Zainol and Mohamed (2009) somewhat concurred with Shamiri, Isa and Hassan (2008) after analyzing KLCI data from August 1990 to August 2005 and concluded that the GARCH models could be used to model KLCI with only fair results.

Some of the models used for volatility modelling and forecasting will be discussed in this section. Generally there are two main groups of volatility models based on time series forecasting models and option based forecasting models. The time series models can be further classified as naïve (random walk, moving average etc) and sophisticated models (GARCH family models).

The time series based models use historical data while the option based models use present prices or best guessed prices.

4.3.1 Naïve models

Some of the more widely used naïve models are random walk (RW), historical mean, moving average (MA), simple regression and exponentially weighted moving average (EWMA).

Random walk (RW) subscribes that the best forecast for the period analysed was the last period's realised volatility.

Historical mean assumes that the conditional expectation of volatility is constant and forecasts volatility as the historical average of the past observed volatilities.

For the moving average model (MA), the forecast is based on all the available observations and each observation, whether it is old or new is given equal weight. This is a simple method in terms of executing analyses. However, this method leads to stale prices affecting the forecasts.

In the simple regression model, regression of the actual volatilities on lagged values is measured. In other words, autoregression is performed on the first part of the data which is selected for estimating the parameter and then estimated parameter obtained are used for forecasting the volatility for the following period.

Exponentially weighted moving average (EWMA) is an adaptive forecasting method that gives greater weight to more recent observations so that the recent memory of the market is better represented. This method of analysis adjusts the forecasts based on past forecast errors and the forecast is calculated as a weighted average of the immediate past observed volatility and the forecasted value for the same period.

4.3.2 Sophisticated models (the GARCH family models)

The ARCH and GARCH are considered the sophisticated models for forecast volatilities. ARCH stands for autoregressive conditional heteroskedasticity and was initially introduced by Engle (1982). The ARCH models capture the volatility clustering phenomena usually observed in financial time series data. In the linear ARCH (q) model, the time varying conditional variance is postulated to be a linear function of the past 'q' squared innovations. In other words, variance is modelled as a constant plus a distributed lag on the squared residual terms from earlier periods. In financial markets the ARCH(1) models are widely used and these are very simple models that exhibit constant unconditional variance and non-constant conditional variance.

The problem with the ARCH model is that it involves estimation of a large number of parameters and if some of the parameters become negative, they lead to difficulties in modelling. Bollerslev (1986) proposed a generalized ARCH or GARCH (p,q) model where volatility at time t depends on the observed data at t-1, t-2, t-3,…t-q as well as on volatilities at t-1, t-2, t-3,…t-p. The advantage of the GARCH formulation is that though recent innovations entered the model, it involves only the estimation of a few parameters and there will be little chance that they deviated. Gaussian distribution is assumed in the model and the simplest member of the GARCH family is the GARCH(1,1) model.

The GARCH model has many advantages, among them is its accuracy and flexibility which makes it the preferred model for risk management and also for trading but its weaknesses are that it requires large observations to produce accurate estimation and the model only analyses the magnitude of the movements of return without the direction (Matei, 2009).

4.3.3 Option based volatility models

Option based volatility models are also commonly used in the financial system. Volatility is often an attribute in the price calculation of an asset, thus given the asset (or option) price, the volatility can be deduced. Contrary to the time series model where historical data is used, option based models do not use historical data. It is believed that the accuracy of option based models are more superior to the time series model since the 'guessed' prices used are derived from experienced participants and therefore contain more information. Option prices are highly related to market expectations about an asset's future value movements. Hence, the forecast based on option prices will include all factors that can influence prices in the future. The crucial assumption is that market behaviour is rational. However, Donaldson and Kamstra (2004) found that the best forecast of volatility can be achieved when option-implied volatility is combined with ARCH models.

4.4 Method of Analysis

The autoregressive conditional heteroskedasticity (ARCH) which was introduced by Engle (1982) together with the extension to it known as generalized-ARCH or GARCH models introduced by Bollerslev (1986) were utilized for analyses.

The ARCH model allowed the variance of the error term to change over time, in contrast to the standard time series regression models which assume a constant variance. Bollerslev (1986) generalized the ARCH process by allowing for a lag structure in the variance.

The generalized ARCH models, i.e. the GARCH models, have been found to be valuable in modelling the time series behaviour of stock returns (Baillie and DeGennaro, 1990). Bollerslev (1986) allowed the conditional variance to be a function of a prior period's squared errors as well as its past conditional variances. Volatility was a key parameter in financial modelling and GARCH modelling was probably the best platform for estimating since it could take into account various characteristics of the data (Fabozzi, Tunaru and Wu, 2004).

GARCH models have been used in many recent studies involving time series models for volatility (Malla and Reddy, 2007).

The Lagrange Multiplier can be used to test hypotheses for both linear and non-linear regression models. According to Gujarati (2003), in large samples the test statistics follows the chi-square distribution. According to Studenmund (2006), the Lagrange Multiplier Serial Correlation test (LMSC) was a method to test for serial correlation by analyzing how well the lagged residuals explain the residual in the original equation. Mala and Reddy (2007) utilised the Lagrange Multiplier test and observed the intensity of volatility of the return data on stocks from the Fiji Stock Exchange.

For the purpose of empirical modelling, daily closing data of 11 indices from the Bursa Malaysia for the period between January 2000 and December 2010 has been used. The indices are Bursa Malaysia Composite Index (FBMKLCI), Construction Sector Index (KLCON), Consumer Sector Index (KLCSU), Finance Sector Index (KLFIN), Industrial Sector Index (KLIND), Industrial Production Sector Index (KLPRO), Mining Sector Index (KLTIN), Plantation Sector Index (KLPLN), Property Sector Index (KLPRP), Service Sector Index (KLSER) and Technology Sector Index (KLTEC). It is understood that the intra-day changes might not be fully captured by this approach. Events such as stock splits or dividend announcements might not be accurately reflected as well. The specific reason for choosing the time-frame was to analyse the behaviour of the market for a reasonable period with the latest possible data since the Bursa Malaysia benchmark indices themselves have evolved in recent years.

The equity return was calculated as the log difference of the market indices, as in the equation below:

$$R_t = \ln (P_t) - \ln (P_{t-1}) \qquad \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

As mentioned previously, the volatility of the Bursa Malaysia indices was modelled as a GARCH process that considers the changing volatility. The relatively simple GARCH(1,1) model was chosen for the process. All calculations were done with the EViews computer software. The process was divided into four steps:

Step 1: The past Bursa Malaysia indices data was estimated with an ARCH and a GARCH term. This process provided an indication if volatility existed in the time series. Computation of the GARCH model involves joint estimation of a mean and a conditional variance equation. The conditional mean is stated as follows:

$$Y_t = x_t'\theta + \varepsilon_t \qquad \ldots\ldots\ldots\ldots\ldots\ldots(4)$$

Where x_t being the vector of exogenous variables. The conditional variance equation GARCH(1,1) model can be stated as follows:

$$\sigma^2 = \omega + \alpha\varepsilon^2_{t-1} + \beta\sigma^2_{t-1} \qquad \text{......................(5)}$$

where ω is a constant term, $\alpha\varepsilon^2_{t-1}$ is the ARCH term and $\beta\sigma^2_{t-1}$ is the GARCH term.

Step 2: EGARCH model analysis was carried out to verify the existence of the leverage effect. Specification of EGARCH is given by the following equation,

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \alpha \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \qquad \text{..............................(6)}$$

The logarithmic of conditional variance implies that the leverage effect is exponential (so the variance was non-negative). The leverage effect is considered present if $\gamma < 0$.

Step 3: Lagrange Multiplier tests were conducted on the volatility model. This process provided insight into the intensity of the volatility of the indices analysed.

Step 4: The eleven-year data was then segregated into yearly data. The analysis of volatility was repeated for the yearly data and the results compared with the eleven-year data.

The study examined the distribution of market returns by dividing the sample period into a few categories. In the first category, the whole series for the eleven years' data was taken for analysis. In the second category, the yearly data from same indices was analysed. The reason for analysing the sub periods was to observe the volatilities of the indices during periods of financial crisis, such as the one experienced from 2007 to 2010.