Nonlinear/Non-Gaussian Time Series Prediction Based on RBF-HMM-GMM Model

Dongqing Zhang\textsuperscript{1}, Yubing Han\textsuperscript{2}, Xueyu Tang\textsuperscript{1}

\textsuperscript{1}College of Engineering, Nanjing Agricultural University, Nanjing, 210031, China
\textsuperscript{2}School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China

*corresponding author, e-mail: zhangdq@njau.edu.cn, hanyb@mail.njust.edu.cn, xueyutang@yahoo.com.cn

Abstract

In order to cope with the nonlinear and non-Gaussian time series, a RBF-HMM-GMM model, which is based on radial basis function (RBF) neural networks with the assumption of measurement noise being hidden Markov model (HMM) and the distribution of each hidden states being approximated by Gaussian mixture models (GMM), is proposed in this paper. In the proposed model, both the orders (numbers of nodes and inputs of RBF network, numbers of hidden states of HMM, numbers of Gaussian mixture component of GMM) and the parameters change over time. Firstly, a scheme for time series forecasts based on RBF-HMM-GMM model is proposed. Then an on-line prediction algorithm based on RBF-HMM-GMM model using sequential Monte Carlo (SMC) methods is developed. At last, the monthly West Texas Intermediate crude oil future price series are analyzed, and experimental results indicate that the RBF-HMM-GMM model is able to predict the time series accurately.

Keywords: on-line prediction, time series, radial basis function neural networks, hidden Markov model, sequential Monte Carlo methods

Copyright © 2012 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

Time series forecasting is a growing field of interest playing an important role in nearly all fields of science and engineering. The forecasting domain has been influenced by linear statistical methods such as ARIMA models for a long time. However, the linear models are not adapted to many real applications, so Billings et al proposed nonlinear autoregressive moving average (NARMA) model [1], which provides a unified representation for a wide class of nonlinear systems. Unfortunately, it is, on most occasions, impossible to derive closed-form analytical expressions to represent the nonlinear function in NARMA model, so many researchers resort to neural networks, such as feed-forward neural networks [2], multi-layer perceptions [3], static radial basis function (RBF) neural networks [4], dynamic RBF models with the variable number of basis functions [3] and so on.

When the nonlinear function in NARMA model is approximated with neural networks, selecting the proper order of input is critical. It is desired that the neural network includes as many inputs as possible so that the information in these inputs will influence the predicted value. On the other hand, it is also desired that the neural network includes as few inputs as possible because the complexity and computation increase as the input order increases. There are many methods for selecting the input order, such as prior knowledge [5], autocorrelation and partial correlation function [6], Akaike information criterion (AIC) [7], trial and error method [8] and so on. The order keeps fixed once it is chosen by the above methods. However, many real-world time series are non-stationary and their structures may change, which means that the number of inputs should change over time automatically. This paper is concerned with on-line selection of input order.

On the other hand, the observations always are corrupted by noise, which is modeled in generalizations of the Gaussian, Cauchy, and beta distributions in [9]. However, it is inappropriate to model the noise by a single distribution in many cases, so some researchers resort to mixture models, such as Gaussian mixture models (GMM), Gaussian-Laplace mixture models, multinomial mixture models [10,11] and so on. Among those mixture models, GMM is...
very appealing. It is proved that when the number of Gaussian distribution in GMM is big enough, it can approximate arbitrary noise. Unfortunately, with the number of the mixture model increasing, the computation complexity also increases. In order to overcome the deficiency, the noise is modeled as hidden Markov models (HMM) and the distribution of each hidden state is approximated by Gaussian mixture models (GMM), which called HMM-GMM model in this paper.

Hidden Markov models were originally introduced as far back as 1957 and they have become increasingly popular in the last several years due to strong mathematical structure and theoretical basis [12]. Order estimation is an important statistical problem, whose essence is the estimation of the dimension of a model, where order refers to the number of the states. There is a great amount of literature dedicated to order estimation by various methods, such as AIC or BIC [13], penalized maximum likelihood estimator [14], trial and error method [15], priori knowledge [16], and so on. But the order is fixed in those existing works after estimation. This issue is strongly related to on-line selection of the HMM order.

The paper is organized as follows. In Section 2, the scheme for time series forecasts based on RBF-HMM-GMM model is proposed. In Section 3, on-line prediction algorithm based on RBF-HMM-GMM model using sequential Monte Carlo (SMC) method is presented. In Section 4, the monthly West Texas Intermediate crude oil future price series are analyzed. Finally, a summary is given in Section 5.

2. Problem statement
2.1. NARMA Model

The general NARMA model can be stated by [1]

\[
\hat{y}_t = f(y_{t-1}, \ldots, y_{t-g}, e_{t-1}, \ldots, e_{t-q}) + e_t
\]  

(1)

where \( \{y_{t-i}, i = 1, \ldots, g\} \) is a historical series, \( \hat{y}_t \) is predictive value, \( f(y_{t-1}, \ldots, y_{t-g}, e_{t-1}, \ldots, e_{t-q}) \) denotes the functional dependency between past and future observations, \( g, q \) are the orders of the autoregressive part and the moving average part, respectively. \( \{e_{t-i}, i = 0, \ldots, q\} \) is measurement noise and can be approximated by \( e_{t-i} = y_{t-i} - \hat{y}_{t-i} \). In this paper, the orders of the autoregressive part and the moving average part vary along with time, so we rewrite (1) as

\[
\hat{y}_t = f(y_{t-1}, \ldots, y_{t-g}, e_{t-1}, \ldots, e_{t-q}) + e_t
\]  

(2)

According to (2), the key of NARMA model for prediction is to choose proper expression of nonlinear function \( f(\cdot) \). However, in many cases, it is impossible to express nonlinear function in exact form, so we resort to RBF networks for their abilities of approximating functions.

![Figure 1. Structure of forecast model based on RBF-HMM-GMM](image)
As far as the measurement noise $e_t$ is concerned, it is modeled as hidden Markov models and the distribution of each hidden state is approximated by Gaussian mixture models (called HMM-GMM model). The structure of forecasting model is given in Figure 1.

2.2. Variable Dimension RBF Neural Network Models for Nonlinear Part

In this section the neural network in Figure 1 will be considered. Because RBF networks tend to be more tractable than other neural networks, so they are chosen to approximate the nonlinear function in NARMA model in this paper.

We shall consider an approximation scheme consisting of a mixture of $k$ RBFs and a linear regression term proposed by Holmes in 1998. More precisely, the linear-RBF model at time $t$ is given by

$$
\tilde{y}_t = \sum_{j=1}^{k} a_{j,t} \phi(||x_t - c_{j,t}||) + \beta_t^T x_t
$$

where $x_t = (y_{t-1}, \ldots, y_{t-q}, \ldots, e_{t-q}, \ldots, e_{t-1})^T$ and $\tilde{y}_t$ are the input and output of RBF networks, respectively, $||.||$ denotes a distance metric (usually Euclidean), $\phi(\cdot)$ denote the basis functions, $c_{j,t}$ denotes the $j$ th RBF centre for a model with $k$ RBFs, $a_{j,t}$ denotes the $j$ th RBF amplitude, $\beta_t = (\beta_{1,t}, \ldots, \beta_{d,t})^T$ denotes the linear regression parameters and $d_t = g_t + a_t$.

As for basis functions, the most common choices are linear, cubic, thin plate spline and Gaussian [18]. In this paper we choose Gaussian basis function $\phi_i(x) = e^{-\lambda x^2}$, where $\lambda$ will vary alone with time. For convenience, the model is expressed in vector-matrix form

$$
\tilde{y}_t = D_t a_t
$$

where $D_t = [y_{t-1}, \ldots, y_{t-q}, \ldots, e_{t-q}, \ldots, e_{t-1}, \phi(x_{t,1}), \ldots, \phi(x_{t,p})]$, $a_t = [\beta_{1,t}, \ldots, \beta_{d,t}, a_{1,t}, \ldots, a_{p,t}]^T$. It is noted that the dimension of RBF networks changes along with time, i.e. $k_t, g_t, q_t$ are variable. For the details, please refer to Section 3.4.

2.3. Variable Dimension HMM-GMM model for Measurement Noise

As for measurement noise $e_t$ in Figure 1, it is assumed to be HMM-GMM model and the order is changed over time in this paper. The variable dimension hidden Markov models are characterized as follows

$$
\begin{cases}
    z_t \sim P(z_t | z_{t-1}, A_t), \\
    e_t \sim f(e_t | z_t, B_t)
\end{cases}
$$

where $z_t \in S$ is hidden state at time $t$, $S = \{S_1, \ldots, S_{Q_t}\}$ is a hidden state set with $Q_t$ elements, which depicts the intrinsic driven mechanism. $z_t \sim P(z_t | z_{t-1}, A_t)$ is state evolution density and depends on state transition matrix $A_t = [a_{ij}]$, where $a_{ij} = P(x_t = S_j | x_{t-1} = S_i)$, $e_t \sim f(e_t | z_t, B_t)$ denotes the observation density, and $B_t = \{g_t(e_t), 1 \leq i \leq n_t\}$ is the conditional probability distribution of observation variable with definition of $g_t(e_t) = P(e_t | z_t = S_i)$, which can be approximated with Gaussian mixture models. In many literatures, the number of Gaussian mixture components keeps fixed once it is estimated [19]. However, the number of mixture components of the GMM changes over time in this paper. It assumes that the density function is given by

$$
g_t(e_t) = \sum_{k=1}^{M_t} v_{ik} N(e_t | \mu_{ik}, \sigma_{ik}^2), i = 1, \ldots, Q_t
$$

where $N(e_t | \mu_{ik}, \sigma_{ik}^2)$ is the probability density belong to the $i$ th Gaussian component density in the mixture, $M_t$ is the number of mixture components at the $i$ th state, $\mu_{ik}, \sigma_{ik}^2, v_{ik}$ are,
respectively, the means, variances and weights of the individual Gaussian component densities at state \( S_i \). Also \( v_{ik} \geq 0 \) and \( \sum_{k=1}^{n_G} v_{ik} = 1 \), \( v_{ik} \geq 0, i = 1, \ldots, Q \).

3. On-line Prediction Algorithm based on RBF-HMM-GMM Model Using SMC Method

In this section, our intention is to predict \( \hat{y}_t \) using the RBF-HMM-GMM model given all available information \( y_{1:t-1} \), where \( y_{1:t-1} = (y_1, \ldots, y_{t-1}) \).

3.1. Measurement Equation and State Transition Equation

From above discussion, the measurement equation of RBF-HMM-GMM model can be written

\[
\hat{y}_t = \tilde{y}_t + e_t = D_ia_i + e_t \tag{6}
\]

And the state transitions equation are given by

\[
\begin{align*}
{k}_t &\sim P(k_t | k_{t-1}) \\
g_t &\sim P(g_t | g_{t-1}) \\
q_t &\sim P(q_t | q_{t-1}) \\
\lambda_t &\sim \lambda_{t-1} + e_\lambda \\
c_{i,t} &\sim c_{i,t-1} + e_c, \quad j = 1, \ldots, k_t \\
a_t &\sim a_{t-1} + e_a \\
Q_t &\sim P(Q_t | Q_{t-1}) \\
M_{i,t} &\sim P(M_{i,t} | M_{i,t-1}), \quad i = 1, \ldots, Q \\
A_t &\sim A_{t-1} + e_A \\
\mu_t &\sim \mu_{t-1} + e_\mu \\
\ln(\sigma_t^2) &\sim \ln(\sigma_{t-1}^2) + e_\sigma \\
v_t &\sim v_{t-1} + e_v \\
z_t &\sim P(z_t | z_{t-1}, A_t) \\
I_t &\sim P(I_t | z_t, v_t)
\end{align*}
\tag{7}
\]

Figure 2. Dynamic graphical model. The circles represents unknowns, while the double circles correspond to the data.

Figure 3. Dynamic revolution of RBF model

where \( P(k | k_{t-1}), P(g_t | g_{t-1}), P(q_t | q_{t-1}), P(Q_t | Q_{t-1}), P(M_{i,t} | M_{i,t-1}), i = 1, \ldots, Q \) are specified discrete distributions, \( \mu_t, \sigma_t^2, v_t \) are the mean vector, variance vector and weight vector of HMM, respectively, while the diffusion processes are sampled from normal distributions.
\( \epsilon_d \sim N(0, \delta^2_d I_{d \times d}), \epsilon_c \sim N(0, \delta^2_c I_{d \times d}), \epsilon_a \sim N(0, \delta^2_a I_{m \times m}), \epsilon_A \sim N(0, \delta^2_A I_{m \times m}), \epsilon_{\mu} \sim N(0, \delta^2_{\mu} I_{n \times n}), \epsilon_v \sim N(0, \delta^2_v I_{n \times n}), \) \\
\( \epsilon_a \sim N(0, \delta^2_a I_{n \times n}), \epsilon_v \sim N(0, \delta^2_v I_{n \times n}), \) where \( m_i = k_i + d_i, \delta^2_d, \delta^2_c, \delta^2_a, \delta^2_A, \delta^2_{\mu}, \delta^2_v, \delta^2_v \) are constants and should be chosen according to priori knowledge, \( I_{d \times d}, I_{m \times m}, I_{n \times n}, J_{n \times n} \) are unit matrix.

\( z_t \sim P(z_t | z_{t-1}) \) is state evolution density and depends on state transition matrix \( A, I \) denotes the indicator of Gaussian mixtures at given state and \( P(I, z, v) \) is a conditional probability distribution depending on \( z_t, v_t \). Figure 2 shows the graphical model representation of the joint distribution for the first four time steps. Figure 3 and Figure 4 depict the dynamic revolution of RBF and HMM-GMM models for the first four time steps, respectively.

![Graphical Model Representation](image)

**Figure 4. Dynamic revolution of HMM-GMM model**

### 3.2. Rao-Blackwellised Particle Filter

Consider the proposed model, the posterior probability distribution function (pdf) \( p(\theta | y_{1:T}) \) should be calculated recursively, where \( \theta = \{z, g, q, \lambda, c, A, v, z, I, Q, M\} \). If the posterior pdf is a linear Gaussian distribution, it can be solved with analytical methods such as Kalman filter. However, on most occasions, it is impossible to derive closed-form analytical expressions to obtain the posterior distribution and its features of interest. To compute these distributions, we need to resort to SMC method and the details of SMC method are given in [20]. However, SMC method in high dimensional state-spaces can be inefficient because a large number of samples are needed to represent the posterior. Therefore, we adopt Rao-Blackwellised particle filter (RBPF) [21], which basic idea is to partition the state-space into a few sub-spaces and to compute some of the variables exactly and sample the rest. In general we will need far fewer particles with RBPF to reach the same accuracy as SMC method [21] while the estimated variance of RBPF can be shown to be smaller than the variance of SMC method [22].

Suppose we partition the state-space \( \theta \) into two sub-spaces, \( r \) and \( s \), i.e. \( \theta = \{r, s\} \), where \( s = \{z\}, r = \{z, g, q, \lambda, c, A, v, z, I, Q, M\} \). By the chain rule of probability, we have
\[ p(r_t, s_t \mid y_{1:t}) = p(r_t \mid y_{1:t}) p(s_t \mid r_t, y_{1:t}) \]  

We sample \( p(r_t \mid y_{1:t}) \) using standard SMC method. It is clear that, conditionally upon \( r_t \), we end up with a linear Gaussian system with states \( s_t \), i.e. \( p(s_t \mid r_t, y_{1:t}) = N(\bar{\alpha}_t, P_t) \), where \( \bar{\alpha}_t = E(\alpha_t) \) and \( P_t = \text{cov}(\alpha_t, \alpha_t) = E((\alpha_t - \bar{\alpha}_t)(\alpha_t - \bar{\alpha}_t)^T) \) correspond to the conditional mean and covariance matrix of \( \alpha_t \), respectively, so we can update \( p(s_t \mid r_t, y_{1:t}) \) analytically and efficiently use Kalman filter.

### 3.3. On-line Prediction Algorithm

As mentioned above, there is \( p(r_{t-1}, s_{t-1} \mid y_{1:t-1}) = p(r_{t-1} \mid y_{1:t-1}) p(s_{t-1} \mid r_{t-1}, y_{1:t-1}) \). Suppose we have

\[ p(r_{t-1} \mid y_{1:t-1}) = \sum_{i=1}^{n_p} w_{i-1}^{(p)} \delta(r_{t-1} - r_{t-1}^{(i)}) \]  
\[ p(s_{t-1} \mid r_{t-1}^{(i)}, y_{1:t-1}) = N(\bar{\alpha}_t^{(i)}, P_t^{(i)}), \quad i = 1, \ldots, n_p \]

where \( \{r_{t-1}^{(i)}\}_{i=1}^{n_p} \) is a set of particles with associated weights \( \{w_{i-1}^{(p)}\}_{i=1}^{n_p} \), \( n_p \) is the number of particles, \( \delta(\cdot) \) denotes Kronecker delta function. Now we will calculate the pdf \( p(\hat{y}_t \mid y_{1:t-1}) \) and \( p(r_t, s_t \mid y_{1:t-1}) \) recursively.

#### 3.3.1. Sampling step

Because

\[ p(r_t \mid y_{1:t-1}) = \int p(r_t \mid y_{1:t-1}) p(r_t \mid r_{t-1}, y_{1:t-1}) dr_{t-1} \]  

Substituting (9) into (11) yields

\[ p(r_t \mid y_{1:t-1}) = \sum_{i=1}^{n_p} w_{t-1}^{(p)} \delta(r_t - r_t^{(i)}) p(r_{t-1} \mid r_{t-1}^{(i)}, y_{1:t-1}) ] dr_{t-1} = \sum_{i=1}^{n_p} w_{t-1}^{(p)} p(r_t \mid r_t^{(i)}, y_{1:t-1}) \]

We adopt \( p(r_t \mid r_{t-1}^{(i)}, y_{1:t-1}) = p(r_t \mid r_{t-1}^{(i)}) \) as the important density function, namely, draw \( k_t^{(i)} \sim p(k_t \mid r_{t-1}^{(i)}) \), \( g_t^{(i)} \sim p(g_t \mid r_{t-1}^{(i)}), q_t^{(i)} \sim p(q_t \mid q_{t-1}^{(i)}), \lambda_t = \lambda_{t-1} + \varepsilon_k, c_{t-1,j} = c_{t-1,j} + \varepsilon_c, A_t^{(i)} = A_{t-1}^{(i)} + \varepsilon_A, I_t \sim \mathcal{N}(I_t \mid z_t, \nu_t) \), \( \mu_t = \mathcal{N}(\mu_t \mid Q_{t-1}, A_{t-1}^{(i)} M_t + M_{t-1}, I_t \sim \mathcal{N}(I_t \mid z_t, \nu_t) \), and obtain new particles \( r_{t-1}^{(i)} = (k_t^{(i)}, g_t^{(i)}, q_t^{(i)}, \lambda_t, c_{t-1,j}^{(i)}, A_t^{(i)}, I_t) \). Thus \( p(r_t \mid y_{1:t-1}) \) can be approximated as

\[ p(r_t \mid y_{1:t-1}) = \sum_{i=1}^{n_p} w_{t-1}^{(p)} \delta(r_t - r_t^{(i)}) \]

#### 3.3.2. Measurement prediction step

In this step, \( p(\hat{y}_t \mid y_{1:t-1}) \) is discussed. By the Chapman-Kolmogorov equation and (13), we have

\[ p(\hat{y}_t \mid y_{1:t-1}) = \int p(r_t \mid y_{1:t-1}) p(\hat{y}_t \mid r_t, y_{1:t-1}) dr_t = \int \sum_{i=1}^{n_p} w_{t-1}^{(p)} \delta(r_t - r_t^{(i)}) p(\hat{y}_t \mid r_t, y_{1:t-1}) dr_t \]

\[ = \sum_{i=1}^{n_p} w_{t-1}^{(p)} p(\hat{y}_t \mid r_t^{(i)}, y_{1:t-1}) \]

Obviously, once \( r_t^{(i)} \) has been drawn, \( D_t^{(i)} \) can be obtained according to its definition in section 2.2. On the other hand, \( \varepsilon_t \) is assumed to be HMM-GMM and it follows Gaussian
distribution given \(z_{i}^{(i)}\) and \(I_{i}^{(i)}\), i.e., \(p(a_{i} | z_{i}^{(i)}, I_{i}^{(i)}) = N(\mu_{i}^{(i)}(z_{i}^{(i)}, I_{i}^{(i)}), \sigma_{a}^{2(i)}(z_{i}^{(i)}, I_{i}^{(i)}))\). Moreover, \(a_{i} = a_{i-1} + e_{a}\) and \(\hat{y}_{i} = D_{i}a_{i} + e_{y}\), so we have

\[
p(a_{i} | y_{1:i-1}, r_{i}^{(i)}) = N(\bar{a}_{1:i-1}, P_{1:i-1} + \delta_{a}^{2}I_{m,n,m,n})
\]

\[
p(\hat{y}_{i} | y_{1:i-1}, r_{i}^{(i)}) = N(D_{i}^{(i)}\bar{a}_{i-1}^{(i)} + \mu_{i}^{(i)}(z_{i}^{(i)}, I_{i}^{(i)}), D_{i}^{(i)}(P_{1:i-1}^{(i)} + \delta_{a}^{2}I_{m,n,m,n}))(D_{i}^{(i)})^T + \sigma_{y}^{2(i)}(z_{i}^{(i)}, I_{i}^{(i)})\)

Substitution (16) into (14) yields

\[
p(\hat{y}_{i} | y_{1:i-1}) = \sum_{i=1}^{n_{p}} w_{i}^{(i)}(D_{i}^{(i)}\bar{a}_{i-1}^{(i)} + \mu_{i}^{(i)}(z_{i}^{(i)}, I_{i}^{(i)}))
\]

In real application, the mean of \(p(\hat{y}_{i} | y_{1:i-1})\) is often used to denote the prediction, that is

\[
\hat{y}_{i} = \sum_{i=1}^{n_{p}} w_{i}^{(i)}(D_{i}^{(i)}\bar{a}_{i-1}^{(i)} + \mu_{i}^{(i)}(z_{i}^{(i)}, I_{i}^{(i)}))
\]

3.3.3. Weight updating and resampling step

When the true measurement \(y_{i}\) is available, we can update the weight according to

\[
w_{i}^{(i)} = w_{i}^{(i)}p(y_{i} | r_{i}^{(i)}, y_{1:i-1})\]

where the likelihood \(p(y_{i} | r_{i}^{(i)}, y_{1:i-1})\) has been presented in (16). Obviously, all particles have different weights. If some particles have too small weights, it will cause degeneracy. In order to avoid the degeneracy phenomenon, systematic resampling is preferred by author. For more details about systematic resampling, refer to [23]. Thus we have

\[
p(r_{i} | y_{1:i}) = \sum_{i=1}^{n_{p}} w_{i}^{(i)} \delta(r_{i} - r_{i}^{(i)})
\]

3.3.4. Exact step

Now we will study \(p(s_{i} | r_{i}, y_{1:i})\). Because \(s_{i}\) is linear Gaussian given \(r_{i}\), it is possible to integrate out \(s_{i}\) thereby reducing the variance of the estimates. The end result corresponds to a bank of Kalman filters for network output:

\[
K_{i}^{(i)} = (P_{i}^{(i)} + \delta_{I}^{2}I_{m,n,m,n})^{-1} [D_{i}^{(i)}(\bar{a}_{i-1}^{(i)} + \mu_{i}^{(i)}(z_{i}^{(i)}, I_{i}^{(i)})) + D_{i}^{(i)}(P_{i}^{(i)} + \delta_{a}^{2}I_{m,n,m,n}))(D_{i}^{(i)})^{-1}\]

\[
\bar{a}_{i}^{(i)} = \bar{a}_{i-1}^{(i)} + K_{i}^{(i)}(y_{i} - D_{i}^{(i)}\bar{a}_{i-1}^{(i)} - \mu_{i}^{(i)}(z_{i}^{(i)}, I_{i}^{(i)}))\]

\[
P_{i}^{(i)} = P_{i-1}^{(i)} + \delta_{a}^{2}I_{m,n,m,n} - K_{i}^{(i)}(D_{i}^{(i)}(P_{i-1}^{(i)} + \delta_{a}^{2}I_{m,n,m,n})\]

where \(K_{i}^{(i)}\) denotes the Kalman gain.

According to (20), we obtain \(\bar{a}_{i}^{(i)}\) and \(P_{i}^{(i)}\), such that

\[
p(s_{i} | r_{i}^{(i)}, y_{1:i}) = p(a_{i} | r_{i}^{(i)}, y_{1:i}) = N(\bar{a}_{i}^{(i)}, P_{i}^{(i)}), \quad i = 1, \ldots, n_{p}\]

From above discussion, we obtain \(\hat{y}_{i}\) and calculate the pdf \(p(r_{i}, s_{i} | y_{1:i})\) recursively. Note that if \(k_{i}, g_{i}, q_{i}, Q_{i}, M_{i,i}\) change, some parameters should be adjusted accordingly. For the details, refer to Section 3.4.

3.4. Parameters Adjustment

In this paper, the order of RBF-HMM-GMM changes dynamically, i.e., \(k_{i}, g_{i}, q_{i}, Q_{i}, M_{i,i}\) vary with time. Once \(k_{i}, g_{i}, q_{i}, Q_{i}, M_{i,i}\) changes, the dimensions of \(D_{i}, a_{i}, P_{i}, \mu_{i}, \sigma_{a}^{2}, \nu_{i}, M_{i,i}\) also change accordingly.
According to (7), we obtain the parameters at time $t$. If $k_t, g_t, q_t, Q_t, M_t$ change, some parameters should be adjusted. In this section, we will discuss the parameters adjustment with $Q_t, M_t$ of HMM-GMM in detail. As for that with $k_t, g_t, q_t$ of RBF networks, please refer to [24].

For ease of exposition, $x_t, c_t, D_t, a_t, P_t$ are represented as

$$
\begin{aligned}
x_t &= ((x_{t,1}), (x_{t,2}))^T \\
x_{t,1} &= (y_{t-\sigma_t}, \ldots, y_{t-1})^T, \\
x_{t,2} &= (e_{t-\sigma_t}, \ldots, e_{t-1})^T \\
c_t &= (c_{t,1}, \ldots, c_{t,k_t}) \\
c_{t,j} &= ((c_{t,j,1}), (c_{t,j,2}))^T, \\
c_{t,j,1} &= (c_{t,j,1}, \ldots, c_{t,j,k_{t-1}})^T, \\
c_{t,j,2} &= (c_{t,j,k_{t-1}+1}, \ldots, c_{t,j,k_t})^T, \\
D_t &= (D_{t,1}, D_{t,2}, D_{t,3}) \\
D_{t,1} &= (x_{t,1})^T = (y_{t-\sigma_t}, \ldots, y_{t-1}) \\
D_{t,2} &= (x_{t,2})^T = (e_{t-\sigma_t}, \ldots, e_{t-1}) \\
D_{t,3} &= (\phi_t(x_t, c_{t,1}), \ldots, \phi_t(x_t, c_{t,k_t})) \\
a_t &= (a_{t,1}, a_{t,2}, a_{t,3})^T \\
a_{t,1} &= (\beta_{t,1}, \ldots, \beta_{t,k_t})^T \\
a_{t,2} &= (\beta_{t,k_{t-1}+1}, \ldots, \beta_{t,k_t})^T \\
a_{t,3} &= (a_{t,1}, \ldots, a_{t,k_t})^T \\
P_t &= \begin{pmatrix}
P_{t,1,1} & P_{t,1,2} & P_{t,1,3} \\
P_{t,2,1} & P_{t,2,2} & P_{t,2,3} \\
P_{t,3,1} & P_{t,3,2} & P_{t,3,3}
\end{pmatrix} \\
P_{t,ij} &= Cov(a_{t,i}, a_{t,j}) = E[(a_{t,i} - \tilde{a}_{t,i})(a_{t,j} - \tilde{a}_{t,j})^T],
\end{aligned}
$$

(22)-(26)

Notice that, we hereafter use the superscripts $b$ to represent the parameters evolved from the previous time, while use the superscripts $a$ to represent the parameters adjusted.

### 3.4.1 Change of the Number of hidden states

If $Q_t$ changes, the dimensions of $A_t, \mu_t, \sigma_t^2, v_t, M_t$ will change accordingly. For simplicity, we consider the following three cases: (i) $Q_t = Q_{t-1}$; (ii) $Q_t = Q_{t-1} + 1$; (iii) $Q_t = Q_{t-1} - 1$. As for case (i), we needn’t adjust parameters any more and study the rest two cases.

1. **The number of hidden states increases**

   If $Q_t$ increases, i.e. $Q_t = Q_{t-1} + 1$, the dimensions of $A_t, \mu_t, \sigma_t^2, v_t, M_t$ will change as follows. At first, we adjust $A_t$. Suppose that $A_t^{(0)}$ is the matrix without adjustment while $A_t^{(a)}$ is the matrix adjusted.
According to \( P(Q_{t-1} + 1 \mid Q_{t-1}) \) and the definition of \( A_t^{(a)} \) should satisfy the following conditions:

**Condition 1**, Inheritance

\[
a_{i,j}^* = \frac{(1-a_{i,Q_i})a_{i-1,j}}{\sum_{j=1}^{Q_{i-1}} a_{i,j}}, \quad 1 \leq i, j \leq Q_{i-1}
\]

\[P^{(a)}(S_t = i) = (1-P^{(a)}(S_t = Q_{t-1} + 1))\sum_{j=1}^{Q_{i-1}} \frac{P^{(b)}(S_{t-1} = j)}{\sum_{j=1}^{Q_{i-1}} P^{(b)}(S_{t-1} = j)}
\]

where \( P^{(b)}(S_{t-1} = i), 1 \leq i \leq Q_{t-1} \) and \( P^{(a)}(S_t = i), 1 \leq i \leq Q_{t-1} + 1 \) are the stable distributions corresponding to before and after adjustment, respectively.

**Condition 2**, State transition probability

\[\sum_{i=1}^{Q_{t-1}} a_{i,Q_t}^* = P(Q_t \mid Q_{t-1})
\]

**Condition 3**, Normalization

\[\sum_{j=1}^{Q_{i-1}} a_{i,j}^* = 1, \quad 1 \leq i \leq Q_{t-1}
\]

\[\sum_{j=1}^{Q_{Q_t}} a_{Q_t,j}^* = 1
\]

Obviously, there are \((Q_{t-1}^2 + Q_{t-1} + 2)\) independent equations. As for \( A_t^{(a)} \), the sum of elements in last line is one and there are \((Q_{t-1} + 1)^2\) unknown variables, therefore, it has definitely infinite solutions and has \((Q_{t-1} - 1)\) degree of freedom. A method is given to adjust and the steps are as follows.

**Step 1**, According to the state transition matrix, we assign values to \( a_{1,Q_t}^*, \cdots, a_{Q_{t-1},Q_t}^* \) randomly, and they should meet \( \sum_{i=1}^{Q_{t-1}} a_{i,Q_t}^* = P(Q_{t}+1 \mid Q_{t-1}) \);

**Step 2**, According to (28) and (31), we have \( a_{i,j}^* = (1-a_{i,Q_t}^*)a_{i-1,j}, \quad 1 \leq i, j \leq Q_{t-1}; \)

**Step 3**, According to (29) and (32), there are \( Q_t \) equations and \( Q_t \) variables, so \( a_{Q_t,1}^*, \cdots, a_{Q_t,Q_t}^* \) is a unique solution.
As for solving the stable distribution, we will describe the solution process for $A^{(b)}_t$.

Suppose that $P^{(b)}(S_{t-1} = i), 1 \leq i \leq Q_{t-1}$ is the stable distribution, then

$$
\begin{bmatrix}
P^{(b)}(S_{t-1} = 1) \\
\vdots \\
P^{(b)}(S_{t-1} = Q_{t-1})
\end{bmatrix}^T = \begin{bmatrix}
P^{(b)}(S_{t-1} = 1) \\
\vdots \\
P^{(b)}(S_{t-1} = Q_{t-1})
\end{bmatrix}^T \begin{bmatrix}
a_{1,t} & \cdots & a_{Q_{t-1},t}
\end{bmatrix}
$$

(33)

To solve (33) and we can obtain $P^{(b)}(S_{t-1} = i), 1 \leq i \leq Q_{t-1}$. The solution of $P^{(b)}(S_{t} = i), 1 \leq i \leq Q_{t-1} + 1$ is similar with $P^{(b)}(S_{t-1} = i), 1 \leq i \leq Q_{t-1}$, so it can be omitted.

Now we will discuss how to adjust $\mathbf{M}_t$. Suppose that the mixture components number of the new added hidden state is $M_{t, Q_t}$, then adjust $\mathbf{M}_t$ as follows.

$$
\mathbf{M}^{(a)}_t = \begin{bmatrix}
\mathbf{M}^{(b)}_{t, Q_t}
\end{bmatrix}
$$

(34)

where $M_{t, Q_t}$ is a positive integer generated randomly. $v_t, \mu_t, \sigma_t^2$ are adjusted as follows.

$$
\begin{bmatrix}
v_t^{(a)} \\
\mu_t^{(a)} \\
\sigma_t^{2(a)}
\end{bmatrix} = \begin{bmatrix}
(v_t^{(b)})^T, (V_{t, Q_t, 1} \cdots V_{t, Q_t, M_{t, Q_t}})^T \\
(\mu_t^{(b)})^T, (\mu_{t, Q_t, 1} \cdots \mu_{t, Q_t, M_{t, Q_t}})^T \\
(\sigma_t^{2(b)})^T, (\sigma_{t, Q_t, 1}^2 \cdots \sigma_{t, Q_t, M_{t, Q_t}}^2)^T
\end{bmatrix}
$$

(35) (36) (37)

where $V_{t, Q_t, 1} \cdots V_{t, Q_t, M_{t, Q_t}}, \mu_{t, Q_t, 1} \cdots \mu_{t, Q_t, M_{t, Q_t}}, \sigma_{t, Q_t, 1}^2 \cdots \sigma_{t, Q_t, M_{t, Q_t}}^2$ are constants generated randomly, also $V_{t, Q_t, j}, \mu_{t, Q_t, j}, \sigma_{t, Q_t, j} \geq 0, j = 1 \cdots M_{t, Q_t}$.

(2) The number of hidden states decreases

If $Q_t$ decreases, i.e. $Q_t = Q_{t-1} - 1$, the dimensions of $A_t, \mu_t, \sigma_t^2, v_t, M_t$ will change as follows. When the number of hidden states adjusted before is one, remain this hidden state. Otherwise, delete 1 hidden state according to $P^{(b)}(S_{t-1} = i)$. Suppose that the $i$ th is deleted, then adjust $A_t$ as follows. First delete the elements of $i$ th line and $j$ th column of $A^{(b)}_t$, then normalize the rest lines of $A^{(b)}_t$ and assign into $A^{(a)}_t$. As for adjustment of $\mu_t, \sigma_t^2, v_t, M_t$, delete the elements corresponding to $i$ th hidden state.

Notice that, if the system is in the $i$ th hidden state at time $t-1$ (i.e. $z_{t-1} = i$) and the $i$ th hidden state happens to be deleted, it is impossible to involve $z_t \sim P(z_t | z_{t-1}, A_t)$. In this case, $z_{t-1}$ can be generated by sampling according to $P^{(a)}(S = i), 1 \leq i \leq Q_{t-1} - 1$.

3.4.2. Change of Gaussian mixture components

If the number of Gaussian mixture components changes, the dimensions of $\mu_t, \sigma_t^2, v_t$ will change accordingly. Suppose that the number of Gaussian mixture components corresponding to the $i$ th hidden state changes while the others remain unchanged. For simplicity, we consider
the following three cases: (i) \(M_{t,i} = M_{t-1,i} + 1\); (ii) \(M_{t,i} = M_{t-1,i} + 1\); (iii) \(M_{t,i} = M_{t-1,i} - 1\). As for case (i), we needn’t adjust parameters any more and only study the rest two cases.

1) The number of Gaussian mixture components increases

If the number of Gaussian mixture components increases, i.e. \(Q_t = Q_{t-1} + 1\), the dimensions of \(v_t, \mu_t, \sigma_t^2\) will be adjusted as follows.

\[
\begin{align*}
\mathbf{v}_t^{(a)} &= \begin{bmatrix} \mathbf{v}_t^{(b)} \\ v_{t,i,M_{t,i}} \end{bmatrix}, \\
\mathbf{v}_t^{(a)} &= \begin{bmatrix} \mathbf{v}_t^{(b)} \\ v_{t,i,M_{t,i}} \end{bmatrix}, \\
\mathbf{\mu}_t^{(a)} &= \begin{bmatrix} \mathbf{\mu}_t^{(b)} \\ \mathbf{\mu}_{t,i,M_{t,i}} \end{bmatrix}, \\
\mathbf{\mu}_t^{(a)} &= \begin{bmatrix} \mathbf{\mu}_t^{(b)} \\ \mathbf{\mu}_{t,i,M_{t,i}} \end{bmatrix}, \\
\mathbf{\sigma}_t^{2(a)} &= \begin{bmatrix} \mathbf{\sigma}_t^{2(b)} \\ \mathbf{\sigma}_{t,i,M_{t,i}}^{2(a)} \end{bmatrix}, \\
\mathbf{\sigma}_t^{2(a)} &= \begin{bmatrix} \mathbf{\sigma}_t^{2(b)} \\ \mathbf{\sigma}_{t,i,M_{t,i}}^{2(a)} \end{bmatrix}, \quad (38)
\end{align*}
\]

where \(v_{t,i,M_{t,i}} \geq 0,\) \(\mathbf{\mu}_{t,i,M_{t,i}}\), \(\sigma_{t,i,M_{t,i}}^2 \geq 0\) are constants generated randomly.

2) The number of Gaussian mixture components decreases

If the number of Gaussian mixture components decreases, i.e. \(M_{t,i} = M_{t-1,i} - 1\), the dimensions of \(v_t, \mu_t, \sigma_t^2\) will change accordingly. When the number of Gaussian mixture components adjusted before is one, remain this Gaussian component. Otherwise, delete 1 Gaussian mixture component randomly according to \(v_t^{(b)} = (v_{t,i,1}, \cdots, v_{t,i,M_{t-1},i})^T\).

Suppose the \(j\)th is deleted, then adjust \(v_t, \mu_t, \sigma_t^2\) as follows.

\[
\begin{align*}
\mathbf{v}_t^{(a)} &= \begin{bmatrix} \mathbf{v}_t^{(b)} \\ v_{t,i,j} \end{bmatrix}, \\
\mathbf{v}_t^{(a)} &= \begin{bmatrix} \mathbf{v}_t^{(b)} \\ v_{t,i,j} \end{bmatrix}, \\
\mathbf{\mu}_t^{(a)} &= \begin{bmatrix} \mathbf{\mu}_t^{(b)} \\ \mathbf{\mu}_{t,i,j} \end{bmatrix}, \\
\mathbf{\mu}_t^{(a)} &= \begin{bmatrix} \mathbf{\mu}_t^{(b)} \\ \mathbf{\mu}_{t,i,j} \end{bmatrix}, \\
\mathbf{\sigma}_t^{2(a)} &= \begin{bmatrix} \mathbf{\sigma}_t^{2(b)} \\ \mathbf{\sigma}_{t,i,j}^{2(a)} \end{bmatrix}, \\
\mathbf{\sigma}_t^{2(a)} &= \begin{bmatrix} \mathbf{\sigma}_t^{2(b)} \\ \mathbf{\sigma}_{t,i,j}^{2(a)} \end{bmatrix}, \quad (41)
\end{align*}
\]

\[
\begin{align*}
\mathbf{v}_t^{(a)} &= \begin{bmatrix} \mathbf{v}_t^{(b)} \\ v_{t,i,j} \end{bmatrix}, \\
\mathbf{v}_t^{(a)} &= \begin{bmatrix} \mathbf{v}_t^{(b)} \\ v_{t,i,j} \end{bmatrix}, \\
\mathbf{\mu}_t^{(a)} &= \begin{bmatrix} \mathbf{\mu}_t^{(b)} \\ \mathbf{\mu}_{t,i,j} \end{bmatrix}, \\
\mathbf{\mu}_t^{(a)} &= \begin{bmatrix} \mathbf{\mu}_t^{(b)} \\ \mathbf{\mu}_{t,i,j} \end{bmatrix}, \\
\mathbf{\sigma}_t^{2(a)} &= \begin{bmatrix} \mathbf{\sigma}_t^{2(b)} \\ \mathbf{\sigma}_{t,i,j}^{2(a)} \end{bmatrix}, \\
\mathbf{\sigma}_t^{2(a)} &= \begin{bmatrix} \mathbf{\sigma}_t^{2(b)} \\ \mathbf{\sigma}_{t,i,j}^{2(a)} \end{bmatrix}, \quad (42)
\end{align*}
\]

\[
\begin{align*}
\mathbf{v}_t^{(a)} &= \begin{bmatrix} \mathbf{v}_t^{(b)} \\ v_{t,i,j} \end{bmatrix}, \\
\mathbf{v}_t^{(a)} &= \begin{bmatrix} \mathbf{v}_t^{(b)} \\ v_{t,i,j} \end{bmatrix}, \\
\mathbf{\mu}_t^{(a)} &= \begin{bmatrix} \mathbf{\mu}_t^{(b)} \\ \mathbf{\mu}_{t,i,j} \end{bmatrix}, \\
\mathbf{\mu}_t^{(a)} &= \begin{bmatrix} \mathbf{\mu}_t^{(b)} \\ \mathbf{\mu}_{t,i,j} \end{bmatrix}, \\
\mathbf{\sigma}_t^{2(a)} &= \begin{bmatrix} \mathbf{\sigma}_t^{2(b)} \\ \mathbf{\sigma}_{t,i,j}^{2(a)} \end{bmatrix}, \\
\mathbf{\sigma}_t^{2(a)} &= \begin{bmatrix} \mathbf{\sigma}_t^{2(b)} \\ \mathbf{\sigma}_{t,i,j}^{2(a)} \end{bmatrix}, \quad (43)
\end{align*}
\]
4. Experiment Researches

The crude oil price is influenced by many factors, and most of these factors interact in very complex ways. For this reason, forecasting it through a fundamentalist approach is a difficult task. An alternative is to use time series methodologies, with which the price's past behavior is conveniently analyzed, and used to predict future movements. In this section, we apply the variable dimension RBF-HMM-GMM model to predict future crude oil price movements.

The following trial used the monthly West Texas Intermediate (WTI) crude oil future price (USD/Barrel) series, from Apr. 1983 to Mar. 2012, obtained from the website (http://www.eia.gov/dnav/pet/pet_pri_fut_s1_m.htm). It contains 348 observations, of which the first 333 (training set) were used to estimate the parameter values and the last 15 (test set) to test forecast performances.

A RBF-HMM-GMM model with $k_0 = 5, q_0 = 1, Q_0 = 2, M_{ij} = 1$ initially were applied to predict the data. $P(k_t | k_{t-1}), P(q_t | q_{t-1}), P(q_t | q_{t-1}), P(Q_t | Q_{t-1}), P(M_{ij} | M_{t-1})$ are 0.025, 0.95 and 0.025, respectively (increase, unchanged, decrease). The number of particles was set to 1000.

For the details of online prediction algorithm, refer to Section 3.

For comparison purpose, we consider the following models: the fixed dimension RBF-HMM-GMM model and the trivial model. When the fixed dimension RBF-HMM-GMM model was applied, $k_0 = 5, q_0 = 1, Q_0 = 2, M_{ij} = 1$ are set by trial-and-error with the training set containing the 333 samples from Apr. 1983 to Dec. 2010. The trivial model simply involves using the current value as the next prediction, namely $\hat{y}_{t+1} = y_t$. Table 1 and Figure 5 show the predicted errors of various methods.

It can be seen that the fixed dimension RBF-HMM-GMM model performs better than the trivial approach, while the variable dimension RBF-HMM-GMM performs better than the fixed one, the reason is that the order and parameters of variable dimension RBF-HMM-GMM model are updated whenever a new data is available, which means that it can seize the dynamic attribute of time series.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean absolute Percentage error (MAPE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable dimension of RBF-HMM-GMM model</td>
<td>0.0424</td>
</tr>
<tr>
<td>Fixed dimension of RBF-HMM-GMM model</td>
<td>0.0431</td>
</tr>
<tr>
<td>Trivial</td>
<td>0.0461</td>
</tr>
</tbody>
</table>

Table 1 Predicted errors of various methods in WTI crude oil price

5. Conclusion

In order to cope with difficult nonlinear and non-Gaussian time series, a RBF-HMM-GMM model, which is based on RBF network with the assumption of measurement noise being HMM and the distribution of each hidden state being approximated by GMM, is proposed in this paper. The RBF-HMM-GMM model has three outstanding advantages as follows: (i) Variable dimension RBF networks model is adopted to depict the nonlinear function. (ii) Variable dimension HMM-GMM is used to model the measurement noise. (iii) Predict time series on-line based on RBF-HMM-GMM model using SMC method. At last, the monthly West Texas Intermediate crude oil future price series are analyzed, and we compare the results against trivial model and the fixed RBF-HMM-GMM model. The experimental results show that the variable dimension RBF-HMM-GMM model proposed in this paper is effective.
Acknowledgement

This work is supported by the Education Department Foundation of Jiangsu Province (2010SJD630061), the National Natural Science Foundation of China (71101072), and the Natural Science Foundation of Jiangsu Province (BK2011652).

References