FORECASTING RETURNS FOR THE STOCK EXCHANGE OF THAILAND INDEX USING MULTIPLE REGRESSION BASED ON PRINCIPAL COMPONENT ANALYSIS

1Nop Sopipan, 2Anchalee Sattayatham and 3Samruam Chongcharoen

1Program of Mathematics and Applied Statistics, Faculty of Science and Technology, Nakhon Ratchasima Rajabhat University, Nakhon Ratchasima, Thailand
2Faculty of Liberal Arts, Mahidol University, Nakhon Pathom, Thailand
3National Institute of Development Administration (NIDA), School of Applied Statistics, Bangkapi, Bangkok, Thailand

Received 2012-12-12, Revised 2013-02-10; Accepted 2013-03-18

ABSTRACT

The aim of this study was to forecast the returns for the Stock Exchange of Thailand (SET) Index by adding some explanatory variables and stationary Autoregressive Moving-Average order p and q (ARMA (p, q)) in the mean equation of returns. In addition, we used Principal Component Analysis (PCA) to remove possible complications caused by multicollinearity. Afterwards, we forecast the volatility of the returns for the SET Index. Results showed that the ARMA (1,1), which includes multiple regression based on PCA, has the best performance. In forecasting the volatility of returns, the GARCH model performs best for one day ahead; and the EGARCH model performs best for five days, ten days and twenty-two days ahead.

Keywords: SET Index, Forecasting, Principal Component Analysis, Multicollinearity, Volatility Models

1. INTRODUCTION

In order to forecast the return $r_t$ for specific purposes, many researchers have made different assumptions for $\mu_t$ as appears in Equation (2). For example, Sopipan et al. (2012) assumes $\mu_t$ to be a constant and Sattayatham et al. (2012) assume $\mu_t$ to be an ARMA process with a one-week delay. Caporale et al. (2011) assume the returns employ both fractional and non-fractional models.

The financial returns $r_t$ ($r_t = \ln(P_t/P_{t-1})$ for $t = 1, 2, \ldots, T-1$, $P$ denoting the financial price at time $t$ depend concurrently and dynamically on many economic and financial variables. Since the returns have a statistically significant autocorrelation themselves, lagged returns might be useful in predicting future returns. In order to model these financial returns, Tsay (2010) assumes that $r_t$ follows a simple time series model such as a stationary ARMA (p, q) model with some explanatory variables $X_{it}$.

In other words, $r_t$ satisfies the following Equation 1:

$$ r_t = \mu_t + \varepsilon_t, $$

$$ \mu_t = \mu_0 + \sum_{i=1}^{n} \beta_i X_{it} + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} $$

Where:

$$ X_s = 100 \ln \left( \frac{P_s}{P_{s-1}} \right) $$

Here $P_s$ denotes the financial price asset $i$ for $i = 1, 2, \ldots, n$ at time $t$, and $r_{t,s}$ is the return at lag $s$-th. $\varepsilon_t$ represents errors assumed to be a white noise series with an i.i.d. mean of zero and a constant variance $\sigma^2$. $\mu_0$, $\beta_i$, $\phi_i$, $\theta_i$, $\theta_m$ are constants and $n$, $p$ and $q$ are positive integers.

Generally, the variance of errors $\varepsilon_t$ in the model (1) is assumed to be a constant; some authors use this assumption in the modeling of gold prices (Ismail et al., 2009). But in this study, we shall consider the case where
the variance of \( \epsilon_t \) is not constant. That is, we shall introduce the heteroskedasticity model to forecast the volatility of returns using GARCH, EGARCH, GJR-GARCH and Markov Regime Switching GARCH (MRS-GARCH) with distribution normal, student-t and General Error Distribution (GED).

The objective of this study is to forecast returns for the Stock Exchange of Thailand (SET) Index by using model (1). We vary the process \( \mu_t \) using four different types and compare the performance of the different types. Moreover, we forecast the volatility of returns with heteroskedasticity models.

In the next section, we present the basics of principal component analysis to remove possible complications caused by the multicollinearity of explanatory variables and the volatility models. The empirical study and formulae for model estimation are given in section 3. The methodology and results are presented in section 4 and the conclusions are presented in section 5.

2. MATERIALS AND METHODS

2.1. Principal Component Analysis

The given is an n-dimensional random variable \( X_t = (X_{1t}, X_{2t}, \ldots, X_{nt})' \) with covariance matrix \( \Sigma_t \), where \((.)'\) denotes the transpose matrix. Principal Component Analysis (PCA) is concerned with using fewer linear combinations of \( X_t \) to explain the structure of \( \Sigma_t \). If \( X_{it} \) denotes returns as appears in (2) for \( i = 1, 2, \ldots, n \), then PCA can be used to study the source of variations of these \( n \) returns.

Let \( (\lambda_1, e_1), \ldots, (\lambda_n, e_n) \) be the eigenvalue-eigenvector pairs of \( \Sigma_t \), with the eigenvalues \( \lambda_i \) set up in decreasing order \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0 \). Then the i-th principal component of \( X_t \) is given by \( Z_i = e_i'X_t = \sum_{j=1}^{n} e_{ij}X_{jt} \) for \( i = 1, \ldots, n \). We note that Equation 3:

\[
\begin{align*}
\text{Var}(Z_i) &= e_i' \Sigma_t e_i = \lambda_i, \\
\text{Cov}(Z_i, Z_j) &= e_i' \Sigma_t e_j = 0, \quad i \neq j, i, j = 1, 2, \ldots, n \\
\end{align*}
\]

where, \( e_i' = (e_{i1}, \ldots, e_{in}) \) is orthonormal vector.

In order to cope with the problem of multicollinearity, we transform the explanatory variables in model (1) into the principal components. Then the new model for forecasting \( r_t \) is Equation 4:

\[
r_t = \mu_t + \sum_{i=1}^{n} \alpha_i Z_i + \sum_{i=1}^{n} \phi_i r_{t-i} - \sum_{i=1}^{n} \theta_i \epsilon_{t-n} + \epsilon_t,
\]

where, \( Z_i \ i = 1, 2, \ldots, n \) are i-th principal components of explanatory variables at time \( t \).

We follow Tsay (2010) by assuming that the asset return series \( r_t \) is a weakly stationary process.

2.2. Volatility Models

Since we aim to find a suitable model for the volatility of \( r_t \) we shall give a brief review of some known volatility models of interest to us. These models are GARCH, EGARCH, GJR-GARCH and MRS-GARCH.

The GARCH (1,1) model is as follows:

\[
\eta_i = \eta_t \sqrt{h_t},
\]

\[
h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}
\]

where \( \eta_t \) is i.i.d. distributed with zero mean and unit variance, \( \alpha_0, \alpha_1 > 0 \) and \( \beta_1 > 0 \) to ensure positive conditional variance. The inequality \( \alpha_1 + \beta_1 < 1 \) must be satisfied for a stationary covariance process of returns.

The Exponential GARCH (EGARCH) model was coped with the skewness often encountered in financial returns. The EGARCH (1, 1) model is defined as:

\[
\ln(h_t) = \alpha_0 + \frac{(\alpha_1 \epsilon_{t-1}^2 + \beta_1 \ln(h_{t-1}) + \xi \epsilon_{t-1})}{\sqrt{h_{t-1}}},
\]

where, \( \xi \) is the asymmetry parameter to capture the leverage effect.

The GJR-GARCH model was accounted for the leverage effect; it is a model that allows for different impacts of positive and negative shocks on volatility. The GJR-GARCH (1,1) model takes the following form:

\[
h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 \left( 1 - I_{[\epsilon_{t-1}>0]} \right)
\]

\[
+ \beta_1 h_{t-1} + \xi \epsilon_{t-1}^2 \left( 1 - I_{[\epsilon_{t-1}>0]} \right)
\]

where, \( I_{[\epsilon_{t-1}>0]} \) is equal to one when \( \epsilon_{t-1} \) is greater than zero and equal to zero elsewhere. The conditions \( \alpha_0 > 0, (\alpha_1 + \xi) > 2 > 0 \) and \( \beta_1 > 0 \) must be satisfied to ensure positive conditional variance.

The Markov Regime Switching GARCH (MRS-GARCH) model has two regimes which can be represented as follows:

\[
r_t = \mu_t + \sum_{i=1}^{n} \alpha_i Z_i + \sum_{i=1}^{n} \phi_i r_{t-i} - \sum_{i=1}^{n} \theta_i \epsilon_{t-n} + \epsilon_t,
\]
\[ \varepsilon_t = \eta_t \sqrt{h_{t-1}} \quad \text{and} \quad h_{t-1} = \alpha_{S_t} + \alpha_{S_t} \varepsilon^2_{t-1} + \beta_{S_t} h_{t-1} \]

where, \( S_t \) = 1 or 2, \( h_{t-1} \) is the conditional variance with measurable functions of \( F_{t-\tau} \) for \( \tau \leq T-1 \). In order to ensure easily the positivity of conditional variance, we impose the restrictions \( a_{0,S_t} > a_{1,S_t} \geq 0 \) and \( \beta_{1,S_t} \geq 0 \). The sum \( a_{1,S_t} + \beta_{1,S_t} \) measures the persistence of a shock to the conditional variance.

3. EMPIRICAL STUDIES AND METHODOLOGY

Naturally, the Thai stock market has unique characteristics, so the factors influencing the price of stocks traded in this market are different from the factors influencing other stock markets (Chaigusin et al., 2008). Examples of factors that influence the Thai stock market and the statistics used by researchers who have studied these factors in forecasting the SET Index are shown in Table 1.

3.1. Data

The data sets used in this study are the daily return closing prices for the SET Index at time \( t \) (dependent variables) and the daily return closing prices for twelve factors (explanatory independent variables). These twelve factors are the following:

- The Dow Jones Index at time \( t-1 \) (DJIA)
- The Financial Times 100 Index at time \( t-1 \) (FSTE)
- The S&P 500 Index at time \( t-1 \) (SP)
- The Nikkei225 Index at time \( t \) (NIX)
- The Hang Seng Index at time \( t \) (HSKI)
- The Singapore Straits Times Industrial Index at time \( t \) (SES009)
- The Taiwan Stock Weighted Index at time \( t \) (TWII)
- The South Korea Stock Exchange Index at time \( t \) (KOSPI)
- The Oil Price in the New York Mercantile Exchange at time \( t \) (OIL)
- The Gold Price in the New York Mercantile Exchange at time \( t \) (GOLD)
- The Currency Exchange Rate in Thai Baht for one US dollar at time \( t \) (THB/USD)

The Currency Exchange Rate in Thai Baht for one Hong Kong dollar at time \( t \) (THB/HKD).

The actual closing prices for these twelve factors were obtained from http://www.efinancethai.com. We used data sets from April 5, 2000, to July 5, 2012. We divided these data into two disjoint sets. The first set, from April 5, 2000, to December 30, 2011, was used as a sample (2,873 observations). The second set, from January 3, 2012, to July 5, 2012, was used as out-of-sample (125 observations). The plot for the SET Index closing prices and returns is given in Fig. 1.

Descriptive statistics and the correlations matrix are given in Table 2 and 3. As can be seen from Table 3, there are highly significant correlations (p<0.01) between the dependent variables and the explanatory variables. Therefore, these explanatory variables were used to predict the SET Index. Also, there are highly significant correlations (p<0.01) among the explanatory variables. These correlations provide a measure for the linear relations between two variables and also indicate the existence of multicollinearity between the explanatory variables. However, multiple regression analysis based on this dataset also shows that there was a multicollinearity problem with the Variance Inflation Factor (VIF > 5.0) as shown in Table 2. One approach to avoid this problem is PCA. Hence, we used twelve explanatory variables to find the principal components and overall descriptive statistics for selected Principal Components (PCs), as shown in Table 4 and 5, respectively.

3.2. Results of Principal Component Analysis

Bartlett’s sphericity test for testing the null hypothesis where the correlation matrix is an identity matrix was used to verify the applicability of PCA. The value of Bartlett’s sphericity test for the SET Index was 18,167.07, which implies that the PCA is applicable to our datasets (Table 2). Moreover, Kaiser’s measure of sampling adequacy was also computed as 0.788, which indicates that the sample sizes were sufficient for us to apply the PCA. The results for PCA (Table 4) indicate that there are twelve Principal Components (PCs) for multiple regression analysis.

3.3. Forecasting the Returns the Set Index By Mean Equations

We forecast the returns for the SET Index (\( r_t = \mu_t + \varepsilon_t \)) using four mean equations (\( \mu_t \)): Constant, ARMA (1, 1), multiple regression based on PCA and ARMA (1,1), which includes multiple regression based on PCA. Afterwards, we compare error using two loss functions, i.e., Mean Square Error (MSE) and Mean Absolute Error (MAE).
Table 1. Impact factors on the Stock Exchange of Thailand Index (SET Index)

| Factors                                             | Researchers |
|-----------------------------------------------------|-------------|
| The Nasdaq Index                                     | X           |
| The Down Jones Index                                 | X X X X X X X X |
| The S&P 500 Index                                    | X           |
| The Nikkei Index                                     | X X X X X X X |
| The Hang Seng Index                                  | X X X X X X X |
| The Straits Times Industrial Index                   | X X X       |
| The Currency Exchange Rate in Thai Baht to one US dollar | X X X     |
| The Currency Exchange Rate in Thai Baht to 100 Japan Yen | X X       |
| The Currency Exchange Rate in Thai Baht to one Hong Kong dollar | X X |
| The Currency Exchange Rate in Thai Baht to one Singapore dollar | X |
| Gold Prices                                          | X X X       |
| Oil Prices                                           | X X X       |
| Minimum Loan Rates                                   | X X X X     |

*X is selected in multiple regression*
### Table 2. Descriptive statistics of the SET Index and explanatory variables

| Variables | Mean   | Std. Deviation | Skewness | Kurtosis | Correlation with SETclose | VIF |
|-----------|--------|----------------|----------|----------|---------------------------|-----|
| SET       | 0.0373 | 1.4644         | -0.690   | 9.194    | 1                         |     |
| DJIA      | 0.0047 | 1.2792         | -0.017   | 7.626    | 0.219**                   | 14.581 |
| FSTE      | -0.0043| 1.3280         | -0.169   | 5.718    | 0.166**                   | 1.527 |
| SP        | -0.0031| 1.3647         | -0.128   | 7.764    | 0.239**                   | 15.197 |
| NIX       | -0.0273| 1.5986         | -0.499   | 7.609    | 0.369**                   | 2.010 |
| HSKI      | 0.0053 | 1.6593         | -0.067   | 8.960    | 0.495**                   | 2.405 |
| SES900    | 0.0122 | 1.3011         | -0.337   | 7.674    | 0.507**                   | 2.150 |
| TWII      | -0.0096| 1.5716         | -0.202   | 3.348    | 0.351**                   | 1.618 |
| KOSPI     | 0.0272 | 1.7733         | -0.867   | 7.973    | 0.410**                   | 2.152 |
| OIL       | 0.0413 | 2.5662         | 0.087    | 7.578    | 0.119**                   | 1.057 |
| TWII/USD  | -0.0063| 0.4258         | 0.511    | 20.223   | -0.152**                  | 2.197 |
| TWII/HKD  | -0.0059| 0.5304         | 0.570    | 32.596   | -0.107**                  | 2.175 |

**Jarque-Bera Normality test in SETclose**

10741.72**

Augmented Dickey-Fuller test in SETclose

-52.76**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy

0.79

Bartlett's sphericity test

Approx. Chi-Square 18167.073

df 66.000

Sig. 0.000

**; Significant at the 0.01 level (2-tailed)

### Table 3. Correlation matrix of the SET Index and explanatory variables

| Correlations | SET | DJIA | FSTE | SP | NIX | HSKI | SES900 | TWII | KOSPI | OIL | GOLD | TWII/USD | TWII/HKD |
|--------------|-----|------|------|----|-----|------|--------|------|-------|-----|------|----------|----------|
| SET          | 1   |      |      |    |     |      |        |      |       |     |      |          |          |
| DJIA         | 0.22** | 1  |      |    |     |      |        |      |       |     |      |          |          |
| FSTE         | 0.17**  | 0.55** | 1   |    |     |      |        |      |       |     |      |          |          |
| SP           | 0.24**  | 0.36** | 0.56** | 1  |     |      |        |      |       |     |      |          |          |
| NIX          | 0.37**  | 0.45** | 0.39** | 0.47** | 1  |      |        |      |       |     |      |          |          |
| HSKI         | 0.50**  | 0.37** | 0.29** | 0.40** | 0.59** | 1  |        |      |       |     |      |          |          |
| SES900       | 0.51**  | 0.33** | 0.20** | 0.35** | 0.53** | 0.70** | 1      |      |       |     |      |          |          |
| TWII         | 0.35**  | 0.30** | 0.23** | 0.32** | 0.45** | 0.49** | 0.47** | 1  |       |     |      |          |          |
| KOSPI        | 0.41**  | 0.31** | 0.26** | 0.34** | 0.59** | 0.61** | 0.57** | 0.57** | 1  |      |      |          |          |
| OIL          | 0.12**  | 0.01  | -0.01 | 0.01 | 0.46** | 0.10** | 0.11** | 0.06** | 0.06** | 1  |      |      |          |          |
| GOLD         | 0.08**  | 0.04* | 0.03  | 0.05** | 0.07** | 0.09** | 0.07** | 0.02  | 0.07** | 0.20** | 1   |      |      |          |          |
| TWII/USD     | -0.15** | -0.07** | -0.05** | -0.08** | -0.12** | -0.12** | -0.10** | -0.13** | -0.04** | -0.13** | 1  |      |      |          |          |
| TWII/HKD     | -0.11** | 0.00  | -0.01 | -0.02 | 0.00  | -0.07** | -0.10** | -0.11** | -0.08** | -0.12** | -0.02 | -0.10** | 1  |          |          |

**; Correlation significant at the 0.01 level (2-tailed)

### Table 4. Descriptive statistics of selected PCs

| Initial eigenvalues | Weight for the PCs |
|---------------------|---------------------|
| PC1                  | DJIA               |
| PC2                  | FSTE               |
| PC3                  | SP                 |
| PC4                  | NIX                |
| PC5                  | HSKI              |
| PC6                  | SES900            |
| PC7                  | TWII              |
| PC8                  | KOSPI             |
| PC9                  | OIL               |
| PC10                 | GOLD              |
| PC11                 | TWII/USD         |
| PC12                 | TWII/HKD         |

**; Significant at the 0.01 level (2-tailed)
Table 5. Mean equations for returns of the SET Index and loss functions

| Model                                  | Mean Equation                             | MSE   | MAE   |
|----------------------------------------|-------------------------------------------|-------|-------|
| 1. Constant mean.                      | $\mu_t = E[r_t]$, $\mu_t = 0.0373$,     | 0.8914| 0.7576|
|                                        | $\mu_t = \phi_0 + \sum_{j=1}^{p} \theta_j r_{t-j} - \sum_{k=1}^{q} \phi_k e_{t-k}$, $\mu_t = 0.4545e_{t-1} + 0.4951e_{t-1}$ | 0.8963| 0.7627|
| 2. ARMA (1,1)                          | $\mu_t = \phi_0 + \sum_{j=1}^{p} \theta_j Z_{t-j}$, $\mu_t = 0.163Z_t - 0.055Z_{t-1} + 0.259Z_{t-2} + 0.499Z_{t-3} + 0.059Z_{t-4} + 0.124Z_{t-5}$ | 0.5444| 0.5963|
| 3. Multiple regressions based on PCA.  | $\mu_t = \phi_0 + \sum_{i=1}^{n} \alpha_i Z_{t-i}$, $\mu_t = 0.162Z_t - 0.054Z_{t-1} + 0.258Z_{t-2} + 0.500Z_{t-3} + 0.059Z_{t-4} + 0.124Z_{t-5} + 0.271Z_{t-6} + 0.214Z_{t-7} + 0.080Z_{t-8} - 0.146Z_{t-9} + 0.405Z_{t-10} - 0.991r_{t-1} - 0.996r_{t-2}$ | 0.5393| 0.5947|
| 4. ARMA (1,1) and multiple regressions | $\mu_t = \phi_0 + \sum_{j=1}^{p} \theta_j Z_{t-j}$, $\mu_t = 0.162Z_t - 0.054Z_{t-1} + 0.258Z_{t-2} + 0.500Z_{t-3} + 0.059Z_{t-4} + 0.124Z_{t-5} + 0.271Z_{t-6} + 0.214Z_{t-7} + 0.080Z_{t-8} - 0.146Z_{t-9} + 0.405Z_{t-10} - 0.991r_{t-1} - 0.996r_{t-2}$ | 0.5393| 0.5947|

Table 6. The ACF of the SET Index returns series squared mean adjusted returns and results for the Engle’s ARCH test

| Lags | ACF of returns | LBQ Test | P-value | ACF of squared mean adjusted | LBQ Test | P-value | Engle’s ARCH test | P-value |
|------|----------------|----------|---------|------------------------------|----------|---------|-------------------|---------|
| 1    | 0.03620        | 3.92450  | 0.0476  | 0.3164                       | 300.2830 | 0       | 262.0482          | 0       |
| 2    | 0.07270        | 19.7643  | 0.0001  | 0.0569                       | 310.0081 | 0       | 273.6911          | 0       |
| 3    | 0.00650        | 19.8900  | 0.0002  | 0.0368                       | 314.0780 | 0       | 274.1944          | 0       |
| 4    | -0.01800       | 20.8621  | 0.0003  | 0.0173                       | 314.9743 | 0       | 274.1831          | 0       |
| 5    | -0.00400       | 20.9113  | 0.0008  | 0.0370                       | 319.0913 | 0       | 274.1547          | 0       |
| 6    | -0.04810       | 27.8598  | 0.0001  | 0.0233                       | 320.7266 | 0       | 274.1156          | 0       |
| 7    | 0.00600        | 27.9695  | 0.0002  | 0.0062                       | 320.8405 | 0       | 274.0736          | 0       |
| 8    | -0.01640       | 27.8705  | 0.0003  | 0.0362                       | 324.7808 | 0       | 274.0363          | 0       |
| 9    | 0.03430        | 32.3206  | 0.0002  | 0.0493                       | 332.0998 | 0       | 274.0368          | 0       |
| 10   | 0.04280        | 37.8204  | 0.0000  | 0.3164                       | 300.2830 | 0       | 262.0482          | 0       |
| 22   | -0.00450       | 62.1468  | 0.0000  | 0.0097                       | 358.6197 | 0       | 273.4667          | 0       |

Table 7. Estimation parameters of volatility models

Panel A. Summary results of the GARCH type models

| Parameters | GARCH     | EGARCH    | GJR-GARCH |
|------------|-----------|-----------|-----------|
| N          | 0.1875    | -0.1098   | 0.2256    |
| t          | 0.0896    | -0.1385   | 0.1053    |
| GED        | 0.113     | -0.1307   | 0.1352    |
| Std. err.  | 0.0166*** | 0.0129*** | 0.0188*** |
| a0         | 0.1129    | 0.2265    | 0.1978    |
| Std. err.  | 0.01175***| 0.2142    | 0.1684    |
| a1         | 0.0216*** | 0.219     | 0.1799    |
| Std. err.  | 0.0216****| 0.0289*** | 0.0557    |
| $\beta_0$  | 0.7923    | -0.1104   | 0.7621    |
| Std. err.  | 0.8346    | -0.0732   | 0.8212    |
| $\beta_1$  | 0.8252    | -0.0657   | 0.8067    |
| Std. err.  | 0.0231*** | 0.0289*** | 0.0557    |
| $\xi$      | 0.8904    | 0.0098*** | 0.0437    |
| Std. err.  | 0.9463    | 0.0139*** | 0.0748    |
| v           | 1.3066    | 0.0153*** | 0.0668    |
| Std. err.  | 0.9316    | 0.0208*** | 0.0748    |
| n           | 7.1941    | 0.0097*** | 0.0668    |
| Std. err.  | 1.3066    | 0.0254*** | 0.0748    |
| Log(L)     | 4982.5400 | 7.5148    | 7.4846    |
| Persistence| 0.9079    | 0.0985    | 0.9324    |
The parameters for mean equations for forecasting the SET Index and the value of loss functions are shown in Table 5. We found that the mean equation ARMA (1, 1) that includes multiple regression based on PCAs (Table 5, No. 4) has the best performance (MSE = 0.5393, MAE = 0.5947). So, we use this mean equation for forecasting the returns for the SET Index.

3.4. Forecasting the Volatility of Returns the Set Index

We applied Ljung and Box to test serial correlation for returns ($r_t$) and squared mean returns adjusted ($r_t - \mu_t$)² where $\mu_t$ is the mean equation in Table 5 (No. 4). We used a specified lag from the first to the tenth lags and we used the twenty-second lag in Table 6. Serial correlation for returns is confirmed as stationary because the Autocorrelation Function (ACF) values decrease very fast when lags increase and this is confirmed by the Augmented Dickey-Fuller Test (-52.76**) as in Table 1. We analyzed the significance of autocorrelation in the squared mean adjusted returns series with the Ljung-Box Q-test and used Engle’s ARCH test to test the ARCH effects. Therefore, the squared mean for the adjusted return is non-stationary, which suggests conditional heteroskedasticity.
| models          | MSE1 | MSE2 | QLIKE | MAD1 | MAD2 | HMSE | MSE1 | MSE2 | QLIKE | MAD1 | MAD2 | HMSE | MSE1 | MSE2 | QLIKE | MAD1 | MAD2 | HMSE |
|-----------------|------|------|-------|------|------|------|------|------|-------|------|------|------|------|------|-------|------|------|------|
| GARCH-N         | 0.7415 | 1.2641 | 0.1159 | 0.7906 | 0.5977 | 3.8836 | 33.7606 | 1.1285 | 1.4994 | 4.1836 | 0.5981 |
| GARCH-t         | 0.7451 | 1.2772 | 0.1189 | 0.7942 | 0.5977 | 3.9473 | 34.5974 | 1.1403 | 1.5133 | 4.2501 | 0.5981 |
| GARCH-GED       | 0.7407 | 1.2632 | 0.1152 | 0.7897 | 0.5977 | 3.9036 | 33.9598 | 1.1329 | 1.5042 | 4.2045 | 0.5976 |
| EGARCH-N        | 0.9302 | 1.9584 | 0.2453 | 0.7283 | 0.9832 | 0.5977 | 3.9036 | 33.7606 | 1.1285 | 1.4994 | 4.1836 | 0.5981 |
| EGARCH-t        | 0.9572 | 2.1035 | 0.2588 | 0.7374 | 1.0111 | 0.5977 | 3.9473 | 34.5974 | 1.1403 | 1.5133 | 4.2501 | 0.5981 |
| EGARCH-GED      | 0.9491 | 2.0255 | 0.2492 | 0.7316 | 0.9946 | 0.5977 | 3.9036 | 33.9598 | 1.1329 | 1.5042 | 4.2045 | 0.5976 |
| GJR-GARCH-N     | 0.7415 | 1.2641 | 0.1159 | 0.7906 | 0.5977 | 3.8836 | 33.7606 | 1.1285 | 1.4994 | 4.1836 | 0.5981 |
| GJR-GARCH-t     | 0.7451 | 1.2772 | 0.1189 | 0.7942 | 0.5977 | 3.9473 | 34.5974 | 1.1403 | 1.5133 | 4.2501 | 0.5981 |
| GJR-GARCH-GED   | 0.7407 | 1.2632 | 0.1152 | 0.7897 | 0.5977 | 3.9036 | 33.9598 | 1.1329 | 1.5042 | 4.2045 | 0.5976 |
| MRS-GARCH-N     | 0.9302 | 1.9584 | 0.2453 | 0.7283 | 0.9832 | 0.5977 | 3.9036 | 33.7606 | 1.1285 | 1.4994 | 4.1836 | 0.5981 |
| MRS-GARCH-t     | 0.9572 | 2.1035 | 0.2588 | 0.7374 | 1.0111 | 0.5977 | 3.9473 | 34.5974 | 1.1403 | 1.5133 | 4.2501 | 0.5981 |
| MRS-GARCH-GED   | 0.9491 | 2.0255 | 0.2492 | 0.7316 | 0.9946 | 0.5977 | 3.9036 | 33.9598 | 1.1329 | 1.5042 | 4.2045 | 0.5976 |

4. EMPIRICAL METHODOLOGY

This empirical section adopts the GARCH type and the MRS-GARCH (1, 1) models to estimate the volatility of the returns on the SET Index. The GARCH type models considered are the GARCH (1, 1), EGARCH (1,1) and GJR-GARCH (1,1). In order to account for the fat-tailed feature of financial returns, we considered three different distributions for the innovations: Normal (N), Student-t (t) and Generalized Error Distributions (GED).

4.1. GARCH Type Models

Panel A of Table 7 presents an estimation of the results for the GARCH type models. It is clear from the table that almost all parameter estimates in the GARCH type models are highly significant at 1%. However, the asymmetry effect term \( \xi \) in the EGARCH models is significantly different from zero which indicates unexpected negative returns, implying higher conditional variance as compared to the same-sized positive returns. All models display strong persistence in volatility ranging from 0.8895 to 0.9572, that is, volatility is likely to remain high over several price periods once it increases.

4.2. Markov Regime Switching GARCH Models

The estimated results and summary statistics for the MRS-GARCH models are presented in Panel B of Table 7. Most parameter estimates in the MRS-GARCH are significantly different from zero at least at the 95% confidence level. But \( \alpha_1, \beta_1 \) are not significantly different in some areas. All models display strong persistence in volatility ranging from 0.6650 to 0.9892, that is,
volatility is likely to remain high over several price periods once it increases.

4.3. In-Sample Evaluation

We used various goodness-of-fit statistics to compare volatility models. These statistics are the Akaike Information Criteria (AIC), the Schwarz Bayesian Information Criteria (BIC) and the Log-Likelihood (LOGL) values. Table 8 presents the results for the goodness-of-fit statistics and loss functions for all volatility models. According to the BIC, the MSE2 and the QLIKE, the GJR model performs best in modeling SET Index volatility. However, the contrast in the AIC, the LOGL, the MSE1, the R2LOG, the MAD2 and the MAD1 suggests that the MRS-GARCH performs best.

4.4. Forecasting Volatility in Out-of-Samples

We investigate the ability of the GARCH, EGARCH, GJR-GARCH and MRS-GARCH models to forecast volatility for the SET Index out-of-sample set. In Table 9, we present the results for loss function for out-of-samples in forecasting volatility for one day ahead, five days ahead (short term), ten days ahead and twenty-two days ahead (long term). We found the GARCH model performs best for one day ahead; the EGARCH model performs best for five days, ten days and twenty-two days ahead.

5. CONCLUSION

We considered the problem of forecasting returns for the SET Index by using a stationary Autoregressive Moving-average order p and q (ARMA (p, q)) with some explanatory variables. After considering four types of mean equations, we found that ARMA (1, 1), which includes multiple regressions based on PCA, has the best performance (MSE = 0.5393, MAE = 0.5947). In forecasting the volatility of the returns for the SET Index, GARCH type models such as GARCH (1, 1), EGARCH (1, 1), GJR-GARCH (1, 1) and MRS-GARCH (1, 1) were considered. We found that the GARCH (1, 1) model performs best for one day ahead and the EGARCH (1, 1) model performs best for five days, ten days and twenty-two days ahead respectively.

6. REFERENCES

Caporale, G.M., J. Cunado and L.A. Gil-Alana, 2011. Forecasting the Spanish stock market returns with fractional and non-fractional models. Am. J. Econ. Bus. Admin., 3: 586-588. DOI: 10.3844/ajebasp.2011.586.588

Chaigusin, S., C. Chirathamjaree and J. Clayden, 2008. Soft computing in the forecasting of the stock exchange of Thailand (SET). Proceedings of the 4th IEEE International Conference on Management of Innovation and Technology, Sept. 21-24, IEEE Xplore Press, Bangkok, pp: 1277-1281. DOI: 10.1109/ICMIT.2008.4654554

Ismail, Z., A. Yahya and A. Shabri, 2009. Forecasting gold prices using multiple linear regression method. Am. J. Applied Sci., 6: 1509-1514. DOI: 10.3844/ajassp.2009.1509.1514

Sattayatham, P., N. Sopipan and B. Premanode, 2012. Forecasting the stock exchange of Thailand uses day of the week effect and markov regime switching GARCH. Am. J. Econ. Bus. Admin., 4: 84-93. DOI: 10.3844/ajebasp.2012.84.93

Sopipan, N., P. Sattayatham and B. Premanode, 2012. Forecasting volatility of gold price using Markov regime switching and trading strategy. J. Math. Finance, 2: 121-131 DOI: 10.4236/jmf.2012.21014

Tsay, R.S., 2010. Analysis of Financial Time Series. 3rd Edn. John Wiley and Sons, ISBN-10: 1118017099, pp: 712.