Expert Systems with Applications 36 (2009) 8107–8114

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A distance-based fuzzy time series model for exchange rates forecasting

Yungho Leu *, Chien-Pang Lee, Yie-Zu Jou

Department of Information Management, National Taiwan University of Science and Technology, 43, Keelung Road, Section 4, Taipei 10607, Taiwan, ROC

ARTICLE INFO

Keywords:
Distance-based fuzzy time series
Fuzzy time series
Exchange rate forecasting

ABSTRACT

Fuzzy time series model has been successfully employed in predicting stock prices and foreign exchange rates. In this paper, we propose a new fuzzy time series model termed as distance-based fuzzy time series (DBFTS) to predict the exchange rate. Unlike the existing fuzzy time series models which require exact match of the fuzzy logic relationships (FLRs), the distance-based fuzzy time series model uses the distance between two FLRs in selecting prediction rules. To predict the exchange rate, a two factors distance-based fuzzy time series model is constructed. The first factor of the model is the exchange rate itself and the second factor comprises many candidate variables affecting the fluctuation of exchange rates. Using the exchange rate data released by the Central Bank of Taiwan, we conducted several experiments on exchange rate forecasting. The experiment results showed that the distance-based fuzzy time series outperformed the random walk model and the artificial neural network model in terms of mean square error.

1. Introduction

Forecasting financial time series such as the stock prices or the exchange rates is important to the investors and the government. A good forecasting of a financial time series requires strong domain knowledge and good analysis tools. Many well-established methods, such as autoregressive (AR), autoregressive moving average (ARMA) and generalized autoregressive conditional heteroscedasticity (GARCH), have been successfully applied for financial forecasting. However, there are some limitations on applying the existing methods. For example, in applying ARMA method one needs to make sure that the residue is normally distributed. Recently, due to the advance in artificial intelligence, many researchers have focused on developing computational intelligence methods for forecasting financial time series. For example, Panda and Narasimhan predicted the exchange rate of Rupee and US dollar by artificial neural networks (Panda & Narasimhan, 2007). Cao et al. used support vector machines to predict the USD/GBP exchange rate (Cao, Pang, & Bai, 2005). Shin and Han proposed an integrated approach (Shin & Han, 2000) which uses genetic algorithms to choose the correct threshold parameters for wavelet so as to produce significant signal for a neural network. The integrated approach was used to predict the exchange rate of Korean won/US dollar. The fuzzy time series method is one of the computational intelligence methods that draw much attention today. The Fuzzy time series model was first proposed by Song and Chissom (1993a, 1993b). Since then, many fuzzy time series models have been developed for forecasting index prices (Cheng, Chen, Teoh, & Chiang, 2008; Lee, Wang, Chen, & Leu, 2006; Yu, 2005) and enrollment of universities (Chen, 1996, 2002; Song & Chissom, 1993a, 1993b). Recently, many extensions on fuzzy time series have been proposed. For example, Lee et al. (2006) proposed to allow more than one factor in a fuzzy time series to improve the forecasting accuracy. Yu (2005) proposed to assign different weights on fuzzy logic relationships (FLRs) based on their recency. The weights on fuzzy logic relationships are then used in defuzzification to determine the forecasted value of TAIEX.

The existing models for exchange rate forecasting use only the historical exchange rate itself to build the forecasting model. However, according to the literatures, there are many factors that affect the exchange rate of two currencies. For example, government policies, expectations of the investors, and the stock indices are all shown to affect the fluctuation of the exchange rate (Benjamin, 2006; Nieh & Lee, 2001). Thus, we propose to use the two-factor fuzzy time series to forecast the exchange rate. In an attempt to use the two-factor fuzzy time series model for exchange rate forecasting, we found that it was very difficult to find matching rules for calculating the forecasted value. To remedy this shortcoming, we propose to extend the fuzzy time series model by using the Euclidean distance between two fuzzy logic relationships to find similar fuzzy logic relationships for calculating the forecasted value. The new fuzzy time series model proposed in this paper is termed as distance-based fuzzy time series (DBFTS). DBFTS is a two-factor high-order fuzzy time series model. The first factor of DBFTS is the historical exchange rate, and the second factor comprises several important variables that have an effect of on the exchange rate.

* Corresponding author.
E-mail address: yhl@cs.ntust.edu.tw (Y. Leu).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.10.034

The contributions of this paper are threefold: (1) we propose to use the distance between two FLRs for selecting prediction rules which makes the fuzzy time series model applicable for real life applications in which the exact match of FLRs is not possible; (2) we propose to dynamically adjust the number of intervals of the second factor to improve the prediction accuracy;(3) we propose to use PCA to derive the second factor, which allows incorporating multiple influential variables into the second factor.

The remainder of this paper is organized as follows. Section 2 reviews several definitions of fuzzy time series models. Section 3 introduces the DBFTS model. Section 4 presents the procedure of using DBFTS in exchange rate forecasting. In Section 5, we report the results of several experiments and compare the performance of DBFTS, random walk model and RBFNN model in terms of mean square error (MSE) and directional symmetry (DS). Finally, we conclude this paper in Section 6.

2. Review of fuzzy time series models

According to the literature, the following definitions are given for fuzzy time series models.

Definition 1. Let $Y(t)$ ($t = \ldots,0,1,2,\ldots$), a subset of R^1, be the universe of discourse on which fuzzy sets $f_i(t)$ ($i = 1,2,\ldots$) are defined. If $F(t)$ is a collection of $f_i(t)$. Then $F(t)$ is called a fuzzy time series defined on $Y(t)$.

Definition 2. If for any $f_j(t) \in F(t)$, there exists an $f_i(t-1) \in F(t-1)$, such that there exists a fuzzy relation $R_{ij}(t,t-1)$ and $f_j(t) = f_i(t-1) \circ R_{ij}(t,t-1)$ where '$\circ$' is the max–min composition, then $F(t)$ is said to be caused by $F(t-1)$ only. Let $R(t,t-1) = \cup_{ij} R_{ij}(t,t-1)$ where '$\cup$' is the union operator. Then $R(t,t-1)$ is called the fuzzy relation between $F(t)$ and $F(t-1)$. We can write $F(t) = F(t-1) \circ R(t,t-1)$, and denote this as $F(t-1) \to F(t)$.

Definition 3. If $F(t)$ is caused by $F(t\text{-}1), F(t-2), \ldots$, and $F(t-n)$, $F(t)$ is called a one-factor n-order fuzzy time series, and is denoted by

$$F(t-n), \ldots, F(t-2), F(t-1) \to F(t).$$

Based on the one-factor n-order fuzzy time series, Lee et al. (2006) proposed the two-factor high-order fuzzy time series as is defined in the following.

Definition 4. If $F_1(t)$ is caused by $(F_1(t-1), F_2(t-1))$, $(F_1(t-2), F_2(t-2)), \ldots, (F_1(t-n), F_2(t-n))$, $F_1(t)$ is called a two-factor n-order fuzzy time series, which is denoted by

$$(F_1(t-n), F_2(t-n)), \ldots, (F_1(t-2), F_2(t-2)), (F_1(t-1), F_2(t-1)) \to F_1(t).$$

Let $F_1(t) = X_t$ and $F_2(t) = Y_t$, where X_t and Y_t are fuzzy sets on day t. Then, a two-factor n-order fuzzy relationship can be expressed as

$$(X_{t-n}, Y_{t-n}), \ldots, (X_{t-2}, Y_{t-2}), (X_{t-1}, Y_{t-1}) \to X_t,$$

where $(X_{t-n}, Y_{t-n}), \ldots, (X_{t-2}, Y_{t-2})$ and (X_{t-1}, Y_{t-1}), are referred to as the left-hand side (LHS) of the relationship, and X_t is referred to as the right-hand side (RHS) of the relationship. A fuzzy relationship is also called a fuzzy logic relationship (FLR) in Chen (1996) and Song and Chissom (1993a, 1993b).

3. Distance-based fuzzy time series

As discussed in Section 1, the fluctuation of exchange rates in an exchange rate market is affected by many important variables. In the beginning, we employed the two-factor high-order fuzzy time series proposed by Lee et al. (2006) to forecast the exchange rates of Taiwan exchange rate market. Unfortunately, there are to too many fuzzy sets at the LHS of an FLR, it was very difficult to find a matched FLR for predicting the exchange rate. This motivates us to propose the DBFTS model which features in selecting forecasting FLRs based on a distance metric defined on the LHS of an FLR, instead of a complete match of the LHS. The details of the DBFTS are explained in the following procedure. Note that different from the Lee's model, our approach contains two more steps for determining the second factor.

3.1. Step 1: Test of correlation coefficient

As defined in Definition 4, the first factor is caused by the first and the second factors. Furthermore, the second factor may be derived from many candidate variables. To decide whether a candidate variable is suitable for a component of the second factor, one has to test the correlation between the candidate variable and the first factor. If the correlation coefficient of the first factor and the candidate variable is not significant, the candidate variable should not be considered in the derivation of the second factor.

3.2. Step 2: Principal components analysis

After testing of correlation coefficient, the second factor can be expressed as a linear combination of the candidate variables through principal components analysis (PCA). For exchange rate forecasting, the second factor is the absolute score value of the first principal component, which is denoted by $|PRIN1|$. Note that usually more than one component is needed for the second factor. However, for exchange rate forecasting, the first component has already explained 50–70% of the proportion of variance; we therefore choose only the first component as the second factor.

3.3. Step 3: Divide the universe of discourse

The universe of discourse of the first factor is defined as $U = [D_{\min} - D_1, D_{\max} + D_2]$, where $D_{\min}$ and $D_{\max}$ are the minimum and maximum values of the first factor, respectively; D_1 and D_2 are two positive real numbers to divide the universe of discourse into n equal length intervals (Song & Chissom, 1994). The universe of discourse of the second factor is defined as $V = [V_{\min} - V_1, V_{\max} + V_2]$, where $V_{\min}$ and $V_{\max}$ are the minimum and maximum values of the second factor, respectively; V_1 and V_2 are two positive real numbers used to divide the universe of discourse of the second factor into m equal length intervals.

3.4. Step 4: Define fuzzy sets

Linguistic term A_i, $1 \leqslant i \leqslant n$, is defined as a fuzzy set on the intervals of the first factor. A_i is defined as follows:

$$\begin{aligned}
A_1 &= 1/u_1 + 0.5/u_2 + 0/u_3 + \cdots + 0/u_{n-2} + 0/u_{n-1} + 0/u_n,\\
A_2 &= 0.5/u_1 + 1/u_2 + 0.5/u_3 + \cdots + 0/u_{n-2} + 0/u_{n-1} + 0/u_n,\\
&\vdots\\
A_{n-1} &= 0/u_1 + 0/u_2 + 0/u_3 + \cdots + 0.5/u_{n-2} + 1/u_{n-1} + 0.5/u_n,\\
A_n &= 0/u_1 + 0/u_2 + 0/u_3 + \cdots + 0/u_{n-2} + 0.5/u_{n-1} + 1/u_n.
\end{aligned}$$

Similarly, linguistic term B_j, $1 \leqslant j \leqslant m$, is defined as a fuzzy set on the intervals of the second factor. B_j is defined as follows: