Chapter 2

Research Methodology

2.1 Data and Sample Selection

This dissertation focuses on examining SET50 Index Options. As TFEX index options are European-style, the basic Black-Scholes option pricing model is used, but it causes bias in the calculated implied volatility. Fleming *et al.* (1995) and Hull and White (1987) have found that the calculation of implied volatilities can eliminate the mis-measurement and bias problem from the near-the-money and close-to-expiry options. Therefore, a total of eight near-the-money close-to-expiry SET50 call and put options prices (four call options and four put options) are used to calculate expected volatility accurately.

Thus, VIX calculation represents the volatility of an hypothetical option that is at-the-money with a constraint 22 trading days (30-day calendar period) to expiration. However, TVIX calculation represents the volatility that is at-the-money with constraint 66 trading days (90-day calendar period) to expiration. For the SEV index, the trading days are used.

Both series are obtained from Bloomberg (account at the Faculty of Economics, Chiang Mai University and Research Institute, the Stock Exchange of Thailand).

The options data used for this dissertation are high-frequency intraday data, which are the data at one-minute intervals between 09.45–12.30 and 14.30–16.55; for a total of 5 hours and 10 minutes each day. For Chapter 3, the sample period is from 27 January 2008 until 31 October 2008. For contract month December 2008, the data are downloaded until 31 October 2008. The one-minute intervals returns are used. In order to estimate TVIX and SEV index and predict for call and put option price, we use the SAS 9.1 software package for the estimation and forecasting of time series data, as it offers a number of features that are not available in traditional econometric software. As the SAS 9.1 software is used, the trading days for each month are counted through the actual trading days at the SET for SEV index since there is trading.

For Chapter 4 and 5, the sample period is from 27 January 2008 until 30 September 2009. The contract months are March, June, September 2008 and 2009 and December 2008. In order to calculate TVIX, we use the SAS 9.1 software package for the calculation as it offers a number of features that are not available in traditional econometric software.

For the estimation in Chapter 4, OxMetrics5 software is used to estimate ARFIMA-FIGARCH and -FIAPARCH on daily returns. For the estimation in Chapter 5, we use daily returns of TVIX to estimate ARMA-GARCH, -EGARCH, -GJR-GARCH, and -PGARCH by using E-Views 6.0 software.

The returns of TVIX at time t are calculated as $\boldsymbol{R_{i,t}} = \mathbf{log}(\boldsymbol{P_{i,t}} / \boldsymbol{P_{i,t-1}})$ where $\boldsymbol{P_{i,t}}$ and $\boldsymbol{P_{i,t-1}}$ are the closing prices of TVIX at time t and t-1, respectively.

2.2 Unit Root Tests

2.2.1 The Augmented Dickey-Fuller (ADF) Test

In order to test stationary of the data, the Augmented Dickey Fuller (ADF) test is used which is as follows:

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i-1} + \varepsilon_t \tag{2.1}$$

where $\gamma = -(1 - \sum_{i=1}^{p} a_0)$

and $\beta_i = -\sum_{j=1}^{p} a_j$ (2.2)

In (2.1), the coefficient of interest is γ, if $\gamma = 0$, the equation is entirely in first difference. Then, y_t has unit root and stationary (Ender (1995)). 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.

2.2.2 The Phillips-Perron (PP) Test

Phillips and Perron (1988) propose an alternative (nonparametric) method of controlling for serial correlation when testing for a unit root. The PP method estimates the non-augmented DF test, and modifies the t-ratio of the α

coefficient so that serial correlation does not affect the asymptotic distribution of the test statistic. The PP test is based on the statistic as follows:

$$\hat{t}_{\alpha} = t_{\alpha}(\frac{\gamma_0}{f_0})^{1/2} - \frac{T(f_0 - \gamma_0)(se(\hat{\alpha}))}{2f_0^{1/2}s} \qquad (2.3)$$

where $\hat{\alpha}$ os the estimate, and t_{α} the t-ratio of α, $se(\hat{\alpha})$ is coefficient standard error, and s the standard error of the test regression. In addition, γ_0 is a consistent estimate of the error variance, calculated as $(T-k)s^2/T$, where k is the number of regressors. The remaining term, f_0, is an estimator of the residual spectrum at frequency zero.

2.3 Volatility Index

The idea of estimating implied volatility from options is relatively simple. There is no straightforward method to extract the information. With the large number of option pricing models, many researchers have applied various methods of estimating implied volatilities from option pricing models, especially the Black-Scholes model from Black and Scholes (1973). The model was originally developed to estimate implied volatility at each exercise price, as in Melino and Turnbull (1990), Nandi (1996), and Bakshi, Cao and Chen (1997).

Option prices is used to calculate implied volatility that represents a market-based estimate of future price volatility, so that implied volatility is regarded as a fear gauge from Whaley (2000). Implied volatilities are reported by investors, financial news services and other finance professionals. The information content and forecast quality of implied volatility is an important topic in financial markets research.

Latane and Rendleman (1976), Chiras and Manaster (1978), Beckers (1981) and Jorion (1995) provided early assessments of the forecast quality of implied volatility. They concluded that implied volatilities outperform historical standard deviations, although perhaps biased, as a good predictor of future volatility. Christensen and Prabhala (1998) found that implied volatility forecasts are biased, but dominate historical volatility in terms of ex ante forecasting power. Fleming (1998) used a similar volatility measure to show that implied volatilities outperform historical information.

Fleming *et al.* (1995) showed that implied volatilities from S&P100 index options yield efficient forecasts of one-month ahead S&P100 index return volatility, and can also eliminate mis-specification problems. Blair *et al.* (2001) concluded that the VIX index provides the most accurate forecasts for low- or high-frequency observations, and are also unbiased.

Dennis *et al.* (2006) found that daily innovations in VIX contain very reliable incremental information about the future volatility of the S&P100 index. Other studies that attempt to forecast implied volatility or use the information contained in implied volatility to trade in option markets include Harvey and Whaley (1992), Noh *et al.* (1994), and Poon and Pope (2000).

2.3.1 VIX from CBOE

VIX measures market expectation of near term volatility conveyed by stock index option prices. The original VIX was constructed using the implied volatilities of eight different S&P100 (OEX) option series so that, at any given time, it represented the implied volatility of an hypothetical at-the-money OEX option with exactly 30 days to expiration from an option-pricing model.

In 2003, the CBOE made two key enhancements to the VIX methodology. The new VIX is based on an up-to-the-minute market estimation of expected volatility that is calculated by using real-time S&P500 Index (SPX) option bid/ask quotes, and incorporates information from the volatility "skew" by using a wider range of strike prices rather than just at-the-money series with the market's expectation of 30-day volatility, and using nearby and second nearby options.

Until 2006, VIX was trading on the CBOE. The VIX options contract is the first product on market volatility to be listed on an SEC-regulated securities exchange. This new product can be traded from an options-approved securities account. Many investors consider the VIX Index to be the world's premier barometer of investor sentiment and market volatility, and VIX options are a very powerful risk management tool. VIX is quoted in percentage points, just like the standard deviation of a rate of return.

2.3.2 New VIX Procedure

The New VIX is more robust because it pools the information from option prices over the whole volatility skew, and not just from at-the-money options. The formula used in the new VIX calculation is given by the CBOE as follows:

$$\sigma^2 = \frac{2}{T}\sum_{i}\frac{\Delta K_i}{K_i^2}e^{RT}Q(K_i) - \frac{1}{T}\left[\frac{F}{K_0}-1\right]^2, \qquad (2.4)$$

where

σ = VIX / 100 (so that VIX = σ x 100),

T = Time to expiration (in minutes),

F = Forward index level, derived from index option prices (based on at-the-money option prices, the difference between call and put prices is smallest).

The formula used to calculate the forward index level is:

F = Strike price (at-the-money) + e^{RT} x (Call price – Put price),

where

R = risk-free interest rate is assumed to be 3.01% (for simplicity, the government T-bills 3 month contract interest rate is used, as the Thailand options contract is a 3 months contract);

T = $\{M_{current\ day} + M_{settlement\ day} + M_{other\ days}\}$/minutes in a year,

where

$M_{current\ day}$ = # of minutes remaining until midnight of the current day,

$M_{settlement\ day}$ = # of minutes from midnight until 9:45 am on the TFEX settlement day,

$M_{\text{other days}}$ = Total # of minutes in the days between the current day and the settlement day;

K_i = Strike price of i^{th} out-of-the-money option; a call if K_i > F and a put if K_i < F;

ΔK_i = Interval between strike prices - half the distance between the strike on either side of K_i: $\Delta K_i = \frac{K_{i+1} - K_{i-1}}{2}$;

K_0 = First strike below the forward index level, F;

$Q(K_i)$ = The midpoint of the bid-ask spread for each option with strike K_i.

(Note: ΔK_i for the lowest strike is simply the difference between the lowest strike and the next higher strike. LikewiseΔ K for the highest strike is the difference between the highest strike and the next lower strike.)

With the adaptation of the VIX calculation to Thailand SET 50 index options, the Thailand expected volatility (TVIX) can be estimated.

2.4 A Simple Expected Volatility Index (SEV Index)

From the apparently complicated expected volatility formula, this paper tries to simplify the VIX formula into an SEV Index to obtain new results about the information content in option prices. The simplified formulae for the expected volatility index are as follows:

$$\boldsymbol{SEV\_1 = \log(\Delta K) / \log(index)}, \quad (2.5)$$

$$SEV\_2 = \Delta K / index , \tag{2.6}$$

$$SEV\_3 = \Delta K / index^2 , \tag{2.7}$$

where ΔK = the difference between the strike prices.

2.5 The Black-Scholes Model

The original Black and Scholes (1973) option-pricing model was developed to value options primarily on equities. The modified Black-Scholes European model that is used at the Thailand Futures Exchange (TFEX) has a number of restrictive assumptions, as follows:

1. The options pay no dividends during the option's life (q = 0);
2. European exercise terms dictate that the option can only be exercised on the expiration date;
3. Returns on the underlying asset are lognormally distributed;
4. No commissions are charged.

From the model given below, SET50 index call and put option prices are used to calculate implied volatility.

The TFEX Black-Scholes options pricing model is as follows:

Call option pricing formula:

$$C = Se^{-qt/365} \cdot N(d1) - Xe^{-rt/365} \cdot N(d2) . \tag{2.8}$$

A call option affords the buyer the right to purchase an underlying asset for a fixed price in the future.

Put options pricing formula:

$$P = Xe^{-rt/365} \cdot (1 - N(d2)) - Se^{-qt/365} \cdot (1 - N(d1)) . \qquad (2.9)$$

A put option affords the buyer the right to sell the underlying asset for a fixed price in the future.

$$d1 = \frac{\ln(S/X) + (r - q + (V^2/2) \cdot (t/365)}{V \cdot \sqrt{t/365}}$$

$$d2 = d1 - V \cdot \sqrt{t/365}$$

where

S	=	price of underlying asset,
X	=	strike price at maturity date,
r	=	risk-free rate (apply zero-coupon bond at 3 month maturity to calculate options with 3 months maturity),
q	=	dividend yield of underlying asset (q = 0),
t	=	time to maturity (days),
N	=	the cumulative normal distribution function,

V = standard deviation of the rate of return during the life of the option (the expected volatility or TVIX).

With the Black-Scholes option pricing model, the expected volatilities are substituted to predict call and put option prices at each strike price and expiration.

2.6 Measures of Statistic Fit

In order to assess the performance of the TVIX and SEV index, the goodness of fit of the model can be evaluated by measuring the descriptive statistics for the volatility index, as follows:

Goodness of Fit	Equations
Mean Square Error	$MSE = \frac{SSE}{n}$
Root Mean Square Error	$RMSE = \sqrt{MSE}$
Mean Absolute Percent Error	$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left\|(y_t - \hat{y}_t)/y_t\right\|$
Mean Absolute Error	$MAE = \frac{1}{n}\sum_{t=1}^{n}\left\|y_t - \hat{y}_t\right\|$
Adjusted R^2	$AdjR^2 = 1 - [(n-1)/(n-k)](1-R^2)$
AIC	$n \ln(MSE) + 2k$
SBIC	$n \ln(MSE) + k\ln(n)$
	where:
	$R^2 = 1 - SSE/SST$
	$SSE = \sum_{t=1}^{n}(y_t - \hat{y}_t)^2$
	$SST = \sum_{t=1}^{n}(y_t - \bar{y})^2$
	y_t = the data at time t
	$\hat{y}$ = the one-step predicted value
	$\bar{y}$ = the series mean
	n = the number of observations
	k = the number of estimated parameters

The mean square error (MSE) uses the one-step-ahead forecasts. Root mean square error (RMSE) is useful for determining how accurately the model might predict future observations. Adjusted R-squared (Adj R^2) is used as a standard model selection criterion. The Akaike information criterion (AIC) (Akaike (1973)) and Schwarz Bayesian Information criterion (SBIC) (Schwarz (1978)) are useful to determine which of several competing nested or non-nested models may fit the data the best. The model with the lowest values of AIC and SBIC is selected as fitting the sample data better.

2.7 ARFIMA Model

ARIMA models are frequently used for seasonal time series (Box and Jenkins, 1976). A general multiplicative seasonal ARIMA model for time series Z_t is as follows:

$$\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D Z_t = \theta(L)\rho(L^s)a_t \qquad (2.10)$$

where:

L = a backshift or lag operator $(B_{zt} - Z_{t-1})$

S = seasonal period

$\phi(L)$ = $(1-\phi_1 L - \ldots - \phi_p L^p)$ is the non-seasonal AR operator

$\Phi(L^s)$ = $(1-\Phi_1 L^s - \ldots - \Phi_s L^s)$ is the seasonal AR operator

$\theta(L)$ = $(1-\theta_1 L - \ldots - \theta_q L^q)$ is the non-seasonal MA operator

$\rho(L)$ = $(1-\rho_1 L^s - \ldots - \theta_0 L^{Q_s})$ is the seasonal MA operator

$(1-L)^d(1-L^s)$ = non-seasonal differencing of order d and seasonal differencing of order D

Granger and Joyeux (1980) and Hosking (1981) proposed an autoregressive fractionally integrated moving-average (ARFIMA) model and proposed the method to fit long-memory data. ARFIMA(p,d,q) is written as follow:

$$\phi(L)\Delta^d y_t = \delta + \theta(L)u_t \tag{2.11}$$

with $\phi(L) = 1-\phi_1 L - \ldots - \phi_p L^p$ and $\theta(L) = 1-\theta_1 L - \ldots - \theta_q L^q$

where:

δ = a constant term

$\theta(L)$ = the MA operator at order q

u_t = an error term

$\phi(L)$ = the AR operator at order p

$\Delta^d y_t$ = the differencing operator at order d of time series data y_t

For d = *(-0.5, 0)*, the process exhibits intermediate memory or long range negative dependence, while d = *(0, 0.5)*, the process exhibits long memory or long range positive dependence. For d = *[0.5, 1)*, the process is mean reverting with no long run impact to future values of the process and the process becomes a short memory when d = 0 corresponding to a standard ARMA process.

2.8 Fractional Integrated Models

2.8.1 FIGARCH Model

The GARCH model by Bollerslev (1986) imposed important limitations, not to capture a positive or negative sign of u_t which both positive and negative shocks has the same impact on the conditional variance, h_t , as follows,

$$u_t = \eta_t \sqrt{\sigma_t}, \tag{2.12}$$

$$\sigma_t^2 = \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 \tag{2.13}$$

where $\omega > 0, \alpha_i \geq 0$ for i = 1,…,p and $\beta_j \geq 0$ for j = 1,…,q are sufficient to ensure that the conditional variance, h_t is non-negative. For the GARCH process to be defined, it is required that $\omega > 0$. Therefore, a univariate GARCH(1,1) model is known as ARCH(∞) model (Engle, 1982) as an infinite expansion in u_{t-i}^2.

Baillie *et al.* (1996) proposed fractionally integrated GARCH (FIGARCH) model to determine long memory in return volatility. The FIGARCH (p,d,q) process is as follow:

$$\phi(L)(1-L)^d u_t^2 = \omega + [1-\beta(L)]v_t, \tag{2.14}$$

where $v_t = u_t^2 - \sigma_t^2$, $0 < d < 1$, $\phi(L) = \sum_{i=1}^{m-1} \phi_i L^i$ is of order m-1, and all the roots of $\phi(L)$ and $[1-\beta(L)]$ lie outside the unit circle. The FIGARCH model is derived from standard GARCH model with fractional difference operator, $(1- L)^d$. The FIGARCH(p,d,q) model is reduced to the standard GARCH when d = 0 and becomes IGARCH model when d = 1.

Baillie *et al.* (1996) claimed with the arguments of Nelson (1990) that the FIGARCH(p,d,m) is ergodic and strictly stationary which is difficult to verify. The degree of persistence of the FIGARCH model operates reversely direction of the ARFIMA process.

Chung (2001) suggested the analysis of the FIGARCH specification

$$\sigma_t^2 = \{1-[1-\beta(L)]^{-1})(1-L)^d \phi(L)\}\varepsilon_t^2 \tag{2.15}$$

2.8.2 FIAPARCH Model

Tse (1998) extended the asymmetric power ARCH (APARCH) model of Ding *et al.* (1993) to fractionally integrated of Baillie *et al.* (1996) which is extended to FIAPARCH model as follows:

$$\sigma_t^\delta = \omega + \left[1-\frac{[1-\phi(L)](1-L)^d}{1-\beta(L)})\right][|u_t|-\gamma u_t]^\delta, \tag{2.16}$$

where $0 < d < 1$, $\omega, \delta > 0, \phi, \beta < 1$, $-1 < \gamma < 1$ and L is the lag operator. When $\gamma > 0$, negative shocks have a higher volatility than positive shocks. The particular value of power term may lead to suboptimal modeling and forecasting performance. Ding *et al.* (1993) found that the closer of d value converge to 1, the larger the memory of the process becomes. The process of FIAPARCH allows for asymmetry. When $\gamma = 0$ and $\delta = 2$, the process of FIAPARCH is reduced to FIGARCH process.

ARFIMA-FIAPARCH generates the long memory property in both the first and (power transformed) second conditional moments and is sufficiently flexible to handle the dual long memory behavior. It can recognize the long memory aspect and provides an empirical measure of real uncertainty that accounts for long memory in the power transformed conditional variance of the process.

2.9 Univariate Models

2.9.1 GARCH Model

GARCH model by Bollerslev (1986) imposes important limitations, not to capture a positive or negative sign of u_t, which both positive and negative shocks has the same impact on the conditional variance, h_t, as follows,

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \qquad (2.17)$$

where $\omega > 0$, $\alpha_i \geq 0$ for i = 1,…,p and $\beta_j \geq 0$ for j = 1,…,q are sufficient to ensure that the conditional variance, σ_t, is non-negative. For the GARCH process to be defined, it is required that $\omega > 0$. Also, a univariate GARCH(1,1) model is known as ARCH(∞) model (Engle, 1982) as an infinite expansion in u_{t-i}^2. The α represents the ARCH effect and β represents the GARCH effect.

2.9.2 Exponential GARCH (EGARCH) Model

Exponential GARCH (EGARCH) model by Nelson (1991) is the logarithm of conditional volatility in order to capture asymmetries between positive and negative shocks that the leverage effect is exponential, as follows,

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^{p} \alpha_i \left|\eta_{t-i}\right| + \sum_{j=1}^{q} \beta_j \log\left(\sigma_{t-j}^2\right) + \sum_{k=1}^{r} \gamma_k \left|\eta_{t-k}\right| \quad (2.18)$$

where $\eta_{t-i} = \frac{u_{t-i}}{\sigma_{t-i}}$ which η_{t-i} and $\left|\eta_{t-i}\right|$ capture the sign and size effects of the standardized shocks. There are no restrictions on the parameters in the model. The moment conditions of the model are also straightforward because the standardized shocks have finite moments. There is an leverage effect when $\gamma < \alpha < -\gamma$. This implies that the negative shocks increase volatility and vice versa.

2.9.3 Glosten, Jagannathan and Runkle (GJR-GARCH) Model

Glosten, Jagannathan and Runkle (GJR-GARCH) model by Glosten *et al.* (1993) is to capture possible asymmetric impacts of positive and negative shocks on the conditional variance for σ_t, as follows,

$$\sigma_t = \omega + \sum_{i=1}^{p} (\alpha_i + \gamma_i I(\varepsilon_{t-i}))\varepsilon_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}, \quad (2.19)$$

where $I(\varepsilon_{t-j})$ is an indicator function that equals to 1 if $\varepsilon_{t-j} < 0$ and 0 otherwise. If p = q = 1, $\omega > 0, \alpha_1 \geq 0, \alpha_1 + \gamma_1 \geq 0$ and $\beta_1 \geq 0$ are sufficient conditions to ensure that the conditional variance h_t is non-negative. $\alpha_1(\alpha_1 + \gamma_1)$ gives the short-run persistence of positive (negative) shocks. If $\gamma_1 \neq 0$, the news impact is asymmetry. If $\gamma > 0$, there is a leverage effect that bad news increases volatility.

Lee and Hansen (1994) derived the log-moment condition for GARCH(1,1) of conditional volatility which is sufficient for the statistical properties of the QMLE to be consistent, as follows,

$$E(\log(\alpha_1 \eta_t^2 + \beta_1)) < 0. \tag{2.20}$$

It is essential to note that the log-moment condition is a weaker regularity condition than the second moment condition. Therefore, the second moment is sufficient condition for consistency and asymptotic normality of the QMLE, as follows,

$$\alpha_1 + \beta_1 < 1. \tag{2.21}$$

Moreover, McAleer *et al.* (2007) established the log-moment and second moment condition for GJR(1,1) as follows,

$$E(\log((\alpha_1 + \gamma_1 I(\eta_t))\eta_t^2 + \beta_1)) < 0 \tag{2.22}$$

and $\alpha + (\gamma / 2) + \beta < 1$ (2.23)

Both moments are the sufficient conditions for the consistency and asymptotic normality of the QMLE for the GJR(1,1).

2.9.4 Power GARCH (PGARCH) Model

Power GARCH (PGARCH) model by Taylor (1986) and Schwert (1989) use the conditional standard deviation as a measure of volatility instead of the conditional variance. This model is generalized by Ding *et al.* (1993) using the PGARCH model as follows:

$$\sigma_t^{\delta} = \omega + \sum_{i=1}^{q} \alpha_i \left(|u_{t-i}| - \gamma_i u_{t-i} \right)^{\delta} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^{\delta} \tag{2.24}$$

where

$\delta > 0, |\gamma_i| \leq 1 \, for \, i = 1,2,\ldots,r \, and \, \gamma_i = 0 \, for \, i > r, and \, r \leq p.$

In the PGARCH model, if $\gamma \neq 0$, this captures asymmetric effects. The PGARCH model reduces to the GARCH model when $\delta = 2$ and $\gamma_i = 0$ for all *i*.