Energy Forward Price Prediction with a Hybrid Adaptive Model

Hang T. Nguyen and Ian T. Nabney

Abstract— This paper presents a forecasting technique for forward electricity/gas prices, one day ahead. This technique combines a Kalman filter (KF) and a generalised autoregressive conditional heteroschedasticity (GARCH) model (often used in financial forecasting). The GARCH model is used to compute next value of a time series. The KF updates parameters of the GARCH model when the new observation is available. This technique is applied to real data from the UK energy markets to evaluate its performance. The results show that the forecasting accuracy is improved significantly by using this hybrid model. The methodology can be also applied to forecasting market clearing prices and electricity/gas loads.

I. Introduction

ACCURATE electricity/gas price forecasting is very important for traders in the energy market, especially energy generators. If an energy generator makes an accurate forecast of the market price, it can develop a strategy to maximise its own profits and minimise risk due to price spikes by appropriate trading in forward contracts. It can also plan its actions to maximise benefits or utilities by reducing/increasing its generation. Energy suppliers can use short-term price forecasts to adjust their bidding strategies to achieve the maximum benefit. In addition, understanding the process of forward price development can help the generators make money on the forward market.

The relevant literature on energy price forecasting issues includes a wide range of methods. Many approaches based on time series models have been used for price forecasting, such as threshold auto-regressive switching (TAR) models [1], AR models [2], autoregressive integrated moving average (ARIMA) models [3], [4], and GARCH [5], [6]. Moreover, neural networks (NNs) are used widely for electricity price forecasting in the literature [7]. Due to the complexity of the environment, the functional relationships we are looking for might be non-linear. Several researchers have proposed additional procedures to improve accuracy. Pre-processing procedures and regularisation methods are used in [7]. Another approach for improving forecasting performance is multiple NNs. The use of a committee of NNs for forecasting is suggested in [8]. Similarly, cascaded neural networks are proposed in [9]. The use of Input/Output Hidden Markov Model (IOHMM) to combine two NNs which estimate underlying market states and observed price respectively are suggested in [10].

However, forecasting energy prices presents a number of challenges because of the volatility characteristic of prices,

Hang T. Nguyen and Ian T. Nabney are with the Neural Computing Research Group, School of Engineering and Applied Science, Aston University, United Kingdom (e-mail: thihangn@aston.ac.uk and i.t.nabney@aston.ac.uk).

especially in competitive environments [5]. To overcome this problem, some researchers have proposed a hybrid form of forecasting models, which are the combination of Kalman filter/Extend Kalman Filter (EKF) and prediction models, such as Radial Basic Function (RBF), Multilayer perceptrons (MLP), linear regression (LR), or financial models. The forecast models are used to forecast next-day price, and KF/EKF updated parameters of these models as the new value of price are observed. Niranjan [11] used EKF algorithm to recursively reestimate parameters of the Black-Scholes model from observations (the Black-Scholes model is well-known financial formula for approximate option price). Nabney et al. [12] showed that an EKF used for online-learning parameters of RBF model give much better tracking of non-stationary data than a fixed RBF model. Some researchers proposed a training method, which used an EKF in order to train MLP network. The simulation results on predicting exchange rate [13], positioning a GSM mobile phone in real-time [14], estimate wind turbine power generation [15], and predicting New England electricity price [16] show that this method is good in the speed of learning and the accuracy of mapping. Parameters of linear models were also estimated using a KF in [17], [18].

This paper uses the combination of KF and GARCH model. This model is applied in forecasting forward electricity/gas price in the UK market. We compare the prediction performance of three methods:

- *Method 1*: a random walk model, which is used as a benchmark.
- *Method 2*: fixed GARCH model, i.e. model whose parameters are fixed after training on a subset of the data set.
- *Method 3*: a combination of KF and GARCH model, where parameters of GARCH model are adapted continuously on the test set using KF.

Compared with previous work, our paper has the following contributions. Firstly, we propose a framework for an adaptive GARCH model and apply it for electricity/gas price forecasting problem. Secondly, besides historical price data, a number of exogenous variables, for example, electricity/gas demand, temperature, exchange rate (USD:GBP) etc, are also considered as candidates for input variables. Some pre-processing procedures are used to chose the relevant input variables for each forecasting model.

This paper is organised as follows. Section II provides the detailed forecasting framework. In Section III, the detail of the hybrid model is presented. The numerical results on data from the UK electricity/gas market are given in Section IV. Section V provides some conclusions.
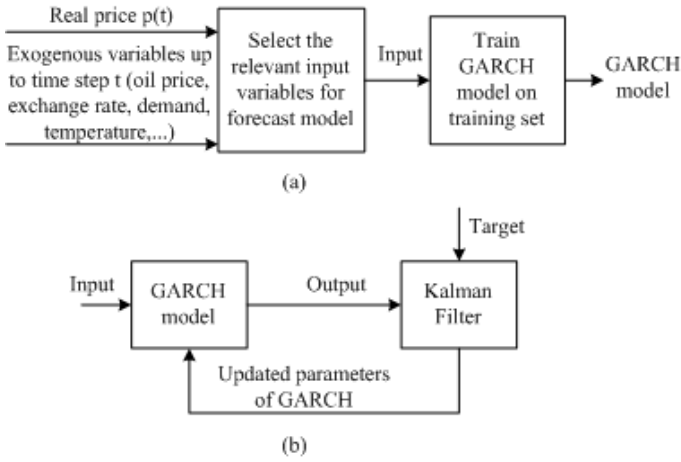
Fig. 1. The combination model of GARCH and KF. (a) Training phase, (b) Test phase.

II. FORECASTING FRAMEWORK

The framework of Method 3 for forecasting electricity/gas prices of forward contracts is shown in Figure 1. A data set is divided into two subsets: (1) a training set to train the models and (2) a test set to evaluate these models by calculating their errors.

- Training phase:

Step 1: Determine the input vectors for GARCH model. In addition to historical price data, there is a large number of observable variables which are potential inputs. However, only some of them are relevant. Using irrelevant variables as inputs might reduce the performance of the forecasting models. Therefore, selecting correct inputs for each type of model is very important. In the this phase, various measures were used to determine the relevant input variables, including the correlation matrix (CM), autocorrelation function (ACF), and partial autocorrelation function (PACF).

Step 2: The training sets are used to estimate parameters of GARCH models. We can use maximum likelihood to train the model with an iterative non-linear optimisation algorithm.

- Test phase: Two steps are recursively repeated.

Step 1: When the new observation is available, the Kalman Filter updates parameters of the GARCH model.

Step 2: Use the GARCH model with the latest estimated parameters to predict the next-day price.

III. COMBINATION MODEL

A. GARCH

In several forecast model, such as MLP, RBF or LR, the errors are assumed to be homoscedastic (i.e., the variance of the residual is assumed to be independent of time). GARCH [19] can be used to model changes in the variance of the errors as a function of time. The GARCH(r, m) model is given by:

$$
\begin{align}
y_t &= \beta_0 + \widehat{\beta} x_t + \varepsilon_t \tag{1} \\
\varepsilon_t &= \sqrt{n_t} e_t, \qquad e_t \sim D(0,1) \tag{2} \\
n_t &= \alpha_0 + \sum_{i=1}^{m} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{r} \gamma_j n_{t-j}, \tag{3}
\end{align}
$$

with constraints:

$$
\alpha_i, \gamma_j > 0, \ \sum_{i=1}^{m} \alpha_i + \sum_{j=1}^{r} \gamma_j < 1, \tag{4}
$$

where x_t, y_t, and ε_t represent the input vector, output vector, and error of the model respectively, n_t is variance of error ε_t, $\beta = \{\beta_0, \widehat{\beta}\}$ is parameters vector for the AR process, m and r are order of ARCH process and AR for the variance n_t respectively.

e_t is i.i.d, with $E(e_t) = 0$ and $var(e_t) = 1$. e_t can be a Gaussian or Student-T distribution. GARCH is a generalisation of a linear time series model with homoscedastic disturbances in which the variance n_t of the noise varies with information about errors and its variance up to time t.

We can train a GARCH model using maximum likelihood. To deal with the constraints (4), we used penalty function method to optimise the model.

B. Kalman Model

1) State space models: The Kalman Filter is based on a state space model; we assume that the observed time series $y_t \in R^p$ is a function of random variables $x_t \in R^k$ which are not observed.

$$
y_t = H_t x_t + v_t, \tag{5}
$$

where $v_t \in R^p$ is assumed to be Gaussian noise $N(0, R)$, y_t is the output of the model, x_t is the hidden state vector. The $p \times k$ matrix H_t is the output matrix. Equation (5) represents the output function. We assume that the dynamics of the hidden state space is given by

$$
x_{t+1} = F_t x_t + w_t, \tag{6}
$$

where the $k \times k$ matrix F_t forms the parameters of the model and w_t is zero-mean Gaussian noise with covariance Q. Equation (6) represents the state transition function. The hidden state vector obeys the Markov independence property (i.e. the current state depends only on the previous state).

x_0 is the system initial condition, modelled as a Gaussian random vector $x_0 \sim N(\pi_1, P_1)$. Equation (6) shows that if $p(x_t)$ is Gaussian, then so is $p(x_{t+1})$.

2) Kalman Filter: The Kalman Filter [20] is a recursive algorithm used to compute the probability of the current hidden state space x_t given the sequence of observations up to time t. One iteration of the KF is composed of the following consecutive steps:

- Prediction:

$$
\begin{align}
x_t^{t-1} &= F_t x_{t-1}^{t-1} \\
P_t^{t-1} &= F_t P_{t-1}^{t-1} F_t' + Q.
\end{align}
$$

- Update

$$
\begin{align}
K_t &= P_t^{t-1} H_t' \left(H_t P_t^{t-1} H_t' + R_t \right)^{-1} \\
x_t^t &= x_t^{t-1} + K_t \left(y_t - H_t x_t^{t-1} \right) \\
P_t^t &= (I - K_t H_t) P_t^{t-1},
\end{align}
$$

where:

$$
\begin{aligned}
x_t^t &= E[x_t | y_1, ... y_t] \\
x_t^{t-1} &= E[x_t | y_1, ... y_{t-1}] \\
P_t^t &= E[(x_t - x_t^t)(x_t - x_t^t)' | y_1, ... y_t] \\
P_t^{t-1} &= E[(x_t - x_t^{t-1})(x_t - x_t^{t-1})' | y_1, ... y_{t-1}],
\end{aligned}
$$

where x_t^{t-1} is the *a priori* state estimate at time step t given knowledge of the process prior to step t, and x_t^t is an *a posteriori* state estimate at time step t given measurement y_t. The matrices P_t^{t-1} and P_t^t are *a priori* estimate error covariance and the *a posteriori* estimate error covariance respectively.

In the prediction phase, the *a priori* state estimate x_t^{t-1} and the *a priori* estimate error covariance P_t^{t-1} are computed forward from time step $t-1$ to step t. In the update phase, we assume that we have measured the process to obtain y_t. Firstly, the Kalman gain K_t is computed. Then the *a posteriori* state estimate x_t^t and the *a posteriori* estimate error covariance P_t^t are calculated. The initial values are $x_1^0 = \pi_1$ and $P_1^0 = P_1$.

C. Combination of GARCH and Kalman filter

In a fixed GARCH model, its parameters are estimated using the training set only, and the test set is not used to adjust parameters. This constraint may reduce the forecast accuracy, especially in predicting non-stationary data. To overcome this, the KF can be used to update parameters of a GARCH model by treating the weights as the states of an unforced linear dynamical system. This can be considered as an estimation problem where the weight values are unknown and have to be estimated. A little noise is added to the existing state estimate. Denote the vector of GARCH parameters which need to be updated by θ, then the evolution equation of GARCH parameters is given by

$$
\theta_t = \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, Q).
$$

The other parameters of the Kalman filter can be derived from the GARCH model. We consider two solutions of combination of GARCH and Kalman Filter as follows.

1) Solution 1: In this solution, all parameters of GARCH model are set on training set, with exception of the bias β_0 which is adjusted on-line on the test set. Denote $\delta = \{\alpha_0, ..., \alpha_m, \gamma_1, ..., \gamma_r\}$

- **On the training set**: we use maximum likelihood to compute GARCH parameters, denoted $\beta(0)$, $\delta(0)$.
- **On the test set**: two steps are recursively repeated.

Step 1: Update parameters of the GARCH model using the KF. We fixed the value $\delta(t) = \delta(0)$ and $\widehat{\beta}(t) = \widehat{\beta}(0)$, and used a KF to update value of $\beta_0(t)$. The KF is given by

$$
\begin{aligned}
\beta_0(t) &= \beta_0(t-1) + \omega(t), \quad \omega(t) \sim N(0, Q) \\
y_t &= \beta_0(t) + \widehat{\beta}(0)x_t + \varepsilon_t \quad \varepsilon_t \sim N(0, n_t) \\
n_t &= \alpha_0(0) + \sum_{i=1}^{m} \alpha_i(0)\varepsilon_{t-i}^2 + \sum_{j=1}^{r} \gamma_j(0)n_{t-j}.
\end{aligned}
$$

Parameters $\Omega = \{Q, P_1\}$ of the Kalman Filter can be estimated by using maximum log likelihood (using the Kalman smoother) [21], or just initialised to relatively small values. Other parameters of KF are given by

$$
\begin{aligned}
\pi_1 &= \beta_0(0) \\
F_t &= I \quad \text{(i.e. identity matrix)} \\
H_t &= 1 \\
R_t &= \alpha_0(0) + \sum_{i=1}^{m} \alpha_i(0)\varepsilon_{t-i}^2 + \sum_{j=1}^{r} \gamma_j(0)R_{t-j}.
\end{aligned}
$$

Step 2: Predict next-day price

$$
y_t = \beta_0(t) + \widehat{\beta}(0)x_t.
$$

2) Solution 2: In this solution, all parameters of GARCH model are set on training set, with exception of $\beta = \{\beta_0, \widehat{\beta}\}$ which is adjusted on-line on the test set. Denote $\delta = \{\alpha_0, ..., \alpha_m, \gamma_1, ..., \gamma_r\}$

- **On the training set**: we use maximum likelihood to compute GARCH parameters, denoted $\beta(0)$, $\delta(0)$.
- **On the test set**: two steps are recursively repeated.

Step 1: Update parameters of the GARCH model using the KF. We fixed the value $\delta(t) = \delta(0)$, and used a KF to update value of $\beta(t)$. The KF is given by

$$
\begin{aligned}
\beta(t) &= \beta(t-1) + \omega(t), \quad \omega(t) \sim N(0, Q) \\
\beta(t) &= \{\beta_0(t), \widehat{\beta}(t)\} \\
y_t &= \beta_0(t) + \widehat{\beta}(t)x_t + \varepsilon_t \quad \varepsilon_t \sim N(0, n_t) \\
n_t &= \alpha_0(0) + \sum_{i=1}^{m} \alpha_i(0)\varepsilon_{t-i}^2 + \sum_{j=1}^{r} \gamma_j(0)n_{t-j}.
\end{aligned}
$$

Parameters $\Omega = \{Q, P_1\}$ of the Kalman Filter can be estimated by using maximum log likelihood (using Kalman smoother) [21], or just initialised to relatively small values. Other parameters of KF are given by

$$
\begin{aligned}
\pi_1 &= \beta(0) \\
F_t &= I \quad \text{(i.e. identity matrix)} \\
H_t &= [1, x_t] \\
R_t &= \alpha_0(0) + \sum_{i=1}^{m} \alpha_i(0)\varepsilon_{t-i}^2 + \sum_{j=1}^{r} \gamma_j(0)R_{t-j}.
\end{aligned}
$$

Step 2: Predict next-day price

$$
y_t = \beta_0(t) + \widehat{\beta}(t)x_t.
$$

IV. EXPERIMENTAL RESULTS

A. Data

Two datasets are used in this work: (1) daily price of monthly electricity baseload forward products and (2) daily price of monthly gas forward products. They are taken from the UK energy market, provided by E.ON UK plc. The monthly electricity/gas product is a forward contract for supplying electricity/gas in a single month in the future. In the UK energy market, it is possible to trade gas from one to six month(s) ahead and to trade electricity from one to four month(s) ahead. There is four/six months of daily price data (approximately 90/130 data points) for each monthly electricity/gas product. For example, the July 2006 gas product
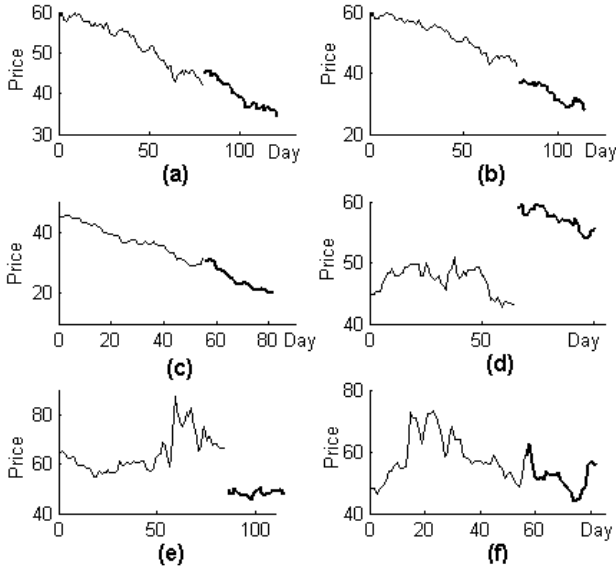
Fig. 2. Six sub-datasets (1-6) of the ELECTRICITY price dataset. The thin lines are the training sets and the bold lines are the test sets.

Fig. 3. Six sub-datasets (7-12) of the GAS price dataset. The thin lines are the training sets and the bold lines are the test sets.

can be traded from 03 Jan 2006 to 30 Jun 2006. To illustrate the behaviour of the proposed methods, six sub-datasets for each datasets are used. Figure 2 and 3 show six sub-dataset of the electricity dataset (denote sub-datset 1 to 6) and gas dataset (denote sub-datset 7 to 12) respectively. The test sets of the first three sub-datasets of each dataset correspond to the beginning, middle and end of a stable monthly product. The test sets of the last three sub-datasets of each dataset correspond to the beginning, middle and end of a volatile monthly product. If a test set of a sub-dataset is the beginning or middle of a monthly gas product trading period, the training set are data of another monthly product. For example, in sub-dataset 8 in the gas price dataset (Figure 3 (b)), the test set is the middle samples of March 2007 product and the training set is November 2006 product. The prices for different products are different even in the same trading day; for example, the price of a colder month is normally higher than that of a warmer month. This is the reason why there is gap between the training sets and test sets of these sub-datasets in Figure 2 and 3.

We have randomly selected the sub-datasets from tables of price of monthly electricity/gas forward products. The input of the forecast models include historical price of monthly products, seasonal products and weekend ahead products (i.e. contracts for the next Saturday/Sunday.). In this paper, the GARCH models with $m = 1$ and $r = 1$ are used.

B. Model evaluation

The random walk (RW) is used as a benchmark model. This model is given by:

$$y(t + 1) = y(t) + \varepsilon(t),$$

where ε is a zero-mean noise. The model predicts that the next value of the time series is the same as the current value.
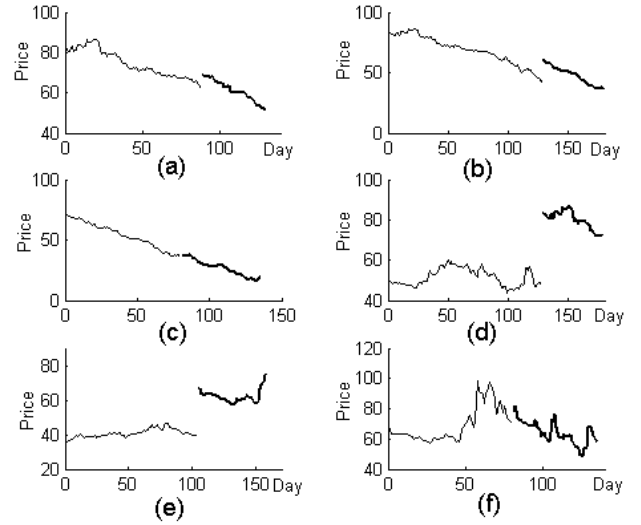
To evaluate the prediction performance of these models, three types of prediction errors of the test sets are computed: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percent Error (MAPE).

$$e_{MSE} = \frac{\sum_1^T \left(y_{\mathrm{real}} - y_{forecast}\right)^2}{\sum_1^T \left(y_{\mathrm{real}} - \overline{y}\right)^2}$$

$$e_{RMSE} = \frac{\sqrt{\sum_1^T \left(y_{\mathrm{real}} - y_{forecast}\right)^2}}{\sqrt{\sum_1^T \left(y_{\mathrm{real}} - \overline{y}\right)^2}}$$

$$e_{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{y_{\mathrm{real}}(t) - y_{forecast}(t)}{y_{\mathrm{real}}(t)} \right|$$

where y_{real} is the real gas price, $y_{forecast}$ is the forecast gas price, $\overline{y}$ is the mean of y_{real}, and T is the number of data points.

We also compute the improvement ratio (IR) of errors of a method comparing to errors of the RW model. Errors here may be MSE, RMSE or MAPE. For example, the IR of MSE of a model M comparing to MSE of the RW is given by:

$$IR_{MSE}(M) = \frac{e_{MSE}(RW) - e_{MSE}(M)}{e_{MSE}(RW)} \times 100\%$$

C. Results

The IR_{MSE}, MSE, MAPE, and RMSE are computed for each sub-dataset and for prediction methods. Their averaged values for the six sub-datasets of electricity and gas dataset of these methods are shown in Table I and Table II respectively. For the purpose of comparison, these errors are computed for four methods, as the follows

- Random walk model, which is used as the bench mark model.
- Fixed GARCH model, whose parameters are computed on the training set.

- Adaptive-β_0 GARCH model, whose all parameters are set on training set, with exception of the bias β_0 which is adjusted on-line on the test set.
- Adaptive-β GARCH model, whose all parameters are set on training set, with exception of β which is adjusted on-line on the test set.

It can be noticed that the errors of adaptive models are reduced in comparison with those of the fixed GARCH model. Therefore, the adaptive models do improve the performance of the GARCH model in these datasets. This improvement proves the usefulness of updating parameters using a Kalman filter. MSE of adaptive GARCH models are 0.2155 and 0.0945 for electricity price dataset and gas price dataset respectively, their MSE improves around 15.9% compared to the MSE of the random walk model. There is not much difference between the results of adaptive-β_0 GARCH and those of adaptive-β GARCH.

TABLE I

AVERAGE ERRORS FOR THE SIX SUB-DATASETS IN THE ELECTRICITY DATASET OF THE PREDICTION METHODS.

Model	IR_{MSE} (%)	MSE	MAPE	RMSE
RW	0.00	0.25623	0.01767	0.43623
Fixed GARCH	8.09	0.23551	0.01654	0.41007
Adaptive-β_0 GARCH	15.89	0.21551	0.01643	0.39828
Adaptive-β GARCH	15.88	0.21554	0.01643	0.39832

TABLE II

AVERAGE ERRORS FOR THE SIX SUB-DATASETS IN THE GAS DATASET OF THE PREDICTION METHODS.

Model	IR_{MSE} (%)	MSE	MAPE	RMSE
RW	0.00	0.11238	0.02168	0.29658
Fixed GARCH	12.41	0.09843	0.01983	0.27508
Adaptive-β_0 GARCH	15.90	0.09451	0.01991	0.27053
Adaptive-β GARCH	15.91	0.09450	0.01990	0.27053

Table III, IV, V and VI show the MSE and Improvement ratio (IR) of these forecast models for sub-datasets. It can be seen that the adaptive GARCH models reduce the prediction errors in comparison to fixed GARCH model in case of volatility data (i.e sub-datasets 4, 5, 6, 10 and 12). For example, in the results for sub-dataset 5, the MSE of the adaptive-β_0 GARCH model and the adaptive-β GARCH model improved 15.76% compared to these in the random walk model while the improvement ratio for the fix GARCH model is only 3.42%. However applying the filter has not much effect on stationary data (i.e. sub-datasets 1, 2, 3, 7, 8 and 9).

The adaptive models were also applied for gas quarterly products and achieved comparable but slightly worse results.

V. CONCLUSION

This paper proposes a new forecast technique which is a combination of GARCH and Kalman filter models. Testing on the UK electricity/gas forward price shows that the GARCH models with online adaptive parameters outperform the normal

TABLE III

MSE OF SIX SUB-DATASETS IN THE ELECTRICITY PRICE DATASET USING DIFFERENT PREDICTION MODELS.

MSE	RW	Fixed GARCH	Adaptive-β_0	Adaptive-β
Sd 1	0.05040	0.04240	0.04206	0.04206
Sd 2	0.09456	0.08856	0.08792	0.08794
Sd 3	0.05300	0.03001	0.02969	0.02970
Sd 4	0.12765	0.12698	0.12550	0.12554
Sd 5	0.92351	0.89189	0.77794	0.77796
Sd 6	0.28827	0.23322	0.22995	0.23003
Average	0.25623	0.23551	0.21551	0.21554

Sd: Sub-dataset
Adaptive-β_0: Adaptive-β_0 GARCH
Adaptive-β: Adaptive-β GARCH

TABLE IV

IMPROVEMENT RATIO OF MSE OF SIX SUB-DATASETS IN THE ELECTRICITY PRICE DATASET USING DIFFERENT PREDICTION MODELS.

IR_{MSE} (%)	Fixed GARCH	Adaptive-β_0	Adaptive-β
Sd 1	15.86	16.55	16.54
Sd 2	6.34	7.02	7.00
Sd 3	43.38	43.97	43.95
Sd 4	0.52	1.68	1.65
Sd 5	3.42	15.76	15.76
Sd 6	19.10	20.23	20.20
Average	8.09	15.89	15.88

Sd: Sub-dataset
Adaptive-β_0: Adaptive-β_0 GARCH
Adaptive-β: Adaptive-β GARCH

GARCH model on volatile data. The framework of this combination model of GARCH can be applied to other prediction models, such as machine learning or financial models.

This paper forecasts daily electricity/gas prices for the monthly forward product as an example to verify the performance of the proposed method. It can be used for forecasting other types of forward products, market clearing prices, and electricity/gas demand forecasts.

TABLE V

MSE OF SIX SUB-DATASETS IN THE GAS PRICE DATASET USING

DIFFERENT PREDICTION MODELS.

MSE	RW	Fixed GARCH	Adaptive-β_0	Adaptive-β
Sd 7	0.05584	0.03580	0.03671	0.03671
Sd 8	0.02649	0.01991	0.02000	0.02000
Sd 9	0.03155	0.02973	0.02961	0.02961
Sd 10	0.05374	0.05582	0.04880	0.04880
Sd 11	0.12435	0.10596	0.10706	0.10706
Sd 12	0.38232	0.34338	0.32488	0.32488
Average	0.11238	0.09843	0.09451	0.09450
Sd: Sub-dataset				
Adaptive-β_0: Adaptive-β_0 GARCH				
Adaptive-β: Adaptive-β GARCH				

TABLE VI

IMPROVEMENT RATIO OF MSE OF SIX SUB-DATASETS IN THE GAS

PRICE DATASET USING DIFFERENT PREDICTION MODELS.

IR_{MSE} (%)	Fixed GARCH	Adaptive-β_0	Adaptive-β
Sd 7	35.88	34.25	34.25
Sd 8	24.85	24.51	24.51
Sd 9	5.77	6.17	6.17
Sd 10	-3.87	9.19	9.19
Sd 11	14.79	13.90	13.90
Sd 12	10.19	15.02	15.02
Average	12.41	15.90	15.91
Sd: Sub-dataset			
Adaptive-β_0: Adaptive-β_0 GARCH			
Adaptive-β: Adaptive-β GARCH			

REFERENCES

[1] Stevenson, M., "Filtering and Forecasting Spot Electricity Price in the Increasingly Deregulated Australian Electricity Market," *The International Institute of Forecaster Conference*, Atlanta, 2001.

[2] Nogales, F., Contreras J., Conejo, A, and Espinola, R., "Forecasting next-day electricity price by time series models," *IEEE Transaction in Power System*, vol. 2, pp. 342-348, 2002.

[3] Contreras, J., Espinola, R., Nogales, F., and Conejo, A., "ARIMA model to predict next-day electricity price," *IEEE Transaction on Power Systems*, 18(3), pp. 1014-1020, 2003.

[4] Conejo, A. J., Plazas, M. A., Espinola, R., and Molina, A. B., "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Transaction in Power System*, vol. 20, pp. 1035-1042, 2005.

[5] Hua, Z., Li, X., and Li-zi, Z., "Electricity price forecasting based on GARCH model in deregulated market," *Power Engineering Conference, The 7th International*, pp. -410, 2005.

[6] Garcia, R., Contreras, J., Van Akkeren, M., and Garcia, J., "A GARCH forecasting model to predict day-ahead electricity prices," *IEEE Transaction in Power System*, vol. 2, pp. 867-874, 2005.

[7] Gao, F., Guan, X., Cao, X., and Papalexopoulos, A, "Forecasting power market clearing price and quantity using neural network," *Proc. Power Engineering Summer Meet*, vol. 4, pp. 2183-2188, 2000.

[8] Guo, J. J. and Luh, P. B., "Improving market clearing price prediction by using committee machine of neural networks," *IEEE Transactions on Power Systems*, 19(4), pp. 1867-1876, 2004.

[9] Zhang, L., Luh, P. and Kasiviswanathan, K., "Energy clearing price prediction and confidence interval estimation with cascaded neural networks," *IEEE Transaction on Power System*, vol. 1, pp. 99-105, 2003.

[10] Gonzalez, A., Roque, A., and Garcia, G. L., "Modelling and forecasting electricity price with input/output hidden Markov models," *IEEE Transactions on Power Systems*, 20(1), pp. 13-24, 2005.

[11] Niranjan, M. "On data driven of option price using neural network," *Forecasting Financial Markets: Advances for Exchange Rate, Interest Rate and Asset Management*, pp. 1-13, 1999.

[12] Nabney, I. T., McLachlan, A., and Lowe, D. "Practical methods of tracking non-stationary time series applied to real world data," *In*

AeroSense '96 : Applications and Science of Artificial Neural Networks, no. 2760. Invited talk, 1996.

[13] Andreou, A. S., Georgopoulos, E. F., and Likothanassis, S. D. "Exchange-rates forecasting: a hybrid algorithm based on genetically optimized adaptive neural networks," *Computational Economics*, vol. 20, 191-210, 2002.

[14] Anne, K. R., Kyamakya, K., Erbas, F., Takenga, C., and Chedjou, J.C. "GSM RSSI-based positioning using extended Kalman filter for training artificial neural networks," *Vehicular Technology Conference, 2004, VTC2004-Fall, 2004 IEEE 30th*, vol. 6, pp. 4141-4145, 2004.

[15] Li, S., Wunsch, D. C., O'Hair, E., and Giesselmann, M. G., "Wind Turbine Power Estimation by Neural Network with Kalman Filter Training on SIMD Parallel Machine," *International Joint Conference on Neural Networks*, vol. 5, pp. 3430-3434, 1999.

[16] Zhang, L. and Luh, P. B., "Power market clearing price prediction and confidence interval estimation with fast neural network learning," *Proc. IEEE 2002 Power Engineering Society Winter Meeting*, vol. 1, pp. 268-273 , 2002.

[17] Patil, S., Srinivasan, S., Prasad, S., Irwin, R., Lazarou, G., and Picone, J., "Sequential state-space filters for speech enhancement," *Proceedings of the IEEE Southeast Con.*, pp. 240-243, 2006.

[18] Labarre, D., Grivel, E., Berthoumieu, Y., Todini, E. and Najim, M. "Consistent estimation of autoregressive parameters from noisy observations based on two interacting Kalman filters," *Signal Processing*, 86(10), pp. 2863-2876, 2006.

[19] T. Bollerslev, "Generalized Autoregressive Conditional Heteroskedasticity, " *Journal of Econometrics*, vol. 31, pp. 307-327, 1986.

[20] Kalman, R.E. "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME - Journal of Basic Engineering*, vol. 82, pp.35-45, 1960.

[21] Ghahramani, Z. and Hilton, G. "Parameter estimation for linear dynamical system," University of Toronto, 1996.