

## Federal Funds Rate Prediction: A Comparison Between the Robust RBF Neural Network and Economic Models\*

CHIEN-CHENG LEE<sup>1</sup>, CHUN-LI TSAI<sup>2</sup> AND YU-CHUN CHIANG<sup>3</sup>

<sup>1</sup>*Department of Communications Engineering*

<sup>3</sup>*Department of Mechanical Engineering and Yuan Ze Fuel Cell Center*

*Yuan Ze University*

*Chungli, 320 Taiwan*

*E-mail: {cclee; ycchiang}@saturn.yzu.edu.tw*

<sup>2</sup>*Department of Economics*

*National Cheng Kung University*

*Tainan, 701 Taiwan*

*E-mail: tchunli@mail.ncku.edu.tw*

Neural network forecasting models have been widely used in the analyses of financial time series during the last decade. This paper attempts to fill this gap in the literature by examining a variety of univariate and multivariate, linear, nonlinear Economics empirical modes and neural network. In this paper, we construct an M-estimator based RBF (MRBF) neural network with growing and pruning techniques. Then we compare the forecasting performances of MRBF with six other time-series forecasting models for daily U.S. effective federal funds rate. The results show that the proposed MRBF network can produce the lowest root mean square errors in one-day-ahead forecasting for the federal funds rate. It implies that MRBF can be one good method for the predictions of some financial time series data.

**Keywords:** federal funds rate, time series prediction, M-estimator, RBF, growing and pruning

### 1. INTRODUCTION

Since interest rate is an important factor to influence the financial market, the prediction of interest rate has been a popular issue in Economics field. The U.S. Fed always implements monetary policy by targeting the effective Federal Funds rate. The effective federal funds rate fluctuates day by day because each day the Trading Desk of the Federal Reserve Bank of New York conducts open market operations designed to move the federal funds rate. Thus, many papers about prediction focus on the high-frequency behavior of the federal funds rate.

The Economics literature found some important variables can provide the prediction power of federal funds rate movements. Krueger and Kuttner [1] found the federal funds rate movement can be explained by the decomposition of the federal funds future rates. Taylor [2] and Rudebusch [3] showed that the federal funds rate target announced by Federal Open Market Committee (FOMC) can provide some prediction information about the forecast of funds rate. Sarno and Thornton [4], Hansen and Seo [5] showed that term structure of interest rate is related to the federal funds rate. Since the prediction of

Received February 1, 2008; accepted November 28, 2008.

Communicated by Yau-Hwang Kuo, Pau-Choo Chung and Jar-Ferr Yang.

\* This work was also supported by the National Science Council of Taiwan, R.O.C. under project No. NSC 95-2415-H-006-005.

funds rate can help financial market hedge the risks on the investments, a range of time-series models in Economics literature are used to forecast the funds rate. Sarno *et al.* [6] did the consensus to examine a variety of univariate and multivariate, linear and nonlinear empirical prediction models of the federal funds rate. They examined the forecasting performance of daily federal funds rate by using both univariate and multivariate, linear and nonlinear models, which are all tried to capture the unknown process that drives federal funds rate movements.

A neural network always can perform well on the forecast because it can be regarded as a universal approximator and possesses self-learning capability. Neural network forecasting models also have been widely used in the analyses of financial time series during the last decade. To date, however, there appears to be few paper on the comparison for the prediction of federal funds rate between Economic prediction models and neural network models. Thus, this paper attempts to fill this gap in the literature by examining a variety of univariate and multivariate, linear, nonlinear Economics empirical models and neural network models. That is, the main goal of this paper is to examine how well artificial neural networks perform the forecasting of daily U.S. effective federal funds rate compared some economics prediction models.

Among the current neural network models, the radial basis function (RBF) neural network is considered as a good candidate for approximation and prediction due to its rapid learning capacity. It has been applied successfully to some nonlinear time series modeling and predictions [8-11]. Although the RBF networks have been used successfully in many applications, there still exist some unsolved problems. First, the Gaussian function is typically selected as the activation function in the RBF networks. A network iteratively adjusts parameters for each node by minimizing the least-mean square (LMS) criterion according to gradient descent algorithm. Nevertheless, when some of the training patterns contain large errors caused by the presence of noise, the network will yield inadequate responses in the neighborhood of the noise due to the LMS criterion. Therefore, the training results of the network might not be good since all training patterns must be interpolated by including the noisy ones. That is, the network is not robust as input training patterns contain noise.

This robustness can be achieved by replacing the MSE criterion with M-estimator, in the case of traditional neural networks, and it has been also applied in many fields [12-15]. In our previous work [16, 17], a robust objective function derived from Hampel's M-estimator was adopted to develop the robust RBF network. Chen *et al.* [12] have used the M-estimator to deal with the noise and develop a fast template matching algorithm. Four different robust estimators (Huber estimator, Tukey estimator, Geman-McClure estimator, and the trimmed mean M-estimator) were used in their experiment to compare the recognition and speedup performances of template matching. Hong *et al.* [13] have introduced a new orthogonal forward regression (OFR) model identification algorithm, in which it uses D-optimality for modeling the structure selection. It is based on an M-estimators of parameter estimates. The proposed approach extended the OFR algorithm on the parsimonious model structure determination, and therefore can deal with noisy data conditions via the derivation of parameter M-estimators with inherent robustness to noise. Chuang *et al.* [14] have adopted the concepts of M-estimator and annealing to develop a feedforward network with robust backpropagation learning algorithm. In addition, some of the robust methods used for calibration in analytical chemistry were reviewed in [15],

such as Huber M-estimator, Tukey estimator, Welsh GM-estimators, fuzzy estimators, *etc.*

On the other hand, the second crucial problem of training the RBF networks is the difficulty of determining the number of hidden nodes in the network. If the number of hidden nodes is underestimated, the capability of the network is limited and the performance therefore might be degraded. However, if the number of hidden nodes is over-determined, a large network is generated and that may lead to increase training time. In the past literature, several methods used to determine the number of hidden nodes in RBF networks have been proposed. Chen *et al.* [22-25] first proposed an orthogonal least squares (OLS) learning algorithm to choose RBF centers. Later, they also combined the genetic algorithm and regularized orthogonal least squares learning to improve the network generalization performance. Huang *et al.* [11, 21] proposed a concept of significance of a neuron, which differed from other methods. This method in Huang *et al.* [11, 21] was also much simpler. The significance is defined as a neuron's statistical contribution to the overall performances of the network, and it is used in growing and pruning strategies. A new neuron is added only if its significance is larger than the chosen threshold. Conversely, if the significance for a neuron is less than the threshold, the neuron is pruned. In this paper, we define the network growing and pruning algorithm by adopting this concept of the significance of a neuron. According to this approach, the network can produce a proper size in a given problem.

While developing the M-estimator based RBF (MRBF) neural network, it can resolve the traditional RBF problems by the following two advantages. The first advantage is the robustness for noise. The second one is the determination of the hidden-node number. In this paper, we use the Welsh M-estimator function [18-20] as the RBF error function, and adopt the ideas of neuron significance [21] to implement the network growing and pruning algorithm. Through the combination of M-estimator with growing and pruning techniques, the MRBF network not only can achieve the optimal network topology but also can have the capability of resisting the outliers.

The federal funds rate is the interest rate at which a depository institution lends immediately available funds to another depository institution overnight. The rate may vary from bank to bank and from day to day. Theoretically, the rate should be determined by the open market. Interest rates tend to increase with an increase in the rate of borrowing and when the economy is flowing. In the other hand, interest rates decrease in situations when economy is not going too well. In order to maintain the stable economy, the FOMC always set a target for this rate to keep the federal funds rate be around its target level. However, government policy, terrorist incident, or some other nonblank financial panics also affect the movements of interest rate. Thus, it has been a difficult task to predict the fluctuations of federal funds rate in a complicated U.S. economy.

In this study, we treat the sudden changes of interest rate as outliers in MRBF network. We compare the forecasting performances of MRBF network and other six economic time-series models for daily U.S. effective federal funds rate. These six economic empirical models include random walk model, autoregressive conditional heteroskedasticity model, Markov-switching  $p$ th-order autoregressive model with M-regimes [32], momentum-threshold autoregressive, bivariate TAR model, and Markov-switching vector error correction model. The experimental results show that the proposed MRBF network model can achieve a better prediction performance for the U.S. daily funds rate.

The rest of this paper is organized as follows. Section 2 introduces MRBF network, section 3 describes other time-series forecasting models. Section 4 describes the data and reports the experimental results. Section 5 concludes.

## 2. M-ESTIMATOR BASED ROBUST RBF NEURAL NETWORK

The basic architecture of an RBF network is a single hidden layer feed forward neural network, as shown in Fig. 1. The output of the RBF network is described by

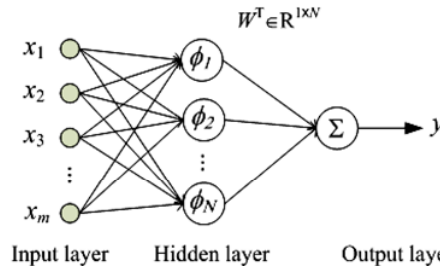


Fig. 1. Basic architecture of RBF neural network.

$$y = f(\mathbf{x}) = \sum_{k=1}^N w_k \phi_k(\|\mathbf{x} - \mathbf{c}_k\|, \sigma_k) \quad (1)$$

where  $y$  is the actual network output,  $\mathbf{x} \in R^{m \times 1}$  is an input vector signal, with individual vector components given as  $x_j$ , for  $j = 1, 2, \dots, m$ , that is,  $\mathbf{x} = [x_1, x_2, \dots, x_m]^T \in R^{m \times 1}$ .  $\mathbf{w} = [w_1, w_2, \dots, w_N]^T \in R^{N \times 1}$  is the vector of the weights in the output layer,  $N$  is the number of neurons in the hidden layer, and  $\phi_k(\cdot)$  is the basis function of the network from  $R^{m \times 1}$  to  $R$ .  $\mathbf{c}_k = [c_{k1}, c_{k2}, \dots, c_{km}]^T \in R^{m \times 1}$  is called the center vector of the  $k$ th node,  $k$  is the bandwidth of the basis function  $\phi_k(\cdot)$ , and  $\|\cdot\|$  denotes the Euclidean distance. For each neuron in the hidden layer, the Euclidean distance between its associated center and the input to the network is computed. The output of the neuron in a hidden layer is a nonlinear function of the distance, and the Gaussian function is most widely selected as the nonlinear basis function. After the computation of the output for each neuron, the output of the network is computed as a weighted sum of the hidden layer outputs.

In the training procedure, the steepest gradient of descent learning process is used to adjust the appropriate settings of the parameters (e.g. weights, centers, and bandwidths), which make the performance of the network mapping optimized. A common optimization criterion is used to minimize the LMS between the actual and desired network outputs. LMS error function is defined as

$$\rho(r_n) = \frac{1}{2} r_n^2 \quad (2)$$

where  $r_n = d(n) - y(n)$  represents the residual error between the desired,  $d(n)$ , and the actual network outputs,  $y(n)$ .  $n$  indicates the index of the series.

The cost function can be defined as an ensemble average error,

$$J(\theta) = E[\rho(r_n)] \tag{3}$$

where  $\theta$  is one of the parameter sets of the network.

According to the gradient descent method, the update equation for the network parameters is given by

$$\theta(n+1) = \theta(n) - \mu_\theta \frac{\partial}{\partial \theta} J(\theta) \approx \theta(n) + \mu_\theta r_n \frac{\partial y}{\partial \theta}. \tag{4}$$

We can define the influence function as

$$\psi(r_n) = \frac{\partial \rho(r_n)}{\partial r_n}, \tag{5}$$

and rewrite Eq. (4) to obtain the generalized update equation as

$$\theta(n+1) = \theta(n) + \mu_\theta \psi(r_n) \frac{\partial y}{\partial \theta}. \tag{6}$$

Recalling Eq. (6), the network updates are proportional to the linear influence function  $\psi(r_n)$ . It would offer the key to an understanding of overcoming the outlier problem. One possible solution for improving this problem is to employ a robust criterion instead of LMS. Among several methods, which deal with the noise problem, M-estimator technique [18-20] is the most robust and has been applied in many applications. The M-estimator uses some cost functions which increase less rapidly than that of least square estimators as the residual departs from zero. When the residual error increases over a threshold, the M-estimator suppresses the response instead. Therefore, the M-estimator based error function is more robust to noise than LMS based error function.

Several M-estimators have been studied in the past. Some commonly used are Huber, Cauchy, Geman-McClure, Welsch, and Tukey. In this paper, we employ the Welsch function as the error function, given by

$$\rho_W(r_n) = \frac{\alpha^2}{2} [1 - \exp(-(r_n/\alpha)^2)] \tag{7}$$

where  $\alpha$  is a scale parameter. The corresponding influence function is given by

$$\psi_W(r_n) = \frac{d\rho_W(r_n)}{dr_n} = r_n \exp(-(r_n/\alpha)^2), \tag{8}$$

and the error function and influence function are graphically depicted in Fig. 2. Observing Fig. 2, we can find that Welsch functions reduce the effects of the large errors.

The update Eq. (6) can also be rewritten as

$$\theta(n+1) \approx \theta(n) + \mu_\theta r_n \exp(-(r_n/\alpha)^2) \frac{\partial y}{\partial \theta}. \tag{9}$$

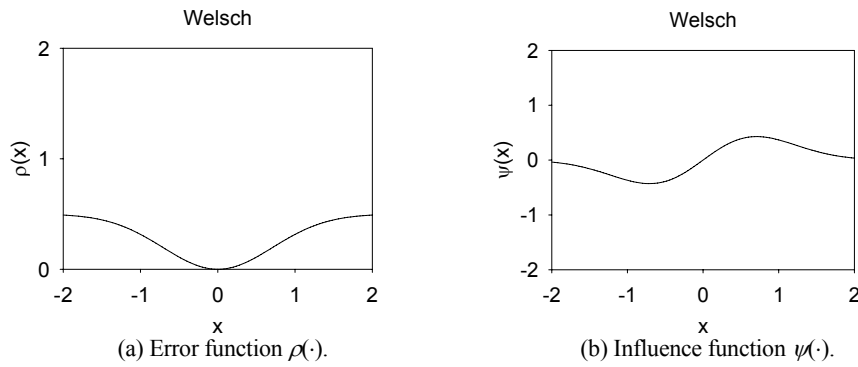


Fig. 2. Graphic representations of Welsch M-estimators.

To find the confidence interval, we let  $d\psi_w(r_n)/dr_n = 0$ , given by

$$\frac{\partial \psi_w(r_n)}{\partial r_n} = (1 - 2(r_n/\alpha)^2) \exp(-(r_n/\alpha)^2) = 0. \tag{10}$$

Obviously, the extreme points are  $\pm 2^{-0.5}\alpha$ , and the confidence interval is the range  $[-2^{-0.5}\alpha, 2^{-0.5}\alpha]$ . The interval depends on the scale  $\alpha$ . When  $\alpha$  is large, noise may not be discriminated from the majority of the data points. Conversely, if  $\alpha$  is small, some of desired data will be treated as noise. To estimate the scale, a median operator is used in this paper. The median operator is simple to understand, easy to calculate, and more robust in the presence of noise values than the mean value. It is given by

$$\alpha = \beta(k) \operatorname{median}_{\substack{i=n+\lfloor \frac{s}{2} \rfloor \\ i=n-\lfloor \frac{s}{2} \rfloor}} (|r_i|) \tag{11}$$

where

$$\beta(k) = \beta_0 \exp\left(-\eta \frac{k}{TotalIteration}\right) \tag{12}$$

$s$  indicates the number of values in the data set of median,  $k$  indicates the iteration index,  $TotalIteration$  is the number of the training epochs, and  $\beta$  is a real number, used to control the confidence interval.

Another major challenge in this design of the robust RBF neural network is to determine the number of the centers. The performance of the network is critically dependent on the number of the centers. If an insufficient number of centers is selected, the capability of the network is limited. The performance therefore might be degraded. On the other hand, if the number of centers is over-determined, this produces a larger network and increases the time for the network process.

In the past, various methods for growing and pruning in RBF networks have been proposed [21, 26-30]. Huang *et al.* [21] have proposed the concept of significance of a

neuron, which is wholly different from and much simpler than other methods. The significance is defined as a neuron’s statistical contribution to the overall performance of the network. It can be used in growing and pruning strategies. A new neuron is added only if its significance is larger than the chosen threshold. Conversely, if the significance for a neuron becomes less than the threshold, then that neuron is pruned. In this paper, we define the network growing and pruning algorithm by adopting the concept of the significance of a neuron.

(1) *Pruning Node*: To define the significance of a neuron for pruning (SNP), we assume the output of a RBF network with  $N$  neurons for an input  $\mathbf{x}$  is given by Eq. (1). If the neuron  $q$  is removed, the output of the RBF network with the remaining  $N - 1$  neurons is

$$y_q = \sum_{k=1}^{q-1} w_k \phi_k(\|\mathbf{x} - \mathbf{c}_k\|, \sigma_k) + \sum_{k=q+1}^N w_k \phi_k(\|\mathbf{x} - \mathbf{c}_k\|, \sigma_k). \tag{13}$$

Therefore, for an input  $\mathbf{x}_i$ , the error caused from removing neuron  $q$  is the absolute difference between  $y$  and  $y_q$ , that is

$$PErr(i, q) = |y - y_q| = |w_q| \phi_q(\|\mathbf{x}_i - \mathbf{c}_q\|, \sigma_q). \tag{14}$$

The significance of a neuron for pruning is defined as the average error for all  $M$  sequentially learned inputs by removing neuron  $q$ , given by

$$SNP(q) = \frac{\sum_{i=1}^M PErr(i, q)}{M} = \frac{|w_q|}{M} \sum_{i=1}^M \phi_q(\|\mathbf{x}_i - \mathbf{c}_q\|, \sigma_q). \tag{15}$$

If  $SNP(q) < TPErr$  ( $TPErr$  is a predefined threshold value), it means the neuron  $q$  does not make significant contribution to the overall performance of the network. Hence, this neuron should be removed.

(2) *Growing Node*: The definition of the significance of a neuron in the process of growing node is similar to the one for pruning. For the newly added neuron  $N + 1$ , the output of the RBF network with the  $N + 1$  neurons is

$$y_{N+1} = \sum_{k=1}^{N+1} w_k \phi_k(\|\mathbf{x} - \mathbf{c}_k\|, \sigma_k). \tag{16}$$

Therefore, for a given input  $\mathbf{x}_i$ , the error produced from the newly added neuron  $N + 1$  is the absolute difference between  $y$  and  $y_{N+1}$ . That is,

$$GErr(i, N + 1) = |y - y_{N+1}| = |w_{N+1}| \phi_{N+1}(\|\mathbf{x}_i - \mathbf{c}_{N+1}\|, \sigma_{N+1}). \tag{17}$$

The significance of a neuron for growing (SNG) is defined as the average error for all  $M$  sequentially learned inputs by the newly added neuron  $N + 1$ , given by

$$SNG(N+1) = \frac{\sum_{i=1}^M GErr(i, N+1)}{M} = \frac{|w_{N+1}|}{M} \sum_{i=1}^M \phi_{N+1}(\|x_i - c_{N+1}\|, \sigma_{N+1}). \quad (18)$$

The parameters associated with the newly neuron  $N+1$  are taken as follows:

$$\begin{aligned} w_{N+1} &= -RMSE / \sum_{i=1}^M \phi_{N+1}(\|x_i - c_{N+1}\|, \sigma_{N+1}) \\ c_{N+1} &= \{x_i \mid \arg \max_i(|\rho(r_i)|)\} \\ \sigma_{N+1} &= \kappa \|x_i - c_{nr}\| \end{aligned} \quad (19)$$

where  $c_{nr}$  is the center which is nearest to  $x_i$ , and  $\kappa$  is a real number,  $\kappa = 0.95$  in this paper.

If  $SNG(N+1) > TGErr$  ( $TGErr$  is a predefined threshold value), it means the neuron  $N+1$  makes significant contribution to the overall performance of the network; hence, this newly neuron should be added into the network. Besides, to control the amount of the newly added neuron, the newly added neuron should not be too closed to the existing neurons. In other words, a newly added neuron must be far away from the existing neurons by at least  $TDist$  ( $TDist$  is a predefined threshold value). That is,  $\min_{k=1..N} (|c_{N+1} - c_k|) > TDist$ .

To avoid network stability problems, frequently caused by the growing and pruning, the check-period of the growing and pruning algorithm is generated as  $T = 2 * \text{Number of Neurons}$ . Initially, the number of neurons increases faster to cope with the complexity of the problem. Afterwards, while the network is approaching the stable state, the interval  $T$  between the checking process of the growing and pruning increases for avoiding the network oscillation. In addition, the extreme node (the most or least significant one) is considered as the check candidate in each growing and pruning. In the following simulations, the parameters are set as  $TGErr = 0.0003$ ,  $TPErr = 0.02$ , and  $TDist = 0.01$ , while all input data is normalized to the range  $[0, 1]$ .

### 3. OTHER SIX TIME-SERIES FORECAST MODELS

We attempt to compare the forecasting performances of MRBF network to other six economic time-series models for daily U.S. effective federal funds rate, including random walk model, autoregressive conditional heteroskedasticity model, Markov-switching  $p$ th-order autoregressive model with M-regimes, momentum-threshold autoregressive, bivariate TAR model, and Markov-switching vector error correction model.

Although some researcher concluded that the interest rate fails to follow a random walk, most studies cannot reject the unit root hypothesis for federal funds rate. Thus, it seems reasonable to consider a simple driftless random walk (RW) model as the first time-series model

$$\Delta i_t^{FF} = u_t. \quad (20)$$

where  $\Delta i_t^{FF}$  is the daily change in the effective federal funds rate.

The second model is a univariate model of the daily federal funds rate – the exponential generalized autoregressive conditional heteroskedasticity (EGARCH) model. An EGARCH ( $p, q$ ) model for the daily change in the federal funds rate is written as

$$\Delta i_t^{FF} = \eta(i_{t-1}^{FF} - i_{t-3}^{FF}) + \sigma_t v_t, \tag{21}$$

$$\log(\sigma_t^2) = w + \sum_{j=1}^p g_j \sigma_{t-j}^2 + \sum_{s=1}^q \left[ \rho_s \left( \left| \frac{v_{t-s}}{\sigma_{t-s}} \right| - E \left| \frac{v_{t-s}}{\sigma_{t-s}} \right| \right) + \xi_s \frac{v_{t-s}}{\sigma_{t-s}} \right] \tag{22}$$

where  $(i_{t-1}^{FF} - i_{t-3}^{FF})$  is the cumulative change of the federal funds rate over the preceding two days, and  $v_t$  denotes independently and identically distributed innovations with zero mean and unit variances. We model the conditional variances  $\sigma_t^2$  by the paper of [7], which consider the asymmetric effect between future conditional variances and current federal funds rate changes, both sign and magnitude of innovations are taken into account.

The third model is a univariate nonlinear model–Markov-switching– $p$ th-order autoregressive model with  $M$ -regimes, MS-AR( $M, p$ ), written as

$$\Delta i_t^{FF} = \sum_{j=1}^p \varphi(z_t) \Delta i_{t-j}^{FF} + \sigma(z_t) \omega_t, \quad z_{t=1,2,\dots,M}. \tag{23}$$

It allows for the autoregressive structure,  $\sum_{j=1}^p \varphi_j(z_t) \Delta i_{t-j}^{FF}$ , and the variance of the error term,  $\sigma^2(z_t)$ , to be shifting over time across regimes. In this model, the variance of the innovations  $\omega_t$  is time-varying but, unlike for the EGARCH, the dynamics is governed by an unobservable variables  $z_t$  which is assumed to follow a first-order Markov chain.

The fourth model is a multivariate model–Momentum-Threshold Autoregressive (M-TAR) model. The model is defined as

$$\Delta y_t = \sum_{j=1}^{p-1} \Lambda_j \Delta y_{t-j} + I_t [\alpha_1 \beta' y_{t-1}] + (1 - I_t) [\alpha_2 \beta' y_{t-1}] + error \tag{24}$$

where  $I_t$  is an indicator,  $I_t = 1$  if  $\beta' y_{t-1} \geq 0$ ,  $I_t = 0$  if  $\beta' y_{t-1} < 0$ , and  $y_t = [i_t^{FF}, i_t^{TB}]$ ,  $i_t^{TB}$  is the 3-month Treasury Bill (TB) rate. This model explicitly takes into account the existence of asymmetries that may occur in the adjustment process along the short-end of the yield curve. Essentially, the adjustment parameters,  $\alpha_1, \alpha_2$ , differ depending on whether the slope of the short-end of the yield curve is positive or negative.

Another multivariate threshold model is bivariate TAR (BTAR) model.

$$\Delta y_t = I_t \left[ \sum_{j=1}^{p-1} \Lambda_{1j} \Delta y_{t-j} + \alpha_1 \beta' y_{t-1} \right] + (1 - I_t) \left[ \sum_{j=1}^{p-1} \Lambda_{2j} \Delta y_{t-j} + \alpha_2 \beta' y_{t-1} \right] + error \tag{25}$$

where  $y_t = [i_t^{FF}, i_t^{TB}]$ ,  $I_t = 1$  if  $\beta' y_{t-1} \geq \bar{k}$ , and  $I_t = 0$  if  $\beta' y_{t-1} < \bar{k}$ .

Although both M-TAR models and the BTAR model belong to the family of threshold autoregressive models, they differ in several respects. First, in the BTAR model the Heaviside indicator function  $I_t$  is equal to zero or unity according to whether the value of

the cointegrating residual is smaller (or larger) than a threshold  $\bar{k}$ , which must be estimated. In contrast, in the M-TAR model the threshold is assumed to be equal to zero. Second, in the BTAR model the whole set of parameters,  $\Lambda_i, \alpha_i$  is shifting over time, while in the M-TAR model only the speed of adjustment parameter, namely  $\alpha_i$  (for  $i = 1, 2$ ) is shifting over time.

The sixth model considered is a Markov-switching vector error correction model (MS-VECM) of the term structure of federal futures rates. The federal funds rate and federal futures rates are assumed to cointegrate with a cointegrating vector  $[1, -1]$ . The joint dynamics of the federal funds rate and the federal funds futures rates are described by a VECM:

$$\Delta y_t = \sum_{j=1}^{p-1} \Gamma_j(z_t) \Delta y_{t-j} + \pi(z_t) y_{t-1} + \Sigma^{1/2}(z_t) \varepsilon_t \quad (26)$$

were  $y_t = [i_t^{FF}, f_t^1, f_t^2]$ , and the long-run impact matrix  $\pi(z_t) = \alpha(z_t) \beta'$ .  $f^1, f^2$  is the 1-month and 2-month federal funds futures rates, respectively. The VECM has been generalized to a Markov-Switching framework where the parameters can shift to take into account the evidence that interest rate changes are heteroskedastic and that their distribution is well approximated by a mixture of normal distribution.

## 4. EXPERIMENTAL RESULTS

In this section, Mackey-Glass chaotic time series prediction and federal funds rate prediction are conducted to examine the performance of the proposed MRBF network. Mackey-Glass chaotic time series is regarded as a benchmark for the problem of time series prediction. In the first experiment, the prediction of Mackey-Glass time series is performed for showing the proposed method is good enough to resolve the time series prediction problem. The results indicate that the proposed MRBF outperforms other prediction models for this time series data, even when data contains noises. In the second experiment, the MRBF is applied to the prediction of federal funds rate. We compare the forecasting performances of the MRBF network and other six economic time-series models for daily U.S. effective federal funds rate. The results show that the MRBF can produce the smaller prediction errors than other models.

### 4.1 Performance Evaluation of MRBF

We examine the performance of the proposed M-estimator in the prediction of time series data. We define a chaotic Mackey-Glass differential delay function as

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \quad (27)$$

where  $\tau > 17$ . In this experiment, 1000 data points are generated with an initial condition  $x(0) = 1.2$  and  $\tau = 17$ . The first 500 data pairs of the series were used as training data, while the remaining 500 were used to validate the model identified. The networks are

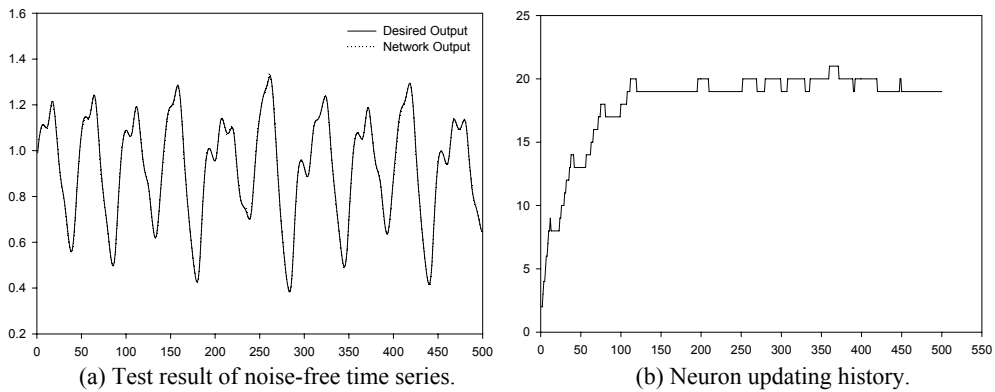


Fig. 3. Mackey-glass chaotic time series prediction.

**Table 1. Comparison results of the Mackey-glass time series prediction error of different methods.**

Method	Prediction Error (RMSE)	Epochs
Our Approach	0.005541	500
	0.003266	3000
Classical RBF (with 23 neurons) [27]	0.0114	
ANFIS and fuzzy system [28]	0.007	500
LLWNN + hybrid [29]	0.0036	3000
Neural tree [30]	0.0069	
Backpropagation NN	0.02	500
Auto Regressive Model	0.19	500
Genetic algorithm and fuzzy system (Ensemble) [31]	0.026243	500

employed to predict the values of the time series at point  $x(t + 6)$  from the four past samples  $[x(t), x(t - 6), x(t - 12), x(t - 18)]$ . Two neurons are given in the beginning of the training, and the corresponding centers are uniformly assigned from data range. The initial weights are randomly selected from  $[-0.3, 0.3]$ . The initial scale parameters are set as  $\beta_0 = 0.8, \eta = -1, s = 7$ . Fig. 3 (a) shows the test result of noise-free time series by using the proposed method. The RMSE is 0.003266 for 3000 iterations. Fig. 3 (b) shows the neuron updating history. The number of neurons dynamically increases from 2 to 19.

Table 1 reports the Mackey-Glass time series prediction performances of noise-free data among different methods including our proposed method. The data of last three rows in Table 1 is obtained from [31]. Some values are blank because these values are not available from the original articles. By the comparison results, we can find that our proposed algorithm can produce the smallest prediction error than other methods.

To further examine the advantage of the robustness in our proposed RBF neural network, different noise levels of the magnitude of the training data have been used to test the capability of resisting noise. Fig. 4 shows the results of the training phase. The corresponding RMSEs are shown in Table 2. The results show that the RMSE is 0.027035

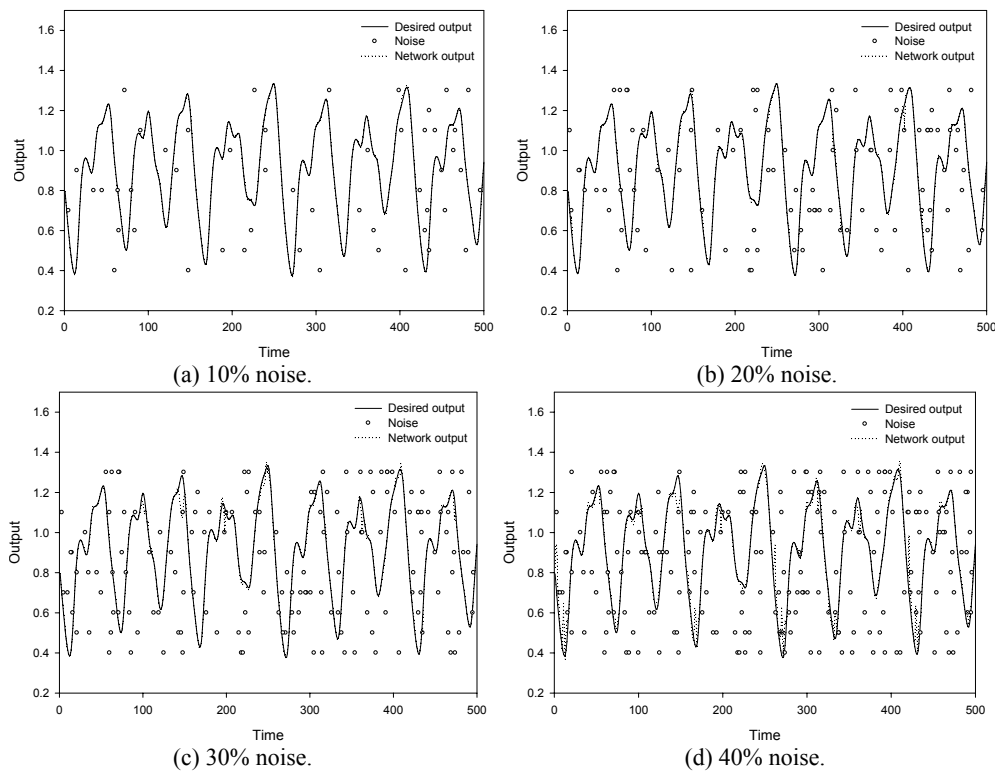


Fig. 4. Training results of different levels of uniform random noise.

**Table 2. Training results of different levels of uniform random noise.**

Level of Noise	Prediction Error (RMSE)	Number of Neurons at the Final Iteration	Iterations
10%	0.007745	22	500
20%	0.013488	22	500
30%	0.027035	22	500
40%	0.047285	28	500

in the case of 30% of the magnitude of the training data. It is nearly similar with the result of genetic algorithm and fuzzy system method (*i.e.* 0.0262431), even without noise in their method.

#### 4.2 Prediction of Federal Funds Rate Data

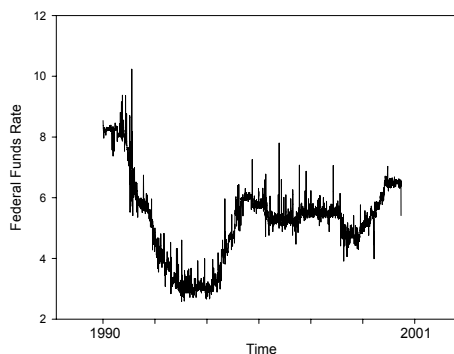
In order to compare the forecast performances for U.S. effective funds rate with [6], we have the data of funds rate over the sample period from January 1, 1990 through December 31, 2000. There are totally 2869 observations, as shown in Fig. 5. Some other variables used to forecast funds rate on these six economic prediction models consist of daily observations on the U.S. effective federal funds rate<sup>1</sup>, the 3-month T-bill  $t_t^{TB}$ , the

**Table 3. Summary statistics.**

	$i_t^{FF}$	$i_t^{TB}$	$f^1$	$f^2$
Mean	5.249	5.056	5.277	5.304
Variance	1.893	1.578	1.771	1.741

**Table 4. Prediction accuracy from Jan. 1 1997 to Dec. 31 2000.**

Models	RMSE
MRBF	0.178
RW	0.196
M-TAR	0.189
EGARCH(1, 1)	0.186
MSAH-AR(2, 1)	0.194
BTAR	0.202
MSIH-VECH(3, 1)	0.216

**Fig. 5. Federal funds rate data from Jan. 1, 1990 through Dec. 31, 2000.**

1-month and 2-month federal funds futures,  $f^1$ ,  $f^2$ . In order to eliminate the effect of the increased volatility on settlement days, the funds rate series used in our paper are already adjusted. Table 3 presents summary statistics for these variables.

We first estimate each prediction model over the sample period January 1, 1990 to December 31, 1996. Next, we calculate one-day-ahead forecasts over the period, January 1, 1997 to December 31, 2000. To compare the predicting performances with these individual economics models of [6], we report the RMSE over the forecast period January 1, 1997 to December 31, 2000 (as the same forecast periods with [6]). The out-of-sample forecasts are constructed according to a recursive procedure, which is conditional only upon information available up to the date of the forecasts and with successive re-estimation as the date on which forecasts and conditioned moves through the data set. The network parameters are set to the same as those used in the Mackey-Glass time series prediction. The prediction result is shown in Table 4. It reports the RMSE for each of the estimated models. Our results show that the MRBF network exhibits the best out-of sample prediction performance: the RMSE obtained is the lowest across all models.

## 5. CONCLUSIONS

Neural network forecasting models have been widely used in the analysis of financial time series during the last decade. This paper attempts to fill this gap in the literature by comparing a variety of univariate and multivariate, linear, nonlinear economics prediction modes and neural network. We compare the forecasting performances between MRBF network to a variety of empirical models for daily federal funds rate. The daily

<sup>1</sup> The federal funds rate is a weighted average of the rates on federal funds transactions of a group of federal funds brokers who report their transactions daily to the Federal Reserve Bank of New York.

data spans over the period from January 1, 1990 through December 31, 2000, yielding a total of 2869 observations. We compare the performances of out-of sample prediction on one-day-ahead federal funds rate over the period January 1, 1997 to December 31, 2000.

Our results indicate that MRBF network can produce the best out-of sample performance: the RMSE obtained for funds rate is the lowest across all other models. Thus, the result shows MRBF network is a good prediction method for the movements of federal funds rate. That is, it implies MRBF network can be one good forecast method for some financial time series data. We conclude that it is because the proposed MRBF have two advantages, nonlinearity and robustness. The federal funds rate varies from bank to bank and from day to day. Theoretically, the rate should be determined by the open market. Since government policy, terrorist incident, or some other nonblank financial panics could affect the movements of interest rate, so some sudden changes of interest rate can be regarded as noise. Other methods introduced earlier are not robust for the data with noise. Therefore, the proposed MRBF which can handle the nonlinear and noisy time series is superior to other methods.

## REFERENCES

1. J. T. Krueger and K. N. Kuttner, "The fed funds futures rate as a predictor of federal reserve policy," *Journal of Futures Markets*, Vol. 16, 1996, pp. 865-879.
2. J. B. Taylor, "Expectations, open market operations, and changes in the federal funds rate," *Federal Reserve Bank of St. Louis Review*, Vol. 83, 2001, pp. 33-47.
3. G. D. Rudebusch, "Federal reserve interest rate targeting, rational expectations, and the term structure," *Journal of Monetary Economics*, Vol. 35, 1995, pp. 245-274.
4. L. Sarno and D. L. Thornton, "The dynamic relationship between the federal funds rate and treasury bill rate: An empirical investigation," *Journal of Banking and Finance*, Vol. 27, 2003, pp. 1079-1110.
5. H. Bruce and B. Seo, "Testing for two-regime threshold cointegration in vector error-correction models," *Journal of Econometrics*, Vol. 110, 2002, pp. 293-318.
6. L. Sarno, D. Thornton, and G. Valente, "Federal funds rate prediction," *Journal of Money, Credit, and Banking*, Vol. 37, 2005, pp. 449-471.
7. D. B. Nelson, "Conditional Heteroskedasticity in asset returns: A new approach," *Econometrica*, Vol. 59, 1990, pp. 347-370.
8. E. S. Chng, S. Chen, and B. Mulgrew, "Gradient radial basis function networks for nonlinear and nonstationary time series prediction," *IEEE Transactions on Neural Networks*, Vol. 7, 1996, pp. 190-194.
9. H. Leung, T. Lo, and S. Wang, "Prediction of noisy chaotic time series using an optimal radial basis functions neural network," *IEEE Transactions on Neural Networks*, Vol. 12, 2001, pp. 1163-1172.
10. C. Li, H. Ye, and G. Wang, "Nonlinear time series modeling and prediction using RBF network with improved clustering algorithm," in *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 4, 2004, pp. 3513-3518.
11. Y. Wang, G. B. Huang, P. Saratchandran, and N. Sundararajan, "Time series study of GGAP-RBF network: Predictions of Nasdaq stock and nitrate contamination of drinking water," in *Proceedings of IEEE International Joint Conference on Neural*

- Networks*, 2005, pp. 3127-3132.
12. J. H. Chen, C. S. Chen, and Y. S. Chen, "Fast algorithm for robust template matching with M-estimators," *IEEE Transactions on Signal Processing*, Vol. 51, 2003, pp. 230-243.
  13. X. Hong and S. Chen, "M-estimator and D-optimality model construction using orthogonal forward regression," *IEEE Transactions on Systems, Man, and Cybernetics – Part B*, Vol. 35, 2005, pp. 155-162.
  14. C. C. Chuang, S. F. Su, and C. C. Hsiao, "The annealing robust backpropagation (ARBP) learning algorithm," *IEEE Transactions on Neural Networks*, Vol. 11, 2000, pp. 1067-1076.
  15. M. C. Ortiz, L. A. Sarabia, and A. Herrero, "Robust regression techniques: A useful alternative for the detection of outlier data in chemical analysis," *Talanta*, Vol. 70, 2006, pp. 499-512.
  16. C. C. Lee, C. L. Tsai, Y. C. Chiang, and C. Y. Shih, "M-estimator based robust radial basis function neural networks with growing and pruning techniques," in *Proceedings of the International Computer Symposium*, 2006, pp. 1065-1069.
  17. C. C. Lee, P. C. Chung, J. R. Tsai, and C. I. Chang, "Robust radial basis function neural networks," *IEEE Transactions on Systems, Man, and Cybernetics – Part B*, Vol. 29, 1999, pp. 674-684.
  18. P. J. Huber, *Robust Statistics*, John Wiley and Sons, New York, 1984.
  19. V. Barnett and T. Lewis, *Outliers in Statistical Data*, John Wiley and Sons, New York, 1994.
  20. P. J. Rousseeuw and S. Verboven, "Robust estimation in very small samples," *Computational Statistics and Data Analysis*, Vol. 40, 2002, pp. 741-758.
  21. G. B. Huang, P. Saratchandran, and N. Sundararajan, "An efficient sequential learning algorithm for growing and pruning RBF (GAP-RBF) networks," *IEEE Transactions on Systems, Man, and Cybernetics – Part B*, Vol. 34, 2004, pp. 2284-2292.
  22. S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks*, Vol. 2, 1991, pp. 302-309.
  23. S. Chen, E. S. Chng, and K. Alkadhim, "Regularized orthogonal least squares algorithm for constructing radial basis function networks," *International Journal of Control*, Vol. 64, 1996, pp. 829-837.
  24. S. Chen, Y. Wu, and B. L. Luk, "Combined genetic algorithm optimization and regularized orthogonal least squares learning for radial basis function networks," *IEEE Transactions on Neural Networks*, Vol. 10, 1999, pp. 1239-1243.
  25. S. Chen, "Local regularization assisted orthogonal least squares regression," *Neuro-Computing*, Vol. 69, 2006, pp. 559-585.
  26. M. J. L. Orr, "Regularization on the selection of radial basis function centers," *Neural Computation*, Vol. 7, 1995, pp. 606-623.
  27. K. B. Cho and B. H. Wang, "Radial basis function based adaptive fuzzy systems their applications to system identification and prediction," *Fuzzy Sets and Systems*, Vol. 83, 1996, pp. 325-339.
  28. J. S. R. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice-Hall, Upper Saddle River, NJ, 1997.

29. Y. Chen, B. Yang, and J. Dong, "Time-series prediction using a local linear wavelet neural network," *Neurocomputing*, Vol. 69, 2006, pp. 449-465.
30. Y. H. Chen, B. Yang, and J. Dong, "Nonlinear system modeling via optimal design of neural trees," *International Journal of Neural Systems*, Vol. 14, 2004, pp. 125-137.
31. D. Kim and C. Kim, "Forecasting time series with genetic fuzzy predictor ensemble," *IEEE Transactions on Fuzzy Systems*, Vol. 5, 1997, pp. 523-535.
32. S. F. Gray, "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, Vol. 42, 1996, pp. 27-62.



**Chien-Cheng Lee (李建誠)** received the Ph.D. degrees in Electrical Engineering from National Cheng Kung University, Tainan, Taiwan in 2003. Dr. Lee is currently an Assistant Professor in the Department of Communications Engineering, Yuan Ze University, Chungli, Taiwan. He has been a research visitor at the Telcordia Inc. (formerly Bellcore), NJ, USA, from Oct. 2007 to Jan. 2008. He is one of the guest editors for a special issue on Signal Processing for Applications in Healthcare Systems for EURASIP Journal on Advances in Signal Processing, 2008. His research interests include image processing, pattern recognition, neural networks, and telemedicine. He is a member of IEEE.



**Chun-Li Tsai (蔡群立)** received the Ph.D. degree in Department of Economics, Texas A&M University, U.S.A., in 2005. Dr. Tsai is currently an Assistant Professor in the Department of Economics, National Cheng Kung University, Tainan, Taiwan. His research interests include monetary economics, financial economics, and intelligent computing. His main projects are the forecast of macroeconomics, the impacts of monetary policy to financial markets, and the international transmission of monetary channel.



**Yu-Chun Chiang (江右君)** received a Ph.D. in Environmental Engineering from National Taiwan University in Taiwan in 1999. She used to be a visiting scholar at University of Delaware in 1998 and at Desert Research Institute, University of Nevada-Reno in 2003. Her primary research area is activated carbon adsorption and synthesis of carbon nanotubes.