Expert Systems with Applications 36 (2009) 7888–7897

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A hybrid multi-order fuzzy time series for forecasting stock markets

Hia Jong Teoh [a,*], Tai-Liang Chen [b], Ching-Hsue Cheng [c], Hsing-Hui Chu [a]

[a] Department of Accounting & Information Technology, Ling Tung University, 1, Ling Tung Road, Nantun, Taichung 408, Taiwan, ROC
[b] Department of Information Management and Communication, WenZao Ursuline College of Language, 900, Min-Tzu 1st Road Kaohsiung, 807 Taiwan, ROC
[c] Department of Information Management, National Yunlin University of Science and Technology, 123, Section 3, University Road, Touliu, Yunlin 640, Taiwan, ROC

ARTICLE INFO

Keywords:
Multi-order fuzzy time series
Rough sets theory
Fuzzy logical relationships (FLR)

ABSTRACT

This paper proposes a hybrid model based on multi-order fuzzy time series, which employs rough sets theory to mine fuzzy logical relationship from time series and an adaptive expectation model to adjust forecasting results, to improve forecasting accuracy. Two empirical stock markets (TAIEX and NASDAQ) are used as empirical databases to verify the forecasting performance of the proposed model, and two other methodologies, proposed earlier by Chen and Yu, are employed as comparison models. Besides, to compare with conventional statistic method, the partial autocorrelation function and autoregressive models are utilized to estimate the time lags periods within the databases. Based on comparison results, the proposed model can effectively improve the forecasting performance and outperforms the listing models. From the empirical study, the conventional statistic method and the proposed model both have revealed that the estimated time lags for the two empirical databases are one lagged period.

1. Introduction

In the area of time series research, Kuo, Chen, and Hwang (2001) has demonstrated that general techniques employed for stock prediction are mathematical and statistical models such as time series analysis (Kendall & Ord, 1990) and multiple regression models. Because the above models are stiffed and complicated to understand by stock investors, other following researchers have presented different forecasting models to solve time series forecasting problems with fuzzy theory. Tanaka (1987) first applied linear programming to model fuzzy regression, Watada (1992) applied fuzzy regression to deal with fuzzy time series forecasting problems, Tseng et al. extended fuzzy regression to autoregressive integrated moving average (ARIMA) analyses (Tseng & Tzeng, 2002; Tseng, Tzeng, Yu, & Yuan, 2000), and Song and Chissom proposed fuzzy time series model for university enrollment forecasting (Song & Chissom, 1993a, 1993b).

Since Song and Chissom (1993) proposed the original model (Song & Chissom, 1994), fuzzy time series model has been employed to deal with various domains forecasting problems, such as university enrollment forecasting (Chen, 1996, 2002; Chen & Chung, 2006; Chen & Hsu, 2004), temperature forecasting (Chen & Hwang, 2000) and stock index forecasting (Chen, Cheng, & Teoh, 2007; Cheng, Chen, & Chiang, 2006; Huarng, 2001a, 2001b; Huarng & Yu, 2003, 2004, 2005, 2006). In stock price forecasting, Huarng (2001a) proposed heuristic models to improve stock index forecasting performance by integrating problem-specific heuristic knowledge with Chen's (1996) model. In Huarng's following research, he claimed that the length of linguistic intervals for the universe of discourse may affect forecasting results and proposed two linguistic interval partitioning approaches, distribution-based and average-based length, to approach this issue (Huarng, 2001). In 2005, Yu proposed a weighted method to tackle recurrence and weighting in fuzzy time series forecasting (Yu, 2005). Yu argued that recurrent fuzzy relationships, which were simply ignored in Chen's (1996) studies, should be considered in forecasting and recommended that different weights should be assigned to various fuzzy relationships. Therefore, Cheng et al. (2006) proposed a trend-weighted model to echo Yu's research.

From the literature above, determining the length of linguistic intervals and mining fuzzy logical relationships (FLR) from time series are considered as important processes influencing the forecasting accuracy of fuzzy time series models. In this paper, we argue that a thoughtful fuzzy time series model should consider non-linear relationships in past periods and linear relationships between recent periods. Therefore, we propose a hybrid model, which employs rough set LEM2 (Grzymala-Busse, 1997) algorithm to mine non-linear relationships from time series and adaptive expectation model to modify forecasting results to enhance forecasting accuracy.

In empirical analysis, various linguistic values are used to evaluate and determine the efficacy one for the proposed model because the length of linguistic intervals may affect forecasting results from Huarng research. Additionally, two large scale of stock databases, the TAIEX (Taiwan Stock Exchange Capitalization

* Corresponding author. Tel.: +886 922599714; fax: +886 423827085.
E-mail address: g9320824@yuntech.edu.tw (H.J. Teoh).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.11.009

Weighted Stock Index), and NASDAQ (National Association of Securities Dealers Automated Quotations system), from 1990 to 1999, are used as experimental datasets To evaluate the performance of the proposed model, two recent fuzzy time series models, Chen's (1996) and Yu's (2005) are employed as comparison models, and, to validate the proposed model, we use two statistic methods, partial autocorrelation function and autoregressive model (Montgomery, Runger, & Hubele, 2004; Yang, 2005) to estimate the time lags periods of two stock markets.

The rest of this paper is organized as follows: Section 2 briefly review the definitions of fuzzy time series and the basic concepts of rough sets theory. Section 3 demonstrates the framework and algorithm of the proposed model. Section 4 presents empirical analyzes and models comparisons, and concluding remarks are given in final section.

2. Related works

This section will briefly review the definitions of fuzzy time series and the basic concepts of rough sets theory.

2.1. Fuzzy time series

Fuzzy time series forecasting emerged as a popular approach for predicting the future values in a situation where neither a trend is viewed nor a pattern in variations of time series are visualized and moreover the information (data) are imprecise and vague. Fuzzy theory, the modern concept of uncertainty, was introduced by Zadeh (1965) to deal with linguistic terms (Zadeh, 1976; Zadeh, Klir, & Yuan, 1996). The membership in a fuzzy set is not a matter of affirmation or denial, but rather a matter of degree.

A fuzzy set is a class of objects with a continuum of grade of membership. Let U be the universe of discourse with $U = \{u_1, u_2, \ldots u_n\}$, where u_i are possible linguistic values of U. A fuzzy set A defined in the universe of discourse U can be represented as follows (Chen, 1996; Chen & Chung, 2006).

Definition 1: $Y(t)$ ($t = \ldots 0, 1, 2, \ldots$) is a subset of a real number, be the universe of discourse defined by the fuzzy sets $f_i(t)$. If $F(t)$ consists of $f_i(t)$ ($i = 1, 2, \ldots$), $F(t)$ is defined as a fuzzy time series on $Y(t)$ ($t = \ldots, 0, 1, 2, \ldots$).

Definition 2: If there exist a fuzzy relationship $R(t-1, t)$, such that $F(t) = F(t-1) \times R(t-1, t)$, where $\times$ represents an operation, then $F(t)$ is said to be caused by $F(t-1)$, The fuzzy relationship between $F(t-1)$ and $F(t)$ is denoted as $F(t-1) \rightarrow F(t)$.

Definition 3: Let $F(t-1) = A_i$ and $F(t) = A_j$. The fuzzy relationship between two consecutive observations, $F(t)$ and $F(t-1)$, referred to as a fuzzy logical relationship (FLR), can be denoted by A_i to A_j, where A_i is called the left-hand side (LHS) and A_j the right-hand side (RHS) of the FLR.

Definition 4: Assume that $F(t)$ is a fuzzy time series and $F(t)$ is caused by $F(t-1)$, $F(t-2)$, and $F(t-n)$, then the FLR can be represented as following: $F(t-1), F(t-2), \ldots, F(t-n) \rightarrow F(t)$. This expression is called the nth-order fuzzy time series forecasting model, where $n \geqq 2$.

2.2. Rough sets theory

Rough sets theory is a non-parametric technique originated by Pawlak (1991). This approach seems to be of fundamental importance to artificial intelligence and cognitive science, especially in the areas of machine learning, decision analysis, credit fraud detection, stock market rule-generation, climate change, expert systems, decision support systems, inductive reasoning, pattern recognition and knowledge discovery from databases (Pawlak, 1982; Pawlak & Skoworn, 2007; Shyng, Wang, Tzeng, & Wu, 2007; Witlox & Tindemans, 2004). Rough sets theory is useful for handling dataset that (1) do not require any preliminary or additional information about data; (2) can work with missing values; (3) offers the ability to handle large amounts of both quantitative and qualitative data; and (4) can model highly non-linear or discontinuous functional relationships provides a powerful method for characterizing complex, multidimensional patterns (Pawlak, 1997; Pawlak & Skoworn, 2007).

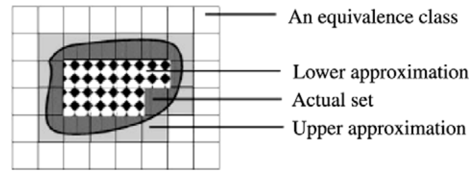

Fig. 1. Basic notions of rough-set.

Table 1
Example of accident cases with describing features.

Case	Driver's age	Vehicle type	Climate	Accident type
1	Young	Motorcycle	Sunny	Off-road
2	Old	Automobile	Sunny	Off-road
3	Young	Motorcycle	Sunny	Rollover
4	Middle-aged	Motorcycle	Sunny	Rollover
5	Middle-aged	Automobile	Rainy	Rollover

Rough sets theory is a new mathematical approach to intelligent data analysis and data mining. It is based on the premise that lowering the degree of precision in the data makes the data pattern more visible, whereas the central premise of rough sets theory philosophy is that the knowledge consists in the ability of classification (Slowinski, 1992). It is founded on the assumption that with every object of the universe of discourse is associated with some information (data, knowledge) and objects that are associated with same information are indiscernible (similar) and belonged to the same class. Concepts such as indiscernibility relations, *lower* and *upper* approximations, and inducts are used to extract classifying rules (Guo & Chankong, 2002). With any rough set, the lower approximation consists of all objects which surely belong to the set and the upper approximation contains all objects which possibly belong to the set. The difference between the upper and the lower approximation constitutes the boundary region of the rough set. Approximations are two basic operations in rough sets theory. The basic notions in rough set are shown in Fig. 1 (Xu, Zhou, & Lu, 2005).

Rough sets theory is a series of logical reasoning procedures for analyzing an information system, an information system can be seen as a decision table denoted by $S = (U, A, C, D)$ where U is universe of discourse, A is a set of primitive features, and $C, D \subset A$ are two subsets of features, assuming that $A = C \cup D$ and $C \cap D = \emptyset$, where C is called condition attribute, and D as decision attribute. Table 1 illustrated an accident occurrence example decision table (Wong & Chung, 2007), there are five cases characterized with three condition attributes: driver's age, vehicle type and climate, and one decision attribute: accident type. Which can be formed into four elementary sets {1,3}, {2}, {4}, and {5}, cases 1 and 3 are indiscernible. Therefore, the rollover accident type is described with the lower approximation set, {4,5}, and the upper approximation set, {1,3,4,5}. Similarly, the concept of the off-road accident type is characterized by its lower approximation set, {2}, and upper approximation set, {1,2,3}.

3. Proposed model

This section consists of two subsections: Section 3.1 describes the concepts of the proposed model, and Section 3.2 details the framework of the proposed model and algorithm.