

# Modeling and Forecasting Stock Market Volatility by Gaussian Processes based on GARCH, EGARCH and GJR Models

PhichHang Ou, Hengshan Wang

**Abstract**— Probabilistic methods called Gaussian processes have been successfully shown as a powerful tool for modeling time series data and prediction problem as they are a Bayesian approach with kernel based learning. In this paper, the Gaussian processes are applied to model and predict financial volatility based on GARCH, EGARCH and GJR. Five different kernels are used to train each of the proposed volatility models. More precisely, the experimental results show that, the nonlinear hybrid models can capture well symmetric and asymmetric effects of news on volatility and yields better predictive performance than the classic GARCH, EGARCH and GJR approaches.

**Index Terms**— Gaussian Process, GARCH, EGARCH, GJR, volatility

## I. INTRODUCTION

Financial time series plays a crucial role in modeling and forecasting volatility of stock markets. The most famous and classic models include GARCH, EGARCH, and GJR models, [1,2,3,4] which cover symmetric and asymmetric effects of news in volatility. Since the volatility is very important for portfolio selection, option pricing and risk management, many researchers have used machine learning techniques, for instance, neural network and support vector machine, to improve the prediction of the financial volatility [5]. The main reason of using neural network and support vector machine is their flexible abilities to approximate any nonlinear functions arbitrarily without priori assumptions on data distribution [6]. Hence these approaches can cope with the situation that stock market is most of the time heavy tailed and violates normality. In addition, the neural network can capture the stylized characteristics of financial returns such as leptokurtosis, volatility clustering, and leverage effects; hence it generates better prediction of GARCH family [7]. For comprehensive discussions of GARCH predictions by the support vector machine approach, it is referred to [8,9,10,11].

Manuscript received in July 23, 2010. This work was supported by Shanghai Leading Academic Discipline Project, Project Number: S30504.

P. H. Ou is with School of Business, University of Shanghai for Science and Technology, Rm 101, International Exchange Center, No. 516, Jun Gong Road, Shanghai 200093, China. (Corresponding author, phone: +86-13661751547; e-mail: phichhang@gmail.com).

H. S. Wang is with School of Business, University of Shanghai for Science and Technology, No. 516, Jun Gong Road, Shanghai 200093, China. (e-mail: wanghs@usst.edu.cn).

Gaussian processes (GP) are one of Bayesian non-parametric methods in machine learning whose functional form are the same as the well-known neural network and support vector machine. Like other kernel based methods such as the SVM, they combine a high flexibility of the model by working in high dimensional feature spaces with the simplicity that all operations are kernelized, i.e., they are performed in the lower dimensional input space using positive definite kernel. See [12] for detailed illustration. Besides, the GP have been applied in various areas including human function [13], real time online model [14], classification [15], non-stationary time series prediction [16], and dynamic model for human motion [17].

Inspired by such efficiently predictive capability of the Gaussian processes and their successful applications, we are very interested in applying it for volatility prediction. Therefore, it is valuable for us to research the problem of whether more accurate forecasting performance could be achieved if we combine the GP with GARCH, EGARCH, and GJR structures to construct new hybrid models of financial volatility, say GP-GARCH, GP-EGARCH and GP-GJR.

In this work, simulated and real data, NASDAQ index, are analyzed to validate the proposed models. Five different kernels (or covariance functions) are applied to each of the volatility models as they are the heart of obtaining the Gaussian processes. The paper is organized as follow. Next section introduces volatility models based on classical GARCH, EGARCH, GJR and hybrid Gaussian Processes with GARCH, EGARCH and GJR. Section 3 presents the experimental results including parameters estimation, Gaussian Processes training and out of sample forecasting performance. The last section is reserved for conclusion.

## II. VOLATILITY MODELING

### A. Parametric models

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

$$y_t = 100(\ln P_t - \ln P_{t-1}) \quad (1)$$

denotes the continuously compounded daily returns of the underlying assets at time  $t$ .

**GARCH(1,1)** is defined as  $y_t = \mu_t + \varepsilon_t$ ,  $\varepsilon_t = \sigma_t z_t$ , and

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

where  $\omega > 0$ ,  $\alpha \geq 0$ ,  $\beta \geq 0$ , and this process is weakly stationary if  $\alpha + \beta < 1$ .  $\{z_t\}$  is a sequence of (iid) independent identically distributed random variables with mean 0 and variance 1. One step ahead forecast of GARCH(1,1) is  $\sigma_{t+1}^2 = \hat{\omega} + \hat{\beta} \sigma_t^2 + \hat{\alpha} \varepsilon_t^2$ .

**GJR(1,1)** is written as the following:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2 + \gamma S_{t-1}^- \varepsilon_{t-1}^2 \quad (3)$$

where  $S_{t-1}^- = 1$  for  $\varepsilon_{t-1} < 0$  and 0 otherwise. Stationary assumption for the model requires  $\alpha + \beta + \frac{1}{2}\gamma < 1$ ,  $\omega > 0$ ,  $\alpha \geq 0$ ,  $\beta \geq 0$ ,  $\alpha + \gamma \geq 0$ . The model assumes that bad news have higher impact than the good news. One step-ahead-forecast for GJR(1,1) is  $\sigma_{t+1}^2 = \hat{\omega} + \hat{\beta}\sigma_t^2 + \hat{\alpha}\varepsilon_t^2$  for  $\varepsilon_t > 0$  and  $\sigma_{t+1}^2 = \hat{\omega} + \hat{\beta}\sigma_t^2 + \hat{\alpha}\varepsilon_t^2 + \hat{\gamma}\varepsilon_t^2$  for  $\varepsilon_t < 0$ .

**EGARCH(1,1)** is represented as logarithm form:  $\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \alpha \left[ \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - E \left( \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} \right) \right] + \gamma \left( \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right)$  (4)

EGARCH model allows good news and bad news to have a different impact on volatility because the level of  $\frac{\varepsilon_{t-1}}{\sigma_{t-1}}$  is included with a coefficient  $\gamma$ . One step ahead forecast by EGARCH(1,1) is

$$\ln(\sigma_{t+1}^2) = \hat{\omega} + \hat{\beta} \ln(\sigma_t^2) + \hat{\alpha} \left[ \frac{|\varepsilon_t|}{\sigma_t} - \sqrt{\frac{2}{\pi}} \right] + \hat{\gamma} \left( \frac{\varepsilon_t}{\sigma_t} \right).$$

### B. Hybrid models of GP-GARCH, GP-EGARCH and GP-GJR

Gaussian Processes based on GARCH, EGARCH and GJR models are represented as

**GP-GARCH:**  $\sigma_t^2 = f(x_{t-1}, x_{t-1}^*)$  (5)

where  $\sigma_t^2$  is treated as dependent variable, and  $x_{t-1} = \sigma_{t-1}^2$ ,  $x_{t-1}^* = \varepsilon_{t-1}^2$  are treated as independent variables. Its one step ahead forecast is  $\sigma_{t+1}^2 = \hat{f}(x_t, x_t^*)$ .

**GP-GJR model:**  $\sigma_t^2 = f(x_{t-1}, x_{t-1}^*, x_{t-1}^{**})$  (6)

where the input variables are defined as  $x_{t-1} = \sigma_{t-1}^2$ ,  $x_{t-1}^* = \varepsilon_{t-1}^2$  and  $x_{t-1}^{**} = S_{t-1}^- \varepsilon_{t-1}^2$ . One step ahead forecast of this model is  $\sigma_{t+1}^2 = \hat{f}(x_t, x_t^*, x_t^{**})$ .

**GP-EGARCH:**  $\ln(\sigma_t^2) = f(x_{t-1}, x_{t-1}^*, x_{t-1}^{**})$  (7)

where  $\ln(\sigma_t^2)$  is treated as response. The regressors, in this case, are denoted as  $x_{t-1} = \ln(\sigma_{t-1}^2)$ ,  $x_{t-1}^* = \left[ \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - E \left( \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} \right) \right]$  and  $x_{t-1}^{**} = \left( \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right)$ . Here the expected value  $E \left( \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} \right)$  is approximated by its mean value. One step ahead forecast based on GP-EGARCH is defined as exponential transformation:  $\sigma_{t+1}^2 = \exp\{ \hat{f}(x_t, x_t^*, x_t^{**}) \}$ . Note, for all models, we take  $y_t^2 = \varepsilon_t^2$  assuming  $\mu_t = 0$  (as usual) and  $\hat{\sigma}_t^2 = \frac{1}{5} \sum_{k=0}^4 y_{t-k}^2$  according to [8] so that we can obtain  $x_{t-1}$ ,  $x_{t-1}^*$ ,  $x_{t-1}^{**}$  and  $\sigma_t^2$  before making analysis. The function  $f$  is approximated by the Gaussian Processes in (10) below to be  $\hat{f}$ .

### C. Gaussian Processes

Gaussian processes are a popular method for nonparametric regression. Gaussian processes are defined by its mean and covariance functions. It is assumed that the observations are normally distributed. Using the kernel matrix as the covariance matrix is a convenient way of extending Bayesian modeling of linear estimators to nonlinear situations. Furthermore, it represents the counterpart of the kernel trick in methods minimizing the regularized risk. Specifically, given a set of training data  $D = \{(\mathbf{x}_i, y_i), i = 1, \dots, n\}$ , the objective is to predict the value of new  $y_*$  given  $\mathbf{x}_*$  via a learning function  $f(\mathbf{x}_i)$  such

that  $y_i = f(\mathbf{x}_i) + \delta_i$  where  $\delta_i$  is normally distributed with mean zero and variance  $\sigma_n^2$ . With the assumption of the Gaussian noise,  $\delta$ , the observed target can be described as  $\mathbf{y} \sim N(0, \mathbf{K}(\mathbf{X}, \mathbf{X}))$  where  $\mathbf{K}(\mathbf{X}, \mathbf{X})$  denotes the covariance matrix and  $\mathbf{X}$  is  $(n \times m)$  matrix of input. The common choice of covariance function is a Gaussian kernel defined as

$$k(\mathbf{x}_p, \mathbf{x}_q) = \sigma_s^2 \exp \left\{ -\frac{1}{2} (\mathbf{x}_p - \mathbf{x}_q)^T \mathbf{W} (\mathbf{x}_p - \mathbf{x}_q) \right\} \quad (8)$$

where  $\sigma_s^2$  denotes the signal variance and  $\mathbf{W}$  are the widths of the Gaussian kernel.

For a new  $\mathbf{x}_*$ , the predictive distribution of the corresponding outputs is simply obtained by conditioning on the training data to obtain  $p(f(\mathbf{x}_*) | \mathbf{x}_*, D)$ . The joint distribution of the observed target values and predicted value for  $\mathbf{x}_*$  is given by

$$\begin{bmatrix} \mathbf{y} \\ f(\mathbf{x}_*) \end{bmatrix} \sim N \left( 0, \begin{bmatrix} \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & \mathbf{k}(\mathbf{X}, \mathbf{x}_*) \\ \mathbf{k}(\mathbf{x}_*, \mathbf{X}) & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix} \right). \quad (9)$$

The conditional distribution is also Gaussian and yields the predicted mean value  $\bar{f}(\mathbf{x}_*)$  with the variance  $V_f(\mathbf{x}_*)$ ,

$$\bar{f}(\mathbf{x}_*) = \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} = \mathbf{k}_*^T \boldsymbol{\alpha} \quad (10)$$

$$V_f(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_* \quad (11)$$

where  $\mathbf{k}_* = \mathbf{k}(\mathbf{X}, \mathbf{x}_*)$ ,  $\mathbf{K} = \mathbf{K}(\mathbf{X}, \mathbf{X})$  and  $\boldsymbol{\alpha}$  denotes the so-called prediction vector.

The predicted mean  $\bar{f}(\mathbf{x}_*)$  is in general used as an estimate of the output  $\hat{f}(\mathbf{x}_*)$  with uncertainty  $V_f(\mathbf{x}_*)$ . The hyper-parameters of Gaussian processes with their kernel and their optimal value for a particular data set can be derived by maximizing the log marginal likelihood using common optimization procedures such as Quasi-Newton methods.

Five different kernels (or covariance functions) used in this study includes

- Gaussian radial basis function
$$k(\mathbf{x}, \mathbf{x}') = \exp(-\sigma \|\mathbf{x} - \mathbf{x}'\|^2) \quad (12)$$

- Linear kernel which is simplest case:
$$k(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle \quad (13)$$

- Polynomial kernel:
$$k(\mathbf{x}, \mathbf{x}') = (\text{scale} \cdot \langle \mathbf{x}, \mathbf{x}' \rangle + \text{offset})^{\text{degree}} \quad (14)$$

- Laplace radial basis kernel:
$$k(\mathbf{x}, \mathbf{x}') = \exp(-\sigma \|\mathbf{x} - \mathbf{x}'\|) \quad (15)$$

- Bessel function:
$$k(\mathbf{x}, \mathbf{x}') = \frac{\text{Bessel}_{(\vartheta+1)}^n(\sigma \|\mathbf{x} - \mathbf{x}'\|)}{(\|\mathbf{x} - \mathbf{x}'\|)^{-n(\vartheta+1)}}. \quad (16)$$

## III. EMPIRICAL ANALYSIS

### A. Artificial Simulated Data

In this section, we investigate the forecasting performance of all models using simulated data. We generate 1000 samples from GARCH(1,1) with coefficients ( $c = 0.0001$ ,  $\omega = 0.01$ ,  $\alpha = 0.1$ ,  $\beta = 0.8$ ) and EGARCH(1,1) ( $c = 0.0001$ ,  $\omega = 0.01$ ,  $\alpha = 0.1$ ,  $\beta = 0.8$ ,  $\gamma = -0.06$ ) as well as GJR(1,1) ( $c = 0.0001$ ,  $\omega = 0.01$ ,  $\alpha = 0.1$ ,  $\beta = 0.8$ ,

$\gamma = 0.06$ ) with, first, Normal innovation and then the Student's t with 6 degree of freedom. The second innovation distribution tries to model the excess of kurtosis that usually appears in real financial time series. Six sets of simulated data are achieved and then fitted by hybrid models based on the GP as in (5), (6), and (7). These simulated data are also fitted again, respectively, by six different models of GARCH, EGARCH and GJR with Normal and Student's innovations to make comparison with the hybrid approaches. 75% of the whole data, 750 observations, are used for estimation or training the models; and the remaining data, 250 points, are reserved for out of sample forecasting. GARCH toolbox in Matlab is used to simulate the return series and to fit GARCH, EGARCH and GJR models; and Kernlab package in R software by [18] is used to fit the volatility models based on the Gaussian processes.

TABLE I  
TRAINING RESULTS BY GAUSSIAN PROCESS WITH FIVE  
DIFFERENT KERNELS (SIMULATED DATA)

Hybrid models	Train. Error	CV error	Kernel Para.
GP-GARCH 1	0.25137	0.00340	5.01863
GP-GARCH 2	0.25866	0.00189	
GP-GARCH 3	0.25473	0.00221	$d=3, a=1, b=1$
GP-GARCH 4	0.20130	0.00470	7.40439
GP-GARCH 5	0.26465	0.00211	$n=1, \theta = 1, \sigma = 1$
GP-EGARCH 1	0.33234	0.22741	0.48660
GP-EGARCH 2	0.34233	0.21404	
GP-EGARCH 3	0.33538	0.22281	3,1,1
GP-EGARCH 4	0.25073	0.24047	0.89982
GP-EGARCH 5	0.35519	0.23249	1,1,1
GP-GJR 1	0.22908	0.00722	5.38091
GP-GJR 2	0.24774	0.00352	
GP-GJR 3	0.24222	0.00417	3,1,1
GP-GJR 4	0.18835	0.00844	5.27188
GP-GJR 5	0.27141	0.00467	1,1,1

Note: 1. GP-GARCH 1, 2, 3, 4, 5 denote Gaussian Process based on GARCH with Gaussian Kernel, Linear kernel, Polynomial kernel, Laplace Kernel, and Bessel kernel respectively.  
2. CV error = 10 fold Cross Validation error.  
3.  $a$  is scale,  $b$  is offset and  $d$  is degree of the polynomial kernel.

TABLE II  
OUT OF SAMPLE FORECASTING PERFORMANCE (SIMULATED DATA)

Simulated	Normal innovation		Student's Innovation	
	NMSE	$R^2$	NMSE	$R^2$
GARCH(1,1)				
GARCH-N,T	0.97330	0.1505	0.97040	0.1499
GP-GARCH1	0.74431	0.2626	0.98552	0.0260
GP-GARCH2	0.70946	0.3022	<b>0.69233</b>	<b>0.3341</b>
GP-GARCH3	<b>0.69732</b>	<b>0.3161</b>	0.70326	0.3216
GP-GARCH4	0.80456	0.1974	0.99819	0.0188
GP-GARCH5	0.69627	0.3226	0.80097	0.2038
Simulated	Normal innovation		Student's Innovation	
EGARC(1,1)	NMSE	$R^2$	NMSE	$R^2$
EGARC-N,T	0.94444	0.0900	0.92210	0.1093
GP-EGARC1	<b>0.70897</b>	<b>0.3300</b>	0.87006	0.1618
GP-EGARC2	0.74054	0.2820	0.79103	0.2599
GP-EGARC3	0.71590	0.3224	<b>0.54061</b>	<b>0.5315</b>
GP-EGARC4	0.71949	0.3177	0.87476	0.1534
GP-EGARC5	0.72364	0.2998	0.85129	0.1936
Simulated	Normal innovation		Student's Innovation	
GJR(1,1)	NMSE	$R^2$	NMSE	$R^2$
GJR-N,T	0.89345	0.1068	0.76465	0.1562
GP-GJR 1	0.83414	0.1630	0.97703	0.0290
GP-GJR 2	0.64595	0.3738	0.66360	0.3601
GP-GJR 3	<b>0.63715</b>	<b>0.3833</b>	<b>0.58039</b>	<b>0.4437</b>
GP-GJR 4	0.87530	0.1280	0.97189	0.0415
GP-GJR 5	0.67596	0.3332	0.94129	0.0603

Note:  $R^2$  is obtained by regression between actual volatilities (square returns) and predicted volatilities. NMSE stands for normalized mean squared error.

In this experiment, 70% of the training data is used for training the GP model to get the optimal hyper-parameter of Kernel (or correlation) function and the remaining data, 30%, is reserved for validation of the model. The optimal results of the training, including the training error, 10 fold cross-validation error and Kernel parameter, are displayed in Table I. The training results of GP from simulated data for GARCH, EGARCH and GJR based returns with Student's t innovation are not shown here to reduce the space. The forecasting performance from each model for out-of-sample data, measured by NMSE and  $R^2$ , are shown in Table II. Basically, the larger  $R^2$  value, the better the model predicts the actual values. The smaller value of NMSE, the more preferred the model. From the Table II, it is seen that most of the hybrid models yield more accuracy than the parametric models.

### B. Real data of NASDAQ index

Now we analyze NASDAQ index,  $P_t$ , which is first downloaded from the Yahoo finance and then transformed into log return as in (1). The whole sample consists of 3274 daily data spanning from 02 Jan. 1996 to 31 Dec. 2008 which covers the financial crisis period. We select subsample of size 3023, dated from 02 Jan. 1996 to 31 Dec. 2007, as the training set for the Gaussian Processes or parameter estimation for classic models. The remaining sample of size 251 daily data, Jan-Dec 2008, is used as the test set or for out-of-sample forecasting. Table III summarizes the descriptive statistics of the index return series along the whole period.

TABLE III  
SUMMARY STATISTIC (NASDAQ)

Sample	3274	Excess Kurtosis	4.57513
Mean	0.0132	LB.Q <sup>2</sup>	3057.27
Std.Dev.	1.7998	JB test	2861.30
Skewness	-0.01079	ARCH(12)	648.367

We remark that these facts suggest a highly competitive and volatile market. There are significant price fluctuations in the markets as suggested by positive standard deviation. The negative skewness indicates that there is a high probability of losses in the market. The excess value of kurtosis suggests that the market is volatile with high probability of extreme events occurrences. Moreover, the rejection of Jarque Bera (JB) test of normality shows that the returns deviate from normal distribution significantly and exhibit leptokurtic. The Ljung Box statistic for squared return and Engle ARCH test prove the exhibition of ARCH effects in the return series. Therefore, it is appropriate to apply GARCH, EGARCH, GJR especially the hybrid approach for modeling volatility to this return.

We, then, apply the return series to (2), (3) and (4) to get parameters of GARCH, GJR and EGARCH models with normal and student's t innovations. The estimation results and diagnosis are shown in Table IV. Among the parametric models, the series best fit to the GJR with student's t innovation, GJR-T, according to highest value of Log likelihood (LL) and smallest value of AIC and BIC. The conditions of asymmetric effects hold for all models according to the significance of leverage effect coefficients  $\gamma$ 's. This implies that the market exhibits two kinds of leverage effects of news impact in volatility.

TABLE IV

ESTIMATION RESULTS BY PARAMETRIC MODELS

Statistics	GARCH-N	EGARCH-N	GJR-N
$\mu$	0.080[0.022]	0.037[0.022]	0.045[0.022]
$\omega$	0.015[0.004]	0.009[0.002]	0.017[0.004]
$\alpha$	0.086[0.007]	0.151[0.012]	0.028[0.007]
$\beta$	0.909[0.007]	0.988[0.002]	0.918[0.007]
$\gamma$		-0.064[0.005]	0.092[0.009]
LL	-4910	-4884	-4885
AIC	9828	9778	9780
BIC	9852	9808	9810

Note: [ ] indicates standard error of estimation.

Statistics	GARCH-T	EGARCH-T	GJR-T
$\mu$	0.084[0.022]	0.101[0.019]	0.059[0.022]
$\omega$	0.011[0.004]	0.001[0.005]	0.013[0.004]
$\alpha$	0.074[0.009]	0.350[0.031]	0.021[0.009]
$\beta$	0.922[0.009]	0.962[0.006]	0.926[0.009]
$\gamma$	--	-0.08[0.014]	0.091[0.013]
$\vartheta$	15.21[2.847]	11.91[2.146]	16.97[3.497]
LL	-4895	-4923	-4874
AIC	9801	9859	9760
BIC	9831	9895	9795

Note: [ ] indicates standard error of estimation.

TABLE V

TRAINING RESULTS BY GAUSSIAN PROCESSES

Hybrid models	Train. Error	CV error	Kernel Para.
GP-GARCH 1	0.20629	16.1661	42.56467
GP-GARCH 2	0.15063	4.00422	
GP-GARCH 3	0.14506	3.71741	3, 1, 1
GP-GARCH 4	0.21857	19.8882	49.10658
GP-GARCH 5	0.18365	6.81695	1, 1, 1
GP-EGARCH 1	0.13332	0.19156	0.64779
GP-EGARCH 2	0.13863	0.18574	
GP-EGARCH 3	0.13386	0.18165	3, 1, 1
GP-EGARCH 4	0.11323	0.18973	0.59678
GP-EGARCH 5	0.13836	0.18820	1, 1, 1
GP-GJR 1	0.21351	17.3607	33.09276
GP-GJR 2	0.14812	3.53482	
GP-GJR 3	0.13935	16.4461	3, 1, 1
GP-GJR 4	0.21810	19.8230	30.89865
GP-GJR 5	0.18494	7.74181	1, 1, 1

Note: 1. GP-GARCH 1, 2, 3, 4, 5 denote Gaussian Process based on GARCH with Gaussian Kernel, Linear kernel, Polynomial kernel, Laplace Kernel, and Bessel kernel respectively.

2. CV error = 10 fold Cross Validation error.

Now we train Gaussian processes to the return series by (5) for GARCH, (6) GJR and (7) for EGARCH. The optimal results of the training, including the training error, 10 fold cross-validation error and Kernel parameter, are displayed in Table V. These volatility models are used to forecast future volatilities in the test set for comparison with the classic models. Table VI report the values of NMSE and  $R^2$  obtained from different models. The Gaussian processes achieve smallest value of NMSE and highest value of  $R^2$ . This implies that the Gaussian Processes method can capture well both symmetric and asymmetric volatilities and also produce improved forecasting performance than the parametric models. Fig. 1,2,3 also plot the superior performance of the Gaussian Processes in modeling and forecasting volatilities based on GARCH (Fig.1) and EGARCH forms (Fig.2) as well as GJR types (Fig.3). It is seen that the forecasting lines by the hybrid models, GP-GARCH, GP-EGARCH and GP-GJR, are more flexible and can capture more extreme points than the corresponding classic models.

TABLE VI

OUT OF SAMPLE FORECASTING PERFORMANCE (NASDAQ)

	NMSE	$R^2$
GARCH-N	0.7925	0.2163
GARCH-T	0.8105	0.1949
GP-GARCH1	0.6995	0.3353
GP-GARCH2	<b>0.5576</b>	<b>0.4513</b>
GP-GARCH3	0.5599	0.4483
GP-GARCH4	0.6965	0.3719
GP-GARCH5	0.6277	0.3847
EGARCH-N	0.8013	0.1961
EGARCH-T	0.6233	0.3934
GP-EGARCH1	0.6741	0.3387
GP-EGARCH2	0.5958	0.4259
GP-EGARCH3	<b>0.5769</b>	<b>0.4392</b>
GP-EGARCH4	0.6521	0.3626
GP-EGARCH5	0.6904	0.3442
GJR-N	0.7782	0.2317
GJR-T	0.7886	0.2188
GP-GJR1	0.7081	0.2899
GP-GJR2	<b>0.5586</b>	<b>0.4500</b>
GP-GJR3	0.5621	0.2965
GP-GJR4	0.6971	0.3701
GP-GJR5	0.6471	0.3693

Note:  $R^2$  is obtained by regression between actual volatilities (square returns) and predicted volatilities. NMSE stands for normalized mean squared error.

#### IV. CONCLUSION

In this paper, we proposed nonlinear volatility models based on Gaussian processes combined with GARCH, EGARCH and GJR models. The empirical analysis of simulated and real data, NASDAQ index, show that these hybrid models can capture well the symmetric and asymmetric effects of news in volatility and generate superior performance of volatility prediction than the corresponding classic models of GARCH, EGARCH and GJR with both Normal and Student's t innovations. Among the five kernels, Linear and Polynomial kernels are most suited for the hybrid Gaussian processes in this study. In addition, the Gaussian process models are simple, practical and powerful Bayesian tools for data analysis.

#### ACKNOWLEDGMENT

We appreciate referees for comments on our work.

#### REFERENCES

- [1] R. F. Engle, "Autoregressive conditional heteroskedasticity with estimates of variance of UK inflation", *Econometrica*, vol. 50, no. 4, pp. 987-1008, 1982.
- [2] T. Bollerslev, "Generalized Auto Regressive Conditional Heteroskedasticity", *Journal of Econometrics*, vol. 31, no. 3, pp 307-327, 1986.
- [3] D. B. Nelson, "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, vol. 59, no. 2, pp. 347-370, 1991.
- [4] L. Glosten, R. Jagannathan, and D. Runkle, "On the relationship between the expected value and the volatility of the nominal excess return on stocks", *Journal of Finance*, vol. 48, no. 5, pp. 1779-1801, 1993.
- [5] R. G. Donaldson, and M. Kamstra, "An artificial neural network-GARCH model for international stock return volatility", *Journal of Empirical Finance*, vol. 4, no. 1, pp. 17-46, Jan. 1997.
- [6] S. Haykin, *Neural networks: a comprehensive foundation*, Englewood cliffs: Prentice Hall, 1999.
- [7] M. Bildirici and O. O. Ersin, "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*, vol. 36, no. 4, pp. 7355-7362, May 2009.

- [8] F. Perez-Cruz, J. A. Afonso-Rodriguez, and J. Giner, “Estimating GARCH models using support vector machines. *Quantitative Finance*, vol. 3, no. 3, pp 163-172, Jun. 2003.
- [9] L. B. Tang, H.Y. Sheng, and L.X. Tang, “Forecasting volatility based on wavelet support vector machine”, *Expert Systems with Applications*, vol. 36, no. 2, pp. 2901-2909, Mar. 2009.
- [10] L. B. Tang, H. Y. Sheng, and L. X. Tang, “GARCH prediction using spline wavelet support vector machine”, *Neural Computing and Application*, vol. 18, no. 8, pp. 913-917, Feb. 2009.
- [11] S. Chen, K. Jeong, & W. Hardle, “Forecasting volatility with support vector machine-based GARCH model”, *Journal of Forecasting*, vol. 29, no. 4, pp. 406-433, Sep. 2009.
- [12] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Massachusetts: the MIT Press, 2006.
- [13] T. L. Griffiths, C. Lucas, J. Williams, and M. L. Kalish, “Modeling human function learning with Gaussian processes”, in *Proc. 22nd Annu. Conf. on Neural Information Processing Systems Foundation*, Berkeley, 2008.
- [14] D. Nguyen-Tuong, M. Seeger and J. Peters, Real-Time Local GP Model Learning, O. Sigaud, J. Peters(Eds): *From Motor Learning to Interaction Learning in Robots, Studies in Computational Intelligence*, vol. 264, pp. 193-207, 2010.
- [15] C. K. I. Williams & D.Barber, “ Bayesian Classification with Gaussian Processes”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1342-1351, 1999.
- [16] S. Brahim-Belhouari, and A. Bermak, “Gaussian process for nonstationary time series prediction”, *Computational Statistics & Data Analysis*, vol. 47, no. 4, pp. 705-712, Nov. 2004.
- [17] J. M. Wang, D. J. Fleet and A. Hertzmann. “Gaussian Process Dynamical Models for Human Motion”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 283-298, Feb. 2008.
- [18] A. Karatzoglou, A. Smola., A. Hornik and A. Zeileis. (2004). An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, [Online].11(9). Available: <http://www.jstatsoft.org/>

APPENDIX

The NMSE is defined as  $NMSE = \frac{1}{ns^2} \sum_{t=1}^n (a_t - p_t)^2$  where  $s^2 = \frac{1}{n-1} \sum_{t=1}^n (a_t - \bar{a}_t)^2$ . Here we take  $a_t = y_t^2$  as actual values, and  $p_t = \hat{\sigma}_t^2$  is treated as the forecasted volatility obtained by each of the competing models.  $n$  is sample size corresponding to each time of forecasting. The square correlation  $R^2$  is a measure of forecasting performance is obtained by regressing squared return on a constant and the forecasted volatility for out-of-sample time point,  $t = 1, 2, \dots, n$ ,  $y_t^2 = C_0 + C_1 \hat{\sigma}_t^2 + e_t$ . In this regression, the constant term  $C_0$  should be close to zero and the slope  $C_1$  should close to 1. The higher value of  $R^2$ , the better the forecasting model.

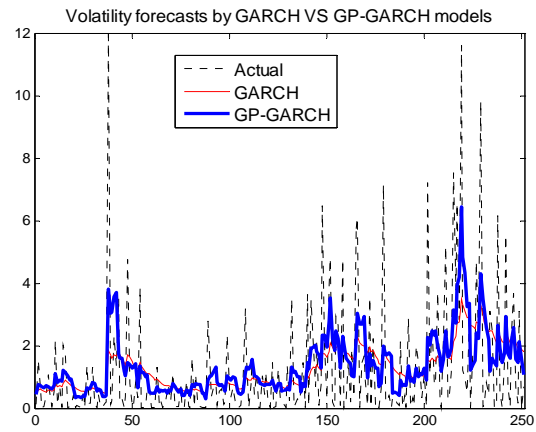


FIGURE 1. . PLOT OF VOLATILITY BASED GARCH VERSUS GP-GARCH

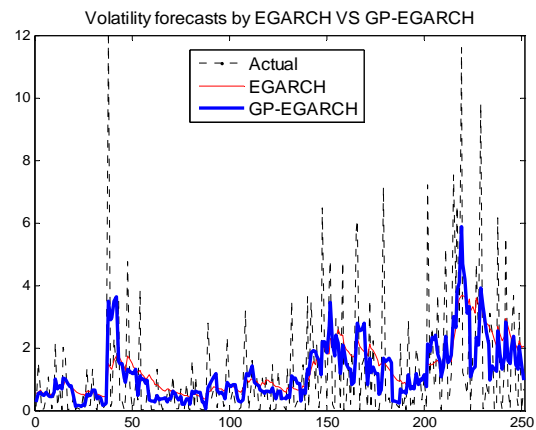


FIGURE 2. . PLOT OF VOLATILITY BASED EGARCH VERSUS GP-EGARCH

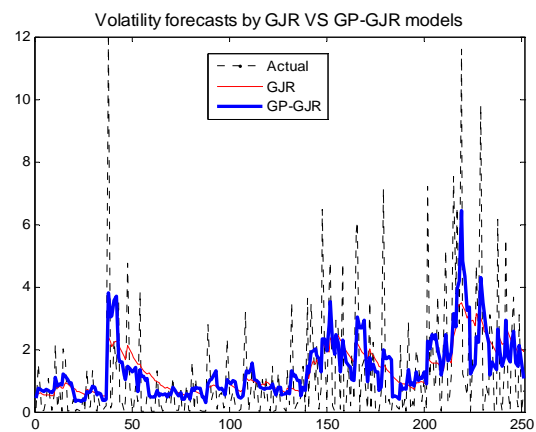


FIGURE 3. PLOT OF VOLATILITY BASED GJR VERSUS GP-GJR