African Journal of Business Management Vol.5 (13), pp. 5188-5198, 4 July, 2011
Available online at http://www.academicjournals.org/AJBM
DOI: 10.5897/AJBM10.1422
ISSN 1993-8233©2011 Academic Journals

Full Length Research Paper

A Novel Price-Pattern Detection Method Based on time series to forecast stock market

Tai-Liang Chen[1]*, Chung-Ho Su[2,3], Ching-Hsu Cheng[2] and Hung-Hsing Chiang[2]

[1]Department of Information Management and Communication, Wenzao Ursuline College of Languages, 900, Mintsu 1st Road, Kaohsiung 807, Taiwan.
[2]Department of Information Management, National Yunlin University of Science and Technology,123, section 3, University Road, Touliu, Yunlin 640, Taiwan.
[3]Department of Digital technology and Game Design, Shu-Te University, 59 Hun Shan Road, Yen Chau, Kaohsiung County 82445, Taiwan.

Accepted 14 January, 2011

In stock markets, many types of time series models such as statistical time series model, fuzzy time series model, and advanced time series model based on artificial intelligence algorithms were advanced by academic researchers to forecast stock price. Some drawbacks are issued for these models as follows: (1) mathematical assumptions are required for statistical time series models; (2) the forecast from fuzzy time series model is a linguistic value that is not as accurate as statistical time series; and (3) a proper threshold is not easy to be produced by advanced time series model and the forecasting algorithm is unintelligible. To deal with these problems, we propose a novel price-pattern detection method to look for certain price-patterns ("price trend" and "price variation") contained in time series variables that can be used to forecast stock market. From model verification using a nine-year period of Taiwan stock market index (TAIEX) as experimental datasets, it is shown that the proposed model outperforms three listing fuzzy time series (Su et al.,2010; Huarng and Yu ,2006; Chen,1996), and statistic time series models (AR(1),AR(2) and ARMA(1,1)).

Key words: Stock forecasting, price-pattern, time series, fuzzy time series.

INTRODUCTION

Many stock investors have made a great loss because they made wrong judgment on stock price trend. In most cases, they have not enough professional knowledge to analyze market trend or useful forecasting tools to analyze stock market. Therefore, for stock market participators, finding a forecasting tool that can predict future trend accurately or a model in which most of stock observations behave has been regarded as a path to make rich.

For academic researchers, stock price forecasting is one of popular research issues and time series model is one of major forecasting approaches to predict stock market. Time series forecasting is to build a model with known past observations to forecast future events. In past literature of stock forecasting, many types of time series models were proposed to forecast stock market and we summarized them as three types of models: (1) statistical time series model (Hanke and Reitsch,1995; Bollerslev, 1986; Engle, 1982); (2) fuzzy time series model (Chen et al., 2008; Chen et al.,2007; Cheng et al., 2006;; Huarng and Yu, 2005; Yu, 2005; Huarng, 2001; Hwang et al.,1998; Chen,1996; Song and Chissom, 1993); and (3) advanced fuzzy time series model based on artificial intelligence algorithm (Su et al., 2010; Cheng et al.,2010; Teoh et al., 2009; Huarng and Yu , 2006; Chen and Chung, 2006; Pai PF and Lin CS, 2004). Statistical time series model is the original model and it employs mathematic formula based on statistical theory to model time series observations. As artificial intelligence (AI) arising in recent scientific research, many advanced theories and algorithms (that is fuzzy set, neural network, genetic algorithm and rough set algorithm) were applied in time series to deal with complex and non-linear relationships within stock market. Fuzzy time series

*Corresponding author: E-mail: 97007@mail.wtuc.edu.tw Tel: +886-920975168.

applied the fuzzy sets (Zadeh, 1975a, b, 1976) to convert numeric observations into linguistic observations and extracted non-linear relationships, fuzzy logical relationship (Song and Chissom, 1993), from historical time series observations as a rule base to forecast future events. Besides, to improve forecasting performance of fuzzy time series, some AI-based algorithms such as neural network (Huarng and Yu, 2006), genetic algorithm (Cheng et al., 2010; Chen and Chung, 2006) and rough set algorithm (Su et al., 2010; Teoh et al., 2009) were applied to refine fuzzy time series model as advanced fuzzy time series model.

After reviewing the forecasting principles of the past models, three disadvantages are addressed: (1) specific statistical assumptions about observations (stock price) are made for statistical time series and the model is represented as mathematic formula that is quite unreadable for common investor; (2) the forecast from fuzzy time series models is a linguistic value and the defuzzification method such as "centroid" method (Chen,1996) to defuzzify the linguistic value usually produce a rough forecast that is not as accurate as statistical time series ; and (3) the optimal threshold to reach the best performance for AI-based time series is generated with a great effort and the forecasting algorithm is unintelligible.

To avoid the disadvantages, this paper has proposed a novel price-pattern detection method to extract "occurred patterns" from time series observations to forecast stock market. The proposed method is to analyze present stock price-pattern, defined as "price trend" and "price variation" in this paper, and measure the similarity with past "occurred patterns" in history stock data. The stock price-patterns with higher similarity are used the pattern basis for forecasting. Additionally, to promote forecasting accuracy, an adaptive model (Chen et al., 2008) is utilized in forecasting process of the proposed model. With the price-pattern detection method and the adaptive model, we argue that the proposed model could produce more "understandable" and "accurate" forecasts than past models.

LITERATURE REVIEW

Statistical time series

Time series forecasting is the use of a model to forecast future events based on known past events: to predict data points before they are measured. An example of time series forecasting in econometrics is predicting the opening price of a stock based on its past performance. Models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are: (1) the *autoregressive* (AR) models, (2) the *integrated* (I) models, and (3) the *moving average* (MA) models.

The simplest correlated stationary processes are autoregressive processes, where Y_t is modeled as a weighted average of past observations plus a white noise "error," which is also called the "noise" or disturbance." Starting with AR (1) processes, the simplest autoregressive processes is given by following statement (Ruppert, 2011).

Let $\varepsilon_1, \varepsilon_2, ...\varepsilon_t$ be white noise with a mean of "0" and a variance of σ_ε . Y_1, Y_2, …,Y_t is an *AR(1)* process by below equation if for some constant parameters μ and ϕ, for all *t.*

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t$$

where the parameter μ is the mean of the process.

A process Y_t is a *moving average process* if Y_t can be expressed as a weighted average (moving average) of the past values of the white noise process ε_t. The *MA* (*1*) (moving average of order 1) process is defined by the equation (Ruppert, 2011):

$$Y_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1}$$

Where as before the white noise with a mean of "0" and a variance of σ_ε .

The three classes of models, the AR models, the *integrated* (I) models, and the MA models, depend linearly (Gershenfeld, 1999) on previous data points. Combinations of these above models produce time series models such as autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA). Box and Jenkins (1976) developed a general linear stochastic model which supposes a time series to be generated by a linear aggregation of random shocks (Box and Jenkins, 1976). The important property of the resulting ARMA models is that it perform forecast at linear stationary condition (Box and Jenkins,1976). An *ARMA* (*p*, *q*) model combines both *AR* and *MA* terms and is defined by the equation, which shows how Y_t depends on lagged values of itself and lagged values of the white noise process (Ruppert, 2011).

$$(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + ... + \phi_p(Y_{t-p} - \mu) + \varepsilon_t + \theta\varepsilon_{t-1} + ... + \theta\varepsilon_{t-q}$$

Nevertheless, many empirical time series behave as through them had no fixed mean. Models that describe such homogeneous non-stationary behavior can be obtained by supposing some suitable difference of the process to be stationary. The properties of the important class of models for which the d -*th* difference is a stationary mixed autoregressive-moving average process. These models are called Autoregressive Integrated Moving Average (ARIMA) processes (Box and Jenkins, 1976). Traditionally, the ARIMA model has been one of the most widely used linear models in time series forecasting.

Non-linear dependence of the level of a series on previous data points is of interest, partly because of the possibility of producing a chaotic time series. Among other types of non-linear time series models, there are models to represent the changes of variance along time that is called "heteroskedasticity". Engle (1982) suggested the ARCH (Autoregressive Conditional Heteroscedasticity) to model time series variables and it has played an important role in financial analysis (Engle,1982). The GARCH (Generalized ARCH) model is generalized from ARCH (Bollerslev, 1986). The collection of ARCH comprises a wide variety of representation such as GARCH, TARCH, EGARCH, FIGARCH, and CGARCH.

Fuzzy time series

Time series models had failed to consider the application of fuzzy sets (Zadeh,1975a,b, 1976) until fuzzy time series was advanced by Song and Chissom (1993). They proposed the definitions of fuzzy time series and methods to model fuzzy relationships among obser-vations of enrollment. For producing better forecasting results, the following researcher, Chen (1996) proposed an arithmetic approach to improve the initial model. After that, many following researchers proposed several fuzzy time series model to improve forecasting accuracy (Su et al., 2010; Cheng et al., 2010; Teoh et al., 2009; Chen et al., 2008; Chen et al., 2007; Huarng and Yu, 2006; Chen and Chung, 2006; Cheng et al., 2006; Huarng and Yu, 2005; Yu, 2005; Huarng, 2001; Hwang et al., 1998).

In this paper, the definitions and forecasting procedures for Song and Chissom's (1993) model and Chen's model (1996), refereed by many following researchers, are used to introduce the main concepts of fuzzy time series.

Song and Chissom's (1993) fuzzy time series

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

Definition 2: If there exists a fuzzy logical relationship $R(t\text{-}1, t)$, such that $F(t) = F(t\text{-}1) \times R(t\text{-}1, t)$, where x represents an operation, then $F(t)$ is said to be caused by $F(t\text{-}1)$. The logical relationship between F(t) and $F(t\text{-}1)$ can be represented as $F(t\text{-}1) \rightarrow F(t)$.

Definition 3: Let $F(t\text{-}1)=A_i$ and $F(t)=A_j$. The relationship between two consecutive observations, $F(t)$ and $F(t\text{-}1)$, referred to as a fuzzy logical relationship (FLR), can be denoted by $A_i \rightarrow A_j$, where A_i is called the Left-Hand Side (LHS) and A_j the Right-Hand Side (RHS) of the FLR.

Definition 4: All fuzzy logical relationships in the training dataset can be further grouped together into different fuzzy logical relationship groups according to the same Left-Hand Sides of the fuzzy logical relationship. For example, there are two fuzzy logical relationships with the same Left-Hand Side (Ai): $A_i \rightarrow A_{j\,1}$ and $A_i \rightarrow A_{j\,2}$. These two fuzzy logical relationships can be grouped into a fuzzy logical relationship group.

Definition 5: Suppose $F(t)$ is caused by $F(t\text{-}1)$ only, and $F(t) = F(t\text{-}1) \times R(t\text{-}1,t)$. For any t, if $R(t\text{-}1,t)$ is independent of t , then $F(t)$ is named a time-invariant fuzzy time series , otherwise a time-variant fuzzy time series.

Definition 6: Assume that $F(t)$ is a fuzzy time series and $F(t)$ is caused by $F(t\text{-}1)$, $F(t\text{-}2)$, . . . , and $F(t\text{-}n)$, then the fuzzy logical relationship can be represented as following: $F(t\text{-}1)$, $F(t\text{-}2)$, . . . , $F(t\text{-}n) \rightarrow F(t)$. This expression is called the n-th order fuzzy time series forecasting model, where $n \geqq 2$(Chen, 1996; Chen and Chung,2006).

Song and Chissom employed six main procedures in time-invariant fuzzy time series and time-variant fuzzy time series models as follows: (1) define and partition the universe of discourse; (2) define fuzzy sets for the observations; (3) partition the intervals; (4) fuzzify the observations; (5) establish the fuzzy relationship, FLR, and forecast; and (6) defuzzify the forecasting results.

Chen's (1996) fuzzy time series

After Song and Chissom's (1993) model, Chen (1996) proposed an arithmetic approach in the procedures of fuzzy time series to enhance the original model. The algorithm of Chen's model is introduced as follows.

Step 1: Define the universe of discourse and intervals for rules abstraction.
Based on the issue domain, the universe of discourse can be defined as: U = [*starting*, *ending*]. As the length of interval is determined, U can be partitioned into several equal length intervals.

Step 2: Define fuzzy sets based on the universe of discourse and fuzzify the historical data.

Step 3: Fuzzify observed rules.
For example, a datum is fuzzified to A_j if the maximal degree of membership of that datum is in A_j.

Step 4: Establish fuzzy logical relationships and group them based on the current states of the data of the fuzzy logical relationships.

For example, $A_1 \rightarrow A_2$, $A_1 \rightarrow A_1$, $A_1 \rightarrow A_3$, can be grouped as: $A_1 \rightarrow A_1, A_2, A_3$.

Step 5: Forecast.
Let $F(t-1) = A_i$.

Case 1: There is only one fuzzy logical relationship in the fuzzy logical relationship sequence. If $A_i \rightarrow A_j$, then $F(t)$, forecast value, is equal to A_j.
Case 2: If $A_i \rightarrow A_i$, A_j , …, A_k, then $F(t)$, forecast value, is equal to A_i , A_j , …, A_k .

Step 6: Defuzzify. Apply "centroid" method to get the results. This procedure (also called center of area, center of gravity) is the most often adopted method of defuzzification.

Proposed concepts

After reviewing literature related to statistical time series and fuzzy time series models, we discovered two disadvantages for past time series models as following statements.

Firstly, some mathematic assumptions for observations are made in statistical time series models. The time series variables are independent each other and identically distribution as normal random variables, and it should be tested whether the variables are stationary or not. Additionally, the time series models are usually represented as complex mathematic equations or formulas that are unintelligible for common stock investors.

Secondly, the forecast generated by fuzzy time series models is usually less accurate than statistical time series models. Fuzzy numbers are used to represent observations and, therefore, the forecast from fuzzy time series models is a linguistic value which has to be defuzzified to produce a numeric forecast by a defuzzification process. The defuzzification method such as "centroid" method (Chen, 1996) usually produce a rough forecast from some specific linguistic values.

In order to enhance fuzzy time series forecasting accuracy and avoid the disadvantages of statistical time series models, a novel forecasting method based on time series (research framework is illustrated in Figure 1) is proposed and three improvement concepts are factored into the proposed method:

(1) the mathematic assump-tions such as stationary and variable independence is ignored; (2) employ objective stock price-pattern, consisting of price trend and variation, contained in history stock data as historical rule basis for forecasting; (3) apply multi-period adaptation model (Chen et al.,2008) in forecasting process to produce self-adapted forecast (Kmenta,1986) to promote forecasting accuracy.

The detailed step-by-step computation processes are proposed in next subsection to crystallize the refined concepts into forecasting.

Proposed algorithm

The proposed algorithm employs the TAIEX of year 2000 processes.

Step 1: Produce price-pattern (price trend and variation) between two consecutive days as historical rule basis.

In this step, price-pattern, which consists of price variation and trend, between consecutive two days are produced as historical rule basis. The variation, $V(t)$, is defined as Equation (1), and the trend is defined as the sign of the variation such as "+" or "-", defined in Equation (2) .

$$V(t) = |P(t) - P(t-1)| \quad 1$$

$$Sign(t) = the\ sign\ of\ P(t) - P(t-1) \quad 2$$

Where $P(t)$ denotes stock index at time *t*.

Take Table 1 as example, the $V(t)$ for the TAIEX on 2000/01/07 (time = 4) is produced by following equation: $V(4) = |P(4) - P(3)| = 76.56$, and the *Sign*(*t*) is '-' which have shown that the price variation and trend between the 3-*th* and 4-*th* day (2000/01/06 and 2000/01/07).

Step 2: Measure difference between present price-pattern and past price-patterns contained in historical rule basis.

In this step, "pattern difference", which consists of pattern distance and sign dissimilarity, between the present price-pattern and all of history price-patterns are produced one by one. In order to compute the pattern difference, the concept of Euclidean distance (Breu et al., 1993) is employed to represent as "pattern distance", which is defined by Equation (3).

$$D(t_a, t_b) = d(V(t_a), V(t_b)) = \sqrt{(V(t_a) - V(t_b))^2} \quad 3$$

Where $V(t_a)$ is the variation at time t_a; $V(t_b)$ is the variation at time t_b; and $D(t_a, t_b)$ is the difference between $V(t_a)$ and $V(t_b)$.

Take Table 2 as example, if the future stock index on 2000/11/01 (time = 225) is used as a predicted object, then the price-pattern on 2000/10/31, *Sign* (224) and $V(224)$, is selected as "present price-pattern". Then, pattern distance and sign dissimilarity between the present pattern and past patterns are measured and compared with one by one (from *V*(2) to *V*(223) and from *Sign*(2) to *Sign*(223)). To simplify computation of this step, the process of distance difference measuring is ignored when the sign of present price-pattern (*Sign* (224)) is not the same as past patterns. For example, $V(224)$ is "114.9" and $V(4)$ is "76.56". The pattern distance, $D(224,4)$, between them is calculated by the following equation: $|V(224) - V(4)| = 38.34$.

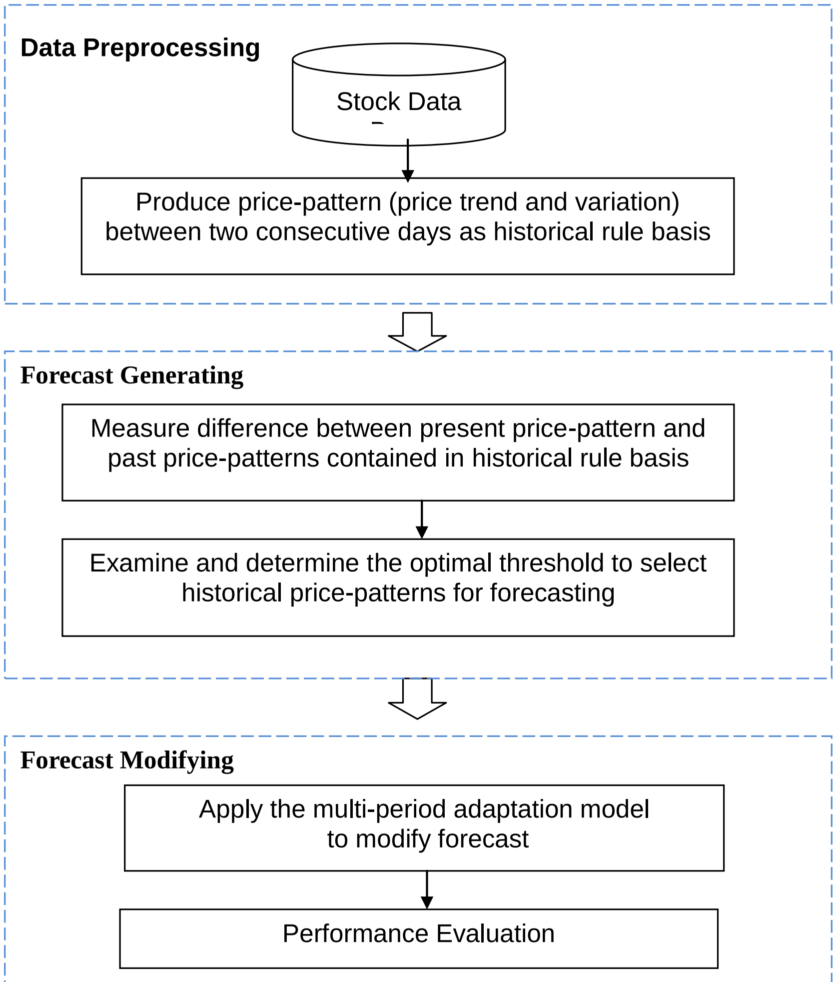

Figure 1. Research framework.

Table 1. Examples of pattern generations in the TAIEX (2000).

Time	Date	P(t)	Price-Pattern	
			***Sign*(*t*)**	***V*(*t*)**
1	2000/01/04	8756.55	N.A.	N.A.
2	2000/01/05	8849.87	+	93.32
3	2000/01/06	8922.03	+	72.16
4	2000/01/07	8845.47	-	76.56
5	2000/01/10	9102.60	+	257.13
6	2000/01/11	8927.03	-	175.57
7	2000/01/12	9144.65	+	217.62
8	2000/01/13	9107.19	-	37.46
222	2000/10/27	5805.17	+	13.79
223	2000/10/30	5659.08	-	71.52
224	2000/10/31	5544.18	-	42.80

Table 2. Examples for producing pattern difference.

Time	Date	P(t)	Price-pattern		Sign dissimilarity	Pattern distance
			***Sign*(*t*)**	***V*(*t*)**	***D*(224, t_b)**	***D*(224, t_b)**
1	2000/01/04	8756.55	N.A.	N.A.	N.A.	N.A.
2	2000/01/05	8849.87	+	93.32	dissimilarity	--
3	2000/01/06	8922.03	+	72.16	dissimilarity	--
4	2000/01/07	8845.47	-	76.56	similarity	33.76
5	2000/01/10	9102.60	+	257.13	dissimilarity	--
6	2000/01/11	8927.03	-	175.57	similarity	132.77
7	2000/01/12	9144.65	+	217.62	dissimilarity	--
8	2000/01/13	9107.19	-	37.46	similarity	5.34
222	2000/10/27	5805.17	+	13.79	dissimilarity	--
223	2000/10/30	5659.08	-	71.52	similarity	28.72
224	2000/10/31	5544.18	-	42.80	similarity	0
255	2000/11/01	?				

"-- "denotes that it is not necessary to measure" pattern difference"

However, $D(224,3)$ is not necessary to be calculated because *Sign* (224) is not the same as *Sign* (3) (*Sign* (224) = "-" and *Sign* (3) = "+").

Step 3: Examine and determine the optimal threshold to select historical price-patterns for forecasting. In this step, the *n* similar price-patterns, $V(t_1)$, $V(t_2)$,…, $V(t_i)$, where *i* = 1 to *n*, with lower difference and the same sign are selected as forecasting rule basis to produce a prediction. The consecutive price-pattern ($V(t_{i+1})$, $Sign(t_{i+1})$) for each similar price-pattern ($V(t_i)$, $Sign(t_i)$) is employed as one forecasting pattern because history stock patterns maybe reoccur. Therefore, *n* forecasting patterns will generated if *n* similar price-patterns are discovered from step 2. In this step, the most important issue is to determine a proper threshold, the amount of *n* similar patterns. For finding out the optimal value, the threshold is selected from 1% to 20% with a stepped value of 1%. The optimal threshold is determined when the forecasting performance reaches the minimum error. The threshold is defined by Equation (4).

$$Threshold = \frac{n}{N} \times 100\% \qquad 4$$

Where the amount of these similar patterns is defined as *n* and the amount of total patterns in training dataset is defined as *N*.

In this step, average method is utilized to produce an initial forecast, and two equations are defined in the method as Equations (5) and (6).

$$Initial\_forecast\_pattern(t+1) = \sum_{i=1}^{n} \frac{sign(t_i+1) \times V(t_i+1)}{n} \qquad 5$$

$$Initial\_forecast(t+1) = Initial\_forecast\_pattern(t+1) + P(t) \qquad 6$$

Where $V(t_i+1)$ is the forecasting variation based on a similar pattern, $V(t_i)$; $Sign(t_i+1)$ is the forecasting sign based on a similar pattern, $Sign(t_i)$; $Initial\_forecast\_pattern(t+1)$ is the average forecasting pattern for the future; *P*(*t*) is the present stock price at time *t*; $Initial\_forecast(t+1)$ is the forecasting value for the future price.

To evaluate forecasting performance, two error indicators, MAPE (defined in Equation (7)) and RMSE (defined in Equation (8)), are used as error indicators.

$$\text{MAPE} = \frac{100}{N} \sum_{t=1}^{N} \left| \frac{d_t - z_t}{d_t} \right| \qquad 7$$

$$\text{RMSE} = \left\{ \frac{1}{N} \sum_{t=1}^{N} (d_t - z_t)^2 \right\}^{0.5} \qquad 8$$

Where N is the number of forecasting periods, d_t is the actual stock price at period t, and z_t is the forecasting stock price at period t.

Tables 3 and 4 list out the forecasting performance based on MAPE and MAPE under the different thresholds

Table 3. MAPE and RMSE for different thresholds (from 1 to 10%).

Threshold	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
MAPE	3.2043	2.5802	2.1457	2.1293	2.123	2.1075	2.1379	2.0477	1.9904	1.9818
RMSE	198.60	164.61	140.59	139.18	142.54	139.29	143.54	138.43	136.27	138.23

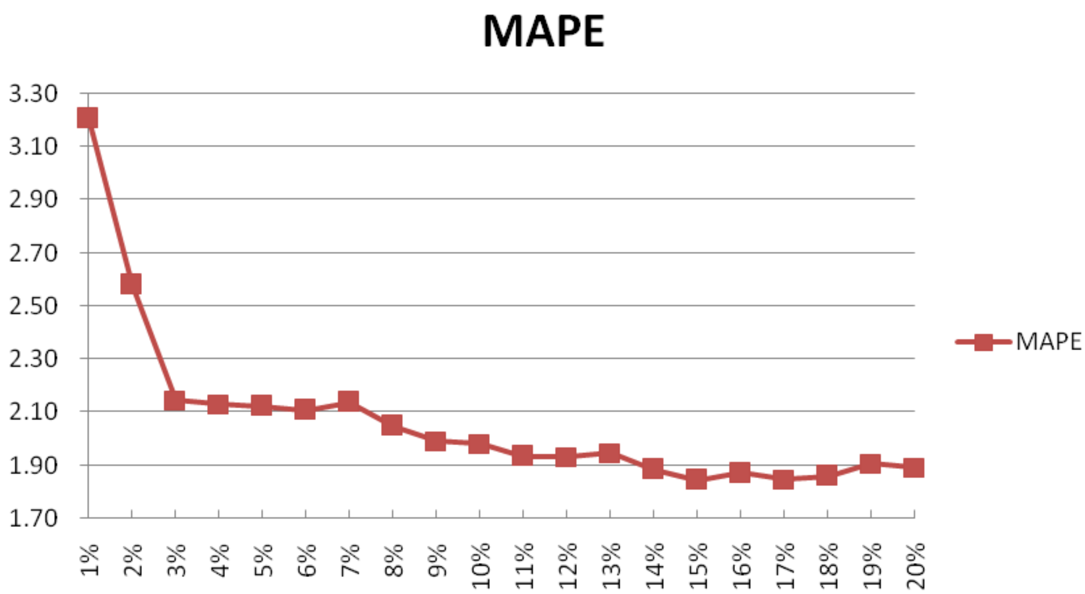

Figure 2 The MAPE for different thresholds in the TAIEX of year 2000.

from 1 to 20%. Figure 2 illustrates the MAPE for the proposed model using different thresholds. Figure 3 illustrates the RMSE for the proposed model using different thresholds. From the performance tables, the optimal threshold is 15% when using MAPE as error indicator and 17% when using RMSE as error indicator. Figure 4 illustrates the initial forecasting results and actual values for the TAIEX of year 2000 with MAPE.

Step 4: apply multi-period adaptation model to modify forecast

In order to promote accuracy of proposed model, in this step, the multi-period adaptation model (Chen et al., 2008) is taken to enhance forecasting performance. The adaptive model is defined in Equation (9)

$$Adapted\_Forecast(t+1) = P(t) + \sum_{i=1}^{k} h_i * \varepsilon_i \qquad 9$$

Where $Adapted\_Forecast(t+1)$ is forecast for the future stock price; $P(t)$ is the present stock price on time t; ε_i is the i-*th* period of forecast error; h_i is a adaptive parameter for ε_i.

In step, one adaptive parameter (referred to 0.01~1.00 with a stepped value 0.01) to adapt to modify initial forecast to reach better forecasting performance. With the adaptive parameter, the initial forecasts (MAPE =1.8450%) are modified as more accurate forecasts (MAPE =1.8389%). Figure 5 illustrates the adapted forecasts and actual values for the TAIEX of year 2000 with MAPE.

Model verification

This section consists of three subsections: (1) introduction for experimental datasets TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index) and performance indicators; (2) performance comparison with fuzzy time series models; and (3) performance comparison with statistic time series models.

Experiment datasets and performance indicators

In this paper, a nine-year period of the TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index), from 1997 to 2005, is selected as experiment datasets, which are collected from the official website of TSEC (Taiwan Stock Exchange Corporation) retrieved from http:// www.tse.com.tw). One-year period of stock

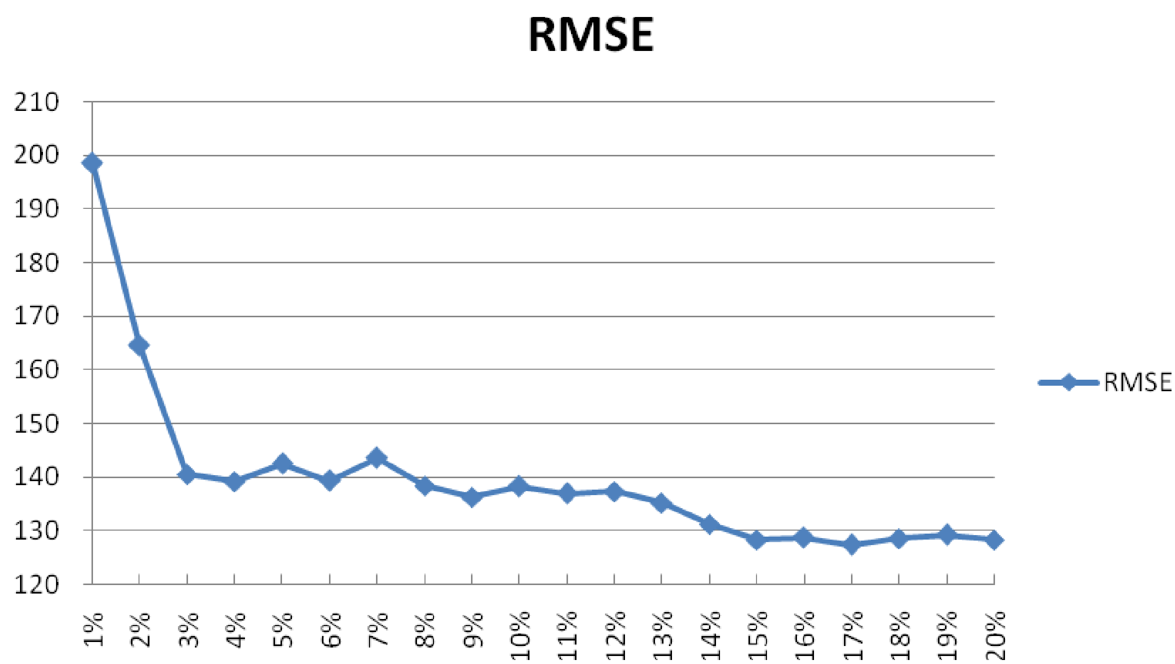

Figure 3 The RMSE for different thresholds in the TAIEX of year 2000.

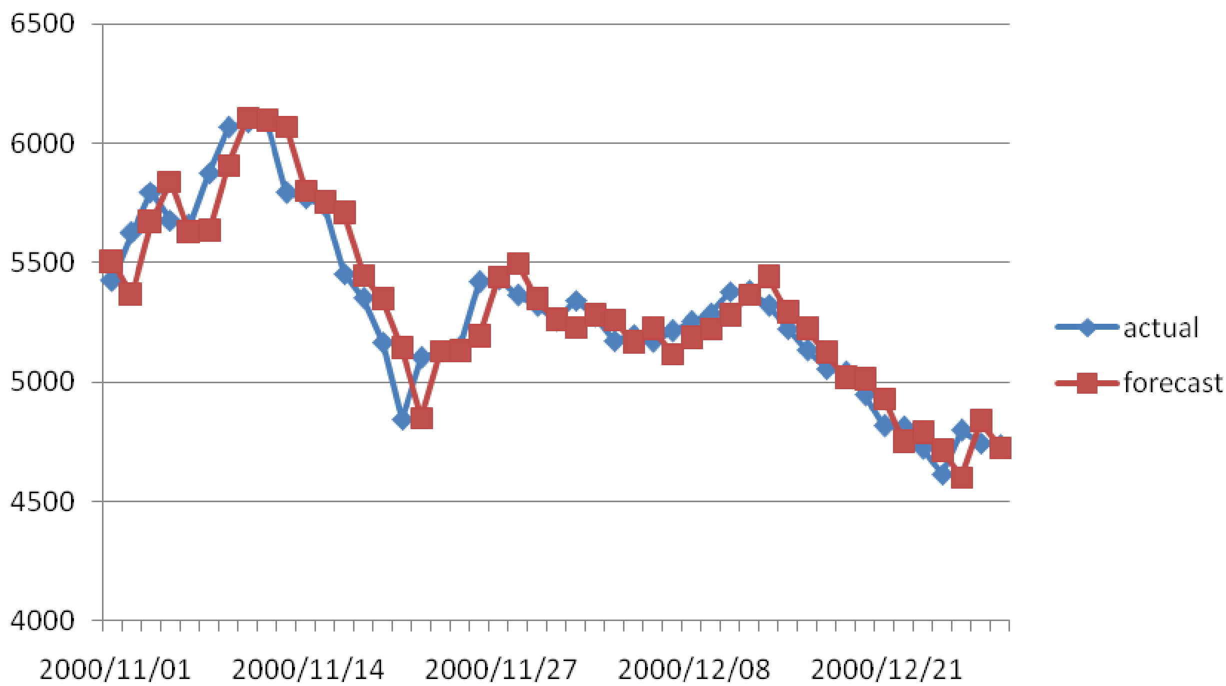

Figure 4 Initial forecasts results with the threshold of 15% (MAPE =1.8450%).

index is used as one unit of experimental dataset. Each unit of dataset is divided into two sub-datasets: (1) previous 10-month period is used as training dataset, and (2) the rest 2-month period, from November to December, is used for testing (Su et al., 2010; Teoh et al., 2009; Chen et al., 2008; Chen et al., 2007). Taking the TAIEX of year 2000 as example, if the cut point date is set on 2000/10/31, the training period is defined from 2000/01/04 to 2000/10/31, and the testing period is from 2000/11/01 to 2000/12/30 (Figure 6).

In the experiment dataset, the training dataset contains 224 observations, and the testing dataset contains 47 observations.

Two error indicators, the RMSE (root mean square error) and the

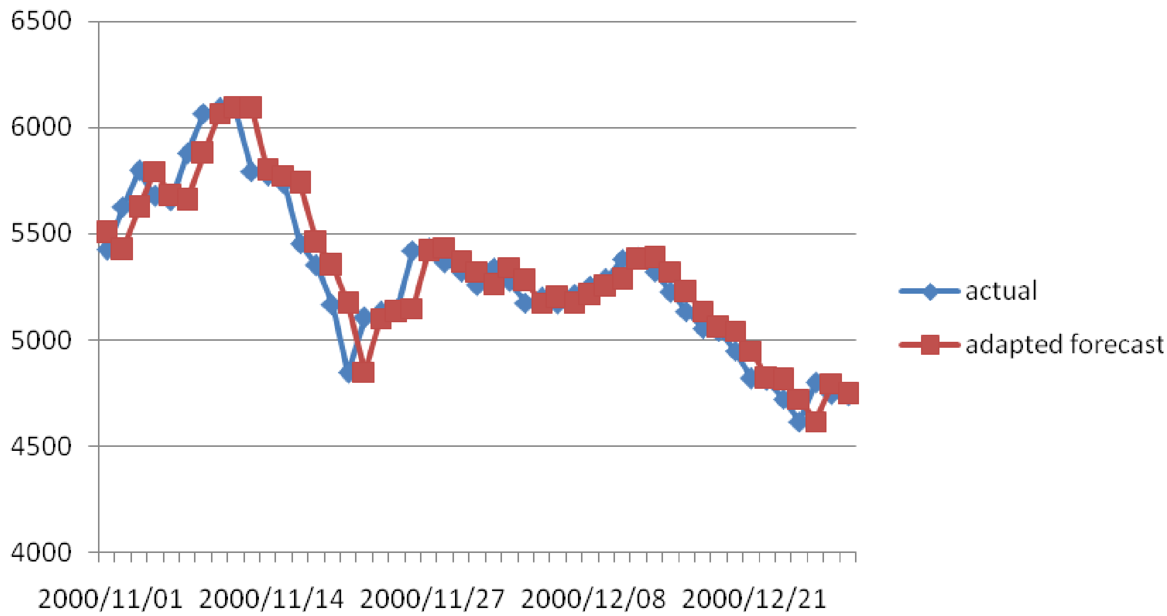

Figure 5. Adapted forecasts based on an adaptive model (MAPE =1.8389%).

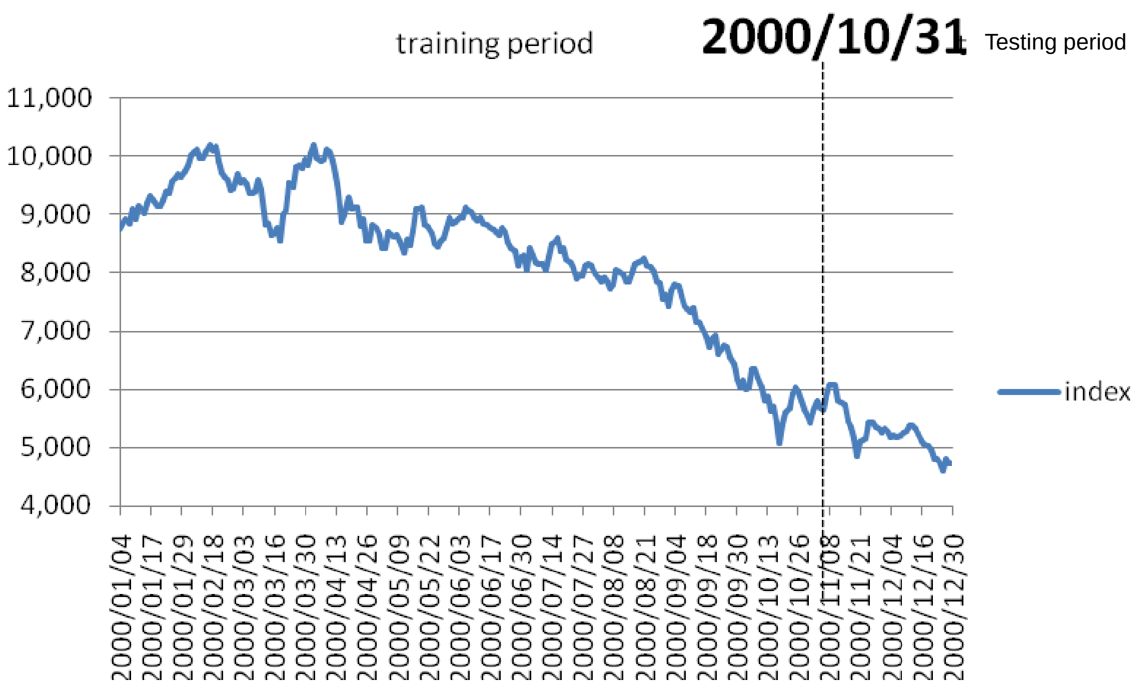

Figure 6. The TAIEX from 2000/01/04 to 2000/12/30.

MAPE (mean absolute percent error), defined in Equation (7) and Equation (8), are employed to evaluate forecasting models. Although the RMSE is a common indicator to measure forecasting performance for fuzzy time series models(Teoh,2009;Chen et al,2008; Chen et al,2007), many researchers used the MAPE for evaluating forecasting accuracy (Hanke,1995; Bowerman et al., 2004). Therefore, in this paper, the MAPE and the RMSE are both used as performance indicators to examine forecasting performance.

Performance comparisons with fuzzy time series

To evaluate the performance of the proposed model carefully, three fuzzy time series models, Chen's (1996),

Table 4. MAPE and RMSE for different thresholds (from 11 to 20%).

Threshold	11%	12%	13%	14%	15%	16%	17%	18%	19%	20%
MAPE	1.934	1.9315	1.9465	1.8848	1.8450*	1.8708	1.8464	1.8588	1.9066	1.8924
RMSE	136.99	137.30	135.23	131.29	128.34	128.71	127.36*	128.55	129.17	128.29

*denotes the minimum among 20 thresholds

Table 5. Performance comparison with fuzzy time series model.

Dataset	TAIEX								
Year	1997	1998	1999	2000	2001	2002	2003	2004	2005
Chen's Model(1996)	154	134	120	176	148	101	74	83	66
Huarng and Yu's Model(2006)	141	121	109	152	130	84	56	79	69
Su et al.'s Model (2010)	-	-	-	-	122	94	55	69	65
Proposed Model	141*	115*	104*	130*	114*	66*	53*	55*	53*

"*" denotes the minimum RMSE among four models.
"-" denotes that performance datum is unavailable.

Table 6. Performance comparison with statistic models (MAPE).

Dataset	TAIEX								
Year	1997	1998	1999	2000	2001	2002	2003	2004	2005
AR(1)	1.43*	1.37*	1.02	1.82	1.91	1.10*	0.70	0.67*	0.68
AR(2)	1.44	1.39	1.01*	1.80*	1.91	1.11	0.71	0.68	0.68
ARMA(1,1)	1.44	1.38	1.01	1.80	1.91	1.11	0.71	0.68	0.69
Proposed Model	1.43*	1.38	1.05	1.84	1.89*	1.10*	0.69*	0.67*	0.66*

* denotes the minimum value

Huarng and Yu's (2006) and Su et al.'s (2010), are employed as comparison models. Chen's (1996) model is a typical case of fuzzy time series model. Huarng and Yu's (2006) model is an advanced model based on neural networks. Su et al.'s (2010) model is also an advanced model using a rough set algorithm. Referred to Huarng and Yu's (2006) and Su et al.'s (2010) research, the datum of forecasting performance for the three models is summarized and listed in Table 5. The performance indicator utilized in their papers is the RMSE. The performance comparison data indicates that the proposed model outperforms the other three models in forecasting accuracy.

Performance comparisons with statistic time series

To validate the superiority of the proposed model, three statistical time series models, AR(1), AR(2) and ARMA(1, 1) are taken as comparison models. Tables 6 and 7 list separately forecasting performance for four models with two performance indicators they also show that the proposed model bears the minimum value of the MAPE and the RMSE in six testing datasets (1997, 2001, 2002, 2003, 2004, and 2005). Although there is little difference in forecasting performance among the four models for each experimental dataset, we still can see that the proposed model outperforms the statistical time series models in most of datasets.

Conclusions

A novel price-pattern detection method based on time series has been proposed in this paper. After delicate model verification, the conclusion is given with confidence that the proposed method can provide more accurate forecast than fuzzy time series and traditional statistical time series models. From performance comparisons above, two major reasons for the superior of the proposed model are summarized as follows.

Firstly, Table 5 has shown that the proposed model outperforms three fuzzy time series models, Chen's (1996), Huarng and Yu's (2006) and Su et al.'s (2010), in forecasting accuracy. The reason may be assumed by that two computation procedures of fuzzy time series models (fuzzify and defuzzify) have made their forecasts less accurate and stable.

Secondly, Tables 6 and 7 clearly indicate that the proposed model performs better than three statistical time

Table 7. Performance comparison with statistic models (RMSE).

Dataset	**TAIEX**								
Year	1997	1998	1999	2000	2001	2002	2003	2004	2005
AR(1)	141*	114*	102*	130	115	66*`	54	55*	54
AR(2)	141*	115	102*	129*	114	67	55	55*	54
ARMA(1,1)	141*	114*	103	129*	115	67	55	55*	54
Proposed Model	141*	115	104	130	114*	66*	53*	55*	53*

* denotes the minimum value

series models, AR(1), AR(2) and ARMA(1, 1). The reason why the proposed model outperforms the statistical models is stated as follows. A specific mathematic formula is built for statistical time series models and the forecasting mechanism is fixed no matter how stock market fluctuates. But the proposed model applies multi-period adaptation model to produce self-adapted predictions to deal with recent price fluctuations in stock market to reduce forecasting error.

After implementing the experiment, three major advantages for the proposed model are issued as follow: (1) no mathematic assumptions about observations are required to form forecasting algorithms and the computer system using the proposed algorithms is easy to build up with lower complexity; (2) the proposed model produce accurate forecasts based on “stock price-patterns” that are understandable for common investors instead of “statistical formula” or “fuzzy logic relations” that are complicated words for common investors; and (3) by using multi-period adaptation model, the proposed method can produce self-modified forecasts to reach better accuracy when stock market go flat and to make smaller loss when stock market fluctuates violently.

In the future works, some suggestions can be offered to improve the proposed model as follows: (1) apply high-order time series (more than two consecutive time series observations) in the price–pattern detection method to evaluate the proposed model; (2) employ other stock market data such as Hong-Kong’s stock market (HSI) and American’s stock market (DJI) as experimental dataset to verify the proposed model; and (3) utilize a investing strategy to buy or sell stock index to evaluate the profit return of the proposed model.

REFERENCES

Bollerslev T (1986). Generalized autoregressive conditional heteroscedasticity. J. Econ., 31: 307-327.

Bowerman BL, O'connell RT, Koehler AB (2004). Forecasting time series and regression: An applied approach. Thomson Brooks/Cole, Belmont,CA.

Box GEP, Jenkins G (1976). Time series analysis: Forecasting and control. Holden-Day, San Francisco, CA.

Breu H, Gil J, Kirkpatrick D, Werman M (1993). Linear Time Euclidean Distance Transform Algorithms. IEEE Trans. Pattern Anal. Machine Intel., 17: 529-533

Chen SM (1996). Forecasting enrollments based on fuzzy time series. Fuzzy Sets Syst., 81:311-319

Chen SM, Chung NY (2006). Forecasting enrollments using high-order fuzzy time series and genetic algorithm. Int. J. Intel. Syst., 21:485-501.

Chen TL, Cheng CH, Teoh HJ (2007). Fuzzy Time series Based on Fibonacci Sequence for Stock Price Forecasting. Phy. A., 380:377-390.

Chen TL, Cheng CH, Teoh HJ (2008). High Order Fuzzy Time series Based on Multi-period Adaptation Model for Forecasting Stock Markets. Physica A., 387: 876-888.

Cheng CH, Chen TL, Chiang CH (2006). Trend-weighted fuzzy time series model for TAIEX forecasting. Lecture notes Comput. Science, 4234: 469-477.

Cheng CH, Chen TL, Wei LY(2010). A Hybrid Model Based on Rough Sets Theory and Genetic Algorithms for Stock Price Forecasting. Inform. Sci., 180: 1610-1629.

Engle RF (1982). Autoregressive conditional heteroscedasticity with estimator of the variance of United Kingdom inflation. Econ., 50: 987-1008.

Gately E (1996). Neural Networks for Financial Forecasting. John Wiley and Sons, New York.

Gershenfeld NA (1999). The nature of mathematical modeling. Cambridge University Press.

Hanke JE, Reitsch AG (1995). Business forecasting. Prentice-Hall, New Jersey.

Huarng K (2001). Heuristic models of fuzzy time series for forecasting. Fuzzy Sets Syst., 123:369-386

Huarng K, Yu HK (2005). A Type 2 fuzzy time series model for stock index forecasting. Phy. A., 353: 445-462.

Huarng K, Yu HK (2006). The application of neural networks to forecast fuzzy time series. Phy. A., 363:481-491.

Hwang JR, Chen SM, Lee CH (1998). Handling forecasting problems using fuzzy time series. Fuzzy Sets Syst., 100: 217-228.

Kmenta J (1986). Elements of econometrics. MacMillan, New York.

Pai PF, Lin CS (2004). A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 33: 497-505.

Ruppert D (2011). Statistics and Data Analysis for Financial Engineering. Springer, New York.

Song Q, Chissom BS (1993). Fuzzy time series and its models. Fuzzy Sets Syst., 54: 269-277.

Su CH, Chen TL, Cheng CH, Chen YC(2010). Forecasting the Stock Market with Linguistic Rules Generated from the Minimize Entropy Principle and the Cumulative Probability Distribution Approaches. Entropy, 12: 2397-2417.

Teoh HJ, Chen TL, Cheng CH, Chu HH (2009). A hybrid multi-order fuzzy time series for forecasting stock markets. Expert Syst. Appl., 36: 7888-7897

Yu HK (2005). Weighted fuzzy time series models for TAIEX forecasting. Phy. A., 349: 609-624.

Zadeh LA (1975a). The concept of a linguistic variable and its application to approximate reasoning I. Inform. Sci., 8: 199-249.

Zadeh LA (1975b). The concept of a linguistic variable and its application to approximate reasoning II. Inform. Sci., 8: 301-357.

Zadeh LA (1976). The concept of a linguistic variable and its application to approximate reasoning III. Inform. Sci., 9: 43-80.