



# Financial Volatility Forecasting by Least Square Support Vector Machine Based on GARCH, EGARCH and GJR Models: Evidence from ASEAN Stock Markets

Phichhang Ou (corresponding author)

School of Business, University of Shanghai for Science and Technology

Rm 101, International Exchange Center, No. 516, Jun Gong Road, Shanghai 200093, China

Fax: 86-21-5527-1502 E-mail: phichhang@gmail.com

Hengshan Wang

School of Business, University of Shanghai for Science and Technology

Box 461, No. 516, Jun Gong Road, Shanghai 200093, China

Tel: 86-21-5527-5971 E-mail: wanghs@usst.edu.cn

*This work is supported by Shanghai Leading Academic Discipline Project, Project Number: S30504.*

## Abstract

In this paper, we aim at comparing semi-parametric method, LSSVM (Least square support vector machine), with the classical GARCH(1,1), EGARCH(1,1) and GJR(1,1) models to forecast financial volatilities of three major ASEAN stock markets. More precisely, the experimental results suggest that using hybrid models, GARCH-LSSVM, EGARCH-LSSVM and GJR-LSSVM provides improved performances in forecasting the leverage effect volatilities, especially during the recently global financial market crashes in 2008.

**Keywords:** Least squared support vector machine, Forecasting Volatility, GARCH, EGARCH, GJR.

## 1. Introduction

Time series method plays a vital role in financial areas, particularly volatility modeling and forecasting. Most of the financial researchers and practitioners are mainly concerned with modeling volatility in asset returns. In this context, volatility is the variability in the asset prices over a particular period of time. It refers to the standard deviation of the continuously compounded returns of a financial instrument with a specific time horizon. It is often used to quantify the risk of the instrument over that time period. Investors want a premium for investing in risky assets. A risk manager must know today the likelihood that his portfolio will decline in the future and he may want to sell it before it becomes too volatile. Therefore, the ability to forecast financial market volatility is important for portfolio selection and asset management as well as the pricing of primary and derivative assets. Researches on time varying volatility using the time series models have been active ever since Engle introduced the ARCH (autoregressive conditional heteroscedasticity) model in 1982. Since its introduction, the GARCH model generalized by Bollerslev (1986) has been extended in various directions. Several extensions of the GARCH model aimed at capturing the asymmetry in the response of the variance to a shock. These extensions recognize that there may be important nonlinearity, asymmetry, and long memory properties in the volatility process as suggested by various researchers based on empirical evidences. The popular approaches can be referred to Exponential GARCH model by Nelson (1991) as well as the GJR model by Glosten, Jagannathan, and Runkle (1993) which both account for the asymmetric relation between stock returns and changes in variance; see Black (1976) the beginning study of the asymmetric effect; Engle and Ng (1993) for further discussion. Other models such as APARCH, AGARCH, TGARCH and QGARCH models have also been developed (by Ding, Granger and Engle (1993); Engle (1990); Zakoian (1994) and Sentana (1995)) for the flexibility of the models. However, all of the models do require specified distribution of innovations in order to estimate the model specification and to appropriately forecast future values. One of the most classic one is Gaussian process and it is widely used in most of literature; but other distributions of innovations are also attracted after the empirical studies of modeling returns have shown the violation of normality conditions. For example, Student's t distribution by Bollerslev (1987), GED in Nelson (1991), Granger and Ding (1995) for the Laplace distribution and Hsieh (1989) for both Student's t and GED as distributional alternative models for innovations. The researches have found that returns usually exhibit empirical regularities including thick tails, volatility clustering, leverage effects (Bollerslev et al,1994).

Semi-parametric approaches do not require any assumptions on data property (i.e. return distribution). These models have been successfully shown for modeling and forecasting time series, including volatility. One of them is NN (neural network) and it is a powerful tool for prediction problems due to their best ability to estimate any function arbitrary with no priori assumption on data property (Haykin, 1999). Donaldson and Kamstra (1997) proposed neural network to model volatility based GJR-GARCH; their hybrid approach captured asymmetric effects of new impact well like parametric model and also generated better forecasting accuracy. Bildirici & Ersin(2009) fitted neural network based on nine different models of GARCH family such as NN-GARCH, NN-EGARCH, NN-TGARCH, NN-GJR, NN-SAGARCH, NN-PGARCH, NN-NGARCH, NN-APGARCH, and NN-NPGARCH to forecast Istanbul stock volatility and most of the hybrid models improved forecasting performance. This indicates that the hybrid model is also able to capture the stylized characteristics of return. Another efficient (semi-parametric) model is SVM (support vector machine) originally introduced by Vapnik (1995). The SVM, a novel neural network algorithm, guarantees to obtain globally optimal solution (Cristianini & Shawe-Taylor, 2000), and hence it solves the problems of multiple local optima in which the neural network usually get trapped into. Perez-Cruz et al (2003) predicted GARCH(1,1) based volatility by SVM and the proposed model yielded better predictive capability than the parametric GARCH(1,1) model for all situation. Chen et al (2008) developed recurrent SVM as a dynamic process to model GARCH(1,1) based volatility. The experimental results with simulated and real data also showed the model generated better performance than MLE (maximum likelihood estimation) based GARCH model. More applications of SVM in GARCH prediction based on different kernels, wavelet and spline wavelet can be referred to Tang et al (2008, 2009).

Another version of SVM is LSSVM (Least squares support vector machine), modified by Suykens et al (1999). The SVM algorithm requires Epsilon insensitive loss function to obtain convex quadratic programming in feature space, while LSSVM just uses least square loss function to obtain a set of linear equations (Suykens, 2000) in dual space so that learning rate is faster and the complexity of calculation in convex programming in SVM is also relaxed. In addition, the LSSVM avoids the drawback faced by SVM such as trade-off parameters ( $C, \sigma^2, \epsilon$ ) selection, instead it requires only two hyper-parameters ( $\gamma, \sigma^2$ ) while training the model. According to Suykens et al (2001), the equality constraints of LSSVM can act as recurrent neural network and nonlinear optimal control. Due to these nice properties, LSSVM has been successfully applied for classification and regression problems, including time series forecasting. See Van Gestel et al (2004) for detailed discussion on classification performance of LSSVM and Ye et al (2004) for predictive capability of LSSVM in chaotic time series prediction. Van Gestel et al (2001) proposed to predict time varying volatility of DAX 30 index by applying Bayesian evidence framework to LSSVM. The volatility model is constructed based on inferred hyperparameters of LSSVM formulation within the evidence framework. The proposed model provided a better predictive performance than GARCH(1,1) and other AR(10) models in term of MSE and MAE.

In this paper, we aim at comparing the LSSVM method with the classical GARCH(1,1), EGARCH(1,1) and GJR(1,1) models to forecast financial volatilities of ASEAN stock markets as a new concept to be investigated. The hybrid models denoted as GARCH-LSSVM, EGARCH-LSSVM, and GJR-LSSVM are constructed by using lagged terms as input and present term as output which corresponds to the parametric models. The hybrid models are not the same as the volatility model proposed by Van Gestel et al (2001) but they are similarly built according with the results by Donaldson & Kamstra (1997) and Bildirici & Ersin(2009) with neural network approach, and Perez-Cruz et al (2003) with SVM method. In our experiment, we consider two stage forecasts for the whole year 2007 as first period and 2008 as the second stage which cover global financial crisis period. Several metrics MAD, NMSE, HR, and linear regression R squared are employed to measure the model performances. The paper is organized as follows. Next section briefly reviews LSSVM formulation. Section 3 discusses volatility modeling of hybrid models based on GARCH, EGARCH and GJR. Section 4 illustrates the experimental results and the final section is about the conclusion.

## 2. Least squared support vector machines

In LSSVM formulation, the data are generated by nonlinear function  $y_i = f(x_i) + e_i$  for  $i = 1, \dots, N$  which may be approximated by another nonlinear function

$$y_i = w^T \phi(x_i) + b + e_i. \quad (1)$$

The model parameter  $w$  is called weight and  $e_i$  is random noise. Output  $y_i \in R$  can be referred as return of an asset or volatility, while the input vector  $x_i \in R^n$  may consist of lagged returns or lagged volatility. Mapping  $\phi(\cdot) : R^n \rightarrow R^n$  is nonlinear function that maps the input vector  $x$  into a higher dimensional feature space. The weight vector  $w$  and function  $\phi(\cdot)$  are never calculated explicitly; instead, Mercer conditioned kernel  $K(x_i, x) = \phi(x_i)^T \phi(x)$  which is symmetric and positive definite is applied. One defines the optimization problem as

objective function

$$\min_{w, b, e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (2)$$

subject to the constraints

$$e_i = y_i - (w^T \phi(x_i) + b) \quad i = 1, \dots, N. \quad (3)$$

Here the equality constraint is used in LSSVM instead of the inequality constraint in SVM. Lagrangian can be defined to solve the above minimization problem as

$$L(w, b, e; \alpha) = J(w, e) - \sum_{i=1}^N \alpha_i (w^T \phi(x_i) + b + e_i - y_i)$$

where  $\alpha_i$  denotes Lagrange multipliers (also called support values). From the Karush-Kuhn-Tucker (KKT) theory, a system of equations is obtained as the following

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N \alpha_i \phi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma_i e_i \quad i = 1, \dots, N \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow b = y_i - w^T \phi(x_i) - e_i, \quad i = 1, \dots, N. \end{array} \right. \quad (4)$$

Note that sparseness is lost from the condition  $\alpha_i = \gamma_i e_i$ . By eliminating  $w$  and  $e_i$ , the following linear system is written as follow

$$\left[ \begin{array}{c|c} 0 & 1_v^T \\ \hline 1_v & \Omega + D_\gamma^{-1} \end{array} \right] \left[ \begin{array}{c} b \\ \alpha \end{array} \right] = \left[ \begin{array}{c} 0 \\ y \end{array} \right] \quad (5)$$

where  $y = [y_1, \dots, y_N]$ ,  $1_v = [1, \dots, 1]$ ,  $e = [e_1, \dots, e_N]$ ,  $\alpha = [\alpha_1, \dots, \alpha_N]$ ,  $D_\gamma = \text{diag}([\gamma_1, \dots, \gamma_N])$ .

Matrix  $\Omega_{ij} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j)$  for  $i, j = 1, \dots, N$  satisfies Mercer's condition and the LS-SVM model for estimating function is obtained as

$$h(x) = w^T \phi(x) + b = \sum_{i=1}^N \alpha_i K(x, x_i) + b. \quad (6)$$

$K(.,.)$  is the Mercer's kernel function representing the high-dimensional feature space that nonlinearly mapped from the input space. In this work, Gaussian kernel or RBF(radial basis function) is used as it tends to give a good performance under general smoothing assumptions. The kernel is defined as  $K(x_1, x_2) = \exp(-\frac{1}{\sigma^2} \|x_1 - x_2\|^2)$ . The kernel and regularized parameters  $(\gamma, \sigma^2)$  are tuned by gridsearch technique to avoid overfitting problem. Matlab toolbox is used in the whole experiment.

### 3. Predictive model of volatility

#### 3.1 Model building

Let  $P_t$  be the stock price at time  $t$ . Then

$$y_t = 100 \cdot \log(P_t / P_{t-1}) \quad (7)$$

denotes the continuously compounded daily returns of the particular stock at time  $t$ .

Let  $F_{t-1}$  be the past information set available up to time  $t-1$ ; this information set contains the realized values of all previous relevant variables. The expected return at time  $t$  after observing the past information up to  $t-1$  defined as

$$\mu_t = E[y_t / F_{t-1}] = f(F_{t-1}). \quad (8)$$

The volatility to investors investing in the particular stock at time  $t-1$  is denoted as follow

$$\sigma_t^2 = \text{Var}[y_t / F_{t-1}] = h(F_{t-1}) \quad (9)$$

where  $f(.)$  and  $h(.)$  are well defined functions with  $h(.) > 0$ .

Then the return of stock  $y_t$  can be modelled

$$y_t = \mu_t + \varepsilon_t \quad (10)$$

and  $\varepsilon_t = \sigma_t z_t$  where  $z_t$  is (iid) independent identically distributed random variables with mean 0 and variance 1. It is common to assume  $\mu_t = 0$  so that square return in (10) is obtained to be the shock squared,

$$y_t^2 = \varepsilon_t^2 = \sigma_t^2 z_t^2. \quad (11)$$

Here, we aim at estimating volatility (or conditional variance of return) in (9) by kernel regression (called semi-parametric method) based on parametric models of GARCH, EGARCH and GJR. One particular approach of the kernel regression is LSSVM(least square support vector machine) presented in the previous section.

LSSVM approximates GARCH(1,1)

$$\sigma_t^2 = \omega + \beta \cdot \sigma_{t-1}^2 + \alpha \cdot \varepsilon_{t-1}^2 \quad (12)$$

by nonlinear function obtained from the LSSVM algorithm,

$$\sigma_t^2 = h(\sigma_{t-1}^2, \varepsilon_{t-1}^2). \quad (13)$$

Similarly, the hybrid model estimates EGARCH(1,1)

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \alpha \cdot [\lvert \varepsilon_{t-1} \rvert / \sigma_{t-1} - E(\lvert \varepsilon_{t-1} \rvert / \sigma_{t-1})] + \gamma \varepsilon_{t-1} / \sigma_{t-1} \quad (14)$$

$$\text{by } \log \sigma_t^2 = h(\log \sigma_{t-1}^2, [\lvert \varepsilon_{t-1} \rvert / \sigma_{t-1} - E(\lvert \varepsilon_{t-1} \rvert / \sigma_{t-1})], \varepsilon_{t-1} / \sigma_{t-1}) \quad (15)$$

where inputs  $\log \sigma_{t-1}^2, [\lvert \varepsilon_{t-1} \rvert / \sigma_{t-1} - E(\lvert \varepsilon_{t-1} \rvert / \sigma_{t-1})], \varepsilon_{t-1} / \sigma_{t-1}$  must be obtained before training LSSVM. The expectation  $E(\lvert \varepsilon_{t-1} \rvert / \sigma_{t-1})$  is estimated by its corresponding average (or mean) value so that specified distribution on innovation (or return in this case) is not required.

Finally, the GJR-LSSVM obtains GJR(1,1)

$$\sigma_t^2 = \omega + \beta \cdot \sigma_{t-1}^2 + \alpha \cdot \varepsilon_{t-1}^2 + \gamma S_{t-1}^- \varepsilon_{t-1}^2 \quad (16)$$

$$\text{as } \sigma_t^2 = h(\sigma_{t-1}^2, \varepsilon_{t-1}^2, S_{t-1}^- \varepsilon_{t-1}^2) \quad (17)$$

where  $S_{t-1}^- = 1$  if  $\varepsilon_{t-1} < 0$  and 0 otherwise. So  $S_{t-1}^- \varepsilon_{t-1}^2$  is the squared value of negative shock at  $t-1$ .

Here  $\hat{\sigma}_t^2 = \frac{1}{5} \sum_{k=0}^4 y_{t-k}^2$  suggested by Perez-Cruz (2003), and the function  $h(\cdot)$  is obtained by LSSVM algorithm in (6).

### 3.2 Forecasting

Out of sample forecasts by the hybrid models are obtained as follow:

$$\sigma_{t+1}^2 = h(\sigma_t^2, \varepsilon_t^2) \quad \text{for GARCH-LSSVM from (13),}$$

$$\sigma_{t+1}^2 = \exp(h(\log \sigma_t^2, [\lvert \varepsilon_t \rvert / \sigma_t - E(\lvert \varepsilon_t \rvert / \sigma_t)], \varepsilon_t / \sigma_t)) \quad \text{for EGARCH-LSSVM from (15) and}$$

$$\sigma_{t+1}^2 = h(\sigma_t^2, \varepsilon_t^2, S_t^- \varepsilon_t^2) \quad \text{for GJR-LSSVM from (17).}$$

## 4. Experimental Analysis

We examine three stock price indexes from three major ASEAN stock markets including Straits time index of Singapore stock market, PSEI of the Philippines and KLCI of Kuala Lumpur stock market. Each stock index price is collected from Yahoo Finance website and is transformed into log return as in (7) before making analysis. Table 1 reports the in-sample and out-of-sample periods of each market for two stages, basic statistics of the data and diagnostics. From the Table 1, we see that mean of all returns is close to zero. Two indexes KLCI and PSEI have positive skewed returns while STI produce negative skewed coefficient. The excess kurtosis appears in all series and the largest is from KLCI (54.152). The Jarque Bera statistics strongly suggest that all returns are non normal. Ljung Box test for squared return at lag 20 and Engle LM test significantly indicate all return series exhibit ARCH effects; that means the homoscedasticity hypothesis is strongly rejected. This shows the presence of volatility clustering and the leverage effects that could be caused by the excess kurtosis. Figure 1 plots price and log return of each index series for the entire sample. Though movement of the index prices of the three markets is almost in similar direction, the returns behave differently. From the plots, we can see some high volatility on log return series after financial crisis in ASEAN in 1997 and during the recent crisis of global market crashes; this is obviously seen that the plots of each stock price fall down sharply during 2008.

#### 4.1 Estimation results

Three parametric models GARCH, EGARCH and GJR are fitted to all return series by (12), (14) and (16) respectively. Each model is estimated twice for each market return as first stage and second stage estimations with updating in-sample.

Table 2.A, 2.B, and 2.C present the model parameters and their corresponding standard errors in brackets. The stationary conditions of the models hold for all series. Furthermore, significance of negative leverage coefficients in EGARCH and positive leverage coefficients of the corresponding GJR indicate the presence of asymmetric effects to the returns for both stages which may be caused by global financial crisis. By log likelihood, AIC and BIC criteria in Table 2.A and 2.C, GJR model is more adequate to the both stage estimations of STI and PSEI returns. For the KLCI return in Table 2.B, the GJR model fits well to the in-sample data at the first stage but the second stage estimation data is favour to EGARCH model according to the log likelihood, AIC and BIC.

Now we proceed to estimation results obtained from training the least square support vector machine. First, return series from all indexes are transformed into input and output format and then get them trained by LSSVM algorithm in (6) so as to get the estimated nonlinear function in (13) for GARCH hybrid, (15) for EGARCH case and (17) for GJR model. The training results are summarized in Table 3.A and 3.B for first and second stages respectively. Each second column of the table 3.A and 3.B shows the costs of training measured by the mean square errors. The third and fourth columns display the optimal regularized parameters and optimal kernel parameters obtained by gridsearch technique while training. The last column tells the bias term of resulted function obtained by the LSSVM.

From the first stage of the cost column in STI index, smallest cost falls to GARCH-LSSVM and the largest value goes to EGARCH-LSSVM. For KLCI series, GJR-LSSVM generates the least cost and the largest cost is from GARCH-LSSVM, but these errors are not far from one another. Finally, PSEI series produces the smallest error to EGARCH-LSSVM and the error is a bit far from the errors driven by GARCH-LSSVM and GJR-LSSVM.

In the second stage, training mean square error for STI and PSEI are analogue to the mean square errors for STI and PSEI in the first stage respectively; that is the STI is in favour with GARCH-LSSVM and PSEI produces the smallest cost while getting trained by EGARCH-LSSVM. For Kula Lumpur stock market, GARCH-LSSVM gives the smallest cost, but EGARCH-LSSVM still produces the highest value of cost like before. In the next section, these hybrid models will be performed to forecast volatility of the three markets and also be compared with the parametric approaches estimated in the previous section.

#### 4.2 Forecasting results

The following Evaluation metrics are used to measure the performance of proposed models in forecasting of the three different stock markets volatilities. They are Mean Absolute Deviation (MAD), Normalized Mean Square Error(NMSE) and Hit Rate (HR) which defined as the following:

$$\text{MAD} = \frac{1}{n'} \sum_{t=1}^{n'} |a_t - p_t|, \quad \text{NMSE} = \frac{1}{s^2 n'} \sum_{t=1}^{n'} (a_t - p_t)^2 \quad \text{where } s^2 = \frac{1}{n'-1} \sum_{t=1}^{n'} (a_t - \bar{a}_t)^2.$$

$$\text{HR} = \frac{1}{n''} \sum_{t=1}^{n''} d_t \quad \text{where } d_t = \begin{cases} 1 & (a_t - a_{t-1})(p_t - p_{t-1}) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Here  $a_t = y_t^2$  actual values and  $p_t = \hat{\sigma}_t^2$  forecasted volatility. Here  $n'' = n' - 1$ .

We also use linear regression to evaluate the forecasting performance of the volatility model. We simply regress square returns on a constant and the forecasted volatility for out-of-sample time point,  $t = 1, 2, \dots, n'$ ,  $y_t^2 = c_0 + c_1 \hat{\sigma}_t^2 + e_t$ . The t-statistic of the coefficients is a measure of the bias and the square correlation  $R^2$  is a measure of forecasting performance. In this regression, the constant term  $c_0$  should be close to zero and the slope  $c_1$  should close to 1. Table 4.A and 4.B illustrate forecasting performances by different models for each market. The MAD, NMSE, HR and R squared with  $c_0$  and  $c_1$  are shown in the second to seventh columns.

*First stage for 2007:*

Beginning with STI series, the hybrid approaches perform better than parametric models for almost all metrics: MAD, NMSE, HR, and R squared. Only R squared criterion is in favour to EGARCH model that generates the highest value. Among all the models, EGARCH-LSSVM is best at a predictive performance because it provides highest HR (0.8273), smallest values of MAD and NMSE and it also satisfies to ( $c_0$  and  $c_1$ ) values which are not far from (0 and 1) respectively. Now by considering Kula Lumpur market, based on MAD and NMSE, the hybrid models are much better but in term of HR and R squared some semi-parametric models especially EGARCH-LSSVM is unable to defeat its counterpart, EGARCH. The KLCI return is well modelled by EGARCH like STI case since it generates least NMSE,

highest R squared and HR among the others. Looking at c0 and c1 criteria, the hybrid models are more satisfied than the parametric approaches. For PSEI, the semi-parametric models are superior to the parametric models for all cases.

#### *Second stage for 2008:*

The STI return series is well forecasted by EGARCH model like its previous performance in the first stage forecast due to the highest values of HR and R squared, which can be seen from the Table 4.B. For the other formed GARCH and GJR, LSSVM is better than the parametric models. From the Table 4.B, the values of c0 and c1 of EGARCH model (with values -3.72, 2.25), though generated best performance, deviate far from the appropriate norm (0, 1) respectively due to the global financial market crashes. However, EGARCH-LSSVM and other LSSVMs are more resistant in forecasting performance to the crashes since their c0 and c1 are not much far from 0 and 1 respectively. For KLCI and PSEI, hybrid approaches beat all parametric models for all criteria and EGARCH-LSSVM is superior among the others. These evidences can argue that LSSVM is more robust than the parametric models in forecasting volatility in spite of the high volatile situation during the global financial market crashes. Figures 2, 3, & 4 plot the out of sample forecasts by parametric models of GARCH, EGARCH and GJR and the corresponding hybrid models for STI, KLCI and PSEI respectively. From the plots, the forecast lines by hybrid models capture more extreme points than the parametric models do and therefore they improve forecasting performance. Noticeably, the LSSVM algorithm here has not been imposed the sparsity and robustness conditions proposed by Suyken et al (2002).

#### **5. Conclusion**

In this paper, we combine Least square support vector machine (LSSVM) with GARCH(1,1), EGARCH(1,1) and GJR(1,1) models as a hybrid approach to forecast leverage effect volatility of ASEAN stock markets. To check the performance of the proposed models, we employ the corresponding parametric models to compare with the hybrid models. The forecasts are conducted twice in which the whole year 2007 is treated as the first stage and the second stage is for 2008 including the recent global financial crisis period. From the experimental results, it is found that the hybrid models are resistant and robust to the high volatile situation of the financial market crashes and hence they generate improved forecasting performance. This supports the general idea that LSSVM is the promising machine learning system which is good at estimating nonlinear function without assumptions on data property in time series applications.

#### **References**

- Bildirici, M. & Ersin O. O. (2009). Improving forecasts of GARCH family models with the artificial neural networks: An application to the daily returns in Istanbul Stock Exchange, *Expert Systems with Applications*, 7355-7362.
- Black, F. (1976). Studies of stock price volatility changes, *Proceedings of the 1976 Meetings of the Business and Economics Statistics Section*, American Statistical Association, 177-181.
- Bollerslev, T., Engle R.F. & Nelson, D.B. (1994). ARCH models, *Handbook of Econometrics*, Volume IV, Elsevier Science B.V.
- Bollerslev, T. (1987) A conditional heteroskedastic time series model for speculative prices and rates of return, *Rev Econ Stat* 69:542-547
- Bollerslev T. (1986). Generalized Auto Regressive Conditional Heteroskedasticity, *Journal of Econometrics*, pp. 307-327.
- Chen, S., Jeong, K. & Hardle, W. (2008). Support Vector Regression Based GARCH Model with Application to Forecasting Volatility of Financial Returns, *SFB 649 Discussion Paper of Economic Risk*, Berlin.
- Cristianini, N., & Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, London: Cambridge University Press.
- Ding, Z., C. W.J. Granger & R. F. Engle. (1993). A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, 1, 83-106 North-Holland.
- Donaldson, R. G. & Kamstra, M. (1997). An artificial neural network-GARCH model for international stock return volatility, *Journal of Empirical Finance*, pp. 17-46.
- Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of variance of UK inflation, *Econometrica* 50, pp. 987-1008.
- Engle, R. F. (1990), Discussion: Stock market volatility and the crash of 87, *Review of Financial Studies*, 3, 103-106.
- Engle, R. F. & Ng, V. K. (1993). Measuring and Testing the Impact of News on Volatility, *Journal of Finance*, pp. 1749-1778.
- Glosten, L., Jagannathan, R. & Runkle, D. (1993). On the relationship between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 46, pp. 1779-801.

- Granger, C. & Ding, Z. (1995) Some properties of absolute return an alternative measure of risk, *Ann Econ Stat* 40:67-91.
- Hsieh, D.A. (1989) The statistical properties of daily foreign exchange rates: 1974-1983, *J. International Economics* 24:129-145.
- Haykin, S. (1999). Neural networks: a comprehensive foundation, *Englewood cliffs*, Prentice Hall.
- Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* 59, pp. 347-370.
- Perez-Cruz, F., Afonso-Rodriguez, J.A. & Giner J. (2003). Estimating GARCH models using support vector machines, *Journal of Quantitative Finance*, pp. 163-172.
- Sentana, E. (1995), Quadratic ARCH models, *Review of Economic Studies*, 62(4), 639-661.
- Suykens, J.A.K, Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters* (9) 293-300.
- Suykens, J.A.K. (2000). Least squares support vector machines for classification and nonlinear modeling, *Neural Network World*, Vol. 10, pp. 29-48.
- Suykens, J.A.K, Vandewalle, J., De Moor, B. (2001). Optimal control by least squares support vector machines. *Neural Networks* (14) 23-35.
- Suykens, J.A.K, & De Brabanter, J., Lukas, L., & Vandewalle, J. (2002). Weighted least squares support vector machines: Robustness sparse approximation. *Neurocomputing*, 48: 1-4, 85-105.
- Tang, L.B., Sheng, H.Y. & Tang, L.X., (2008). Forecasting volatility based on wavelet support vector machine, *Expert Systems with Applications*.
- Tang, L.B., Sheng, H.Y. & Tang, L.X. (2009). GARCH prediction using spline wavelet support vector machine. *Journal of Neural Computing and Application*, Springer-Verlag London.
- Van Gestel, T.V., Suykens, J.A.K, Baestaens, D.E., Lambrechts, A., Lanckriet G., Vandaele B., De Moor B. & Vandewalle, J.(2001). Financial Time Series Prediction using Least Squares Support Vector Machines within the Evidence Framework, *IEEE Transactions on Neural Networks*, (12) 809-821.
- Van Gestel, T., Suykens, J.A.K, Baesens, B., Viaene, S., Vanthienen, J., Dedene, G., Moor B.D., & Vandewalle, G. (2004). Benchmarking Least Squares Support Vector Machine Classifiers, *Machine Learning*, (54) 5-32.
- Vapnik, V.N. (1995). The nature of statistical learning theory, *Springer-Verlag*, New York.
- Ye, M.Y. & Wang, X.D. (2004). Chaotic time series prediction using least squares support vector machine, *J. Chinese Physics*, IOP Publishing Ltd. IP address: 130.120.82.1000.
- Zakoian, J.M. (1994). Threshold Heteroscedastic Models, *Journal of Economic Dynamics and control*, 18, 931-955.

### Acknowledgement

We would like to thank Dr. Nathalie Villa-Vialaneix from Institut de Mathématiques de Toulouse, Toulouse, France for technical comments on this research.

Table 1. Descriptive statistics of each return series

	<b>STI</b>	<b>KLCI</b>	<b>PSEI</b>
In-sample (1 <sup>st</sup> stage)	1/2/1998 - 12/29/2006	1/2/1998 - 12/29/2006	1/2/1998 - 12/29/2006
Out of sample(1 <sup>st</sup> stage)	1/3/2007 - 12/31/2007	1/3/2007 - 12/31/2007	1/2/2007 - 12/28/2007
In-sample (2 <sup>nd</sup> stage)	1/4/1999 - 12/31/2007	1/4/1999 - 12/31/2007	1/4/1999 - 12/28/2007
Out of sample(2 <sup>nd</sup> stage)	1/2/2008 - 12/31/2008	1/2/2008 - 12/31/2008	1/2/2008 - 12/24/2008
Total sample size	2756	2709	2715
Minimum	-9.215	-24.15	-13.08
Maximum	12.87	20.81	16.177
Mean	0.005	0.015	-0.0003
Median	0.022	0.022	-0.013
Variance	2.219	2.395	2.573
Stdev	1.489	1.547	1.604
Skewness	-0.060	0.555	0.456
Kurtosis	7.1018	54.152	11.503
JB <sup>a</sup>	5805	331658	15091
Q <sup>2</sup> (20) <sup>b</sup>	1443.1	1426.1	227
ARCH-LM <sup>c</sup>	372.30	777.28	95.02

**Note:**<sup>a</sup>JB is the Jarque Bera test for normality<sup>b</sup>Q<sup>2</sup>(20) is the Ljung-Box test for squared returns<sup>c</sup>ARCH-LM is the Engle's Lagrange Multiplier test for conditional heteroskedasticity with 12 lags

Table 2.A. MLE estimation of the Parametric models for Straits times index

STI	First stage			Second stage		
	GARCH	EGARCH	GJR	GARCH	EGARCH	GJR
$\mu$	0.066[0.01]*	0.079[0.02]*	0.044[0.02]**	0.071[0.01]*	0.105[0.02]**	0.052[0.02]*
$\omega$	0.016[0.00]*	0.064[0.00]*	0.016[0.00]*	0.016[0.00]*	0.048[0.00]*	0.017[0.00]*
$\alpha$	0.123[0.00]*	0.416[0.02]**	0.076[0.01]*	0.109[0.00]*	0.349[0.03]*	0.073[0.01]*
$\beta$	0.876[0.00]*	0.915[0.00]*	0.880[0.00]*	0.886[0.00]*	0.918[0.01]*	0.886[0.00]*
$\gamma$		-0.067[0.01]*	0.082[0.01]*		-0.051[0.01]*	0.070[0.01]*
LL	-3550	-3589	-3.538	-3406	-3427	-3397
AIC	7109	7189	7.087	6820	6864	6805
BIC	7132	7218	7.116	6843	6893	6833

**Note:** Values in bracket [ ] indicates standard error of estimates; LL denotes Log likelihood values.

\* significant at the 1% level , \*\* significant at 5% level.

Table 2.B. MLE estimation of the Parametric models

KLCI	First stage			Second stage		
	GARCH	EGARCH	GJR	GARCH	EGARCH	GJR
$\mu$	0.043[0.01]*	0.011[0.02]	0.032[0.01]	0.052[0.01]*	0.041[0.01]*	0.041[0.01]*
$\omega$	0.010[0.00]*	0.074[0.00]*	0.010[0.00]*	0.008[0.00]*	0.006[0.00]*	0.009[0.00]*
$\alpha$	0.124[0.00]*	0.535[0.00]**	0.090[0.00]*	0.105[0.00]*	0.249[0.01]*	0.080[0.00]*
$\beta$	0.875[0.00]*	0.921[0.00]*	0.876[0.00]*	0.893[0.00]*	0.977[0.00]*	0.888[0.00]*
$\gamma$		-0.013[0.01]*	0.065[0.01]*		-0.045[0.00]*	0.059[0.01]*
LL	-3148	-3191	-3140	-2862	-2842	-2855
AIC	6305	6392	6291	5733	5694	5720
BIC	6327	6421	6319	5756	5722	5749

**Note:** Values in bracket [ ] indicates standard error of estimates; LL denotes Log likelihood values.

\* significant at the 1% level , \*\* significant at 5% level.

Table 2.C. MLE estimation of the Parametric models

PSEI	First stage			Second stage		
Statistics	GARCH	EGARCH	GJR	GARCH	EGARCH	GJR
$\mu$	0.017[0.02]	0.018[0.02]	0.003[0.02]	0.030[0.02]	0.029[0.02]	0.018[0.02]
$\omega$	0.125[0.01]*	0.038[0.00]*	0.094[0.01]*	0.214[0.03]*	0.050[0.00]*	0.212[0.03]*
$\alpha$	0.130[0.00]*	0.217[0.00]*	0.074[0.00]*	0.125[0.02]*	0.204[0.00]**	0.067[0.00]*
$\beta$	0.821[0.01]*	0.964[0.00]*	0.853[0.00]*	0.769[0.01]*	0.936[0.01]*	0.782[0.02]*
$\gamma$		-0.036[0.00]*	0.073[0.01]*		-0.040[0.00]*	0.086[0.01]*
LL <sup>e</sup>	-3873	-3870	-3866	-3764	-3760	-3757
AIC	7755	7751	7743	7537	7530	7525
BIC	7778	7780	7771	7560	7559	7.553
Q <sup>2</sup> (20) <sup>f</sup>						

**Note:** Values in bracket [ ] indicates standard error of estimates; **LL** denotes Log likelihood values.  
\* significant at the 1% level, \*\* significant at 5% level.

Table 3.A. Training results by LSSVM for 1<sup>st</sup> stage

STI	Cost (MSE)*	Optimal Gamma**	Optimal Sigma2**	b***
GARCH-LSSVM	0.2744	1089.0387	4.7055891	7.1706
EGARCH-LSSVM	0.5879	82.465645	1302.1078	4.8840
GJR-LSSVM	0.2904	4515.4671	79.336258	23.9096
KLCI				
GARCH-LSSVM	0.7295	2220.5867	40.007163	7.0168
EGARCH-LSSVM	0.7528	164.14558	1671.9517	4.0787
GJR-LSSVM	0.7002	3735.476	556.81488	12.7939
PSEI				
GARCH-LSSVM	0.5201	199.5209	599.4858	6.6127
EGARCH-LSSVM	0.2476	39.5453	11.5295	-0.0515
GJR-LSSVM	0.5571	170.6817	513.466	6.4879

\*Cost of estimation by MSE measure.

\*\* Optimal parameters (Gamma and Sigma2) selected by gridsearch technique.

\*\*\* b is the intercept value of the function estimated by LSSVM.

Table 3.B. Training results by LSSVM for 2<sup>st</sup> stage

STI	Cost (MSE)*	Optimal Gamma**	Optimal Sigma2**	b***
GARCH-LSSVM	0.3687	913.3114	1129.7975	-5.9924
EGARCH-LSSVM	0.4471	67.14207	1595.4792	4.6035
GJR-LSSVM	0.4119	1058.6892	1505.9648	-8.9199
KLCI				
GARCH-LSSVM	0.1524	884.44825	1783.175	-4.9433
EGARCH-LSSVM	0.5327	81.229819	1362.0289	3.8441
GJR-LSSVM	0.1898	1097.5405	1726.0707	1.1409
PSEI				
GARCH-LSSVM	1.6928	53.014923	1372.0333	3.6325
EGARCH-LSSVM	0.2825	7966.3458	203.42274	-16.0069
GJR-LSSVM	1.8074	57.086861	1591.2148	3.0683

\*Cost of estimation by MSE measure.

\*\* Optimal parameters (Gamma and Sigma2) selected by gridsearch technique.

\*\*\* b is the intercept value of the function estimated by LSSVM.

Table 4.A. Forecast performances of ASEAN stock volatilities by different models for 2007

STI	MAD	NMSE	$c_0$	$c_1$	$R^2$	HR
GARCH	1.78136	0.74299	-0.81[-2.41]	1.35[09.60]	0.2710	0.76305
GARCH-LSSVM	1.66240	0.72658	-0.04[-0.15]	1.01[09.51]	0.2676	0.79919
EGARCH	1.61913	0.59971	-1.85[-6.52]	1.80[15.66]	0.4970	0.78714
EGARCH-LSSVM	1.51584	0.60361	-0.56[-2.32]	1.44[13.80]	0.4347	0.82730
GJR	1.77206	0.73001	-0.82[-2.50]	1.35[09.94]	0.2850	0.73493
GJR-LSSVM	1.63353	0.62262	-0.42[-1.66]	1.21[12.44]	0.3843	0.78714

KLCI	MAD	NMSE	$c_0$	$c_1$	$R^2$	HR
GARCH	1.05754	0.69884	-0.39[-2.03]	1.26[10.54]	0.3100	0.70161
GARCH-LSSVM	0.91449	0.50993	-0.15[-1.11]	1.06[15.36]	0.4887	0.79435
EGARCH	1.00633	0.46063	-0.91[-6.34]	1.36[19.31]	0.6020	0.83467
EGARCH-LSSVM	0.83636	0.49243	-0.27[-2.03]	1.34[17.05]	0.5407	0.81451
GJR	1.06043	0.66277	-0.33[-1.87]	1.19[11.34]	0.3430	0.73387
GJR-LSSVM	0.96051	0.56356	-0.05[-0.37]	1.00[13.71]	0.4323	0.68548

PSEI	MAD	NMSE	$c_0$	$c_1$	$R^2$	HR
GARCH	2.67548	1.01961	-2.61[-4.16]	1.97[11.42]	0.3490	0.74590
GARCH-LSSVM	2.47743	0.58031	-0.74[-1.61]	1.27[13.69]	0.4355	0.79508
EGARCH	2.63377	1.03119	-3.38[-4.39]	2.33[09.70]	0.2790	0.74590
EGARCH-LSSVM	2.36726	0.60253	-0.70[-1.52]	1.40[13.54]	0.4303	0.77868
GJR	2.63297	1.03151	-2.47[-3.82]	1.90[10.73]	0.3220	0.71721
GJR-LSSVM	2.49653	0.60887	-0.58[-1.22]	1.22[12.72]	0.3999	0.78688

Note: higher R squared and HR is preferred, while smaller values of MAD and NMSE indicate the forecasted volatility is closer to the actual values. The coefficients of  $c_0$  and  $c_1$  should be close to (0, 1) respectively showing small forecasting errors.

Table 4.B. Forecast performances of ASEAN stock volatilities by different models for 2008

STI	MAD	NMSE	$c_0$	$c_1$	$R^2$	HR
GARCH	4.40025	0.69094	-1.08[-1.34]	1.22[10.51]	0.3142	0.73140
GARCH-LSSVM	3.63287	0.51329	-1.42[-2.34]	1.45[16.72]	0.5371	0.76446
EGARCH	3.72500	0.60189	-3.72[-5.72]	2.25[18.38]	0.5839	0.78099
EGARCH-LSSVM	3.63715	0.58018	-1.32[-2.08]	1.62[15.47]	0.4985	0.77685
GJR	4.46939	0.66002	-1.05[-1.35]	1.14[11.16]	0.3407	0.72727
GJR-LSSVM	3.65591	0.52521	-1.37[-2.21]	1.43[16.20]	0.5213	0.77272

KLCI	MAD	NMSE	$c_0$	$c_1$	$R^2$	HR
GARCH	2.05590	0.81824	-1.38[-2.45]	1.65[8.01]	0.2079	0.71951
GARCH-LSSVM	1.72023	0.77390	-0.35[-0.77]	1.28[8.59]	0.2316	0.78861
EGARCH	1.90105	0.78319	2.05[9.89]	-1.98[-3.69]	0.2857	0.78455
EGARCH-LSSVM	1.62305	0.66531	-0.75[-1.89]	1.64[12.56]	0.3918	0.80081
GJR	2.14153	0.78378	1.42[8.60]	-1.21[-2.30]	0.2323	0.73170
GJR-LSSVM	1.69754	0.74259	-0.57[-1.27]	1.39[9.60]	0.2735	0.78455

PSEI	MAD	NMSE	$c_0$	$c_1$	$R^2$	HR
GARCH	3.89602	0.72265	-4.17[-4.66]	2.55[13.90]	0.4421	0.72244
GARCH-LSSVM	3.79703	0.58720	-1.26[-1.66]	1.37[13.89]	0.4417	0.78775
EGARCH	4.07293	0.79128	-6.55[-6.10]	3.31[13.04]	0.4107	0.76326
EGARCH-LSSVM	3.43509	0.53528	-2.32[-3.51]	1.84[18.64]	0.5875	0.75102
GJR	3.89767	0.68178	-3.95[-4.69]	2.35[14.84]	0.4744	0.72653
GJR-LSSVM	3.86604	0.73799	-0.19[-0.21]	1.14[9.29]	0.2613	0.78775

Note: higher R squared and HR is preferred, while smaller values of MAD and NMSE indicate the forecasted volatility is closer to the actual values. The coefficients of  $c_0$  and  $c_1$  should be close to (0, 1) respectively showing small forecasting errors.

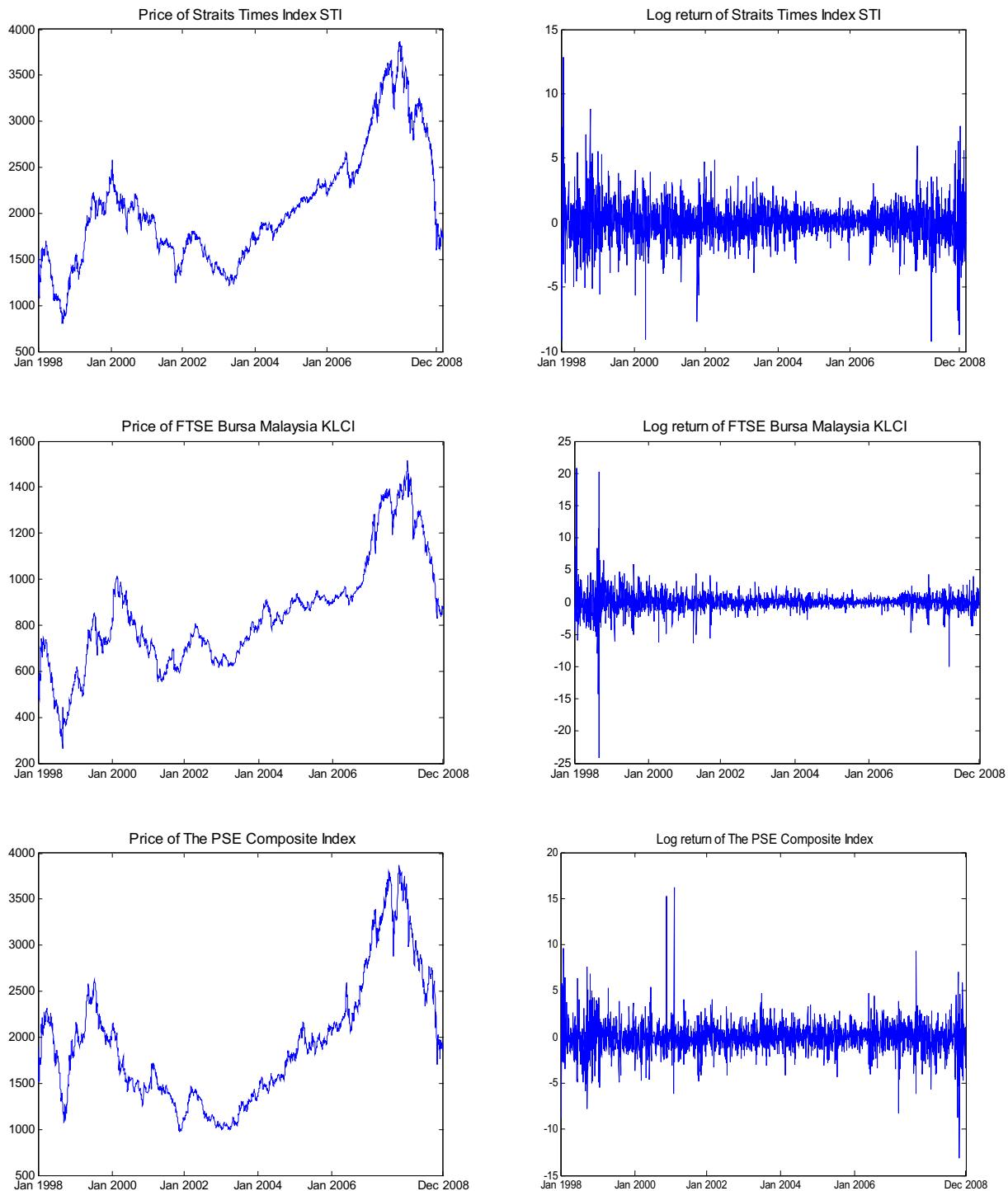


Figure 1. Plots of Prices and log returns of each market index

Plots of each index price (left) and log return (right) for the whole sample. From the left sides, we can see that all index prices movement are almost similar direction but the returns behave differently. The price series of each market falls down sharply at the last period due to global financial crisis. The log return plots exhibit high breaks at 1998 (after ASEAN financial crisis 1997) and in 2008 (the recent financial market crashes).

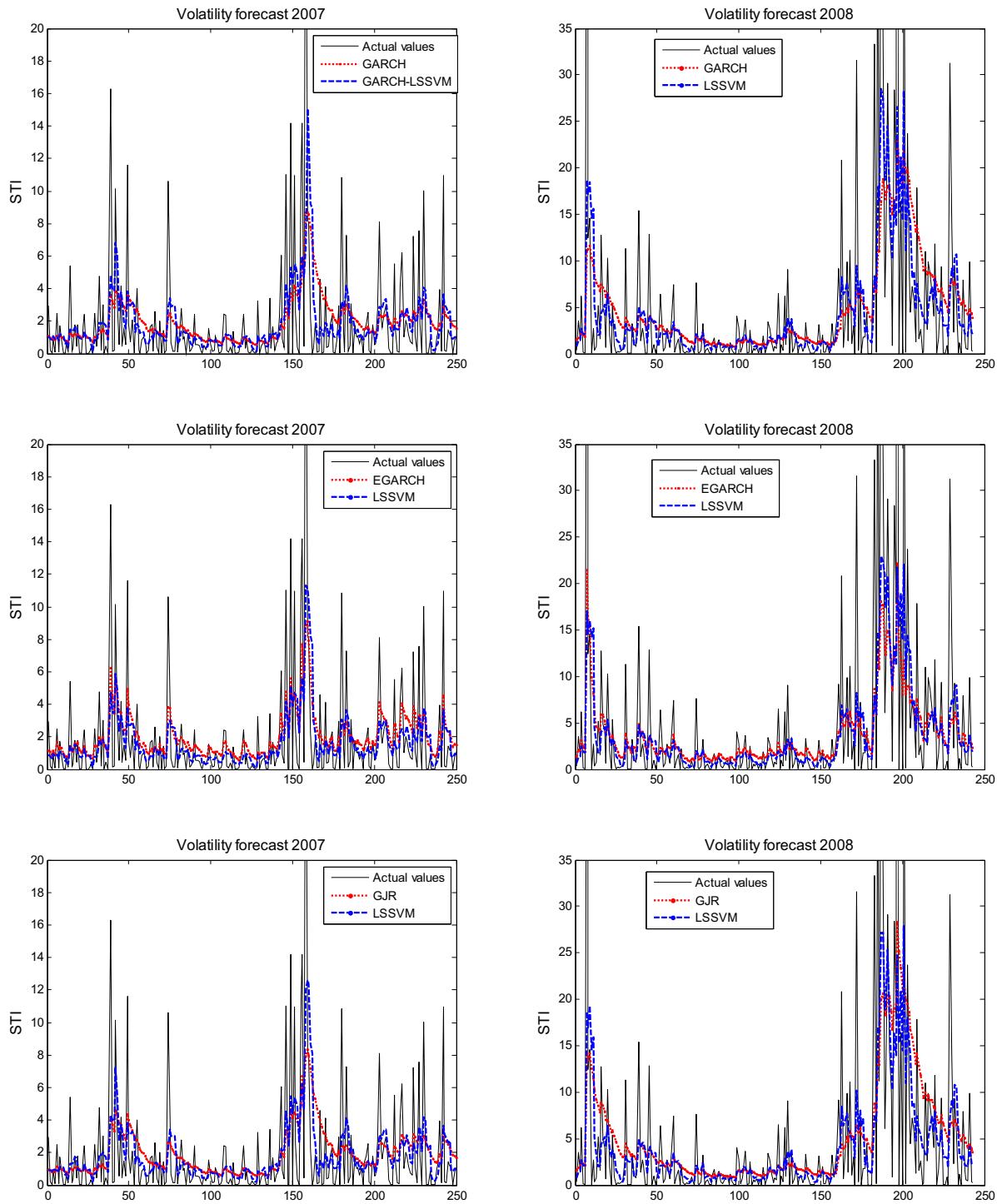


Figure 2. Volatility Forecasts of Singapore Stock Market (STI).

Note: Plots in left part are referred to the First stage forecast in 2007 (before crisis) and plots in the right side are referred to the second stage forecast for whole 2008 (during financial crisis). Small dot line is forecasted by parametric models (GARCH, EGARCH and GJR) while dash line is obtained by hybrid approaches.

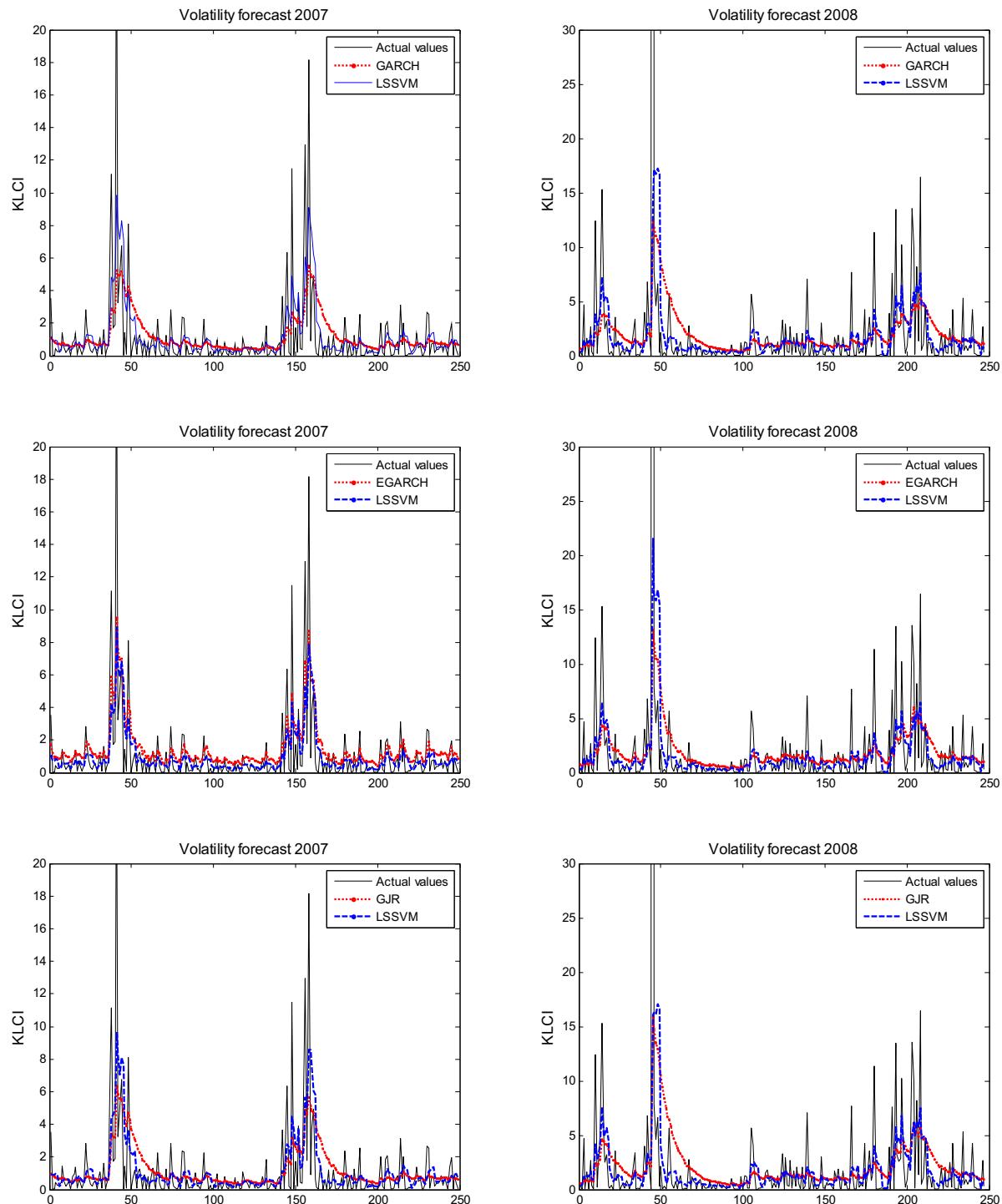


Figure 3. Volatility Forecasts of Kuala Lumpur Stock Market (KLCI).

Note: Plots in left part are referred to the First stage forecast in 2007 (before crisis) and plots in the right side are referred to the second stage forecast for whole 2008 (during financial crisis). Small dot line is forecasted by parametric models (GARCH, EGARCH and GJR) while dash line is obtained by hybrid approaches.

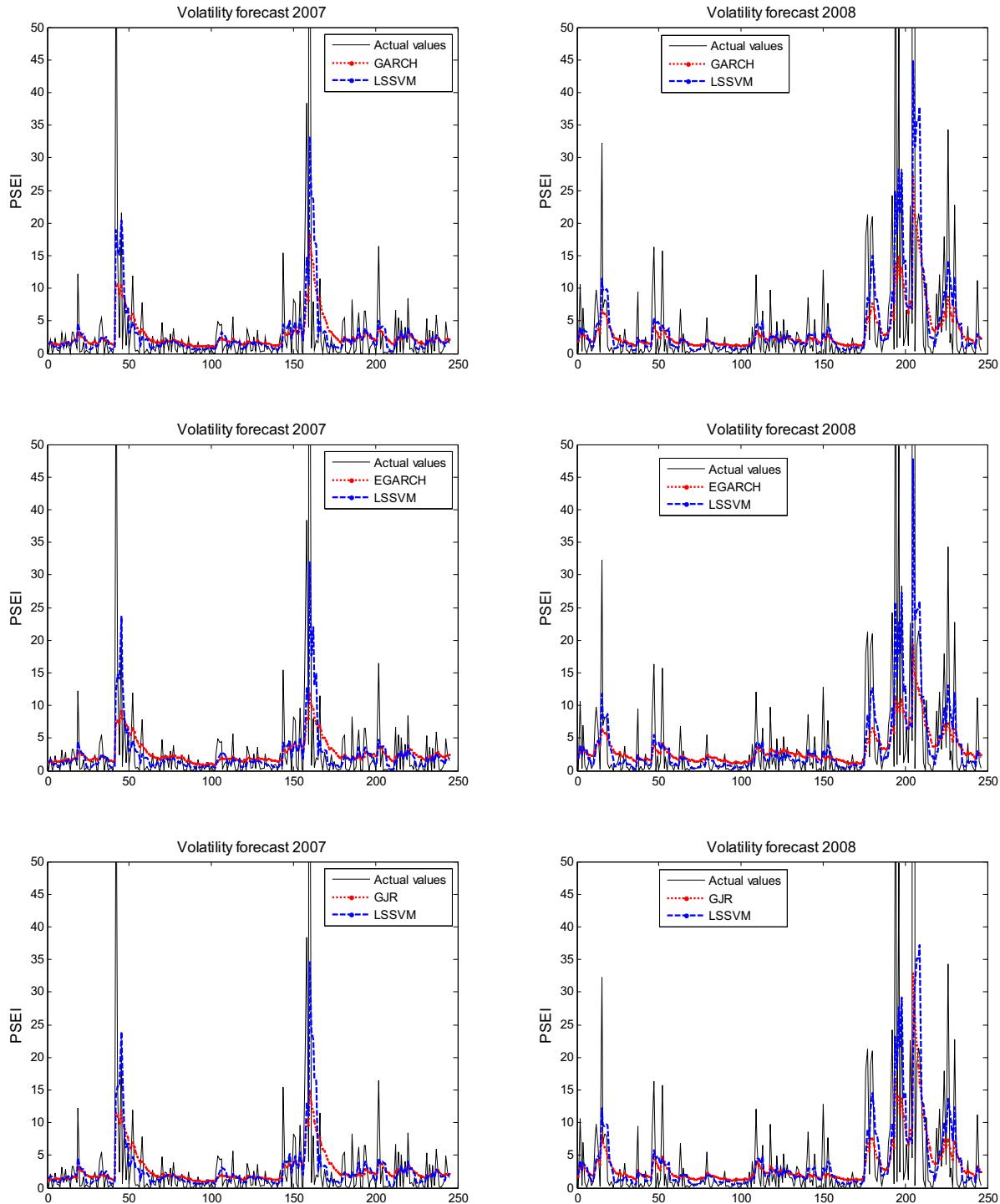


Figure 4. Volatility Forecasts of the Philippines stock market (PSEI).

Note: Plots in left part are referred to the First stage forecast in 2007 (before crisis) and plots in the right side are referred to the second stage forecast for whole 2008 (during financial crisis). Small dot line is forecasted by parametric models (GARCH, EGARCH and GJR) while dash line is obtained by hybrid approaches.