Expert Systems with Applications 36 (2009) 7424–7434

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A new hybrid approach based on SARIMA and partial high order bivariate fuzzy time series forecasting model

Erol Egrioglu [a], Cagdas Hakan Aladag [b,*], Ufuk Yolcu [a], Murat A. Basaran [c], Vedide R. Uslu [a]

[a] Department of Statistics, Ondokuz Mayis University, Samsun 55139, Turkey
[b] Department of Statistics, Hacettepe University, Ankara 06800, Turkey
[c] Department of Mathematics, Nigde University, Nigde 51000, Turkey

ARTICLE INFO

Keywords:
Bivariate
Box-Jenkins method
Feed forward neural networks
Forecasting
High order
Seasonal fuzzy time series

ABSTRACT

In the literature, there have been many studies using fuzzy time series for the purpose of forecasting. The most studied model is the first order fuzzy time series model. In this model, an observation of fuzzy time series is obtained by using the previous observation. In other words, only the first lagged variable is used when constructing the first order fuzzy time series model. Therefore, this model can not be sufficient for some time series such as seasonal time series which is an important class in time series models. Besides, the time series encountered in real life have not only autoregressive (AR) structure but also moving average (MA) structure. The fuzzy time series models available in the literature are AR structured and are not appropriate for MA structured time series. In this paper, a hybrid approach is proposed in order to analyze seasonal fuzzy time series. The proposed hybrid approach is based on partial high order bivariate fuzzy time series forecasting model which is first introduced in this paper. The order of this model is determined by utilizing Box-Jenkins method. In order to show the efficiency of the proposed hybrid method, real time series are analyzed with this method. The results obtained from the proposed method are compared with the other methods. As a result, it is observed that more accurate results are obtained from the proposed hybrid method.

1. Introduction

Zadeh (1965) firstly introduced fuzzy set theory. Based on this paper, fuzzy set theory has found many application areas in science. Fuzzy time series approach based on fuzzy set theory was introduced as an alternative method for conventional time series models. Recently, fuzzy time series has got much attention. Song and Chissom (1993a, 1993b) first introduced fuzzy time series. These studies have been inspired by knowledge presented in the papers Zadeh (1973, 1975). Song and Chissom proposed a method based on matrix operations in order to forecast fuzzy time series (1993a, 1993b, 1994). Chen (1996) introduced a new method, which is computationally easier than the method of Song and Chissom, by employing the fuzzy logical relationship table. Huarng and Yu (2006) proposed another method in which the fuzzy logical relationships are determined by artificial neural networks. The other important studies related to fuzzy time series are conducted by Sullivan and Woodall (1994); Hwang, Chen, and Lee (1998); Chen and Hwang (2000); Huarng and Yu (2005); and Yu (2005a, 2005b). All of these studies are based on first order fuzzy time series models. Taking into account the definition of the first order fuzzy time series model, it should be only used for AR(1) structured time series (Yu & Huarng, 2008). However, the time series faced in real life can not be expected to have only AR(1) structure. Many time series can be forecasted using high order relations such as AR(2), AR(3) and MA(2), MA(3). Moreover, it is possible to model seasonal time series when high order models are used. These high order models generally includes lagged variables which are corresponding to the period number. Chen (2002) proposed a high ordered time series method to analyze time series with AR(1) structured. In Chen's method, high order models include all lagged fuzzy variables. That is, in the analysis of fourth ordered time series, first, second, third, and fourth ordered fuzzy lagged variables exist in the model. Thus, the high order fuzzy time series model proposed by Chen can model the AR(p) ($p > 1$) structured time series, too. It is known that models containing lagged variables, which are defined with respect to the number of period, are quite successful in real seasonal time series. In Chen's method, for the time series whose period is 12, if 12th order model is employed, the number of the inputs of the model increases. In addition, there is no method for the selection of order of the model in Chen's method. Chen determined the order by trail and error. On the other hand, some methods have been proposed to determine the order of the model in the time series literature. Song (1999) proposed a method to analyze seasonal fuzzy time series. Let the period of

* Corresponding author. Tel.: +90 312 2977900; fax: +90 312 2977913.
E-mail address: chaladag@gmail.com (C.H. Aladag).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.09.040

the time series be m, the input and the output are $F(t-m)$ and $F(t)$, respectively, in Song's model. The lagged variable corresponding to the number of period is used in Song's model. Therefore, the model resembles to the first order seasonal autoregressive SAR(1) model. Real time series data rarely bare this structure. There exist more complicated time series data so that analyzing of them requires models such as either SAR(2), SAR(3), ... or SMA(1), SMA(2), ext.

A new hybrid method is proposed in order to analyze fuzzy seasonal time series in this paper. The proposed model contains terms of AR, SAR structure and MA, SMA structure. Namely, the inputs of model include both the terms of time series (X_t) and the terms of error (a_t) series. This situation requires that bivariate case should be employed. Bivariate case approach was previously introduced by Huarng (2001); Hsu, Tse, and Wu (2003) and Yu and Huarng (2008). The method employed in this study is similar to one used in Yu and Huarng (2008). The first order lagged variable is only used in Huarng's method. However, the method proposed inheres added the high ordered terms in the model. Yu and Huarng (2008) employed feed forward artificial neural network to establish the fuzzy relationships. In this paper, feed forward artificial neural network is employed to determine the fuzzy relationships. The proposed approach has two phases. The first phase includes analyzing the time series using Box-Jenkins method. In the second phase, partial high order bivariate fuzzy time series model is constructed based on using the inputs exist in the SARIMA model determined in the previous phase. Partial high order bivariate fuzzy time series forecasting model is firstly introduced in this study. The advantages of this hybrid proposed method can be summarized as follows:

- The proposed method produces high order seasonal fuzzy time series.
- Determination of order of the model is done systematically.
- The proposed model is the first fuzzy time series model containing MA term.
- The proposed method produces much more accurate forecasts.

In the next section, the fundamental fuzzy time series definition is given. Section 3 gives brief knowledge about artificial neural networks and SARIMA models. In addition, the proposed fuzzy time series methods for forecasting are summarized. The new hybrid proposed method is given in detail in Section 4. Based on two real time series, the results of the implementation of the proposed method are given. Section 6 is the last part that gives the summarized results and discussion.

2. Fuzzy time series

The definition of fuzzy time series was firstly introduced by Song and Chissom (1993a, 1993b). In contrast to conventional time series methods, various theoretical assumptions do not need to be checked in fuzzy time series approach. The most important advantage of fuzzy time series approach is to be able to work with a very small set of data and not to require the linearity assumption. General definitions of fuzzy time series are given as follows:

Let U be the universe of discourse, where $U = \{u_1, u_2, \ldots, u_b\}$. A fuzzy set A_i of U is defined as $A_i = f_{A_i}(u_1)/u_1 + f_{A_i}(u_2)/u_2 + \cdots + f_{A_i}(u_b)/u_b$, where f_{A_i} is the membership function of the fuzzy set A_i; $f_{A_i} : U \to [0,1]$. u_a is a generic element of fuzzy set A_i; $f_{A_i}(u_a)$ is the degree of belongingness of u_a to A_i; $f_{A_i}(u_a) \in [0,1]$ and $1 \leqslant a \leqslant b$.

Defnition 1. Fuzzy time series. Let $Y(t)(t = \ldots, 0, 1, 2, \ldots)$, a subset of real numbers, be the universe of discourse by which fuzzy sets $f_j(t)$ are defined. If $F(t)$ is a collection of $f_1(t), f_2(t), \ldots$ then $F(t)$ is called a fuzzy time series defined on $Y(t)$.

Defnition 2. Fuzzy time series relationships assume that $F(t)$ is caused only by $F(t-1)$, then the relationship can be expressed as: $F(t) = F(t-1) * R(t, t-1)$, which is the fuzzy relationship between $F(t)$ and $F(t-1)$, where $*$ represents as an operator. To sum up, let $F(t-1) = A_i$ and $F(t) = A_j$. The fuzzy logical relationship between $F(t)$ and $F(t-1)$ can be denoted as $A_i \to A_j$ where A_i refers to the left-hand side and A_j refers to the right-hand side of the fuzzy logical relationship. Furthermore, these fuzzy logical relationships can be grouped to establish different fuzzy relationships.

The definition of the first order seasonal fuzzy time series model for forecasting proposed by Song (1999) is given as follows:

Defnition 3. Let $F(t)$ be a fuzzy time series. Assume there exists seasonality in $\{F(t)\}$, first order seasonal fuzzy time series forecasting model:

$$F(t-m) \to F(t) \tag{1}$$

where m denotes the period.

The high order fuzzy time series model proposed by Chen (2002) is given as follows:

Defnition 4. Let $F(t)$ be a fuzzy time series. If $F(t)$ is caused by $F(t-1), F(t-2), \ldots,$ and $F(t-n)$, then this fuzzy logical relationship is represented by

$$F(t-n), \ldots, F(t-2), F(t-1) \to F(t) \tag{2}$$

and it is called the nth order fuzzy time series forecasting model.

Bivariate fuzzy time series model defined by Yu and Huarng (2008) is given as follows:

Defnition 5. Let F and G be two fuzzy time series. Suppose that $F(t-1) = A_i$, $G(t-1) = B_k$ and $F(t) = A_j$. A bivariate fuzzy logical relationship is defined as $A_i, B_k \to A_j$, where A_i, B_k are referred to as the left hand side and A_j as the right hand side of the bivariate fuzzy logical relationship.

Therefore, first order bivariate fuzzy time series forecasting model is as follows:

$$F(t-1), G(t-1) \to F(t)$$

3. Preliminary

In this section, artificial neural networks and SARIMA models are briefly given as basis of the proposed method. Also, some important fuzzy time series methods available in the literature are summarized.

3.1. Artificial neural networks

Artificial neural networks are algorithms which mimic the features of brain of human being. These features are generating new knowledge and exploring by learning. In other words, artificial neural networks are synthetic networks which imitate biological neural networks. Artificial neural networks are much more different than biological ones in terms of structure and ability, (Zurada, 1992). Artificial neural networks compose of a mathematical model (Zhang, Patuwo, & Hu, 1998).

The fundamental elements of artificial neural networks are given below:

Architecture structure: The structure of multilayer feed forward artificial neural networks is basically given in Fig. 1. Multilayer feed forward artificial neural networks as illustrated in the figure consist of three parts such as input, hidden, and output layers. Each layer consists of neurons. The architecture structure is determined based on deciding the number of neuron in each layer. These neu-