Application of VPRS model with enhanced threshold parameter selection mechanism to automatic stock market forecasting and portfolio selection

Kuang Yu Huang*

Department of Information Management, Ling Tung University, #1 Ling Tung Road, Taichung City 408, Taiwan

A R T I C L E   I N F O

Keywords:
Variable Precision Rough Set
Fuzzy theory
Fuzzy C-Means
ARX model
Grey relational analysis
Stock portfolio

A B S T R A C T

This study proposes a technique based upon Fuzzy C-Means (FCM) classification theory and related fuzzy theories for choosing an appropriate value of the Variable Precision Rough Set (VPRS) threshold parameter (β) when applied to the classification of continuous information systems. The VPRS model is then combined with a moving Average Autoregressive Exogenous (ARX) prediction model and Grey Systems theory to create an automatic stock market forecasting and portfolio selection mechanism. In the proposed mechanism, financial data are collected automatically every quarter and are input to an ARX prediction model to forecast the future trends of the collected data over the next quarter or half-year period. The forecast data are then reduced using a GM(1,N) model, classified using a FCM clustering algorithm, and then supplied to a VPRS classification module which selects appropriate investment stocks in accordance with a pre-determined set of decision-making rules. Finally, a grey relational analysis technique is employed to weight the selected stocks in such a way as to maximize the rate of return of the stock portfolio. The validity of the proposed approach is demonstrated using electronic stock data extracted from the financial database maintained by the Taiwan Economic Journal (TEJ). The portfolio results obtained using the proposed hybrid model are compared with those obtained using a Rough Set (RS) selection model. The effects of the number of attributes of the RS lower approximation set and VPRS β-lower approximation set on the classification are systematically examined and compared. Overall, the results show that the proposed stock forecasting and stock selection mechanism not only yields a greater number of selected stocks in the β-lower approximation set than in the RS approximation set, but also yields a greater rate of return.

1. Introduction

Predicting stock prices in today's volatile markets is notoriously difficult and represents a major challenge for traditional time-series-based forecasting mechanisms. Consequently, a requirement exists for robust forecasting schemes capable of accurately predicting the future behavior of the stock market in order to provide investors with a reliable indication as to when and where they should invest their money in order to maximize their rate of return.

Many applications have been presented in recent decades for predicting market trends. Typical mechanisms include the use of genetic algorithms (GAs) to choose optimal portfolios (Bauer, 1994; Hassan, Nath, & Kirley, 2007 analysis algorithms (Bazan & Szczuka, 2000; Shen & Loh, 2004; Wang, 2003), Variable Precision Rough Set models (Beynona & Dri1eldb, 2005; Huang & Jane, 2009), and so forth.

However, the success of RS techniques in correctly classifying a dataset relies upon all the collected data being correct and certain. In other words, performing a classification function with a controlled degree of uncertainty or misclassification error falls outside the realm of the RS approach (Ziarko, 1993). In an attempt to extend the applicability of the RS method, Ziarko developed the Variable Precision Rough Set (VPRS) theory (Ziarko, 1993). In contrast to the original RS model, the aim of the VPRS approach is to analyze and identify data patterns which represent statistical trends rather than functional patterns (Ziarko, 2001). VPRS deals with uncertain classification problems through the use of a precision parameter, β. Essentially, β represents a threshold value which determines the portion of the objects in a particular conditional class which are assigned to the same decision class. (Note that in conventional RS theory, β has a value of one.) Thus, determining an appropriate value of β is essential in deriving knowledge from partially-related data objects.

Many VPRS studies presented in the literature discuss the problem of β-reducts or approximate reducts. Ziarko (1993) stated that β-reducts, i.e. a subset P of the set of conditional attributes C with respect to a set of decision attributes D, must satisfy the following conditions: (i) the subset P offers the same quality of classification (subject to the same β value) as that achieved using the whole set of conditional attributes C; and (ii) no attribute can be eliminated from the subset P without affecting the classification quality (subject to the same β value). Despite the fundamental role played by the β-reducts in determining the VPRS classification results, the literature lacks any formal empirical evidence to support the choice of any particular method of β-reducts selection (Beynon, 2000). Ziarko (1993) suggested that the β value could simply be determined at the discretion of the decision maker. Beynon (2000) proposed two methods for selecting an appropriate β-reducts based upon the assumption that the β value fell within an interval whose range was determined by the desired level of classification performance (Beynon, 2001). Beynon (2002) introduced the concept of a $l;u$ -quality graph to map the quality of the VPRS classification performance to the values assigned to the l and u bounds. Su and Hsu (2006) proposed a method for determining a suitable value of the precision parameter based on an assessment of the minimum acceptable upper bound of the misclassification error.

Ziarko (1993) argued that the β value represented a classification error and suggested that its value should be confined to the domain [0.0, 0.5). Conversely, An, Shan, Chan, Cercone, and Ziarko (1996) and Beynon (2001) used the β parameter to denote the proportion of correct classifications, and argued that its value should fall within the range (0.5, 1.0). Many researchers have investigated the problem of assigning an appropriate value to β using some form of β-reducts approach (Beynon, 2000, 2001; Mi, Wu, & Zhang, 2004). Basically, these β-reducts methods utilize the principle of the "extent of classification correctness" to determine an appropriate β value under the premise that changing the number of attributes has no effect on the classification results. However, such approaches have the major drawback that the value of β can only be determined once the classification results have been obtained. In general, information systems are characterized by continuous attribute values, and thus their processing requires some form of clustering technique. However, when clustering continuous data, errors are liable to arise as a result of the fuzzy nature of the data. As a result, selecting an appropriate value of β for the classification of continuous datasets using a β-reducts approach is problematic. Accordingly, the current study presents a method for determining β using the Fuzzy C-Means (FCM) clustering method and related fuzzy set theories.

As discussed above, a requirement exists for reliable forecasting schemes capable of predicting future stock market trends in order to assist investors in determining where best to invest their funds so as to increase their rate of return on investment. However, the use of a VPRS approach to extract information from the stock market is problematic since financial data and market indices vary sequentially over time and have the form of a correlated signal system. In other words, the stock market is essentially a dynamic process. In general, such processes can best be described using a discrete-time series model such as an autoregressive (AR) model, a moving average (MA) model, or an autoregressive with exogenous inputs (ARX) model (Bhardwaj & Swanson, 2006; Ljung, 1999; Marcellino, Stock, & Watson, 2006; Sohn & Lim, 2007). Accordingly, the present study develops an algorithm for predicting the future behavior of the stock market and selecting an appropriate stock portfolio based upon an ARX prediction model, the multivariate GM(1,N) model (Deng, 1989), the FCM clustering technique, VPRS theory, grey relational analysis, and the investment guidelines prescribed by Buffett (Hagstrom, Miller, & Fisher, 2005). The feasibility and effectiveness of the proposed approach are demonstrated using the case of electronic stocks for illustration purposes. The performance of the proposed algorithm is verified by comparing the rate of return on the selected stock portfolio with that of a portfolio chosen using a conventional RS-based hybrid model (Huang & Jane, 2009).

The remainder of this paper is organized as follows. Section 2 presents the fundamental principles of the ARX model, VPRS theory and Grey theory, respectively. Section 3 describes the integration of these concepts to create an automatic stock market forecasting and portfolio selection scheme. Section 4 compares the performance of the proposed scheme with that of the conventional RS-based model. Finally, Section 5 presents some brief concluding remarks and indicates the intended direction of future research.

2. Review of related methodologies

2.1. ARX model

In general, the objective of ARX prediction models is to minimize a positive function of the prediction error. The predictions of an ARX model are based on the assumption that the system changes only gradually over time and can be described by a set of parameters. When constructing the ARX model, these parameters are generally estimated using a system identification process. The basic ARX model has the form

$$\begin{aligned} y\ t \quad & a_1 y\ t \times 1 \qquad a_{n_a} y\ t \times n_a \\ & b_1 u\ t \times 1 \qquad b_{n_b} u\ t \times n_b \quad e\ t\ ; \end{aligned}$$

where $y\ t$ is the output, $u\ t$ is the input and $e\ t$ is a white-noise term.

In the application considered in the present study, i.e. the forecasting of time-series financial data, the most important aspect of the ARX model is its one-step-ahead predictor, which has the form $\hat{y}\ t;\theta \quad \theta^T \varphi\ t$, where $\theta \quad a_1;\ldots a_{n_a};b_1;\ldots;b_{n_b}\ ^T$ is obtained using the least squares method and yields $\min_\theta \frac{1}{N}\sum_{t\ 1}^{N}\ y\ t \times\ \hat{y}\ t;\theta\ ^2$, and $\varphi^T\ t \times \quad y\ t \times 1\ ;..$

2.2. Variable Precision Rough Set (VPRS) theory

VPRS theory operates on what may be described as a knowledge-representation system or an information system (Su & Hsu, 2006). The basic principles of information systems and the application of VPRS theory to the processing of information systems are described in the following sections.

(1) Information systems

A typical information system has the form $S \quad U; A; V_q; f_q$, where U is a non-empty finite set of objects and A is a non-empty finite set of attributes describing each object. Assume that the attributes in set A can be partitioned into a set of conditional attributes $C \neq \phi$ and another set of decisional attributes $D \neq \phi$, $A \quad C \cup D$ and $C \cap D \quad \phi$. If one lists the values of all the attributes (both conditional and decisive) and makes a table, such a table is known as a decision table. For each attribute $q \in A; V_q$ represents the domain of q, i.e. $V \quad \bigcup V_q$. Finally, $f_q : U \quad A \rightarrow V$ is the information function such that $f\ x; q \in \ V_q$ for $\forall q \in A$ and $\forall x \in U$.

For every subset P, where $P \subseteq C$ and $P \neq \phi$, the equivalence relation is given by $I \{ \quad x; y \in \ U \quad U : f\ x; q \quad f\ y; q\ \forall q \in P\}$. The corresponding equivalence class is denoted by $I\ P$ and indicates the objects in a particular conditional class which are assigned to the same decisional class. In VPRS theory, the value assigned to β represents a threshold value which governs the proportion of objects in a particular conditional class which fall within the same equivalent class.

(2) β-Lower and β-upper approximation sets

In processing the information system $S \quad U; A; V_q; f_q\ , A \quad C \cup D$, $X \subseteq U, P \subseteq C$ using a VPRS model with $0{:}5 < \beta \leqslant 1$, the data analysis procedure hinges on two basic concepts, namely the β-lower and β-upper approximations of a set. The β-lower approximation of sets $X \subseteq U$ and $P \subseteq C$ can be expressed as follows:

$$\underline{R}_P^\beta\ X \ \cup \quad \left\{ I\ P \ : \frac{|I\ P \cap \ X|}{|I\ P|} \geqslant \beta \right\}:$$

Similarly, the β-upper approximation of sets $X \subseteq U$ and $P \subseteq C$ is given by

$$\overline{R}_P^\beta\ X \ \cup \quad \left\{ I\ P \ : \frac{|I\ P \cap \ X|}{|I\ P|} > 1 \times \beta \right\}:$$

In the case where $\beta \quad 1$, $\underline{R}_P^\beta\ X$ and $\overline{R}_P^\beta\ X$ are equivalent to the lower and upper approximation sets in RS theory, respectively, i.e. the VPRS model reverts to the original RS model.

Ziarko (1993) defined the following alternative expressions for the β-negative region and β-boundary region of X in S:

$$NEG_P^\beta\ X \quad \bigcup \left\{ I\ P \ : \frac{|I\ P \cap \ X|}{|I\ P|} \leqslant 1 \times \beta \right\};$$

$$BND_P^\beta\ X \quad \bigcup \left\{ I\ P \ : 1 \times \beta < \frac{|I\ P \cap \ X|}{|I\ P|} \leqslant \beta \right\};$$

where $| \quad |$ denotes the cardinality of a set.

In VPRS theory, the quality of the classification result is quantified using the metric

$$\gamma^\beta\ P; D \ \cup \left| \ \left\{ I\ P \ : \frac{|X \cap I\ P|}{|I\ P|} \geqslant \beta \right\} \right| \Big/ |U|:$$

Essentially, the value of $\gamma^\beta\ P; D$ indicates the proportion of objects whose probability of belonging to equivalence class $I\ P$ is greater than or equal to β. In other words, the $\gamma^\beta\ P; D$ involves combining all the β-positive regions and aggregating the number of objects involved within each of the combined region. In practice, the $\gamma^\beta\ P; D$ is used operationally to define and extract suitable reducts to facilitate the application of RS theory and the VPRS model to data mining and rule construction applications.

2.2.1. Determination of suitable threshold value β

This study proposes a mechanism for determining a suitable value of the VPRS precision parameter β based on the Fuzzy C-Means clustering method and the general principles of fuzzy algorithms and fuzzy set theory. The various components of the proposed mechanism are introduced in the sections below.

(1) Fuzzy C-Means (FCM) clustering method (Cox, 2005; Glackina, Maguirea, McIvorb, Humphreysb, & Hermana, 2007; Ramze Rezaee, Lelieveldt, & Reiber, 1998).

FCM, first developed by Dunn (1973) and later refined by Bezdek (1981), is an unsupervised clustering algorithm with multiple applications, ranging from feature analysis, to clustering and classifier design. Fuzzy clustering techniques differ from hard clustering algorithms such as the K-means scheme in that a single data object may be mapped simultaneously to multiple clusters rather than being assigned exclusively to a single cluster. FCM clustering consists of two basic procedures, namely (1) calculating the cluster centers and assigning the data points to these centers on the basis of their Euclidean distance, and (2) determining the cluster memberships of each sample point. The first procedure is repeated iteratively until the cluster centers remain stable from one iteration to the next. However, before the first iteration can be performed, it is necessary to select an initial set of membership values. In doing so, FCM imposes the following direct constraint on the fuzzy membership function associated with each point:

$$\sum_{j \ 1}^{p} \mu_j\ x_i \quad 1; \quad i \quad 1; 2; 3 \ldots; k;$$

where p is the number of specified clusters, k is the number of objects, x_i is the ith object, and $\mu_j\ x_i$ returns the membership value of x_i in the jth cluster, i.e. the degree of belongingness of x_i to the jth cluster. Clearly, the sum of the cluster membership values of each object from one specific attribute must be equal to one (Bezdek, 1981).

The objective of FCM is to minimize a standard loss function expressed as the weighted sum of the squared error within each cluster, i.e.

$$l \quad \sum_{j \ 1}^{p} \sum_{i \ 1}^{n} \mu_j\ x_i \ ^{m'} \|x_i \times c_j\|^2; \quad 1 < m' < \infty;$$

where p is the number of specified clusters, n is the number of objects, $\mu_j\ x_i$ is the membership value of object x_i in the jth cluster, x_i is the ith object, m' is the fuzzification parameter, and c_j is the center of the jth cluster.

It has been shown by Bezdek (1981) that if $| \ |_i x \times c_j| \ | > 0$ for all i and j, then the loss function is minimized when $m' > 1$. Under this condition, the corresponding cluster center value can be computed in accordance with

$$c_j \quad \frac{\sum_i \mu_j\ x_i \ ^{m'} x_i}{\sum_i \mu_j\ x_i \ ^{m'}} \quad \text{for } 1 \leqslant j \leqslant p:$$

Having calculated the cluster centers, the second step in the FCM procedure is to determine the cluster memberships of each sample point. To do so, it is necessary to determine the distance from each point x_i to each of the cluster centers $\ c_1; c_2; \ldots; c_p$. In practice, this is achieved by computing the Euclidean distance between the point and the cluster center in accordance with

$$d_{ji} \ \| \quad x_i \times c_j\|^2;$$

where d_{ji} is the distance of x_i from the center of cluster c_j.

Since the FCM algorithm constrains the total cluster membership value of a single point to one, one can recalculate the membership of a point to a particular cluster is expressed as a fraction of all the total possible memberships assigned to that point. In other words, the membership of a particular object x_i to the jth cluster is given by

$$\mu_j(x_i) = \frac{\left(\frac{1}{d_{ji}}\right)^{\frac{1}{m'-1}}}{\sum_{k=1}^{p}\left(\frac{1}{d_{ki}}\right)^{\frac{1}{m'-1}}} = \frac{1}{\sum_{k=1}^{p}\left(\frac{d_{ji}}{d_{ki}}\right)^{\frac{1}{m'-1}}} \quad \text{for } 1 \leqslant j \leqslant p; \quad 1 \leqslant i \leqslant n;$$

where d_{ji} is the distance metric of x_i from the center of cluster c_j, m' is the fuzzification parameter, p is the number of specified clusters, and d_{ki} is the distance metric of x_i from the center of cluster c_k.

Having computed the value of $\mu_j(x_i)$, it is used in place of the original value of $\mu_j(x_i)$ in the first step of the FCM procedure. This two-step procedure is repeated iteratively until the centers of all the clusters within the dataset converge.

(2) Index function $I_{\max}$.

Suppose that there exists only one conditional attribute for object x_i. Furthermore, assume that this attribute can be divided into p groups such that each object owns p membership functions $\mu_j(x_i)$; $j = 1, 2, \ldots, p$. The index function $I_{\max}$ is defined as

$$I_{\max}(\mu_j(x_i)) = \text{Index}(\max(\mu_j(x_i))) = C(x_i) = k; \quad 1 \leqslant k \leqslant p;$$

where $I_{\max}(\mu_j(x_i))$ is a clustering index in which the maximum membership value of point x_i corresponds to the fuzzy cluster; and $C(x_i)$ is determined by the maximum membership value and indicates the cluster to which the object x_i belongs. For example, suppose that the attribute can be divided into three clusters and the values of membership functions of the first object x_1 are given by $\mu_1(x_1) = 0.35$, $\mu_2(x_1) = 0.63$ and $\mu_3(x_1) = 0.02$, respectively. By definition, $C(x_1) = I_{\max}(\mu_j(x_1)) = 2$, and thus the first object belongs to the second cluster.

The example described above is easily extended to the case of multiple attributes. For example, if every object has m conditional attributes and the lth attribute a_l can be divided into p_l clusters, then $C_l(x_i)$ gives the index of the cluster to which the lth attribute a_l of object x_i belongs. Here $C_l(x_i)$ is given by

$$C_l(x_i) = I_{\max}(\mu_j(x_i, a_l)) = \text{Index}(\max(\mu_j(x_i, a_l)))$$
$$\text{for } 1 \leqslant l \leqslant m; \quad 1 \leqslant i \leqslant n.$$

where $I_{\max}(\mu_j(x_i, a_l)) =$ index of the cluster corresponding to the maximum value of the membership functions x_i associated with the lth attribute.

(3) Implication relations (Tsoukalas & Uhrig, 1997).

Fuzzy *if/then* rules are conditional statements describing the dependence of one (or more) linguistic variable(s) on another variable. The underlying analytical form of an *if/then* rule is a fuzzy relationship known as an implication relation. More than 40 different forms of implication relation have been reported in the literature (Lee, 1990a, 1990b). Typically, these implication relations are acquired by inputting the left-hand side (LHS) and right-hand side (RHS) of a rule into an implication operator, ϕ. The choice of an appropriate implication operator is a significant step in the overall development of a fuzzy linguistic description and reflects both the application-specific criteria and the logical and intuitive considerations relating to the interpretation of the connectives *AND*, *OR* and *ELSE*. Note that an extensive discussion of the most common implication relations can be found in Refs. Lee (1990a, 1990b), Mizumoto (1988) & Ruan & Kerre (1993).

Consider a generic rule involving two linguistic variables, one on each side of the rule expression, i.e.

if x is A then y is B;

in which the linguistic variables x and y take values of A and B, respectively.

The underlying analytical form of this rule is given by the following implication relation:

$$R(x, y) = \int_{x,y} \mu(x, y)/(x, y);$$

where $\mu(x, y)$ is the membership function of the implication relation $R(x, y)$ (Tsoukalas & Uhrig, 1997).

Several options exist for acquiring the membership function of an implication relation. For the rule given above, the implication operator ϕ takes the membership functions of the antecedent and consequent parts, i.e. $\mu_A(x)$ and $\mu_B(y)$, respectively, and generates an output $\mu(x, y)$, i.e.

$$\mu(x, y) = \phi[\mu_A(x), \mu_B(y)].$$

Typical implication operators include:

(a) Zadeh max–min implication operator

$$\phi_m[\mu_A(x), \mu_B(y)] = \mu(x, y) = (\mu_A(x) \wedge \mu_B(y)) \vee (1 - \mu_A(x)).$$

(b) Mamdani min implication operator

$$\phi_c[\mu_A(x), \mu_B(y)] = \mu_A(x) \wedge \mu_B(y).$$

(4) Fuzzy algorithms.

Fuzzy algorithms are essentially an automated procedure for interpreting a linguistic statement formulated as a collection of fuzzy rules. All of the rules within the collection are formed by the same attributes and are connected by the connective *ELSE*. The final result may be interpreted as either a union or an intersection of the rules depending on the implication operator used in the individual rules. For example, consider the following set of rules:

if x is A_1 then y is B_1 ELSE
if x is A_2 then y is B_2 ELSE
...
if x is A_n then y is B_n

Recall that each rule above is represented analytically by an implication relation $R(x, y)$, where the form of $R(x, y)$ depends on the choice of implication operator. In this study, the interpretations of the connective *ELSE* are given as follows:

Implication: ϕ_m (Zadeh), Interpretation of *ELSE* : *AND*$(\wedge)$
Implication: ϕ_m (Mamdani), Interpretation of *ELSE* : *OR*$(\vee)$

The relation describing the entire collection of rules given above is known as the algorithmic relation and has the form

$$R_\alpha(x, y) = \int_{x,y} \mu_\alpha(x, y)/(x, y).$$

Note that R_α is either a union $(\vee)$ or an intersection $(\wedge)$ of the implication relations of the individual rules depending on the chosen implication operator.

In an elementary fuzzy algorithm, each implication rule has just one variable in the antecedent side and one variable in the consequent side. However, in general applications, the linguistic descriptions have more than one variable on either side and the corresponding algorithm is referred to as a multivariate fuzzy algorithm. The interpretation of the connective *ELSE* in a multivariate fuzzy algorithm is the same as that in an elementary fuzzy algorithm. Consider an rule of the form

if x_1 is A_1 AND x_2 is A_2 ... AND x_m is A_m then y is B;

where $x_1;\ldots;x_m$ are the antecedent linguistic variables and have $A_1;\ldots;A_m$ as their corresponding fuzzy values, and y is the consequent linguistic variable with a fuzzy value B. The connectives *AND* in the LHS of the rule given above can be modeled analytically as either min ($\wedge$) or product ( ). In such a case, the proposition in the LHS of the rule can be combined through min ($\wedge$) and an appropriate implication operator ϕ can then be applied to acquire the membership function of the implication relation. Thus, it can be shown that

$$\mu(x_1;x_2;\ldots;x_m;y) = \phi[\mu_{A_1}(x_1) \wedge \mu_{A_2}(x_2) \wedge \cdots \wedge \mu_{A_m}(x_m);\mu_B(y)];$$

in which ϕ is an appropriate implication operator (as described in the previous section).

2.2.2. Detailed procedure for determining β value

It seems reasonable to speculate that the occurrence of errors when classifying continuous systems can be attributed at least in part to inaccuracies in the fuzzy clustering phase performed before the actual classification procedure. Accordingly, this study presents a five-step procedure for determining an appropriate value of the threshold parameter β in the VPRS model.

Step 1: Fuzzify the attributes of the information system using the FCM method.

A continuous-value information system can only be converted into an equivalent fuzzy information system under the condition once a classified fuzzy set has been provided. Assume that each attribute of the information system can be divided into an arbitrary number of clusters, e.g. the notation $a_1 = \{\widetilde{A}_{11};\widetilde{A}_{12}\}$ indicates that the attribute a_1 has a certain degree of belonging to two different clusters.

In the FCM clustering method, the interval values $[\alpha;\beta]$ of the lth attribute a_l can be divided into p_l fuzzy clusters, and the continuous-value information system $(U;A;V_q;f_q)$ can then be converted into the fuzzy information system $(U;A;\Phi;d)$, where $\Phi = \{\widetilde{A}_{lj}| l \leqslant m; j \leqslant p_l\}$, in which $\widetilde{A}_{lj} = \mu_j(x_i(a_l))$. Here, $\mu_j(x_i(a_l))$ represents the values of the membership functions associated with the lth conditional attribute a_l of the ith object.

Step 2: Find the cluster(s) conditional attribute.

Utilizing the index function $I_{\max}(\mu_j(x_i(a_l))) = \text{Index} \max(\mu_j(x_i(a_l)))$ for $1 \leqslant l \leqslant m; 1 \leqslant i \leqslant n$, identify the cluster(s) corresponding to the lth attribute (a_l) of every object $x_i(a_l)$.

Step 3: Arrange the implication relations.

As discussed above, the use of different implication operators results in the construction of different implication relations. The present study utilizes the Zadeh max–min implication operator, which yields the following implication relation:

$$\phi_M[\mu_A(x);\mu_B(y)] = \mu(x;y) = [\mu_A(x) \wedge \mu_B(y)] \vee [1 \times \mu_A(x)]:$$

Suppose that there exist m conditional attributes and one decision attribute for each object. The implication relation for the ith object (x_i) is therefore given as

$$\phi_M[\mu_{a_1}(x_i(a_1)) \wedge \mu_{a_2}(x_i(a_2)) \wedge \cdots \wedge \mu_{a_m}(x_i(a_m));\mu_d(x_i(d))]$$
$$= \mu(x_i(a_1);x_i(a_2);\ldots;x_i(a_m);x_i(d));$$

where $\mu_j(x_i(a_l))$ represents the jth membership function of the lth conditional attribute (a_l) of the ith object (x_i), and $\mu_j(x_i(d))$ is the jth membership function of the decision attribute (d) of the ith object (x_i).

If the decision attribute is crisp (i.e. it has a value of 0 or 1), the implication relation given above has the form $\phi_M[\mu_{a_1}(x_i(a_1)) \wedge \mu_{a_2}(x_i(a_2)) \wedge \cdots \wedge \mu_{a_m}(x_i(a_m));1] = \mu(x_i(a_1);x_i(a_2);\ldots;x_i(a_m);1)$.

For example, suppose that two conditional attributes $(a_1;a_2)$ and one decision attribute (d) are to be classified using the FCM approach. Suppose further that the objective of the clustering process is to divide the conditional attributes into three clusters and the decision attribute into two clusters. Finally, suppose that the ith object's class data set is given by (1, 3, 2), where (1, 3) denotes the clustering indexes of the two conditional attributes $(a_1;a_2)$, respectively, and (2) denotes the clustering index of the decision attribute (d). In general, there exist a total of m classes for classified objects for which the decision attribute index has a value of (2). Suppose that there exist three objects complying with the class (1, 3, 2). Suppose further that the serial numbers of these three objects are 4, 7 and 9, respectively. The implication relation of the object with a serial number of "4" can be expressed as

$$\phi_M[\mu_{a_1}(x_4(a_1)) \wedge \mu_{a_2}(x_4(a_2));\mu_d(x_4(d))] = \mu(x_4(a_1);x_4(a_2);x_4(d)):$$

Step 4: Determine the value of the VPRS β parameter using a fuzzy algorithm.

Taking the implication relation example given in Step 3 for illustration purposes, the objects, which belong to the same lower approximation of RS, can be related to the linguistic descriptions in a fuzzy algorithm by means of the connective operator *ELSE*, i.e.

$$\begin{aligned}
&\textit{if } a_1 \textit{ is } \mu_1(x_4(a_1)) \textit{ and } a_2 \textit{ is } \mu_3(x_4(a_2)) \textit{ then } d \textit{ is } \mu_2(x_4(d)) \textit{ ELSE};\\
&\textit{if } a_1 \textit{ is } \mu_1(x_7(a_1)) \textit{ and } a_2 \textit{ is } \mu_3(x_7(a_2)) \textit{ then } d \textit{ is } \mu_2(x_7(d)) \textit{ ELSE};\\
&\textit{if } a_1 \textit{ is } \mu_1(x_9(a_1)) \textit{ and } a_2 \textit{ is } \mu_3(x_9(a_2)) \textit{ then } d \textit{ is } \mu_2(x_9(d));
\end{aligned}$$

where $\mu_j(x_i(a_l))$ represents the jth membership function of the lth conditional attribute (a_l) of the ith object (x_i), and $\mu_j(x_i(d))$ represents the jth membership function value of the decision attribute (d) of the ith object (x_i). Note that as described above, the interpretation of the connective operator *ELSE* in the fuzzy algorithm varies in accordance with the particular choice of implication operator, i.e. *AND*$(\wedge)$ for the Zadeh implication operator or *OR*$(\vee)$ for the Mamdani implication operator.

Finally, the β value can be obtained directly using a fuzzy algorithm as follows:

$$\beta = \mu_q(x_i(a_1);x_i(a_2);\ldots;x_i(a_m);x_i(d)) \quad \text{where } x_i \in X_q:$$

Here, $\mu_q(x_i(a_1);x_i(a_2);\ldots;x_i(a_m);x_i(d))$ is the threshold value obtained from the fuzzy algorithm for a certain q classified objects comprising a set X_q given a fixed decision attribute. In other words, a unique value of β exists for each set of classified objects.

Step 5: Identify VPRS regions.

Having determined the value of β using the procedure outlined above, it is then possible to determine whether the classified objects belong to the β-lower approximation of the set, the β-upper approximation of the set, or the β-boundary regions as defined by $\underline{R}_P^\beta(X)$, $\overline{R}_P^\beta(X)$, and $BND_P^\beta(X)$, respectively (see Section 2.2).

2.3. GM(1,N) model

Imagine a system described by the sequences $x_i^{(0)}(k); i = 1;2;3;\ldots;n$, in which $x_1^{(0)}(k)$ describes the main factor of interest and sequences $x_2^{(0)}(k);x_3^{(0)}(k);\ldots;x_n^{(0)}(k)$ are the factors which affect this main factor. Such a system can be analyzed using the following multivariate GM(1,N) Grey model:

$$x_1^{(0)}(k) + az^{(1)}(k) = \sum_{j=2}^{N} b_j x_j^{(1)}(k); \quad \text{where } k = 2;3;\ldots;n;$$

in which $x_j^{(1)}(k) = \sum_{i=1}^{k} x_j^{(0)}(i)$ and $z_1^{(1)}(k) = 0.5x_1^{(1)}(k) + 0.5x_1^{(1)}(k \times 1)$; $k \geqslant 2$.

Substituting all possible $x_j^{(1)}(k)$ terms into above equation yields a matrix of the form

$$X_N = \begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(n) \end{bmatrix} = \begin{bmatrix} \times z_1^{(1)}(2) & x_2^{(1)}(2) & \ldots & x_n^{(1)}(2) \\ \times z_1^{(1)}(3) & x_2^{(1)}(3) & \ldots & x_n^{(1)}(3) \\ & \vdots & & \\ \times z_1^{(1)}(n) & x_2^{(1)}(n) & \ldots & x_n^{(1)}(n) \end{bmatrix} \begin{bmatrix} a \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = B\hat{a}:$$

The values of $b_j; j = 2; 3; \ldots; N$ can be found by applying the matrix operation $\hat{a} = (B^T B)^{-1} B^T X_N$. The relative influence exerted on the major sequence by each influencing sequence can then be determined by inspecting the b_j value of the corresponding sequence.

2.4. Grey relational analysis

In grey relational analysis (GRA), data characterized by the same set of features are regarded as belonging to the same series. The relationship between two series can be determined by evalu-

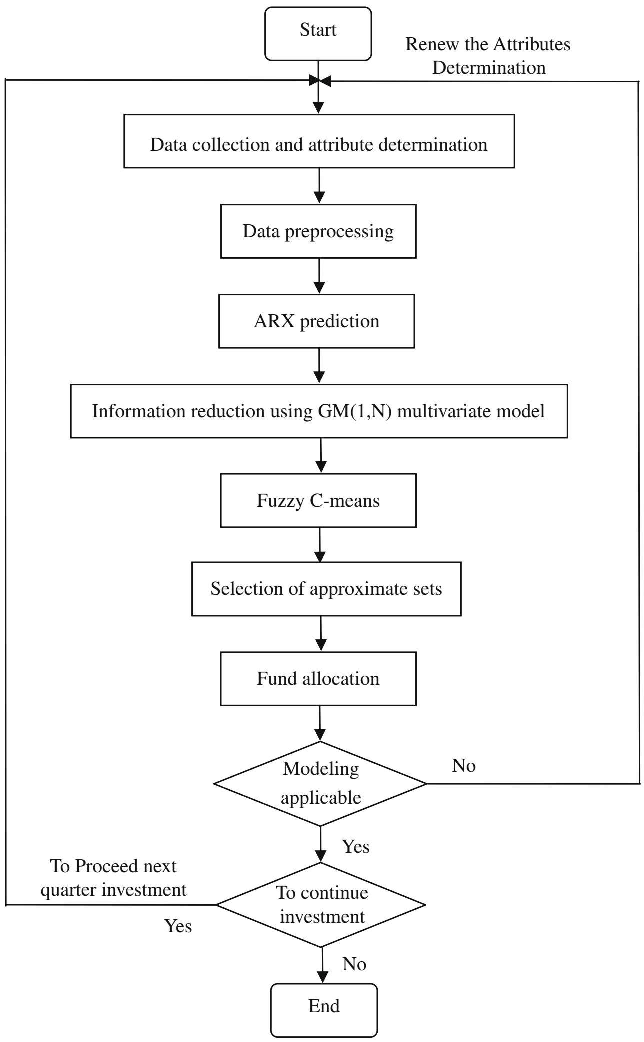

Fig. 1. Flow chart of proposed forecasting and stock selection model.

proposed model, this pruning operation is performed by using a GM(1,N) multivariate scheme to identify the top ten influential sequences (i.e. conditional attributes) and then eliminating the remainder.

Step 5: Fuzzy C-Means clustering

Prior to submission to the VPRS stock selection mechanism, the forecast values of the conditional attributes $C_1 \quad C_n$ are clustered into three groups using a FCM clustering algorithm.

Step 6: Selection of approximate sets

Having clustered the forecast data, the VPRS method is applied to determine the β-lower approximate set. The appropriate investment stocks are derived as the β-lower approximate set is acquired.

Step 7: Fund allocation

Having identified suitable stocks in which to invest, appropriate weights should be assigned to each stock in order to optimize the overall rate of return on the portfolio. In the proposed hybrid model, this fund allocation problem is processed using a Grey Relation Sequence algorithm based on the metric *stock weight i* $\quad \frac{n\times i \;\; 1}{\sum_{i \;\; n} i}$, where i is the grey relation order of each stock item and n is the total number of invested stocks.

Having completed all of the steps described above, a check is made on the overall rate of return on the investment. If the rate of return is acceptable, a decision is made as to whether or not the model should be run for a further quarter using the existing attributes. However, if the rate of return is deemed unacceptable, the suitability of the conditional attributes is reviewed and the choice of attributes amended as appropriate.

4. Evaluation of proposed hybrid model using electronic stock data

4.1. Data extraction

The feasibility of the proposed forecasting and stock selection mechanism was evaluated using electronic stock data extracted from the New Taiwan Economy database (TEJ). The data collection period extended from the first quarter of 2003 to the fourth quarter of 2006, giving a total of 16 quarters in all. Meanwhile, the forecasting period extended from the first quarter in 2003 to the first quarter in 2007, giving a total of 17 quarters.

In general, the financial statements for a particular accounting period are subject to a considerable delay before they are actually published. For example, annual reports are published after four months, half-yearly reports after two months, and first and third quarterly reports (without notarization) after around one month. The submission deadlines for the financial statements maintained in the TEJ database are as follows:

(1) Annual report: the submission deadline laid down by the Security Superintendence Commission is 4 months after the closing balance day. However, companies listed in previous years (TSE and OTC) are permitted to delay filing until 5/31.
(2) Half-yearly report: the submission deadline laid down by the Security Superintendence Commission is 2 months after the closing balance day. However, companies listed in previous years (TSE and OTC) can delay filing until 9/21.
(3) First-quarter report: the submission deadline laid down by the Security Superintendence Commission is 1 month after the closing balance day. However, companies listed in previous years (TSE and OTC) can delay filing until 5/31.
(4) Third-quarter report: the submission deadline laid down by the Security Superintendence Commission is 1 month after the closing balance day. However, companies listed in previous years (TSE and OTC) are permitted to delay filing until 11/15.

Since the financial report for the final quarter of any year is not available until May 31st the following year, the financial data can not be used by the ARX model to predict the financial trends over the first quarter of the year. In other words, the forecasting and stock selection process proposed in this study can only be performed three times every year, i.e. 5/31–09/22, 9/22–11/15 and 11/15–05/31 the following year. In addition, in the decision-making rules used in the VPRS stock selection process, the Return on Equity (ROE) and constant EPS indicators are based on the full 12 months of the previous year. Thus, the forecasting period for investment purposes is reduced from the initial time frame of the first quarter in 2003 to the first quarter in 2007 to the second quarter in 2004 to the fourth quarter in 2006.

4.2. Performance evaluation of VPRS-based stock forecasting and selection mechanism

In this study, the performance of the proposed VPRS-based stock forecasting and selection mechanism was benchmarked against that of a RS-based selection scheme. Tables 1a and 1b summarizes the number of indiscernible classes and the number of companies selected by the RS- and VPRS-based models as a function of the

Table 1a
Comparison of number of indiscernible classes and selected companies for different numbers of conditional attributes and a single decision attribute. Note that each conditional attribute is associated with three fuzzy clusters and the cluster code of the decision attribute is specified as "1"(⩾0).

Number of conditional attributes	2	3	4	5	9	10
Number of indiscernible classes as computed using RS model	5	10	17	21	31	33
(a) Number of companies selected by RS model	22	23	29	34	35	35
Number of indiscernible classes as computed using VPRS model	7	11	18	22	31	33
(b) Number of companies selected by VPRS model	39	33	35	37	35	35
(b)–(a)	17	10	6	3	0	0

Table 1b
Comparison of code(s) of indiscernible classes and selected companies for different numbers of conditional attributes and a single decision attribute. Note that each conditional attribute is associated with three fuzzy clusters and the cluster code of the decision attribute is specified as "1"(⩾0).

Two conditional attributes			Three conditional attributes		
Codes of the conditional attributes	Company code(s) of indiscernible classes as computed using RS model	Company code(s) of indiscernible classes as computed using VPRS model	Codes of the conditional attributes	Company code(s) of indiscernible classes as computed using RS model	Company code(s) of indiscernible classes as computed using VPRS model
{2, 2}	{22, 58, 250, 272}	{22, 58, 250, 272}	{2, 2, 3}	{22}	{22}
			{2, 2, 2}	{58, 250}	{58, 250}
			{2, 2, 1}	{272}	{272}
{1, 2}	{75, 77, 95, 99, 116, 134, 171, 201, 246, 262, 269, 275}	{75, 77, 95, 99, 116, 134, 171, 201, 246, 262, 269, 275}	{1, 2, 2}	{75, 77, 99, 116, 134, 171, 201, 246, 262, 269, 275}	{75, 77, 99, 116, 134, 171, 201, 246, 262, 269, 275}
			{1, 2, 1}	{95}	{95}
{1, 1}	{132, 188}	{132, 188}	{1, 1, 1}	{132, 188}	{132, 188}
{3, 1}	{133, 243}	{133, 243}	{3, 1, 1}	{133}	{133}
			{3, 1, 3}	{243}	{243}
{2, 3}	{242, 259}	{242, 259}	{2, 3, 3}	{242, 259}	{242, 259}
{3, 3}		{49, 61, 179, 239, 251}	{3, 3, 1}	{61}	{61}
[a]{3, 3}		{54, 81}			
{2, 1}		{55, 126, 137, 153, 192, 225, 234, 257, 263}	{2, 1, 1}		{55, 126, 137, 153, 192, 225, 234, 257, 263}
[a]{2, 1}		{156}	[a]{2, 1, 1}		{156}

[a] Indicates the cluster code of the decision attribute is specified as "2"(<0).

Table 2a
Rate of return in 4th quarter 2006 using proposed VPRS-based model.

Listed Company Code	Call(11/15)	Number	Put(5/31) next year	Return rate (%)
2368	19.93	16	24.15	21.17
3037	36.20	7	45.58	25.91
6192	38.82	5	59.11	52.27
2483	23.15	5	22.63	×2.25
2423	16.76	3	32.99	96.83
Total return rate: 30.11%				

Table 2b
Rate of return in 4th quarter 2006 using RS-based model.

Listed Company Code	Call(11/15)	Number	Put(5/31) next year	Return rate (%)
2368	19.93	16	24.15	21.17
3003	27.36	9	31.17	13.92
3037	36.20	5	45.58	25.91
6166	29.45	3	36.55	24.10
3042	46.63	1	58.99	26.51
Total return rate: 20.81%				

number of conditional attributes. As shown, the number of indiscernible classes increases with an increasing number of conditional attributes in both the RS- and the VPRS-based models. However, in both models, the number of objects within each indiscernible class reduces as the number of conditional attributes increases. From inspection, it is evident that for a given number of conditional attributes, the number of indiscernible classes in the VPRS-based scheme is either higher than or equal to that in the RS-based model. VPRS performs a classification function with a controlled degree of misclassification error falling outside the realm of the RS approach. For the case where the number of conditional attributes is specified as {2, 3, 4, 5}, the VPRS model increases the number of indiscernible classes by {2, 1, 1, 1} relative to that in the RS scheme. However, when a greater number of conditional attributes are considered, i.e. {9, 10}, the VPRS and RS models yield an identical numbers of indiscernible classes. Furthermore, it is observed that for a lower number of conditional attributes, i.e. {2, 3, 4, 5}, the VPRS model yields a greater number of selected companies, while for a higher number of attributes, i.e. {9, 10}, the two schemes select an identical number of companies. The probability of which two objects are

Table 3
Comparison of rates of return obtained using proposed VPRS-based model and RS-based model.

Investment period		Forecasting method			
		RS		VPRS	
		Quarter	Yearly	Quarter	Yearly
		Rate of return			
2004	Second quarter	×13.68%	×2.66%	×13.68%	×2.66%
	Third quarter	1.36%		1.36%	
	Fourth quarter	9.66%		9.66%	
2005	Second quarter	14.44%	56.29%	14.44%	53.17%
	Third quarter	3.34%		3.34%	
	Fourth quarter	38.51%		35.39%	
2006	Second quarter	×3.97%	28.17%	×3.97%	37.46%
	Third quarter	11.33%		11.33%	
	Fourth quarter	20.81%		30.11%	
Average yearly rate of return		27.27%		29.33%	
Accumulated rate of return over 3 years		81.80%		87.98%	

indiscernible due to identical clustering indexes would be reduced by increasing the number of conditional attributes or increasing the number of clusters in each attribute.

Tables 2a and 2b summarize the rates of return in the 4th quarter of 2006 obtained from the stock companies selected using the RS-based model and the proposed VPRS-based model, respectively. It can be seen that the two models recommend a slightly different selection of companies for investment purposes. The VPRS model yields a greater number of selected companies and results in a slightly different selection of companies. It can also be seen that the rate of return provided by the companies selected using the VPRS-based model is higher than that obtained from the RS-based model.

Table 3 summarizes the quarterly and yearly rates of return obtained over the nine investment periods between 2004 and 2006 using the proposed VPRS-based fusion model and the RS-based model, respectively. Note that in both models, the results are obtained using five conditional attributes and one decision attribute. As shown, the fusion model achieves a higher average yearly rate of return than the RS-based model.

5. Conclusions and discussions

This study has presented a mechanism for predicting stock market trends and selecting the optimal stock portfolio based upon an ARX prediction model, Grey Systems theory and VPRS theory. The performance of the VPRS-based approach has been compared with that of a RS-based method using electronic stock data extracted from the New Taiwan Economy database (TEJ). The major findings and contributions of this study can be summarized as follows:

(1) In contrast to the β-reducts method used in previous studies, this study determines an appropriate value of the VPRS threshold parameter β using a Fuzzy C-Means clustering approach. Suppose that each object possesses a number of attributes which can each be assigned to multiple clusters during the fuzzy classification process. According to the definition of the index function $I_{\max}$, the membership function of a particular attribute is represented by the value of the maximum membership function of the attribute amongst all of the clusters with which it is associated. The value of the VPRS threshold parameter β is then derived using the Zadeh implication operator and the Zadeh fuzzy algorithm operator (*AND*).

(2) The probability of two objects being indiscernible as the result of having an identical clustering index reduces as the number of conditional attributes considered in the clustering process increases. The number of objects in any indiscernible class of the RS lower approximation set may to be reduced to one. For the case where the number of objects in a particular classified group is equal to one, the lower RS approximation for that classification is identical to the β-lower approximation obtained using the VPRS model irrespective of the value assigned to the β threshold parameter.

(3) The VRPS β-lower approximation set contains a greater number of selected stocks than the RS lower approximation set, and therefore results in a different ranking of the selected stocks by the GRA scheme. The portfolio recommendations obtained using the proposed VPRS-based model differ from those obtained from a RS-based hybrid model and yield an improved rate of return.

(4) The generalized rules extracted by the β-lower approximate set are all found to be recognized rules or relationships in the investment industry, which indicates that the VPRS method not only provides a feasible means of identifying top stock performers by classifying the contributions of the attributes, but is also helpful in constructing decision rules with which to evaluate new stocks. In other words, the VPRS model removes the requirement for the blind, haphazard stock selection methods typically used by investors in the past.

VPRS is essentially an extension of RS in that setting an appropriate value of the threshold value β eases the strict definition of the approximation boundary in the RS model. As a result, the indiscernible classes in the VPRS model include all of the indiscernible classes in the RS model. However, as the number of conditional attributes or number of clusters associated with each attribute increases, the number of objects in each class decreases to one. Under this condition, both the number of indiscernible classes and the object in each class are identical in the VPRS and RS models.

Overall, the results presented in this study have confirmed that the proposed VPRS-based model provides a promising tool for stock portfolio management purposes. Furthermore, the structure of the proposed model represents a suitable foundation for a broad range of derivatives in other research fields. In the current model, the fund allocation task is performed using the Grey relational analysis method proposed by Huang and Jane (2008). However, a future study will investigate the feasibility of improving the rate of return obtained from the selected stocks by integrating the fusion model developed in this study with a heuristic scheme designed to optimize the fund weight distribution subject to the constraint of minimizing the portfolio risk.

References

Abraham, A., Nath, B., & Mohanthi, P. K. (2001). Hybrid intelligent systems for stock market analysis. In V. N. Alexandrov et al. (Eds.), *Computational science* (pp. 337–345). Germany, San Fransisco, USA: Springer-Verlag.

Aiken, M., & Bsat, M. (1999). Forecasting market trends with neural networks. *Information Systems Management, 16*(4), 42–48.

An, A., Shan, N., Chan, C., Cercone, N., & Ziarko, W. (1996). Discovering rules for water demand prediction: An enhanced rough-set approach. *Engineering Application and Artificial Intelligence, 9*, 645–653.

Blaszczyński, J., Greco, S., & Slowiński, R. (2007). Multi-criteria classification-A new scheme for application of dominance-based decision rules. *European Journal of Operational Research, 181*, 1030–1044.

Bauer, R. J. Jr., (1994). *Genetic algorithms and investment strategies*. New York: Wiley.

Bazan, J. G., & Szczuka, M. (2000). RSES and RSESlib – A collection of tools for rough