Appl Intell (2010) 32: 364–377
DOI 10.1007/s10489-008-0150-y

Using rough set to support investment strategies of real-time trading in futures market

Suk Jun Lee · Jae Joon Ahn · Kyong Joo Oh · Tae Yoon Kim

Published online: 21 October 2008

Abstract Finding proper investment strategies in futures market has been a hot issue to everyone involved in major financial markets around the world. However, it is a very difficult problem because of intrinsic unpredictability of the market. What makes things more complicated is the advent of real-time trading due to recent striking advancement of electronic communication technology. The real-time data imposes many difficult tasks to futures market analyst since it provides too much information to be analyzed for an instant. Thus it is inevitable for an analyst to resort to a rule-based trading system for making profits, which is usually done by the help of diverse technical indicators. In this study, we propose using rough set to develop an efficient *real-time rule-based trading system (RRTS)*. In fact, we propose a procedure for building RRTS which is based on rough set analysis of technical indicators. We examine its profitability through an empirical study.

Keywords Rule-based system · Real-time trading · Rough set · Futures market · Technical indicator

S.J. Lee · J.J. Ahn · K.J. Oh (✉)
Department of Information and Industrial Engineering, Yonsei University, 134 Shinchon-Dong, Seodaemun-Gu, Seoul 120-749, South Korea
e-mail: johanoh@yonsei.ac.kr

S.J. Lee
e-mail: lsj77@yonsei.ac.kr

J.J. Ahn
e-mail: redsamzang@yonsei.ac.kr

T.Y. Kim
Department of Statistics, Keimyung University, Daegu 704-701, South Korea
e-mail: tykim@kmu.ac.kr

1 Introduction

Finding proper investment strategies in the futures market is quite difficult since many factors such as political events, general economic conditions, and investors' expectations influence the market in a complicated fashion [3]. To resolve this difficulty, quite a few researchers in finance and investment firms have used expert systems with artificial intelligence technologies as proper technical tools [2, 9, 14, 16, 27]. Their approach basically employs case-based reasoning approach, i.e., it tries to find a similar case from the past to a given situation. Recently, real-time trading becomes popular as various real-time data become available throughout the market [13]. Problems with real-time data are that they provide too much information to be analyzed for an instant and hence use of the case-based reasoning approach for real-time data appears to be limited. In other words, it is hard to figure out a similar case from the past to a given situation at every moment. Thus, most analysts in the market start searching a proper rule applicable to real-time trading and they found various technical indicators to be of great use for finding proper rules [1, 7, 17, 18].

This study proposes a procedure for constructing an efficient real-time investment strategies or *real-time rule-based trading system (RRTS)* which uses technical indicators and rough set analysis. Rough set analysis is known to be quite valuable for extracting trading rules from huge data such as real-time data since it can be used to discover dependences in the data while reducing the effect of superfluous factors in the data [24]. Through empirical studies, we will show that rough set analysis could yield an efficient RRTS against various situations of futures market. Our work is anew in the sense that the rough sets analysis is applied to real-time data in stock futures market for the first time. Until now most studies in stock futures market have focused finding simple rules from technical indicator analysis for historical data [4, 15, 28].

The rest of the paper is organized as follows. Section 2 briefly reviews the futures market, technical indicators, real time data, and the rough set theory. Section 3 describes RRTS construction procedure, particularly rule base generation procedure. In Sect. 4, empirical studies are given to illustrate RRTS construction procedure and verify its efficiency against various market situations. Finally, the concluding remarks are presented in Sect. 5.

2 Background

A futures market is a central financial market where people can trade standardized futures contracts; that is, a contract to buy specific quantities of a commodity or financial instrument at a specified price with delivery set at a specified time in the future. Stock futures market is a futures market that deals with stock price movement as commodity. In stock futures market, one may take a marginal profit by buying contract to buy (long position) when the bull market is ahead and buying contract to sell (or short position) when the bear market is ahead. Thus it is a market based on the directivity of stock price movement which offers the opportunity of making profit in both bull and bear markets. There are various types of trading rules developed for making profits in stock futures market. For instance, ordinal trading rule is concerned about making profit from predicting bullish or bearish stock market ahead (see also discussion of technical analysis below) while cardinal trading rule provides some rules that every aspiring trader must learn in order to succeed (e.g., observe before decide to participate, or cut loses on losing propositions, etc.). In addition, arbitrage trading is another type of trading in stock futures market. For example, index arbitrage trading is a strategy designed to profit from temporary discrepancies between the prices of the stocks comprising an index and the price of a futures contract on that index. By buying either the stocks or the futures contract and selling the other, an investor can sometimes exploit market inefficiency for a profit. Like all arbitrage opportunities, index arbitrage opportunities disappear rapidly once the opportunity becomes well-known and many investors act on it.

Stock prices are basically determined by intrinsic value of the underlying stock. However, since stock futures market is built for the purpose of risk hedging via directivity prediction of the stock prices movement, it is quite essential for proper trading in futures market to predict the stock price fluctuation accurately. Thus, technical analysis for prediction is more popular than intrinsic value evaluation in the stock futures market. Technical analysis basically studies the historical data relevant to price and volume movements of the stock by using charts as a primary tool to forecast future price movements [18]. There are various kinds of technical indicators used in futures market including the initial one dating back to the 1800s [5]. For more detailed references, see Murphy [17, 18], Achelis [1], Colby [7] or Appendix in this paper. Technically, real-time data denotes information that is delivered immediately after collection. There is no delay in the timeliness of the information provided. On the other hand, delayed or historical data is delivered after some delay, usually 10 to 30 minutes, from its real exchange, and historical data is usually adjusted after being combined with real-time data. Croushore and Stark [8] researched into a real-time data set for macroeconomists and Gerberding et al. [11] studied a real-time data set for German macroeconomic variables.

Rough sets theory, introduced by Pawlak [20, 21] originally, is a mathematical approach to managing uncertain data. This theory utilizes "rules" for managing such data. Using this theory, one can handle various types of huge or high dimensional data efficiently and, in this sense, it is a data mining tool [6, 26]. The rough sets theory is built on the notion of indiscernibility, which refers to the inability to distinguish objects by a given set of attributes (or a given set of categories). In fact, objects featured by the same attributes are indiscernible in view of a given set of attributes. Thus, if one calls an elementary set as any set of indiscernible objects, then objects in a class of elementary sets form those that can be clearly distinguished in terms of a given set of attributes. Since in practice it is very likely to have sets of objects that will not be expressed clearly in terms of elementary sets, objects will have to be approximated roughly through a pair of sets which give lower and upper approximations. A rough set is any subset defined through its lower and upper approximation. Figure 1 quoted from Øhrn [19] gives an illustration of this concept where each indiscernibility set is displayed by a pixel. The subset of objects that needs to be approximated is drawn as a dashed line that crosses pixel boundaries, and cannot be defined in a crisp manner. The lower and upper approximations are drawn as thick gridlines.

In the rough set theory, the set of attributes comprises condition attributes and decision attributes. In large data sets, some attributes may be redundant and hence may be eliminated without losing essential classificatory information. The reduct and core are two additional fundamental rough set concepts needed for the elimination of attributes. Reduct is the minimal subset of the attributes that can provide the same object classification as with the full set of attributes. The intersection of all reducts is called the core. Accordingly, the core becomes the class of all indispensable attributes. There are some well-known technical advantages that the rough sets approach has. Indeed, it can deal with a

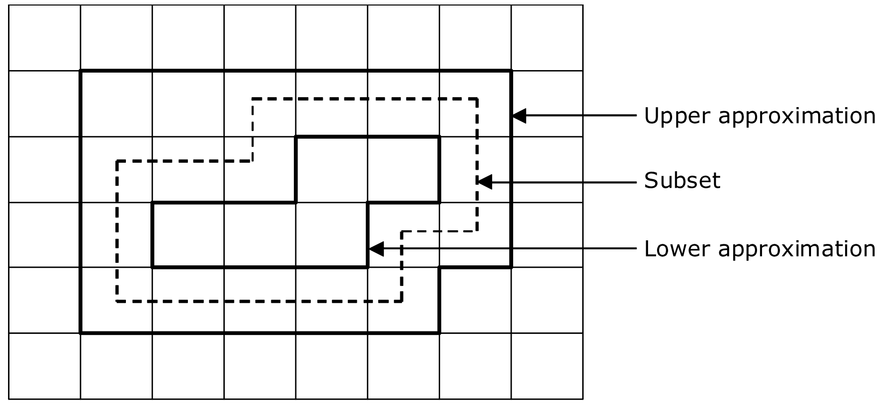


Fig. 1 Lower and upper approximations of sets [19]

set of inconsistent examples, i.e., objects indiscernible by condition attributes but discernible by decision attributes. More importantly, it not only provides useful information for the role of particular attributes and their subsets in the approximation of decision classes but also prepares the foundation for generation of decision rules involving relevant attributes.

Using rough sets, we generate the decision rule often presented in an 'IF condition(s) THEN decision(s)' format. The decision rule reflects a relationship of a set of condition attributes with a set of decision attributes. Thus the generation of decision rules means combining the reducts with the values of the data. Note that the generated decision rules play a critical role for summarizing information. More detailed discussion about the process of the rough sets theory can be referred to Slowinski [25], Dimitras [10], and Jackson et al. [12].

3 Construction procedure of RRTS

In this section a detailed description of construction procedure of RRTS is given. In order to handle real time data for RRTS, suppose that stock market data are collected at certain frequency $\theta > 0$, i.e.,

$$\{x_{t\theta} \in R^d : t = 1, 2, \dots\} \tag{1}$$

where $x_{t\theta}$ is a d-dimensional vector observed at time $t\theta$. For example, $d = 5$ when $x_{t\theta}$ consists of open, high, low, close price and trading volume during time interval $(\theta(t-1), \theta t]$. Without loss of generality it is assumed that $\theta = 1$ below. From here and subsequently, the return rates mean the *yearly profit rates* which are calculated by the ratio of the current capital value to the initial capital value after trading of one year. Also the yearly profit is defined as the yearly gross profit minus transaction costs and slippages (refer to Sect. 4) where the yearly gross profit is the yearly short position minus the yearly long position. Before going to the main issue, note that Fig. 2 illustrates the architecture of RRTS succinctly.

Phase 1: Case-wise input data generation

At this phase we generate input data for each of $i = 1, \dots, I$ cases (or different stock market conditions). Here we assume that the I cases are defined by distinctive behavior of data over a certain period. In practice these I cases might represent various classes of situations, depending on the data available and the opinion of the stock market experts. For example, in our empirical study we use $I = 6$ cases and characterize each of them by its own distinctive trend, i.e., short-term ascending trend (*SAT*), short-term descending trend (*SDT*), long-term ascending trend (*LAT*), long-term descending trend (*LDT*), flat top (*FT*), and flat bottom (*FB*) (see Table 1 for its detailed description).

Generation of input data is done by independently obtaining input data for each case (or the ith case from $i = 1, 2, \dots, I$). Let us start with finding the training data for the ith case from the past, i.e.,

$$\mathbf{x}_i = \{x_{t_i+s} \in R^d : s = 0, \dots, n_i - 1\} \tag{2}$$

where $\{t_i + s : s = 0, 1, \dots, n_i - 1\}$ belongs to the past and n_i denotes the number of data in $\mathbf{x}_i$. Next, application of J different technical indicators (say, $\tau_1, \dots, \tau_J$) are made to given $\mathbf{x}_i$ to produce $n_i \times J$ matrix

$$\mathbf{T}_i^{(0)} = \{T_{i1}, \dots, T_{iJ}\} \tag{3}$$

where $T_{ij} = \{\tau_j(x_{t_i}), \tau_j(x_{t_i+1}), \dots, \tau_j(x_{t_i+n_i-1})\}^T$ and $\tau_j : R^d \to R$ (see, e.g., Appendix). Here note that T_{ij} is a $n_i \times 1$ column vector obtained from application of jth technical indicator to $\mathbf{x}_i$. Now application of "the trading rule designed for τ_j" to $\mathbf{x}_i$ (or to T_{ij} equivalently) for $j = 1, \dots, J$ yields J return rates (refer to Appendix for the specific trading rules designed for each technical indicator). Then one can pick out best $J_0 (\le J)$ ones among the J technical indicators in terms of the calculated return rates, say $\tau_{(1)|i}, \dots, \tau_{(J_0)|i}$. Here $\tau_{(j)|i}$ denotes the jth best technical indicator in terms of profitability when J technical indicators are applied to

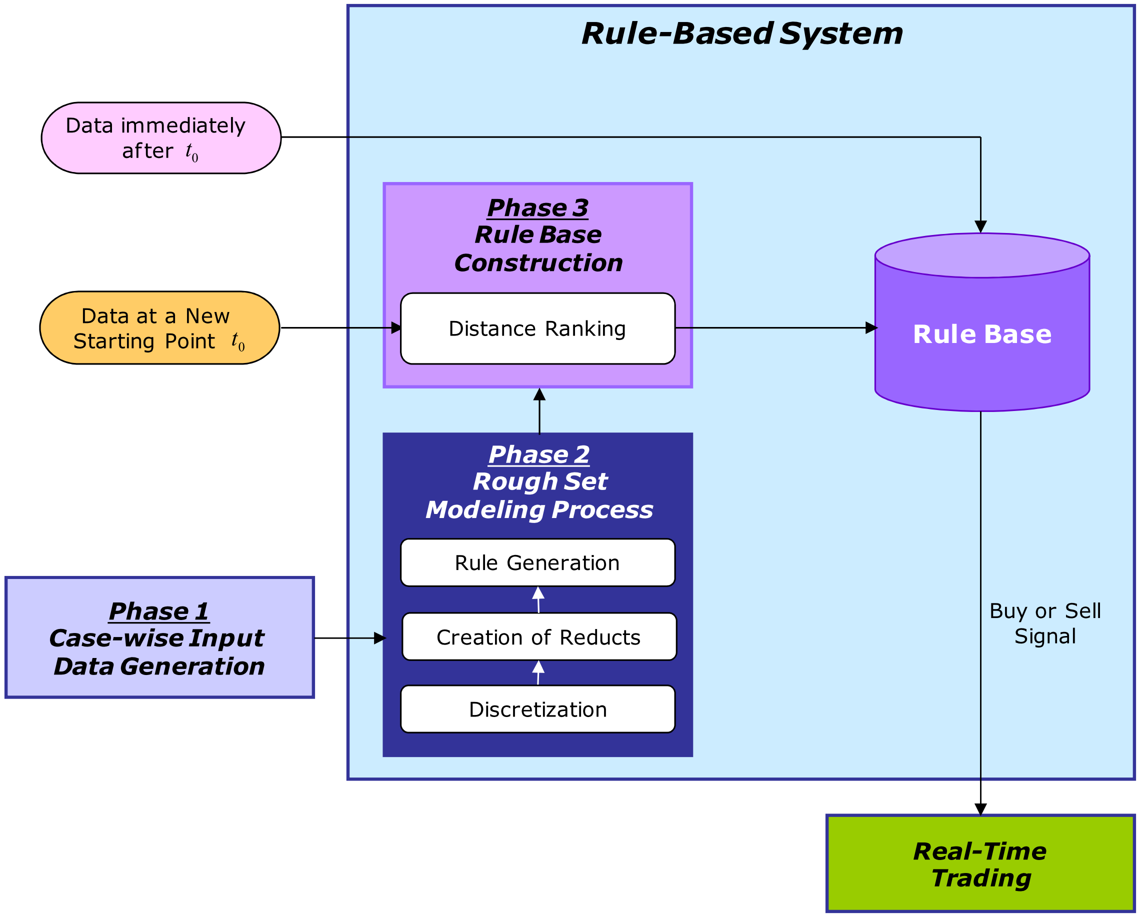


Fig. 2 Architecture of RRTS

the given training data $\mathbf{x}_i$. Now we have input data for the ith case as follows:

$$\mathbf{T}_i = \{T_{i,(j)|i} : j = 1, \dots, J_0\} \tag{4}$$

where $T_{i,(j)|i} = \{\tau_{(j)|i}(x_{t_i}), \dots, \tau_{(j)|i}(x_{t_i+n_i-1})\}^T$ and hence $\mathbf{T}_i$ is $n_i \times J_0$ matrix. Repeating the above procedure for $i = 1, \dots, I$ completes generation of input data $\mathbf{T}_1, \dots, \mathbf{T}_I$.

Phase 2: Rough set modeling process

This phase consisting of three steps is mainly concerned about rough set modeling process applied to $\mathbf{T}_1, \dots, \mathbf{T}_I$. As with the Phase 1, we consider each $\mathbf{T}_i$ independently for completing the process. In the first step, data cleaning with data exploration is conducted. This includes removing obvious outliers and data completion (replacing or deleting blanks). To improve the overall quality of the discovered information, further data transformation is usually conducted by means of *discretization*, which basically corresponds to making the set of attributes smaller. At this moment one may consider several discretization tools such as discernibility preservation, entropy minimization, equal frequency binning and various naive methods. Note that after application of this first step (say, application of transformation Δ) we have the transformed input data set, say, $n_i \times J_0$ matrix $\Delta\mathbf{T}_i$.

The second step is *creation of reducts* which is the computation of the reducts. From the given $n_i \times J_0$ matrix $\Delta\mathbf{T}_i$, $k(\le J_0)$ columns (or the corresponding k technical indicators $\tau_{(i_1(m))|i}, \dots, \tau_{(i_k(m))|i}$) are randomly selected among J_0 columns (or J_0 technical indicators, $\tau_{(1)|i}, \dots, \tau_{(J_0)|i}$) in order to create the corresponding reduct R_m. Here it is assumed that $1 \le i_1(m), \dots, i_k(m) \le J_0$ and $m = 1, \dots, k_1$, which indicates that k_1 reducts out of possible $\frac{J_0!}{k!(J_0-k)!}$ reducts (say, $R_1, \dots, R_{k_1}$) are created at this step. Various methods are available for creation of reduct, e.g., genetic algorithms, manual reducer, dynamic reducts, approximate hitting set approaches, etc. Also note that the generated reducts could be filtered according to some criteria such as coverage and accuracy, attribute cost, advanced quality measures or classificatory performance on external holdout data sets [19].

The final step is *rule generation* for the given case. Based on the k_1 reducts made in Step 2, patterns could be generated in the form of 'IF–THEN' production rules by combining the condition values with the decision values. In general, the connection of condition values (or input variables) and decision values (or output variables) is based on conjunction. An exemplary form of the generated pattern could be for fixed m

$$\begin{aligned}
&\textbf{\textit{IF}}\ \textit{value of technical indicator}\ \tau_{(i_1(m))|i}\\
&\quad \textit{belongs to}\ A_{i_1(m)|i}\\
&\quad \textit{AND value of technical indicator}\ \tau_{(i_2(m))|i}\\
&\quad \textit{belongs to}\ A_{i_2(m)|i} \qquad (5)\\
&\quad \textit{AND}\ldots\textit{AND value of technical indicator}\ \tau_{(i_k(m))|i}\\
&\quad \textit{belongs to}\ A_{i_k(m)|i},\\
&\textbf{\textit{THEN}}\ \textit{BUY (or SELL).}
\end{aligned}$$

Here the sets $A_{i_1(m)},\ldots,A_{i_k(m)}$ (or the range of the corresponding technical indicator) are established from the previous steps. Thus, at this moment, we establish group of trading rules for the ith case, i.e.,

$$\begin{aligned}
G_i = \{&\text{decision rules generated by}\\
&\tau_{(i_1(m))|i},\ldots,\tau_{(i_k(m))|i}\ \text{for}\ m=1,\ldots,k_1\}.
\end{aligned} \tag{6}$$

For practical application of the decision rules created above, one may consider factor(s) such as successive application of the rules or the number of position to hold. For example, one may employ the following implementation rule, say p, to limit the number of position to hold to one.

$$\begin{aligned}
&\textbf{\textit{IF}}\ \textit{signal at time}\ t\ \textit{is BUY And}\\
&\textbf{\textit{IF}}\ \textit{signal at time}\ t-1\ \textit{is BUY}\\
&\textbf{\textit{THEN}}\ \textit{HOLD}\ \textbf{\textit{ELSE}}\ \textit{SELL,}\\
&\textbf{\textit{IF}}\ \textit{signal at time}\ t\ \textit{is SELL And}\\
&\textbf{\textit{IF}}\ \textit{signal at time}\ t-1\ \textit{is SELL}\\
&\textbf{\textit{THEN}}\ \textit{HOLD}\ \textbf{\textit{ELSE}}\ \textit{BUY.}
\end{aligned} \tag{7}$$

Note that this type of rule could be attached to G_i and let G_{ip} denote the G_i with the implementation rule p. Now repeating this procedure for $i=1,\ldots,I$, we build I groups of trading rules, $G_{1p},\ldots,G_{Ip}$.

Phase 3: Rule base construction

At this phase we form the rule base for real-time trading after given (or current) time t_0. In fact, a rule base is formed from utilizing $I_0(\le I)G_{ip}$'s among $G_{1p},\ldots,G_{Ip}$. Selection of I_0G_{ip}'s is made as follows. Let

$$\mathbf{x}(t_0,l) = \{(x_{t_0-l+1},\ldots,x_{t_0})|x_t\in R^d\} \tag{8}$$

be a recent l data segment at given time t_0. Here we assume that $l\le n_i$ for all $i=1,\ldots,I$. Then, for each i, build $l\times J_0$ matrix $\mathbf{T}_i(t_0,l)$, i.e.,

$$\mathbf{T}_i(t_0,l) = \{T_{i,(j)|i}(t_0,l) : j=1,\ldots,J_0\} \tag{9}$$

where $T_{i,(j)|i}(t_0,l)=\{\tau_{(j)|i}(x_{t_0-l+1}),\ldots,\tau_{(j)|i}(x_{t_0})\}^T$ is $l\times 1$ vector. Recalling that $\tau_{(1)|i},\ldots,\tau_{(J_0)|i}$ are the best J_0 technical indicators obtained for training data for the ith case (see Phase 1), one may easily see that $T_{i,(j)|i}(t_0,l)$ is obtained simply by applying the J_0 technical indicators to $\mathbf{x}(t_0,l)$. In a similar fashion, for each i, one may build another $l\times J_0$ matrix $\mathbf{T}_i(l)$ by applying the J_0 technical indicators to $\mathbf{x}_i(l)=\{x_{t_i},\ldots,x_{t_i+l-1}\}$, an initial l data segment of $\mathbf{x}_i$. Indeed

$$\mathbf{T}_i(l) = \{T_{i,(j)|i}(l) : j=1,\ldots,J_0\} \tag{10}$$

where $T_{i,(j)|i}(l)=\{\tau_{(j)|i}(x_{t_i}),\ldots,\tau_{(j)|i}(x_{t_i+l-1})\}^T$. Define a distance

$$\begin{aligned}
d_i &= \|\mathbf{T}_i(t_0,l)-\mathbf{T}_i(l)\|\\
&= \sum_{j=1}^{J_0}\left|T_{i,(j)|i}(t_0,l)-T_{i,(j)|i}(l)\right|
\end{aligned} \tag{11}$$

where $|\cdot|$ is the Euclidean distance in R^l. Calculating d_i for $i=1,\ldots,I$, one may obtain best $I_0(\le I)$ cases in minimizing the distance and then choose $G_{i_{(1)}p},\ldots,G_{i_{(I_0)}p}$ accordingly. Finally, we have the rule base

$$RB(t_0,I_0) = \{G_{i_{(1)}p},\ldots,G_{i_{(I_0)}p}\} \tag{12}$$

which will be used for real-time trading after given (or current) time t_0. At the end of Sect. 4 discusses about selection of I_0.

4 Empirical study

As real-time data become available on a large scale, the Korean futures market has become more volatile and mercurial. Thus, traders need more powerful technical supporters in their investment decision since human capability in analyzing the huge real-time data is limited. This empirical study for constructing RRTS is done by taking the Korea stock price index 200 (KOSPI 200) as the underlying asset (or base index) in the Korean futures market. The underlying asset is the asset for which the price of derivative is derived. For an empirical example of RRTS construction for the derivative, we consider the period from July 1996 to December 2006 and divide it into training period (July 1996 to June 2000) and testing period (July 2000 to December 2006). Figure 3 depicts the stream of KOSPI 200 for the entire period. Throughout this section, we use $I=6$

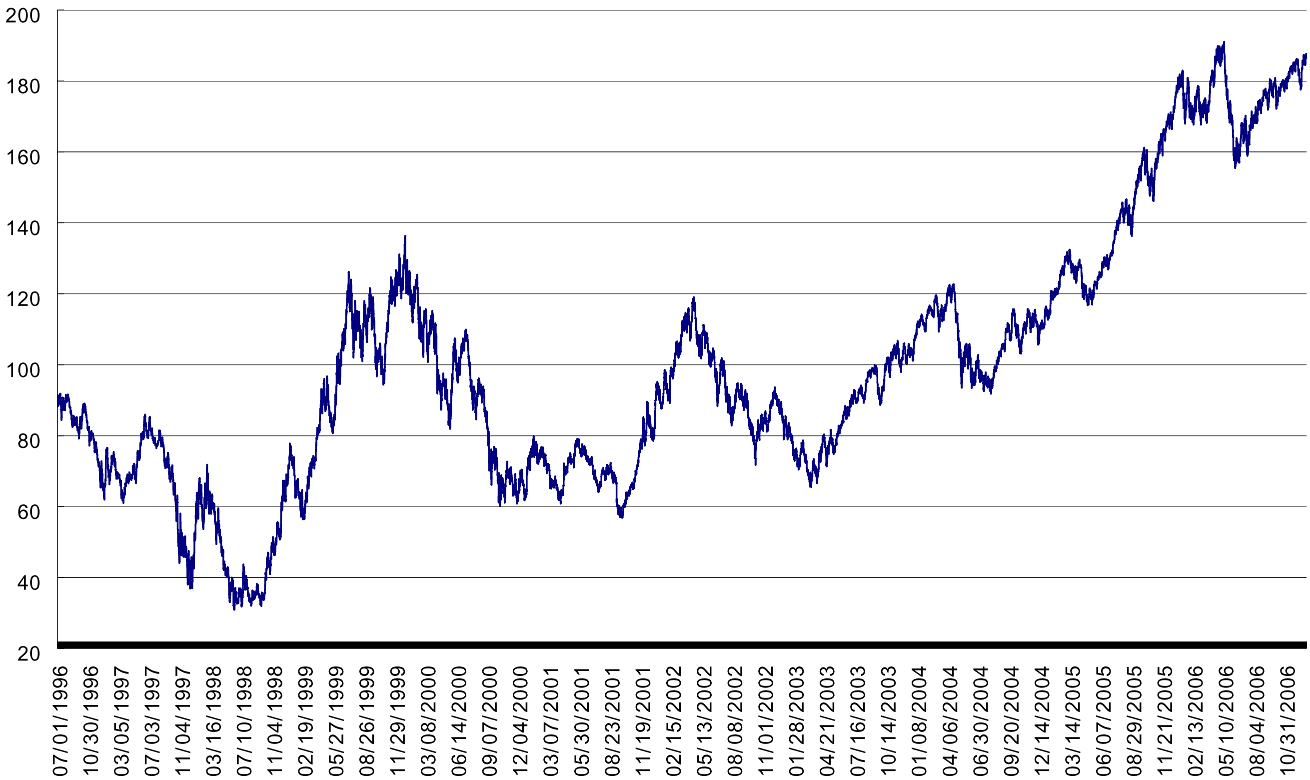


Fig. 3 Overall flow of KOSPI 200 from July 1996 to December 2006 (mm/dd/yyyy)

Table 1 Case description

Trend (case)		Description
SAT	Short-term ascending trend	Daily close prices rise continuously for two to three weeks
SDT	Short-term descending trend	Daily close prices fall continuously for two to three weeks
LAT	Long-term ascending trend	Continuous rise of daily close prices for over 6 months
LDT	Long-term descending trend	Continuous fall of daily close prices for over 6 months
FT	Flat top	After an ascending trend, daily close prices remain unmoved
FB	Flat bottom	After a descending trend, daily close prices remain unmoved

Table 2 Average return rates (%) at various frequencies

Frequency	Return rates[a]	Technical indicators employed
10-minute	7.16	CO, DMI, TRIX, William's %R, PO
30-minute	186.38	CO, NCO, ROC, SMI, Momentum
60-minute	133.53	SMI, NCO, ROC, CO, SMA
Daily	28.57	MI, VROC, DMI, Band %b, PO

[a]Return rates (%) are averaged over the five best technical indicators when 24 technical indicators are applied to real time data during July 1996 to December 2006

cases as mentioned earlier, i.e., short-term ascending trend ($SAT, i = 1$), short-term descending trend ($SDT, i = 2$), long-term ascending trend ($LAT, i = 3$), long-term descending trend ($LDT, i = 4$), flat top ($FT, i = 5$), and flat bottom ($FB, i = 6$) (see Table 1 for its detailed description).

For real-time data, "10, 30, 60-minute and daily" time interval (or frequency) data are considered, i.e., to each frequency the corresponding data set composed of open price, high price, low price, close price, and (trading) volume are considered. As a default condition for system trading, the initial capital is set to ₩1,000,000 (equal to $1,000), interest rates to 5.00%, transaction cost to ₩10,000, and slippage to ₩25,000 where ₩ is the Korean won and slippage cost is an additional amount set for preventing missed trading. For evaluation of trading system, return rates are calculated for the underlying asset. The set of these conditions are applied throughout this empirical study.

We start with selection of an appropriate frequency for real time data for RRTS (or selection of θ at (1)). Examining Table 2 providing average return rates at different frequencies, one may see that 10-minute and daily intervals yield

Table 3 Training and testing periods obtained from 1996.7–2006.6 for each of six distinctive trend periods

Trend	Training period		Testing period	
	Starting date	Ending date	Starting date	Ending date
SAT	Dec. 26, 1997	Mar. 05, 1998	Dec. 26, 2000	Jan. 22, 2001
SDT	Mar. 06, 1998	Jun. 15, 1998	Jul. 14, 2000	Sep. 22, 2000
LAT	Oct. 01, 1998	Jul. 09, 1999	Apr. 02, 2003	Apr. 23, 2004
LDT	Sep. 09, 1996	Dec. 24, 1997	Apr. 24, 2002	Apr. 01, 2003
FT	Jul. 12, 1999	Feb. 03, 2000	Apr. 28, 2004	Dec. 30, 2004
FB	Jun. 16, 1998	Sep. 30, 1998	Sep. 25, 2000	Dec. 22, 2000

relatively lower return rates. It appears that the lower rates of 10-minute are due to frequent trading while the lower rates of daily interval are due to the daily gap (or difference of today's open price and the previous day's close price). Since 30-minute frequency produces higher average return rates (186.38%) than 60-minute frequency, the 30-minute is used as an appropriate frequency θ for real-time trading throughout this study. At this point it should be mentioned that data with θ less than 10-minute are not taken into account because too frequent trading based on them clearly incurs loss via huge transaction fee.

For construction and evaluation of the rule-based trading system, two periods for each $i = 1, \dots, 6$ are obtained from the training and testing period respectively by the help of the futures market experts. Indeed Table 3 provides the training and the testing periods for each of the six trends (or cases). Note that the six periods from the training period yield $\mathbf{x}_i$ with its own starting point t_i and cardinality n_i for $i = 1, \dots, 6$. Since there are 13 30-minute interval data available for each day, we have $n_1 = 546, n_2 = 897, n_3 = 2431, n_4 = 4043, n_5 = 1820, n_6 = 949$. Moreover, training data for *SAT* ($i = 1$) is $\mathbf{x}_1 = \{x_{t_1}, x_{t_1+1}, \dots, x_{t_1+545}\}$ where $x_t \in R^5$ consists of open, high, low, close price and trading volume during time interval $(t-1, t]$, for example.

At Phase 1 case-wise input data are generated. Application of $J = 24$ technical indicators (i.e., application of $\tau_1, \dots, \tau_{24}$ of Appendix with its own trading rule) to each of $\mathbf{x}_1, \dots, \mathbf{x}_6$ allows us to choose $J_0 = 5$ technical indicators from $\tau_1, \dots, \tau_{24}$ which are the best five to each of $\mathbf{x}_1, \dots, \mathbf{x}_6$ in terms of their return rates. In the process of selecting five best technical indicators to each of $\mathbf{x}_1, \dots, \mathbf{x}_6$, top 5% and bottom 5% indicators are intentionally excluded because of possible strong dependence of the indicators on a given specific training period [22]. Note that $J_0 = 5$ is chosen from our trial and error experiments and the return rates of each indicator are routinely calculated by a system trading tool, *Tradestation 2000i*. Table 4 lists the five most proper indicators to each of $\mathbf{x}_1, \dots, \mathbf{x}_6$. Thus we have, for instance, $\tau_{(1)|1} = RSI, \tau_{(2)|1} = WMA, \tau_{(3)|1} = SAR, \tau_{(4)|1} = CCI, \tau_{(5)|1} = SMA$ for $i = 1$. As a result we have obtained the input data to each trend, $\mathbf{T}_1, \dots, \mathbf{T}_6$ in (4). For example, $\mathbf{T}_1 = \{T_{1,(1)|1}, T_{1,(2)|1}, \dots, T_{1,(5)|1}\}$ where $T_{1,(1)|1} = \{RSI(x_{t_1}), \dots, RSI(x_{t_1+545})\}^T$ and so on.

Table 4 Five most profitable indicators to each trend and their return rates (%)

Indicators	Return rates[a]	Indicators	Return rates[a]
(a) *SAT*		(d) *LDT*	
RSI	623.24	*SMI*	863.78
WMA	479.86	*WMA*	630.24
Parabolic *SAR*	271.98	*EMA*	500.00
CCI	155.05	*MACD Osc*	480.94
SMA	134.18	*CCI*	394.89
(b) *SDT*		(e) *FT*	
NCO	536.95	*SMI*	333.78
ROC	536.95	*RSI*	161.83
Momentum	458.84	*ROC*	146.23
Parabolic *SAR*	210.74	*NCO*	146.23
DMI	152.04	*Momentum*	56.98
(c) *LAT*		(f) *FB*	
Momentum	570.66	*RSI*	242.73
MACD Osc	258.33	*SMI*	135.42
ROC	254.86	*Stochastic*	95.91
NCO	254.86	*ROC*	17.88
SMI	249.28	*NCO*	17.88

[a]Return rates are calculated for the training period of each trend

With input data $\mathbf{T}_1, \dots, \mathbf{T}_6$, we generate a set of decision rules to each trend by rough set modeling process (Phase 2). For this, $k = 3 (\le J_0 = 5)$ technical indicators were randomly selected to produce $k_1 = 3 (\le \frac{5!}{3!2!} = 10)$ reducts via manual reducer for each trend. This in turn generates a set of rules proper to each trend as in the form of (5), and hence $\{G_1, \dots, G_6\}$ given by (6) was produced. Here $\{\tau_{(i_1(m))|1}, \tau_{(i_2(m))|1}, \tau_{(i_3(m))|1}\}$ with $m = 1$ at (5) is $\{RSI, WMA, SMA\}$ (refer to Fig. 4(b)), for instance. Moreover, the rule mentioned in (7) is installed across G_i's, which yields $\{G_{1p}, \dots, G_{6p}\}$. For implementing the Phase 2, *ROSETTA*-software was employed and Fig. 4 illustrates

Sheet1$, discretized

	RSI	Parabolic	WMA	SMA	CCI	Actual
1	[47,75, 58,29)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[-66,70, 42,67)	1
2	[47,75, 58,29)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	−1
3	[*, 47,75)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[-66,70, 42,67)	1
4	[*, 47,75)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[-66,70, 42,67)	1
5	[*, 47,75)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[-66,70, 42,67)	−1
6	[*, 47,75)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[-66,70, 42,67)	−1
7	[*, 47,75)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[*, −66,70)	−1
8	[*, 47,75)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[*, −66,70)	1
9	[*, 47,75)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[*, −66,70)	1
10	[*, 47,75)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[-66,70, 42,67)	1
11	[*, 47,75)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[*, −66,70)	1
12	[47,75, 58,29)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[104,08, *)	−1
13	[47,75, 58,29)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	1
14	[47,75, 58,29)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	1
15	[47,75, 58,29)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[104,08, *)	1
16	[58,29, 66,59)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[104,08, *)	1
17	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[104,08, *)	−1
18	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[104,08, *)	−1
19	[58,29, 66,59)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	−1
20	[58,29, 66,59)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	1
21	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	1
22	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	−1
23	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	1
24	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[104,08, *)	−1
25	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[104,08, *)	1
26	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[104,08, *)	1
27	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[104,08, *)	−1
28	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	0
29	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	0
30	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	1
31	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	−1
32	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[-66,70, 42,67)	1
33	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[42,67, 104,08)	−1
34	[58,29, 66,59)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[-66,70, 42,67)	1
35	[58,29, 66,59)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[*, −66,70)	1
36	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[104,08, *)	−1
37	[66,59, *)	[*, 115,09)	[*, 115,72)	[*, 115,42)	[104,08, *)	−1

(a) Discretization

1

	Reduct	Support	Length
1	{RSI, WMA, SMA}	1	3

2

	Reduct	Support	Length
1	{RSI, Parabolic, CCI}	1	3

3

	Reduct	Support	Length
1	{RSI, SMA, CCI}	1	3

(b) Creation of reducts ($k = 3, k_1 = 3$)

No name

	Rule	LHS Support	RHS Support	RHS Accuracy	LHS Coverage
1	RSI([47,75, 58,29)) AND WMA([*, 115,72)) AND SMA([*, 115,42)) => Actual(−1) OR Actual(1)	25	10, 15	0,4, 0,6	0,047619
2	RSI([*, 47,75)) AND WMA([*, 115,72)) AND SMA([*, 115,42)) => Actual(1) OR Actual(−1)	11	7, 4	0,636364, 0,363636	0,020952
3	RSI([58,29, 66,59)) AND WMA([*, 115,72)) AND SMA([*, 115,42)) => Actual(1) OR Actual(−1) OR Actual(0)	37	26, 8, 3	0,702703, 0,216216, 0,081081	0,070476
4	RSI([66,59, *)) AND WMA([*, 115,72)) AND SMA([*, 115,42)) => Actual(−1) OR Actual(1) OR Actual(0)	58	26, 28, 4	0,448276, 0,482759, 0,068966	0,110476
5	RSI([66,59, *)) AND WMA([115,72, 120,79)) AND SMA([115,42, 120,78)) => Actual(1) OR Actual(−1)	14	9, 5	0,642857, 0,357143	0,026667
6	RSI([58,29, 66,59)) AND WMA([115,72, 120,79)) AND SMA([115,42, 120,78)) => Actual(−1) OR Actual(1) OR Actual(0)	18	8, 8, 2	0,444444, 0,444444, 0,111111	0,034286
7	RSI([47,75, 58,29)) AND WMA([115,72, 120,79)) AND SMA([115,42, 120,78)) => Actual(1) OR Actual(−1) OR Actual(0)	36	14, 18, 4	0,388889, 0,5, 0,111111	0,068571
8	RSI([66,59, *)) AND WMA([120,79, 124,16)) AND SMA([115,42, 120,78)) => Actual(1) OR Actual(−1)	3	2, 1	0,666667, 0,333333	0,005714
9	RSI([66,59, *)) AND WMA([120,79, 124,16)) AND SMA([120,78, 124,15)) => Actual(−1) OR Actual(1)	11	7, 4	0,636364, 0,363636	0,020952
10	RSI([58,29, 66,59)) AND WMA([120,79, 124,16)) AND SMA([120,78, 124,15)) => Actual(1) OR Actual(−1)	27	9, 18	0,333333, 0,666667	0,051429
11	RSI([47,75, 58,29)) AND WMA([120,79, 124,16)) AND SMA([120,78, 124,15)) => Actual(0) OR Actual(1) OR Actual(−1)	42	2, 18, 22	0,047619, 0,428571, 0,52381	0,08
12	RSI([*, 47,75)) AND WMA([120,79, 124,16)) AND SMA([120,78, 124,15)) => Actual(−1) OR Actual(1) OR Actual(0)	38	20, 17, 1	0,526316, 0,447368, 0,026316	0,072381
13	RSI([*, 47,75)) AND WMA([115,72, 120,79)) AND SMA([120,78, 124,15)) => Actual(1) OR Actual(−1)	7	3, 4	0,428571, 0,571429	0,013333
14	RSI([*, 47,75)) AND WMA([115,72, 120,79)) AND SMA([115,42, 120,78)) => Actual(1) OR Actual(−1) OR Actual(0)	55	33, 18, 4	0,6, 0,327273, 0,072727	0,104762
15	RSI([58,29, 66,59)) AND WMA([120,79, 124,16)) AND SMA([115,42, 120,78)) => Actual(1) OR Actual(−1)	4	2, 2	0,5, 0,5	0,007619
16	RSI([66,59, *)) AND WMA([124,16, *)) AND SMA([124,15, *)) => Actual(−1) OR Actual(1) OR Actual(0)	45	22, 20, 3	0,488889, 0,444444, 0,066667	0,085714
17	RSI([58,29, 66,59)) AND WMA([124,16, *)) AND SMA([124,15, *)) => Actual(1) OR Actual(0) OR Actual(−1)	42	21, 3, 18	0,5, 0,071429, 0,428571	0,08
18	RSI([47,75, 58,29)) AND WMA([124,16, *)) AND SMA([124,15, *)) => Actual(1) OR Actual(−1)	23	8, 15	0,347826, 0,652174	0,04381
19	RSI([58,29, 66,59)) AND WMA([120,79, 124,16)) AND SMA([124,15, *)) => Actual(1)	1	1	1,0	0,001905
20	RSI([*, 47,75)) AND WMA([124,16, *)) AND SMA([124,15, *)) => Actual(−1) OR Actual(1) OR Actual(0)	16	12, 2, 2	0,75, 0,125, 0,125	0,030476
21	RSI([*, 47,75)) AND WMA([120,79, 124,16)) AND SMA([124,15, *)) => Actual(−1) OR Actual(1)	4	3, 1	0,75, 0,25	0,007619
22	RSI([47,75, 58,29)) AND WMA([115,72, 120,79)) AND SMA([120,78, 124,15)) => Actual(1)	1	1	1,0	0,001905
23	RSI([58,29, 66,59)) AND WMA([124,16, *)) AND SMA([120,78, 124,15)) => Actual(1)	3	3	1,0	0,005714
24	RSI([47,75, 58,29)) AND WMA([124,16, *)) AND SMA([120,78, 124,15)) => Actual(−1) OR Actual(1)	2	1, 1	0,5, 0,5	0,00381
25	RSI([47,75, 58,29)) AND WMA([120,79, 124,16)) AND SMA([124,15, *)) => Actual(−1)	1	1	1,0	0,001905
26	RSI([47,75, 58,29)) AND WMA([120,79, 124,16)) AND SMA([115,42, 120,78)) => Actual(1)	1	1	1,0	0,001905

(c) Rule generation

Fig. 4 An illustration of discretization, creation of reducts, and rule generation for trend SAT using *ROSETTA*-software

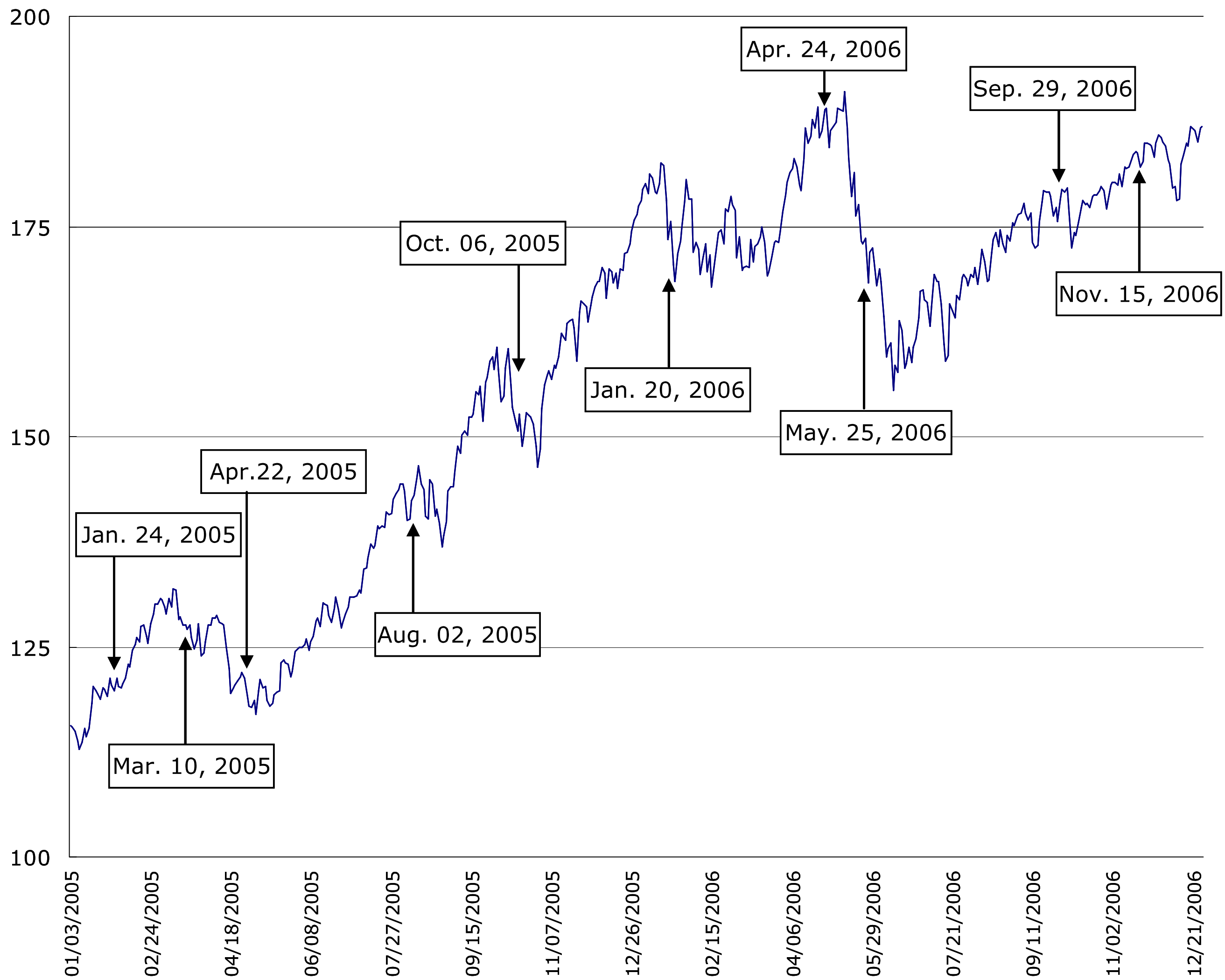


Fig. 5 Ten starting points on KOSPI 200 from January 2005 to December 2006 (mm/dd/yyyy)

the three steps of Phase 2. Note here that $k = 3$ and $k_1 = 3$ are chosen from our trial and error experiments.

For implementation and experimentation of Phase 3, ten starting points (or ten t_0's) are randomly selected from the testing period during Jan. 3, 2005 to Dec. 28, 2006 as shown in Fig. 5. Then, given t_0, we consider recent 20 days (and hence $l = 20 \times 13 = 260$) to construct $\mathbf{x}(t_0, l = 260) = \{x_{t_0-259}, \ldots, x_{t_0}\}$. Recalling $\mathbf{x}_i(260) = \{x_{t_i}, \ldots, x_{t_i+259}\}$, $\mathbf{T}_i(t_0, l = 260)$ in (9) and $\mathbf{T}_i(260)$ in (10) are easily constructed by respective application of the J_0 technical indicators selected for each i at Phase 2 to $\mathbf{x}(t_0, 260)$ and to $\mathbf{x}(260)$. For example, the first column of $\mathbf{T}_1(t_0, l = 260)$ is $T_{1,(1)|1}(t_0, 260) = \{RSI(x_{t_0-259}), \ldots, RSI(x_{t_0})\}$ and the first column of $\mathbf{T}_1(260)$ is $T_{1,(1)|1}(260) = \{RSI(x_{t_1}), \cdots, RSI(x_{t_1+259})\}$. With $\mathbf{T}_i(t_0, 260)$ and $\mathbf{T}_i(260)$ at hand, Table 5 calculates distance d_i in (11) for each i. Now, using Table 5, we have generated $RB(t_0, I_0)$ of (10) to each of 10 t_0's for various choices of $I_0 (\leq 6)$. The average return rates of $RB(t_0, I_0)$ for various combinations of t_0 and I_0 are given in Table 6. For example, the return rates of $RB(t_0, I_0)$ with $t_0 =$ Jan. 24, 2005 and $I_0 = 3$ is an average over three return rates *SAT* $(i = 1)$, *LAT* $(i = 3)$ and *SDT* $(i = 2)$ having 3 smallest distances for $t_0 =$ Jan. 24, 2005 as calculated from Table 5.

As seen in Table 6, selection of I_0 is critical for RRTS. For example, Table 6 shows that as I_0 increases the "absolute" value of average return rates tends to decrease to zero. Theoretically this is so because a small value of I_0 increases dependence of RRTS on the particular case(s) among the I cases while a large value of I_0 yields a robust RRTS (robust in the sense that it performs reasonably against various situations). For practical selection of I_0, the estimated Sharpe ratio is employed for evaluating performance of *RB* as a rule portfolio where the Sharpe ratio is defined as the ratio of the expected difference between return rates of a given portfolio and the risk-free asset over

Table 5 Results of calculation of d_i $(i = 1, \ldots, 6)$ for each starting point

t_0	d_i's in the increasing order from left to right					
Jan. 24, 2005	SAT (894.09)	LAT (897.38)	SDT (897.60)	FB (897.63)	FT (897.91)	LDT (898.85)
Mar. 10, 2005	SDT (931.81)	FB (932.53)	SAT (933.04)	LAT (933.20)	FT (934.01)	LDT (934.77)
Apr.22, 2005	FB (867.59)	LAT (869.63)	SAT (870.62)	SDT (870.69)	LDT (871.18)	FT (873.73)
Aug. 02, 2005	SDT (930.85)	FB (931.11)	SAT (932.44)	FT (933.10)	LAT (933.22)	LDT (933.50)
Oct. 06, 2005	SDT (941.96)	SAT (942.47)	FB (943.07)	LDT (943.25)	FT (944.29)	LAT (945.41)
Jan. 20, 2006	SAT (977.11)	FB (978.22)	LAT (978.67)	FT (981.89)	LDT (992.72)	SDT (994.45)
Apr. 24, 2006	SAT (940.83)	SDT (944.66)	LAT (945.40)	FB (945.54)	LDT (945.70)	FT (948.16)
May. 25, 2006	SAT (940.83)	SDT (944.66)	LAT (945.40)	FB (945.54)	LDT (945.70)	FT (948.16)
Sep. 29, 2006	SDT (916.35)	LAT (916.60)	SAT (916.93)	FB (918.95)	LDT (919.14)	FT (919.49)
Nov. 15, 2006	SDT (935.18)	SAT (936.50)	FB (937.20)	LDT (937.52)	LAT (938.67)	FT (938.94)

Table 6 Average return rates (%) of $RB(t_0, I_0)$ for various combinations of I_0 and t_0

t_0	I_0					
	1	2	3	4	5	6
Jan. 24, 2005	8.01	5.67	4.32	2.81	1.64	−0.36
Mar. 10, 2005	7.56	6.25	21.54	−5.36	−2.86	−0.96
Apr. 22, 2005	6.32	13.62	11.77	0.84	−5.43	−10.37
Aug. 02, 2005	15.07	9.35	11.09	19.13	−2.17	6.33
Oct. 06, 2005	12.50	12.87	14.00	15.99	1.23	2.92
Jan. 20, 2006	15.34	15.40	16.84	9.65	0.67	−3.13
Apr. 24, 2006	7.26	5.95	7.97	−0.29	−6.47	−0.40
May. 25, 2006	19.04	17.86	21.52	−5.61	−7.05	−2.31
Sep. 29, 2006	6.09	16.06	7.08	0.02	1.24	−4.09
Nov. 15, 2006	5.11	6.38	10.84	−4.20	5.12	−1.53

I_0 is the number of groups used for $RB(t_0, I_0)$ given by (12). Average return rates in each cell are averaged over I_0 return rates for given t_0

Table 7 Average return rates (%) and Sharpe ratio of $RB(\cdot, I_0)$

Measure of portfolio	I_0					
	1	2	3	4	5	6
Average return rates	10.23	10.94	12.70	3.30	−1.41	−1.39
Sharpe ratio	1.23	1.42	1.47	−0.15	−1.56	−1.43

Average return rates in each cell are averaged over t_0 return rates

the standard deviation of the difference [23]. Table 7 compares the average return rates of $RB(t_0, I_0)$ over 10 t_0's with its Sharpe ratios for $I_0 = 1, \ldots, 6$. Here, the return rates of risk-free asset used for Sharpe ratio calculation are those of Treasury bills with 3 years' maturity (their average return rates from 2005 to 2006 are 4.55%). From Table 7, one may see that both Sharpe ratio and average return rates increase as I_0 increases from 1 to 3 and then plunge dramatically for $I_0 = 4, 5, 6$. Consequently, it appears that $I_0 = 3$ is optimal choice for $RB(t_0, I_0)$ across t_0. Note that $RB(\cdot, 3)$ registered the average return rates 12.70%, which is a sizable profit compared to the average 5% of the open market interest rates.

5 Concluding remarks

Our study demonstrates usefulness of the rough sets theory for constructing RRTS of the stock futures market. In particular, it is to be stressed that the developed RRTS using the rough sets theory works against various situations. Since the real-time data or trading tends to make the market more volatile and mercurial than ever before, our RRTS against various situations provides a desirable robust solution. In the meantime, we envisage that this study will give an impetus for further studies of RRTS, since it only incorporates some basic tools within rough sets. A more elaborate procedure of RRTS might be developed by the help of other information reduction techniques.

Acknowledgements This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2006-B00422). T.Y. Kim's research was supported by KRF C00112 in 2007.

Appendix: Technical indicators

In the below the subscript t indicates that the related quantities are defined on time interval $(t-1,t]$, unless stated otherwise. For example, C_t, H_t, L_t and V_t denote the closing price, the high price, the low price and the trading volume recorded on $(t-1,t]$, respectively. Further define $y_{t,\max}(n) = \max(y_t, \ldots, y_{t-n+1})$, $y_{t,\min}(n) = \min(y_t, \ldots, y_{t-n+1})$ and $\bar{y}_t(n) = \frac{1}{n}\sum_{i=1}^{n} y_{t-n+i}$. Also always assume $m \le n$.

Technical Indicators	Formula	Trading rule used in empirical study
SMA (Simple Moving Average)	$SMA_t(n) = \bar{C}_t(n) = \frac{1}{n}\sum_{i=1}^{n} C_{t-n+i}$	If $SMA_t(9)$ crosses over C_t, then Buy If $SMA_t(9)$ crosses below C_t, then Sell
EMA (Exponential Moving Average)	$EMA_t(y,n) = \frac{\sum_{i=1}^{n} a^{n-i} y_{t-n+i}}{\sum_{i=1}^{n} a^{n-i}}$ where a is an exponential factor $(0 < a < 1)$.	If $EMA_t(C,9)$ crosses over C_t, then Buy If $EMA_t(C,9)$ crosses below C_t, then Sell
WMA (Weighted Moving Average)	$WMA_t(y,n) = \frac{\sum_{i=1}^{n} w_i \cdot y_{t-i+1}}{\sum_{i=1}^{n} w_i}$ where w_i is a weight factor $(w_1 \ge w_2 \ge \cdots \ge w_n)$.	If $WMA_t(C,9)$ crosses over C_t, then Buy If $WMA_t(C,9)$ crosses below C_t, then Sell
OBV (On Balance Volume)	$OBV_t = \begin{cases} OBV_{t-1} + V_t, & C_t > C_{t-1}, \\ OBV_{t-1} - V_t, & C_t < C_{t-1}, \\ OBV_{t-1}, & C_t = C_{t-1} \end{cases}$	If OBV_t crosses over $OBV_{t-1,\max}$ (15) and OBV_{t-1} crosses over $OBV_{t-2,\max}$ (15), then Buy If OBV_t crosses over $OBV_{t-1,\min}$ (15) and OBV_{t-1} crosses over $OBV_{t-2,\min}$ (15), then Sell
SMI (Stochastic Momentum Index)	$SMI_t(n) = 100 \times \left(\frac{2C_t}{(H_{t,\max}(n) - L_{t,\min}(n))} - 1\right)$	If $SMI_t(6)$ crosses over $EMA_t(SMI, 6)$, then Buy If $SMI_t(6)$ crosses below $EMA_t(SMI, 6)$, then Sell
AD (Accumulation Distribution)	$AD_t = AD_{t-1} + \frac{(C_t - L_t) - (H_t - C_t)}{(H_t - L_t)} V_t$	If $H_{t,\max}(20) < \bar{H}_t(20)$ and $AD_{t,\max}(20) > \overline{AD}_t(20)$, then Buy If $L_{t,\min}(20) > \overline{L}_t(20)$ and $AD_{t,\min}(20) < \overline{AD}_t(20)$, then Sell
CO (Chaikin's Oscillator)	$CO_t(m,n) = EMA_t(AD,m) - EMA_t(AD,n)$	If $CO_t(3, 13)$ crosses over 0, then Buy If $CO_t(3, 13)$ crosses below 0, then Sell
VO (Volume Oscillator)	$VO_t(n,m) = \frac{\overline{V}_t(m) - \overline{V}_t(n)}{\overline{V}_t(n)} \times 100$	If $V_t(12, 26)$ crosses over 0, then Buy If $V_t(12, 26)$ crosses below 0, then Sell
William's %R	$William's\,\%R_t(n) = \frac{H_{t,\max}(n) - C_t}{H_{t,\max}(n) - L_{t,\min}(n)} \times 100$	If $-100 \le William's\%\ R_t(10) \le -80$, then Buy If $-20 \le William's\%\ R_t(10) \le 0$, then Sell.
NCO (Net Change Oscillator)	$NCO_t(n) = C_t - C_{t-n}$	If $NCO_t(12)$ crosses over 0, then Buy If $NCO_t(12)$ crosses below 0, then Sell
ROC (Rate of Change)	$ROC_t(n) = \left(\frac{C_t}{C_{t-n}} - 1\right) \times 100$	If $ROC_t(12)$ crosses over 0, then Buy if $ROC_t(12)$ crosses under 0, then Sell

Technical Indicators	Formula	Trading rule used in empirical study
VROC (Volume Rate of Change)	$VROC_t(n) = \left(\frac{V_t}{V_{t-n}} - 1\right) \times 100$	If $VROC_t(14)$ crosses over 0, then Buy If $VROC_t(14)$ crosses under 0, then Sell
CCI (Commodity Channel Index)	$CCI_t(n) = \frac{M_t - \overline{M}_t(n)}{d_t(n) \times 0.015}$ where $M_t = (H_t + L_t + C_t)/3$ and $d_t(n) = \frac{1}{n}\sum_{i=1}^{n} \lvert M_{t-n+i} - \overline{M}_t(n)\rvert$.	If $CCI_t(10)$ crosses above 0, then Buy If $CCI_t(10)$ crosses below 0, then Sell
Parabolic *SAR* (Stop and Reversal)	$SAR_t = SAR_{t-1} + af \cdot (xp - SAR_{t-1})$ where $af = 0.02$ (acceleration factor) and $xp = 0.2$ (extreme point).	If C_t crosses over SAR_{t+1}, then Buy If C_t crosses below SAR_{t+1}, then Sell
MI (Mass Index)	$MI_t(n) = \sum_{i=1}^{n} \frac{EMA_{t-n+i}(r, 9)}{EMA^2_{t-n+i}(r, 9)}$ Where $r_t = H_t - L_t$ and $EMA^j_t(y, n)$ denotes the j times repeated applications of *EMA* to $\{y_{t-n+1}, \dots, y_t\}$.	If *RB* occurs and $EMA^r_i(9)$ decreases, then Buy If *RB* occurs and $EMA^r_i(9)$ increases, then Sell where *RB* (reversal bulge) means $MI_t(n)$ crosses down 26.5 just after it crosses over 27
PO (Price Oscillator)	$PO_t(m, n) = \frac{SMA_t(m) - SMA_t(n)}{SMA_t(m)}$	If $PO_t(5, 10)$ cross over 0, then Buy If $PO_t(5, 10)$ cross under 0, then Sell
Band %b	$Band\ \%b_t(n) = \frac{C_t - L_t(n)}{U_t(n) - L_t(n)}$ $U_t(n) = \overline{C_t}(n) + \left[\alpha \times \sqrt{\frac{\sum_{i=1}^{n}(C_{t-n+i} - \overline{C_t}(n))^2}{n}}\right]$ $L_t(n) = \overline{C_t}(n) - \left[\alpha \times \sqrt{\frac{\sum_{i=1}^{n}(C_{t-n+i} - \overline{C_t}(n))^2}{n}}\right]$ where α is a pre-assigned number.	If $Band\ \%b_t(9)$ crosses over $Band\ \%b_{t-1}(9)$, then Buy If $Band\ \%b_t(9)$ crosses below $Band\ \%b_{t-1}(9)$, then Sell
TRIX	$TRIX_t(n) = \frac{EMA^3_t(C, n) - EMA^3_{t-1}(C, n)}{EMA^3_{t-1}(C, n)}$	If $TRIX_t(18)$ crosses over 0, then Buy If $TRIX_t(18)$ crosses below 0, then Sell
Momentum	$Momentum_t(n) = \frac{C_t}{C_{t-n}} \times 100$	If $Momentum_t(10)$ crosses over 0, then Buy If $Momentum_t(10)$ crosses below 0, then Sell
Stochastic	$Stochastic_t(n) = \frac{C_t - L_{t-n}}{H_{t-n} - L_{t-n}} \times 100$	If $Stochastic_t(14)$ crosses below 20, then Buy If $Stochastic_t(14)$ crosses over 80, then Sell
DMI (Directional Movement Indicators)	$+DM_t = \begin{cases} H_t - H_{t-1} \text{ for } H_t - H_{t-1} > 0, \\ \quad H_t - H_{t-1} > L_{t-1} - L_t, \\ 0, \\ \quad \text{otherwise} \end{cases}$ $-DM_t = \begin{cases} L_{t-1} - L_{t-1} \text{ for } L_{t-1} - L_t > 0, \\ \quad H_t - H_{t-1} > L_{t-1} - L_t \\ 0, \\ \quad \text{otherwise} \end{cases}$ $TR_t = MAX(H_t - L_t, \lvert C_{t-1} - C_t\rvert, \lvert C_{t-1} - L_t\rvert)$	If $\overline{+DM}_t(14)$ crosses over $\overline{-DM}_t(14)$, then Buy If $\overline{+DM}_t(14)$ crosses below $\overline{-DM}_t(14)$, then Sell

Technical Indicators	Formula	Trading rule used in empirical study
ADX (Average Directional Index)	$ADX_t(n) = \overline{DX}_t(n)$ where $DX_t = \frac{\lvert(+DI_t) - (-DI_t)\rvert}{(+DI_t) + (-DI_t)} \times 100$ $+DI_t = +DM_t/TR_t$ $-DI_t = -DM_t/TR_t$ and $+DI_t(n) = \overline{+DM}_t(n)/\overline{TR}_t(n)$ $-DI_t(n) = \overline{-DM}_t(n)/\overline{TR}_t(n)$	If $+DI_t(7)$ and $ADX_t(7)$ above $-DI_t(7)$, then Buy If $-DI_t(7)$ and $ADX_t(7)$ above $+DI_t(7)$, then Sell
RSI (Relative Strength Index)	$RSI_t(n) = 100 - \frac{100}{1 + RS_t(n)}$ where $RS_t(n) = \sum_{i=1}^{n} U_{t-n+i} / \sum_{i=1}^{n} D_{t-n+i}$ where $U_t = \begin{cases} C_t - C_{t-1}, & C_t \geq C_{t-1}, \\ 0, & \text{otherwise} \end{cases}$ and $U_t = \begin{cases} D_{t-1} - D_t, & D_t \leq D_{t-1}, \\ 0, & \text{otherwise} \end{cases}$	If $RSI_t(14)$ crosses over 70, then Buy If $RSI_t(14)$ crosses below 30, then Sell
MACD (Moving Average Convergence-Divergence)	$MACD_t(m,n) = EMA_t(C,m) - EMA_t(C,n)$	If $MACD_t(12,26)$ crosses over 0, then Buy If $MACD_t(12,26)$ crosses below 0, then Sell

References

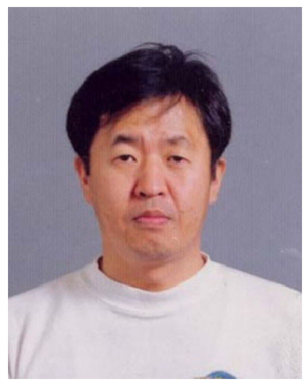
1. Achelis SB (1995) Technical analysis from A to Z. Probus Publishing, Chicago
2. Armano G, Marchesi M, Murru A (2005) A hybrid genetic-neural architecture for stock indexes forecasting. Inf Sci 170(1):3–33
3. Bao D (2007) A generalized model for financial time series representation and prediction. Appl Intell doi:10.1007/s1048900700631
4. Blume L, Easley D, O'Hara M (1994) Market statistics and technical analysis: the role of volume. J Finance 49(1):153–181
5. Brock W, Lakonishok J, Lebaron B (1992) Simple technical trading rules and the stochastic properties of stock returns. J Finance 47(5):1731–1764
6. Chan C (1998) A rough set approach to attribute generalization in data mining. J Inf Sci 107:169–176
7. Colby RW (2003) The encyclopedia of technical market indicators. McGraw-Hill, New York
8. Croushore D, Stark T (2003) A real-time data set for macroeconomists: Does the data vintage matter? Rev Econ Stat 85:605–617
9. David E, Suraphan T (2005) The use of data mining and neural networks for forecasting stock market returns. Expert Syst Appl 29(4):927–940
10. Dimitras AI, Slowinski R, Susmaga R, Zopounidis C (1999) Business failure prediction using rough sets. Eur J Oper Res 144(2):263–280
11. Gerberding C, Kaatz M, Seitz F, Worms A (2005) A real-time data set for German macroeconomic variables. J Appl Soc Sci Stud 125:337–346
12. Jackson AG, Pawlak Z, LeClair SR (1998) Rough sets applied to the discovery of materials knowledge. J Alloys Compd 279:14–21
13. Kim K (2007) Electronic and algorithmic trading technology: The complete guide (Complete technology guides for financial services). Academic Press, Burlington
14. Lee KH, Jo GS (1990) Expert system for predicting stock market timing using a candlestick chart. Expert Syst Appl 16(4):357–364
15. Lo AW, Mamaysky H, Wang J (2000) Foundation of technical analysis: Computational algorithms, statistical inference, and empirical implementation. J Finance 55(4):1705–1765
16. Mochón A, Quintana D, Sáez Y (2007) Soft computing techniques applied to finance. Appl Intell doi:10.1007/s1048900700515
17. Murphy JJ (1986) Technical analysis of the futures markets: A comprehensive guide to trading methods and applications. Prentice-Hall, New York
18. Murphy JJ (1999) Technical analysis of the financial markets. New York Institute of Finance, Paramus, NJ
19. Øhrn A (1999) Discernibility and rough sets in medicine: Tools and applications. PhD Thesis. Norwegian University of Science and Technology, Trondheim
20. Pawlak Z (1982) Rough set. Int J Comput Inf Sci 11(5):341–356
21. Pawlak Z (1997) Rough set approach to knowledge-based decision support. Eur J Oper Res 99:48–57
22. Pruitt G, Hill JR (2003) Building wining trading systems with tradestation. Wiley, New Jersey

23. Sharpe WF (1994) The Sharpe ratio. J Portfolio Manag 21:49–58
24. Shen L, Loh HT (2003) Applying rough sets to market timing decision. Decis Support Syst 37:583–597
25. Slowinski R, Zopounidis C, Dimitras AI (1997) Prediction of company acquisition in Greece by mean of the rough set approach. Eur J Oper Res 100:1–15
26. Wang Y (2003) Mining stock price using fuzzy rough set system. Expert Syst Appl 24:13–23
27. Wei H, Yoshiteru N, Shou-Yang W (2005) Forecasting stock market movement direction with support vector machine. Comput Oper Res 32(10):2513–2522
28. William L, Naval M, Russell P, Tom R (2002) Stock market trading rule discovery using technical charting heuristics. Expert Syst Appl 23:155–159

Suk Jun Lee is a Ph.D. degree student in the Department of Information and Industrial Engineering at Yonsei University, South Korea. He received his M.E. (2007) degree at Yonsei University. He has submitted some technical articles to international journals and conferences. His research interest includes algorithm trading in futures market, volatility trading in options market, and financial information systems.

Jae Joon Ahn is a Ph.D. degree student in the Department of Information and Industrial Engineering at Yonsei University, South Korea. He received his M.E. (2008) degree at Yonsei University. His research fields are volatility in options market, commodities trading in financial market, and artificial intelligence in finance.

Kyong Joo Oh is an Associate Professor in the Department of Information and Industrial Engineering at Yonsei University, South Korea. He received his B.A. (1991) and M.A. (1993) degree in Applied Statistics at Yonsei University and Ph.D. degree (2000) in Management Information Engineering at Korea Advanced Institute of Science and Technology (KAIST). Dr. Oh served as a Researcher at KAIST Techno-Management Institute (2000) and Research Center of Hyundai Securities Co. (2001). Also, he worked as an Assistant Professor in the Division of Business Administration at Hansung University (2002–2004), South Korea. He has published over 15 technical articles in some international journals, such as *Expert Systems with Applications, Neurocomputing, Expert Systems, Intelligent Data Analysis, Asia-Pacific Journal of Operational Research*, etc. His research fields are financial engineering, algorithm trading in futures and options market, financial information systems, etc.



Tae Yoon Kim is a Professor in the Department of Statistics at Keimyung University, South Korea. He got his Ph.D. in statistics at University of Illinois at Urbana-Champaign (1990). His research interest includes nonparametric function estimation, bootstrap, neural network and time series analysis of financial markets. He served as visiting research scholar at The Bank of Korea (2001) and Department of Statistics at Rice University (1998).