Journal of Mathematics and Statistics 9 (1): 29-37, 2013
ISSN 1549-3644

doi:10.3844/jmssp.2013.29.37 Published Online 9 (1) 2013 (http://www.thescipub.com/jmss.toc)

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

[1]Nop Sopipan, [2]Anchalee Sattayatham and [3]Samruam Chongcharoen

[1]Program of Mathematics and Applied Statistics, Faculty of Science and Technology,
Nakhon Ratchasima Rajabhat University, Nakhon Ratchasima, Thailand
[2]Faculty of Liberal Arts, Mahidol University, Nakhon Pathom, Thailand
[3]National 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 μ_t as appears in Equation (2). For example, Sopipan *et al.* (2012) assumes μ_t to be a constant and Sattayatham *et al*. (2012) assume μ_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 = 100. In (P_t/P_{t-1}) for t = 1, 2, …, T-1, P_t 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:

$$r_t = \mu_t + \varepsilon_t,$$
$$\mu_t = \mu_0 + \sum_{i=1}^{n} \beta_i X_{it} + \sum_{s=1}^{p} \phi_s r_{t-s} - \sum_{m=1}^{q} \theta_m \varepsilon_{t-m} \quad (1)$$

Where:

$$X_{it} = 100.\ln\left(\frac{P_{it}}{P_{i(t-1)}}\right) \quad (2)$$

Here P_{it} denotes the financial price asset i for i = 1, 2,…, n at time t, r_{t-5} S = 1, 2, …, P is the returns at lag s-th, ε_t represents errors assumed to be a white noise series with an i.i.d. mean of zero and a constant variance σ_ε^2, μ_0, β_I, ϕ_1, θ_m are constants and n, p and q are positive integers.

Generally, the variance of errors ε_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

Corresponding Author: Nop Sopipan, Program of Mathematics and Applied Statistics, Faculty of Science and Technology, Nakhon Ratchasima Rajabhat University, Nakhon Ratchasima, Thailand

the variance of ε_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 μ_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 $X_t = (X_{1t}{}' X_{2t},\ldots,X_{rt})'$ with covariance matrix Σ_x, where (.)' denotes the transpose matrix. Principal Component Analysis (PCA) is concerned with using fewer linear combinations of X_i to explain the structure of Σ_x. If X_{it} denotes returns as appears in (2) for i = 1,2,..,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 Σ_x, with the eigenvalues λ_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 $Z_{it} = e_i' X_t = \sum_{j=1}^{n} e_{ij} X_{jt}$ for i = 1,..,n. We note that Equation 3:

$$\begin{aligned} &\mathrm{Var}(Z_{it}) = e_i'\Sigma_x e_i = \lambda_i, \\ &\mathrm{Cov}(Z_{it}, Z_{jt}) = e_i'\Sigma_x e_j = 0, \quad i \neq j, i, j = 1, 2, \ldots, n \end{aligned} \tag{3}$$

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:

$$\begin{aligned} r_t &= \mu_0 + \sum_{i=1}^{n} \alpha_i Z_{it} + \sum_{s=1}^{p} \phi_s r_{t-s} \\ &- \sum_{m=1}^{q} \theta_m \varepsilon_{t-m} + \varepsilon_t \end{aligned} \tag{4}$$

where, Z_{it} i = 1,2,…,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: $\varepsilon_t = \eta_t \sqrt{h_t}$, $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$ where η_t is i.i.d. distributed with zero mean and unit variance, $a_0, a_1 > 0$ and $\beta_1 > 0$ to ensure positive conditional variance. The inequality $a_1+\beta_1<1$ 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:

$$\ln(h_t) = \alpha_0 + \alpha_1 \left| \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right| + \beta_1 \ln(h_{t-1}) + \xi \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}},$$

where, ξ 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:

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \left(1 - I_{\{\varepsilon_{t-1}>0\}}\right) \\ &+ \beta_1 h_{t-1} + \xi \varepsilon_{t-1}^2 \left(I_{\{\varepsilon_{t-1}>0\}}\right) \end{aligned}$$

where, $I_{\{\varepsilon t-1>0\}}$ is equal to one when ε_{t-1} is greater than zero and equal to zero elsewhere. The conditions $a_0>0, (a_1+\xi)/2>0$ 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:

$$\varepsilon_t = \eta_t \sqrt{h_{t,S_t}} \text{ and } h_{t,S_t} = \alpha_{0,S_t} + \alpha_{1,S_t}\varepsilon_{t-1}^2 + \beta_{1,S_t} h_{t-1}$$

where, S_t = 1 or 2, h_{st} 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 $a_{0,st} > a_{1,st} \geq 0$ and $\beta_{1,st} \geq 0$. The sum $a_{1,st} + \beta_{1,St}$ 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 (μ_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).

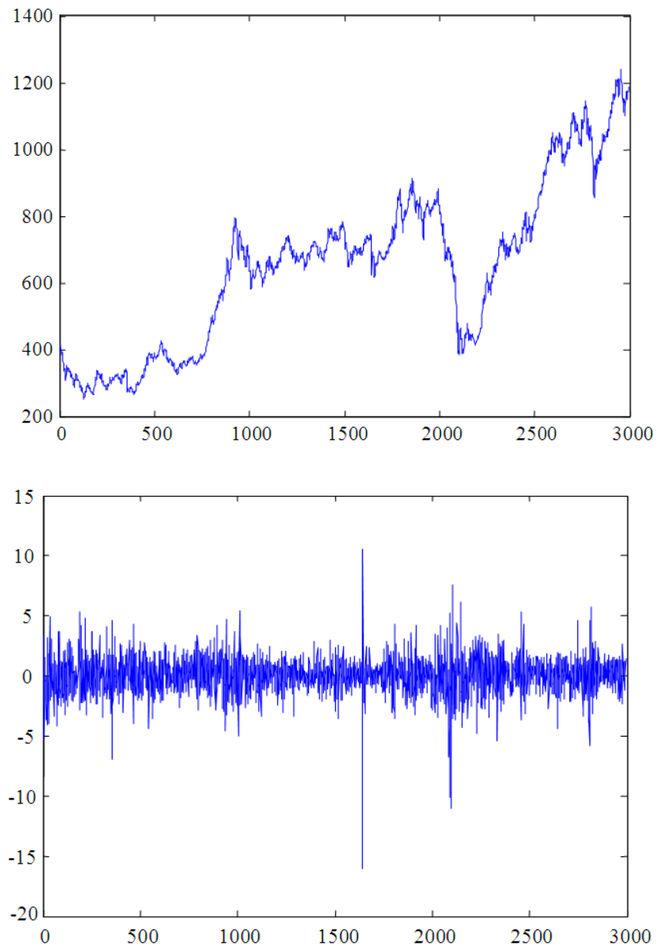

Fig. 1. Graph of the SET index and returns of the SET Index

Table 1. Impact factors on the Stock Exchange of Thailand Index (SET Index)

	Researchers							
Factors	1	2	3	4	5	6	7	8
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
The Hang Seng Index	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	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					
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

*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	9.737	0.410**	2.152
OIL	0.0413	2.5662	0.087	7.578	0.119**	1.057
GOLD	0.0581	1.1831	0.137	6.383	0.077**	1.068
THB/USD	-0.0063	0.4258	0.511	20.223	-0.152**	2.197
THB/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	THB/USD	THB/HKD
SET	1												
DJIA	0.22**	1											
FSTE	0.17**	0.55**	1										
SP	0.24**	0.96**	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	0.06**	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		
THB/USD	-0.15**	-0.07**	-0.05**	-0.08**	-0.08**	-0.12**	-0.12**	-0.10**	-0.13**	-0.04*	-0.13**	1	
THB/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						
PCs	Total	(%) of Variance	Cumulative (%)	DJIA	FSTE	SP	NIX	HSKI	SES900	TWII	KOSPI	OIL	GOLD	THB/ USD	THB/ HKD
1	4.266	35.548	35.548	0.732	0.572	0.753	0.775	0.777	0.727	0.656	0.742	0.104	0.119	-0.224	-0.212
2	1.734	14.453	50.001	0.257	0.224	0.249	0.071	-0.036	-0.068	-0.020	-0.065	-0.158	-0.279	0.854	0.842
3	1.482	12.347	62.348	-0.538	-0.451	-0.518	0.155	0.329	0.377	0.320	0.372	0.220	0.039	0.262	0.294
4	1.151	9.595	71.944	0.097	0.069	0.097	-0.029	-0.037	-0.057	-0.149	-0.126	0.731	0.714	0.126	0.153
5	0.789	6.575	78.519	-0.058	0.009	-0.047	0.071	0.042	-0.014	-0.050	0.059	-0.614	0.622	0.050	0.068
6	0.607	5.056	83.576	-0.145	0.367	-0.147	0.015	-0.231	-0.361	0.462	0.167	0.046	0.043	-0.011	0.007
7	0.570	4.749	88.325	0.232	-0.453	0.214	-0.255	-0.148	-0.030	0.410	-0.008	-0.037	0.079	0.002	0.025
8	0.448	3.736	92.061	-0.051	0.264	-0.055	-0.458	0.197	0.256	0.141	-0.193	-0.033	0.026	-0.001	0.021
9	0.355	2.960	95.020	0.040	0.014	0.038	-0.297	0.017	-0.080	-0.199	0.465	0.016	-0.009	0.023	0.000
10	0.299	2.494	97.514	-0.009	0.066	-0.016	0.017	-0.406	0.339	-0.023	0.072	-0.005	0.018	0.068	-0.060
11	0.264	2.198	99.712	-0.002	-0.012	-0.001	-0.005	0.066	-0.048	0.024	-0.021	0.003	0.012	0.359	-0.357
12	0.035	0.288	100.000	0.130	0.001	-0.133	0.003	0.003	0.000	-0.001	0.001	0.000	0.001	0.000	0.000

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
2. ARMA (1,1)	$\mu_t = \phi_0 + \sum_{j=1}^{p}\phi_j r_{t-j} - \sum_{k=1}^{q}\theta_k\varepsilon_{t-k}\mu_t$, $= 0.5454r_{t-1} + 0.4951\varepsilon_{t-1}$	0.8963	0.7627
3. Multiple regressions based on PCA.	$\mu_t = \phi_0 + \sum_{i=1}^{n}\alpha_i Z_{it}$, $\mu_t = 0.163Z_1$-$0.055Z_2$ $+0.259Z_3+0.499Z_4+0.059Z_5+0.124Z_6$ $+0.272Z_7+0.215Z_8+0.077Z_9-0.146Z_{10}+0.410Z_{11}$	0.5444	0.5963
4. ARMA (1,1) and Multiple regressions based on PCA.	$\mu_t = \phi_0 + \sum_{i=1}^{n}\alpha_i Z_{it} + \sum_{j=1}^{p}\phi_j r_{t-j} - \sum_{k=1}^{q}\theta_k\varepsilon_{t-k}$ μ_t=$0.162Z_1-0.054Z_2+0.258Z_3+0.500Z_4$ $+0.059Z_5+0.124Z_6$ $+0.271Z_7+0.214Z_8+0.080Z_9-0.146Z_{10}$ $+0405Z_{11}-0.991r_{t-1}-0.996\varepsilon_{t-1}$	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

	ACF of returns			ACF of squared mean adjusted			Engle’s ARCH test	
Lags	ACF	LBQ Test	P-value	ACF	LBQ Test	P-value	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	28.7805	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

	GARCH			EGARCH			GJR-GARCH		
Parameters	N	t	GED	N	t	GED	N	t	GED
a_0	0.1875	0.0896	0.113	-0.1098	-0.1385	-0.1307	0.2256	0.1053	0.1352
Std.err.	0.0166***	0.0175***	0.0216***	0.0129***	0.0176***	0.0194***	0.0188***	0.0194***	0.0244***
a_1	0.1129	0.1184	0.1182	0.2265	0.2142	0.219	0.1978	0.1684	0.1799
Std.err.	0.0124***	0.0161***	0.0185***	0.0211***	0.0250***	0.0289***	0.0227***	0.0235***	0.0258***
β_0	0.7923	0.8346	0.8252	-0.1104	-0.0557	-0.0732	0.7621	0.8212	0.8067
Std.err.	0.0184***	0.0193***	0.0231***	0.0098***	0.0139***	0.0153***	0.0209***	0.0208***	0.0254***
ξ				0.8904	0.9463	0.9316	0.0437	0.0748	0.0668
Std.err.				0.0097***	0.0104***	0.0127***	0.0114***	0.0186***	0.0200***
v		7.1941	1.3066		7.5148	1.3312		7.4846	1.3279
Std.err.		0.5994***	0.0245***		0.6422***	0.0259***		0.6346***	0.0249***
Log(L)	-4982.5400	-4832.1100	-4863.5300	-4957.5400	-4824.0900	-4853.3000	-4957.1400	-4822.7000	-4852.1600
Persistence	0.9079	0.9572	0.9458	0.8895	0.9462	0.9313	0.8908	0.9460	0.9324

Panel B. Summary results of the MRS-GARCH models

Parameters State I	MRS-GARCH N i = 1	N i = 2	t i = 1	t i = 2	2t i = 1	2t i = 2	GED i = 1	GED i = 2
a_0^i	6.0608	0.0776	0.0463	1.1969	0.0455	1.4481	0.0733	1.2588
Std.err.	1.0569***	0.0211***	0.0177***	0.4680**	0.0202***	0.8083***	0.0211***	1.0171**
a_1^i	0.1888	0.0600	0.0670	0.2957	0.0688	0.4171	0.0679	0.0413
Std.err.	0.1170	0.0176***	0.0201	0.1072	0.0179***	0.1925**	0.0179***	0.1340
β_1^i	0.0000	0.8354	0.8829	0.3593	0.8725	0.3001	0.8463	0.9479
Std.err.	0.3649	0.0192***	0.0220***	0.1754**	0.0211***	0.2544	0.0202***	0.4533**
p		0.5712		0.9829		0.9798		0.9904
Std.err.	0.1437***		0.0077***		0.0092***		0.0054***	
q	0.9834		0.9070		0.8231		0.4467	
Std.err.	0.0043***		0.0403***		0.0687***		0.2190**	
v^i			8.2587		11.1122	3.694	1.5245	
Std.err.			0.9680***		2.6917***	0.8177***	0.0606***	
Log(L)	-4847.8700		-4808.7100		-4815.9800		-4812.8100	
σ^2	7.4724	0.7419	0.9242	3.4693	0.7751	5.1206	0.8543	6.5556
π	0.4914	0.5086	0.1571	0.8429	0.1025	0.8975	0.0171	0.9829
Persistence	0.8954	0.8954	0.9499	0.6550	0.9413	0.7172	0.9142	0.9892

Note: *** and **; refer to the significance at 99 and 95% confidence level respectively

Table 8. In-sample evaluation results

Models	N	PERS	AIC	BIC	LOGL	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE
GARCH-N	4	0.9079	3.4713	3.4796	-4982.54	1.0845	49.0269	1.63019	8.9160	2.3231	0.8021	20.0635
GARCH-t	5	0.9572	3.3673	3.3777	-4832.11	1.1038	49.8378	1.64870	8.6848	2.3526	0.7921	37.6330
GARCH-GED	5	0.9458	3.3892	3.3995	-4863.53	1.0994	49.4764	1.63897	8.7811	2.3483	0.7972	30.0802
EGARCH-N	5	0.8895	3.4546	3.4650	-4957.54	1.0643	48.0251	1.61546	8.8305	2.2849	0.7943	16.5328
EGARCH-t	6	0.9462	3.3624	3.3749	-4824.09	1.0675	48.7977	1.63144	8.6435	2.2846	0.7820	30.5723
EGARCH-GED	6	0.9313	3.3827	3.3952	-4853.30	1.0691	48.5741	1.62242	8.7227	2.2878	0.7875	24.4238
GJR-N	5	0.8908	3.4543	3.4647	-4957.14	1.0662	47.4778	1.61467	8.8410	2.3084	0.7959	17.6206
GJR-t	6	0.9460	3.3614	3.3739	-4822.70	1.0868	48.7559	1.63050	8.6378	2.3362	0.7868	31.3575
GJR-GED	6	0.9324	3.3819	3.3944	-4852.16	1.0830	48.3094	1.62152	8.7255	2.3340	0.7916	25.2428
MRS-GARCH-N	10	0.9000	3.3817	3.4025	-4847.87	1.0533	48.1202	1.64076	8.6361	2.2815	0.7824	32.0252
MRS-GARCH-t2	12	0.9362	3.3559	3.3808	-4808.71	1.1006	50.5999	1.63915	8.6654	2.3426	0.7908	34.4366
MRS-GARCH-t	11	0.9578	3.3602	3.3831	-4815.98	1.0930	49.8728	1.65543	8.5980	2.3316	0.7843	42.2318
MRS-GARCH-GED	11	0.9525	3.3580	3.3809	-4812.81	1.0591	48.1754	1.65112	8.6063	2.2850	0.7801	41.6225

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 μ_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.

Table 9. Result loss function of out-of-samples with forecasting volatility

Panel A. Out of sample for one day ahead and five days ahead (short term)

	One day ahead						Five days ahead					
Models	MSE1	MSE2	QLIKE	MAD1	MAD2	HMSE	MSE1	MSE2	QLIKE	MAD1	MAD2	HMSE
GARCH-N	0.7415	1.2641	0.1159	0.6511	0.7906	0.5977	3.8836	33.7606	1.1285	1.4994	4.1836	0.598
GARCH-t	0.745	1.2772	0.1189	0.6527	0.7942	0.5977	3.9473	34.5974	1.1403	1.5133	4.2501	0.5981
GARCH-GED	0.7407	1.2632	0.1152	0.6507	0.7897	0.5977	3.9036	33.9598	1.1329	1.5042	4.2045	0.598
EGARCH-N	0.9302	1.9584	0.2453	0.7283	0.9835	0.598	3.2743	24.8546	1.0222	1.3713	3.5466	0.5975
EGARCH-t	0.9572	2.1035	0.258	0.7374	1.011	0.598	3.3348	26.0173	1.0302	1.3823	3.6092	0.5976
EGARCH-GED	0.9411	2.0255	0.2492	0.7316	0.9946	0.598	3.2884	25.2675	1.0225	1.3729	3.5609	0.5975
GJR-GARCH-N	0.9545	2.2116	0.2496	0.7335	1.008	0.598	4.8281	54.5298	1.2442	1.6629	5.1587	0.5983
GJR-GARCH-t	0.975	2.3467	0.2587	0.7401	1.0289	0.598	4.9383	57.3429	1.2568	1.6813	5.2726	0.5983
GJR-GARCH-GED	0.9649	2.2874	0.2535	0.7366	1.0186	0.598	4.877	55.8927	1.2494	1.6707	5.2091	0.5983
MRS-GARCH-N	0.9081	1.8532	0.2213	0.7176	0.9612	0.5979	4.6593	48.7044	1.2164	1.6319	4.9848	0.5982
MRS-GARCH-2t	0.8816	1.9865	0.1983	0.7028	0.9338	0.5979	3.9904	37.6904	1.1375	1.5134	4.2912	0.598
MRS-GARCH-t	0.9355	1.9129	0.2486	0.7316	0.9898	0.598	4.7992	50.2663	1.2429	1.663	5.1306	0.5983
MRS-GARCH-GED	0.8809	2.0507	0.1926	0.7006	0.9328	0.5979	4.0297	39.6146	1.1382	1.5169	4.331	0.598

Panel B. Out of sample for ten days ahead and twenty-two days ahead (long term)

	Ten days ahead						Twenty-two days ahead					
Models	MSE1	MSE2	QLIKE	MAD1	MAD2	HMSE	MSE1	MSE2	QLIKE	MAD1	MAD2	HMSE
GARCH-N	8.2190	145.3781	1.5890	2.1916	8.8564	0.5981	19.8861	808.0251	2.1318	3.4305	21.4175	0.5982
GARCH-t	8.4279	150.9889	1.6067	2.2225	9.0746	0.5981	20.5593	852.9743	2.1539	3.4927	22.1186	0.5983
GARCH-GED	8.2991	147.1451	1.5966	2.2042	8.9404	0.5981	20.1516	823.4314	2.1414	3.4563	21.6946	0.5983
EGARCH-N	4.6671	51.194	1.2514	1.6404	5.1406	0.5965	5.8010	82.6762	1.4123	1.8343	6.6157	0.5939
EGARCH-t	4.7374	53.0202	1.2585	1.6515	5.2141	0.5965	5.9276	86.2470	1.4241	1.8540	6.7511	0.5940
EGARCH-GED	4.6693	51.5573	1.2501	1.6395	5.1425	0.5965	5.8032	83.0537	1.4115	1.8337	6.6176	0.5939
GJR-GARCH-N	9.8631	216.3128	1.6864	2.3909	10.5545	0.5983	22.4507	1046.1885	2.1975	3.6368	24.0700	0.5984
GJR-GARCH-t	10.0906	225.8278	1.7007	2.4194	10.7903	0.5984	22.9203	1082.1067	2.2114	3.6777	24.5578	0.5984
GJR-GARCH-GED	9.9488	219.9585	1.6919	2.4017	10.6433	0.5983	22.5586	1051.8348	2.2014	3.6474	24.1826	0.5984
MRS-GARCH-N	9.6705	204.7504	1.6641	2.3597	10.3530	0.5982	22.6584	1072.9266	2.1928	3.6401	24.2780	0.5983
MRS-GARCH-2t	7.3914	120.5061	1.5284	2.0773	7.9936	0.5979	14.8147	453.4988	1.9669	2.9661	16.1352	0.5977
MRS-GARCH-t	9.9301	210.7154	1.6875	2.3997	10.6249	0.5983	23.0639	1093.9530	2.2080	3.6806	24.7026	0.5984
MRS-GARCH-GED	7.5142	127.357	1.5338	2.0901	8.1196	0.5979	15.1286	477.2495	1.9768	2.9942	16.4612	0.5977

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 ξ 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 a_1,β_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.