Forecasting Stock Market Trends
Using Rough Set

Naglaa Ragaa Saeid Hassan

Assistant Professor Of Operations Research,

Institute of Statistical Studies and Researches

Cairo University

naglaa777subkiii@yahoo.com

Seham Fathy Mohamed Ibrahim

Lecturer of Economics

Faculty of Commerce – Ismailia

Suez Canal University

Sfmik_1@yahoo.com

Abstract

Predicting financial markets and trends is an important issue, either on the microeconomic level for enterprise managers, investors, creditors and supervisors, or on the macroeconomic level. Complexity of financial decisions has increased rapidly, thus highlighting the importance of developing and implementing sophisticated and efficient quantitative analysis techniques for supporting and aiding financial decision making. Rough Set Theory RST is an emerging Automatic Target Recognition (ATR) methodology for determining features and then classifiers from a training data set. RST guarantees that once the training data has been labeled all possible classifiers (based on that labeling) will be generated. Different investors groups present different investment behavior and of course their trades may relate to asset prices in different manners. This study examines whether some investors groups are associated with concurrent or subsequent market-wide price movements through the market trend using rough set theory as one of an artificial intelligence techniques, and comparing the result with that of applying econometric analysis. The experimental results of this study prove on the following (1) the ability of the rough set approach to discover facts hidden in data which represent the market trend. (2) The efficiency of the suggested classifying method to manage the discretization reaching to the general market trend. (3) Using the generalization to interpret the extracting trading rules which represent the market trend to proof with self evident on general known trading rules such as in case of selling the price is downward trend in the market and vice versa.

Key words: The rough set approach, Trading rule, Stock market trends, stock market prediction, econometric analysis, econometric model.

1. Introduction

Most researchers have used artificial intelligence (AI) techniques to predict the price of every trading day, week, and month. However, it is more important to determine stock market best time; when to buy and when to sell stocks, than to predict the price movement for everyday while

investors in stock market generally do not trade every day. If investors trade their stocks every day, they would be charged a tremendous amount of trade fees. Besides, the decision to invest is initially influenced by the trend and then by the prices. In this paper the aggregate purchases and sales of equities are divided among five major investment groups in the economy: Egyptian, Arabian, foreigners, individuals and institutions.

Market and Business prediction is a scientific field which many scholars, credit institutions, clients, economic policymakers, and financial organizations have been working for. Financial ratios methodology was developed for the business prediction. These ratios have been long considered as objective indicators (Beaver, 1966; Courtis, 1978; Altman, 1993). Altman proposed to use the discriminant analysis (a discriminant function with five financial ratios has been assessed) for predicting the business risk. Subsequently, the use of this method has continued to spread out so that it can be found discriminant models of predicting. At the same time, however, the generalization of the discriminant analysis resulted in several studies (Eisenbeis, 1977; Ohlson, 1980; Dimitras et al., 1996).

Since the work of Altman (1968), several studies proposing other methods have been used to overcome some disadvantages of the discriminant analysis and to provide higher prediction accuracy. Among these studies, we can cite the study of Ohlson (1980) using logit analysis and the study of Zmijewski (1984) using probit analysis. Friedman et al. (1985) first employed the recursive partitioning algorithm while Gupta et al. (1990) used mathematical programming methods for the business failure prediction problem. Other methods used were survival analysis by Luoma and Laitinen (1991), expert systems by Messier and Hansen (1988), neural networks by Altman et al. (1994) and multi-factor model by Vermeulen et al. (1998). Moreover, several methods were developed based on multicriteria decision aid methodology (MCDA). They classified firms into categories according to their business failure risk (Zopounidis, 1987; Siskos et al., 1994; Dimitras et al., 1995). The use of multicriteria decision aid methods in the prediction of business failure circumvents many of the problems that exist when using discriminant analysis (see Eisenbeis, 1977 and Dimitras et al., 1996 among others).

Most of the methods mentioned above have already been investigated in the field of comparative studies (Scott, 1981; Zavgren, 1983; Altman,

1984; Jones, 1987; Keasey and Watson, 1991 among others). Dimitras et al. (1996) gave a complete review of methods used for the prediction of business failure and of new trends in this area. Recently, new methods of predicting business trends have been developed. Due to the advancement of computer and information science, they offered significant aid in the evaluation and selection of viable firms for financing.

This paper presents a new method called the rough set approach for the analysis and prediction of business and market trends, and compare it with econometric model and analysis, The concept of a rough set, introduced by Pawlak (1982), proved to be an effective tool for the analysis of information tables (financial information tables) describing a set of objects firms) by a set of multi-valued attributes (financial ratios).

The aim of the this study is to test the ability of the rough set approach in predicting business and market trends, and to determine whether its analysis is compatible with the economic analysis and theory or not. This is important to provide and support economic analysis with new techniques and analysis tools.

Rough Set Approach is proposed to the extraction of trading rules for discriminating between five items in the stock market that are: Egyptian, Arabs, Foreigns, Institutions, and Individuals. Rough set theory is quite valuable for extracting trading rules, because it can be used to discover dependences in data while reducing the effect of superfluous factors in noisy data. In addition, it does not generate a signal to trade when the pattern of the market is uncertain because the selection of reducts and the extraction of rules are controlled by the strength of each reduct and rule. The experimental results are encouraging and show the usefulness of the rough set approach for stock market analysis with respect to profitability.

Rough set theory was introduced by Pawlak(1982). It has attracted the attention of many researchers and practitioners all over the world, who contributed to its development and applications during the last decade (see e.g. Pawlak, 1991; 264 A.I. Dimitras et al. / European Journal of Operational Research 114 (1999) 263±280 Slowinski, 1992; Ziarko, 1994; Lin and Wildberger, 1995; Pawlak et al., 1995; Pawlak and Slowinski, 1994; Slowinski, 1995).

Just to name a few possible uses of rough set theory, this theory may be used to describe dependencies between attributes, to evaluate significance

of attributes, and to deal with inconsistent data. As an approach to handling imperfect data (uncertainty and vagueness), it complements other theories that deal with data uncertainty, such as probability theory, evidence theory, fuzzy set theory, etc.

The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). Objects characterized by the same information are indiscernible in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis for the rough set theory. Any set of all indiscernible objects is called an elementary set, and forms a basic granule of knowledge about the universe.

Any set of objects being a union of some elementary sets is referred to as crisp (precise) ± otherwise the set is rough (imprecise, vague). Consequently, each rough set has boundary-line cases, i.e. objects which cannot be classified with certainty as members of the set or of its complement. Therefore, a rough set can be represented by a pair of crisp sets, called the lower and the upper approximation. The lower approximation consists of all objects which certainly belong to the set and the upper approximation contains objects which possibly belong to the set.

The aim of this paper is to provide an in-depth presentation of the contributions of rough set in the field of financial economics, focusing on the methods used and their real-world applications.

Large number of researchers is undertaking the behavior study of stock markets all round the world. Each researcher work throws new insight owing to the dynamic nature of the capital market. There is never end for that going on debate to tell whether the stock markets are predictable or not. Many research studies have shown empirical evidence stating that the stock markets are not predictable and therefore efficient, and there is equal number of research works showing the predictability of stock market prices.

Before computers, people traded stocks and commodities depending primarily on intuition. As the level of investing and trading grew, people searched for tools and methods that would increase their gains while minimizing their risk. Statistics, technical analysis, fundamental analysis,

and linear regression are all used to attempt to predict and benefit from the market's direction. None of these techniques has proven to be the consistently correct prediction tool that is desired and many analysts argue about the usefulness of many of the approaches.

Traders in the market often use technical tools to guide their investment decisions. They often rely on more than one method to predict the future price of shares in the market. The main assumption behind the use of technical analysis is that history repeats itself, and therefore by identifying the recurring patterns, future price movements are predictable. Widely used technical indicators are the Moving Average (MA), Momentum (M), and Stochas..etc.

In the recent years, investors have started to show interest in trading on stock market indices as it provides an opportunity to hedge their market risk, and at the same time it offers a good investment opportunity for speculators and arbitrageurs. Where, the decision to invest is initially influenced by the trend and then by the prices. Investors adopt unique trading strategies depending on the trend prevailing in the stock market. Thus, the possible trend that would prevail in the stock market, could be helpful for investors to adopt a suitable investment strategy**.**

Typically stock market indices are the performance indicators for the entire market. They act as barometers, which enable us to get an idea about the performance of the entire market in general. Interested investors all around the globe keep track of the movement of various stock market indices to get an idea about how the global markets are moving.

Thus, financial investment decision making, especially in stock exchanges, requires the accurate prediction of the future states based on the identification of hidden patterns in the available historical data. In addition, the stock exchanges historical data are in the format of a multi attribute information table which suits the rough sets model very well.

This paper focuses on recognizing the market trend using rough set theory within a generalization concept of generating trading rules for market timing. Different investors groups present different investment behavior and of course their trades may relate to asset prices in different manners. This paper analyzes the relation between stock market INDEX

EGX 30 and cash flows from a broad array of investor groups (Egyptian, Arabian, foreigners, individuals and institutions) using rough set as one of an artificial intelligence techniques, comparing with regression analysis as an econometric model. So, the stock market's movements are predicted in order to retrieve knowledge that could guide investors on when to buy and sell. The experimental results of this study shows that rough set able to discover important facts hidden in data so, obtaining the market trend while eliminating superfluous factors in noisy stock market data to extract these trading rules.

2. METHODOLOGY

2.1 Rough sets: Foundations

Rough sets theory was introduced by Pawlak in 1982. It was developed based on mathematical tool to deal with vagueness and uncertainty in the classification of objects in a set. The rough set philosophy is founded on the assumption that there is some information regarding features which can be associated with every object of the universe. In rough sets, the data is organized in a table called decision table, which are flat tables containing attributes as columns and data elements as rows. The class label is called as decision attribute, the rest of the attributes are the condition attributes. For rows, rough set theory employs the notion of indiscernible class to group similar topless together; while for columns, it employs the notion of indiscernible attribute to identify significant attributes.

We call the class label a decision feature, the rest of the features are conditional. Let o, F denote a set of sample objects and a set of functions representing object features, respectively. Assume that $B \subseteq F, x \in O.$ Further, let $[x]_B$ denote:

$$[x]_B = \{ y : x \sim B \ y \}$$

Rough set theory define three regions based on the equivalent classes induced by the feature values: lower approximation $\underline{B}X$, upper approximation $\overline{B}X$, and boundary $BND_B(X)$. A lower approximation of a set X contains all equivalence classes $[x]_B$ that are subsets of X, and upper approximation contains all equivalence classes $[x]_B$ that have objects in common with X, while the boundary is the set of all objects in $\overline{B}X$ that are not contained in $\underline{B}X$. So, we can define a rough set as any set with a non-empty boundary [5]. The indiscernibility relation $\sim B$ (or by Ind_B) is a fundamental principle of rough set theory. Informally, $\sim B$ is a set of all objects that have matching descriptions. Based on the selection of B, $\sim B$ is an equivalence relation partitions a set of objects O into equivalence classes. The set of all classes in a partition is denoted by $O/\sim B$ (also by O/ Ind_B). the set O/ Ind_B is called the quotient set. Affinities between objects of interest in the set $X \subseteq O$ and classes in a partition can be discovered by identifying those classes that have objects in common with X. Approximation of the set X begins by determining which elementary sets $[X]_B \in O/\sim B$ are subsets of X. For a detailed review of the basic material, reader may consult sources such as [3, 6-9].

2.2 Prediction model using rough sets:

As in [10], Fig. 1 illustrates the summery of the overall steps in the proposed Rough Set Prediction Model (RSPM) of stock market data. Functionally, RSPM can be partitioned into four distinct phases:

- **Data collection phase**
- **Pre-processing phase**: This phase includes tasks such as extra variables

addition and computation, decision classes, assignments, data cleansing, completeness, correctness, attribute creation, attribute selection and discretization.

- **Applying rough set model for Analysis and Rule Generating Phase**: This phase includes the generation of preliminary knowledge, such as computation of object reducts from data, derivation of rules from reducts, rule evaluation and prediction processes.

- **Classification and Prediction phase and validation tests**: This phase utilize the rules generated from the previous phase to predict the stock price movement.
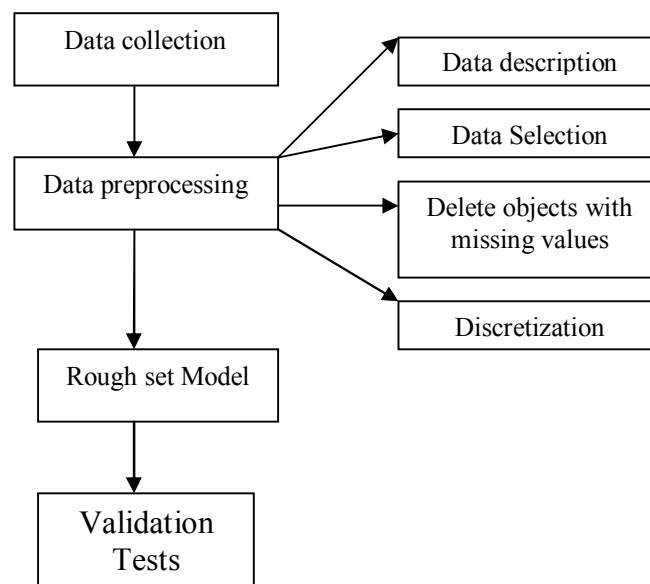

Fig. 1 The process of stock market data prediction and analysis using Rough Set Model

3. Trading rules extraction and the market trend

3.1 Case study

3.1.1. Data collection phase:

The primary source of data in this paper is from "The Egyptian's company for information publishing" which reports purchases, sells and net of major assets by investor groups in the Egyptian economy. The data

covers approximate one year time period daily; from 12 Aug.2008 to 15 Oct. 2009. This study use total purchases and sells of equities for five major investors groups: Egyptian investors, Arab investors, foreign investors, institutions investors and individual investors besides, the data also include the daily opening, closing index values of stock markets and the difference between close and open EGX

3.1.2. Preprocessing phase:

In this phase, the decision table required for rough set analysis is created. In doing so, a number of data preparation tasks such as data conversion, data cleansing, data completion checks, conditional attribute creation, decision attribute generation, discretization of attributes are performed. Data splitting is also performed which created two randomly generated subsets, one subset for analysis containing %75 of the objects in the data set and one validation containing the remainder %25 of the objects. It must be emphasized that data conversion performed on the initial data must generate a form in which specific rough set tools can be applied.

Often, real world data contain missing values. Since rough set classification involves mining for rules from the data, objects with missing values in the data set may have undesirable effects on the rules that are constructed. The aim of the data completion procedure is to remove all objects that have one or more missing values.

Incomplete data or information systems exist broadly in practical data analysis, and approaches to complete the incomplete information system

through various completion methods in the preprocessing stage are normal in data mining and knowledge discovery. However, these methods may result in distorting the original data and knowledge, and can even render the original data to be unable.

In this paper, it is intended to delete some data from original data set to study the effect of that deleted data on extracting knowledge from the rules and generality. To overcome the shortcomings inherent in the traditional methods, we eliminate all incomplete information system (i.e. decision table).

3.1.2.1. Data description:

The decision table of this study contains (216 Objects, 18 Attributes, 15 Conditional attribute and 3 decision attributes) refer to as data set 1 (DS1) and because of their often exist conditional attributes that don't provide (almost) any additional information about the objects. These attributes need to be removed in order to reduce the complexity.

3.1.2.2. Data selection:

The data selection depending on the optimality criterion associated with attributes of DS1 which affects the final generated rules because the rough sets model can only extract the knowledge inherent with existing data set or table, which it means that if the attributes are not represented in the related information system, then the generated rules would be dispersed and become insignificant. After deep looking inside the data contents of DS1, it found that it is possible to work only with the net attributes as it summarizes the total effect of buyers and sellers. Applying

this action of selection on DS1, obtaining a new decision table DS2 which contains (216 Objects, 6 Attributes, 5 Conditional attribute and 1 decision attribute). It is also noticed that according to a relation between the conditional attributes, which:

Egyptians + Arabs + Foreigners = Individuals + Institutions

So, DS2 could be split into two decisions tables and study each one alone as in (Table 1-1 and Table 1-2).

Table 1-1 Description of data (part I)

Item	Type	Description
Date	The universe of obiects	Date of transaction
EgyptianN	Conditional	Net number of Egyptian buyers and
ArabN	Conditional	Net number of Arab buyers and
ForeignerNs	Conditional	Net number of foreigner buyers and
EGX30D	Decision attribute	Difference between close and open EGX

Table 1-2 Description of Data (Part II)

Item	Type	Description
Date	The universe of	Date of transaction
IndividualsN	Conditional attribute	Net number of individual buyers and sellers
InstitutionsN	Conditional attribute	Net number of institution buyers and sellers
EGX30D	Decision attribute	Difference between close and open EGX stock market index

Notice that:

The Table 1-1 is containing (216 Objects, 4 Attributes, 3 Conditional attribute and 1 decision attribute) and Table 1-2 is containing (216 Objects, 3 Attributes, 2 Conditional attribute and 1 decision attribute).

As it is mentioned before, in the Preprocessing phase, the reliability data should be considered where the objects with missing or wrong value will be deleted according to the following relation where:

(Egyptians + Arabs + Foreigners = Individuals + Institutions)

Thus, the raw data set of this project is DS2 and after eliminating the missing and wrong data which includes 132 wrong data objects out of the original 216 objects, So, the rest objects appear in the new decision tables (Table 1-1 and Table 1-2) which contains only 84 objects.

3.1.2.3. Discretization:

In rough sets theory and other induction learning systems, discretization is an important algorithm and can be viewed as a process of information generalization (or abstraction) and data reduction [9]. Discretization is a process of grouping the values of the attributes in intervals in such a way that the knowledge content or the discernibility is not **lost** [11-12]. Many discretization approaches have been developed so far. Nguyen S. H had given some detailed description about discretization in rough set as in reference [3].

Before applying analysis on data set using rough set software or package, first, it must be discretized. As a result of discretization, the precision of the original data will be decreased but its generality will be increased. Of course, the efficiency of the prediction process is increased if applying a suitable discertization algorithm on data set

The discretization of real value attributes is one of the important problems

to be solved in data mining, especially in rough set. We know, when the value set of any attribute in a decision table is a continuous value or a real number, and then it is likely that there will be few objects that will have the same value of the corresponding attributes. In such a situation the number of equivalence classes based on that attribute will be large and there will be very few elements in each of such equivalence class, which it will lead to the generation of a large number of rules in the classification of rough set, thereby making rough set theoretic classifiers inefficient.

As the major objective of stock market analyses is to determine whether the stock market trend is upward, downward or hold. In this study, It is used the values of attribute (EGX 30D) to be (1) if increasing, (-1) if decreasing and (0) otherwise. The same process is applied on all attributes (EgyptianN, ArabN and ForeignersN) in Table 1-1 and all attributes (IndividualsN and InistitutionsN) in Table 1- 2.

3.1.3. Applying rough sets model

After the preprocessing step, the rough sets model is applied as analysis tool to extract rules from decision tables to predict the stock market trend. In validation step, the rules are tested for their validation. The tested sample is selected randomly and classified only once, with its size being set arbitrarily to a constant percentage (25%) of all objects.

3.1.4. Results, analysis and interpretation

Applying rough set software (packages), the rule derived from re-ducts can be used to classify the data. The set of rules is referred to as a classifier and can be used to classify new and unseen data.

Table 1-1 includes {EgyptianN, ArabN, ForeignersN} and after applying the reduction action, it is generated 9 rules which only 2 are consistent (rules 4 and 9) as shown in Fig 2.

	Rule
1	EgyptianN(-1) AND ArabN(1) AND ForeignersN(-1) => EGX 30D(-1) OR EGX 30D(1)
2	EgyptianN(1) AND ArabN(-1) AND ForeignersN(-1) => EGX 30D(-1) OR EGX 30D(1)
3	EgyptianN(1) AND ArabN(-1) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
4	EgyptianN(-1) AND ArabN(-1) AND ForeignersN(-1) => EGX 30D(-1)
5	EgyptianN(-1) AND ArabN(-1) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
6	EgyptianN(1) AND ArabN(1) AND ForeignersN(-1) => EGX 30D(1) OR EGX 30D(-1)
7	EgyptianN(0) AND ArabN(-1) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
8	EgyptianN(-1) AND ArabN(1) AND ForeignersN(1) => EGX 30D(1) OR EGX 30D(-1)
9	EgyptianN(1) AND ArabN(-1) AND ForeignersN(0) => EGX 30D(1)

Fig 2 The rules generate from re-duct

The rules are generated from re-duct using ROSETTA software contains the following knowledge:

Rule 4: EgyptianN(-1) AND ArabN(-1) AND ForeignersN(-1) => EGX 30D(-1)

- Which means "IF Egyptians, Arabs and foreigners are sellers Then the market trend is going downward"

Rule 9: EgyptianN(1) AND ArabN(-1) AND ForeignersN(0) => EGX 30D(1)

- Which means "IF Egyptians are buyers, Arabs are sellers and foreigners are even Then the market trend is going upward"

Otherwise (rules 1, 2, 3, 5, 6, 7 and 8) the market trend may go upward or downward.

So, the Table 1-1 shows a set of the generated rules. These obtained rules

are used to build the prediction system for investors.

- After splitting Table 1-1 into learning (75%) and validation (25%)

The results of learning, obtaining only one reduct where it is including (EgyptianN, ForeignersN) but attribute ArabN is dispensable.

The reduct generated only six rules; two of them are consistent (rules 1 and 6) as in Fig 3.

	Rule
1	EgyptianN(-1) AND ForeignersN(-1) => EGX 30D(-1)
2	EgyptianN(1) AND ForeignersN(-1) => EGX 30D(-1) OR EGX 30D(1)
3	EgyptianN(1) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
4	EgyptianN(-1) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
5	EgyptianN(0) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
6	EgyptianN(1) AND ForeignersN(0) => EGX 30D(1)

Fig 3 The rules generate from reduct

Notices:

Rule 1: EgyptianN(-1) AND ForeignersN(-1) => EGX 30D(-1)

- Which means that "IF Egyptians are sellers and foreigners are sellers Then the market trend is going downward"

Rule 6: EgyptianN(1) AND ForeignersN(0) => EGX 30D(1)

- Which means "IF Egyptians are buyers and foreigners are even Then the market trend is going upward"

- Otherwise (rules 2, 3, 4, and 5) the market trend may go upward or downward.

- When tested on validation set, the accuracy is equal to 0.714286.

Comment: The results are compared with the actual observed trend, and it is seen that the rough set model has predicted the trend accurately. Whereas, if the accuracy is higher, the classification is less ambiguous and the quality of rule is better, therefore, the classification is unambiguous and acceptable

From Table 1-2, it is obtained only 1 reduct including (IndividualsN, InstitutionsN) and generated 3 rules from which only 1 is consistent (rule 3) as in Fig 4 and the same result in learning

	Rule
1	IndividualsN(-1) AND InstitutionsN(1) => EGX 30D(-1) OR EGX 30D(1)
2	IndividualsN(1) AND InstitutionsN(-1) => EGX 30D(-1) OR EGX 30D(1)
3	IndividualsN(-1) AND InstitutionsN(-1) => EGX 30D(-1)

Fig 4. The rules generate as reduct

Notice that:

Rule 3: IndividualsN(-1) AND InstitutionsN(-1) => EGX 30D(-1)

- Which means "IF Individuals are sellers and Institutions are sellers Then the market trend is going downward"
- Otherwise (rules 1 and 2) the market trend may go upward or downward.

- When tested on validation set the accuracy is equal to 0.761905.

3.1.5. Prediction phase

In this phase, the generated rules are utilized to predict the market trend. Combining all previous results, it is noticed that:

1- The attribute ArabN is dispensable.

2- The stock market goes upward only IF Egyptians are buyers and foreigners are even

3- The stock market goes downward

       IF Egyptians and foreigners are sellers

   Or/ and

       IF Individuals and Institutions are sellers

And this rule as proof with self evident on general known trading rules such as in case of selling the price is downward trend in the market and vice versa

4- Different investors groups present different investment behavior and of course their trades may relate to asset prices in different manners.

5- Some investors groups are associated with concurrent or subsequent market-wide price movements through the market trend.

6- So, the market trend is predicted using the rough set model. When the results are compared with the actual data and observed trend, it is seen that the rough set model has predicted the market trend with high accuracy. So, these extracting trading rules could be used to guide investors whether to buy, sell or hold on the stock.

4. Comparison of the rough set approach with econometric analysis:
4.1 Related Literatures:
Several studies documented that individuals and various types of institutions exhibit different trading behaviors. Examples include Del Guercio (1996), Cohen (1999), Dennis and Strickland (2002) and Barber, Lee, Liu and Oden (2007). They found that foreigners tend to be momentum investors, while domestic investors, particularly individuals, tend to be contrarian.

Other studies investigated the relation between both individual and institutional trades and stock prices. Gompers a Metrick (2001), Chakravarty (2001), Campbell, Ramadorai and Vuolteenaho (2005), Kaniel, Saar, and Titman (2008), Nofsinger and Sias (1999), Wermers (1999), Griffin, Harris and Topaloglu (2003), and Cai and Zheng (2004) among those studies.

4.2 procedures:
The primary data source for this study is the "Egypt Company for Information Publishing". Statistical release issued by this company which reports holdings, purchases, and sells of major assets by investor groups in the Egyptian Economy. The Data in this study covers a 4 Years time

period daily data, from the first of January 2007 to the end of 2010. Data presented in this paper covers purchases and sells of equities for five major investor groups that are: Egyptian, Arab, Foreigners, Individuals and Institutions, the same groups that are used in the Rough Set Technique.

4.3 Methodology and results:
4.3.1 augmented Dickey–Fuller test (ADF):

In econometrics, an **augmented Dickey–Fuller test (ADF)** is a test for a unit root in a time series sample. It is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models. The augmented Dickey–Fuller (ADF) statistic, used in the test, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence.

The testing procedure for the ADF test is applied to the model

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

where α is a constant, β the coefficient on a time trend and p the lag order of the autoregressive process. Imposing the constraints $\alpha = 0$ and $\beta = 0$ corresponds to modeling a random walk and using the constraint $\beta = 0$ corresponds to modeling a random walk with a drift. Consequently, there are three main versions of the test, analogous to the ones discussed on the Wikipedia page for the Dickey-Fuller test. See that page for a discussion on dealing with uncertainty about including the intercept and deterministic time trend terms in the test equation.

By including lags of the order p the ADF formulation allows for higher-order autoregressive processes. This means that the lag length p has to be determined when applying the test. One possible approach is to test down from high orders and examine the t-values on coefficients..

The unit root test is then carried out under the null hypothesis $\gamma = 0$ against the alternative hypothesis of $\gamma < 0$. Once a value for the test statistic

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

is computed it can be compared to the relevant critical value for the Dickey–Fuller Test. If the test statistic is less (this test is non symmetrical

so we do not consider an absolute value) than (a larger negative) the critical value, then the null hypothesis of $\gamma = 0$ is rejected and no unit root is present.

4.3.2 Correlation And Regression Analysis:

The Following , the derived models expressing the relation between stock market index and purchases and sales from investors groups, using a correlation test and regression analysis.

As for purchases, It is found that the relation between stock market index and purchases flows is positive and significant for purchases of individuals, foreign investors, and Egyptian. However it is negative and significant for institution purchases.

However with sales, it is found that the relation between stock market index and sales is positive and significant with sales of Egyptian and Arab and foreign, while negative with sales of individuals and institution.

Thus the Econometric model expressing the relation between EGX30 and purchases is:

$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + b_5 x_5$
$Y = 3471 + 1.9\ x_1 + 5.25\ x_2 - 1.72\ x_3 + 2.29\ x_4 - 1.25\ x_5$
However:
Y = EGX 30 (Stock market index)
X1 = Egyptian purchases
X2 = Arab purchases
X3 = Foreign purchases
X4 = individual purchases
X5 = institution purchases
R square = 52.8% indicating that independent variables are responsible for about 53% from changing in the dependant variable.

Econometric model expressing the relation between EGX30 and sales is:
$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + b_5 x_5$
$Y = 4102 + 3.73\ x_1 + 10.25\ x_2 + 4.73\ x_3 - 1.42\ x_4 - 2.91\ x_5$
However:

Y = EGX 30 (Stock market index)
X1 = Egyptian sales
X2 = Arab sales
X3 = Foreign sales
X4 = individual sales

X5 = institution sales

R square = 50.9 % indicating that independent variables are responsible for about 51% from changing in the dependant variable.

The derived equations from regression analysis refer to that in the case of purchasing, the EGX30 is affected positively by all groups of investors, except purchases of institutions that have a negative relation with EGX30. However, in the case of sales, EGX30 is affected positively by sales of Egyptian, Arab, foreigners, but negatively by sales of individuals and institutions.

5. Conclusion

Different investors groups present different investment behavior and of course their trades may relate to asset prices in different manners. In this paper, the relation between stock market INDEX EGX 30 and cash flows from a broad array of investor groups (Egyptian, Arabian, foreigners, individuals and institutions) is analyzed using rough set as one of the artificial intelligence techniques.

This paper proves on the following: First, the ability of the rough set approach to discover facts hidden in data which represent the market trend as an index. Second, this study is a self evident on general known trading rules such as in case of selling the stocks, the price is downward trend of the market and vice versa by generalization the extracting trading rules representing the market trend.

It has been demonstrated that suggested model is a successful and promising forecasting tool for predicting the general future trend in the market where the experimental results are encouraging and prove that the usefulness of suggested method obtaining a higher strength of prediction in terms of the significance of rules which means it could be used to

guide investors whether to buy, sell or hold a stock.

Thus, rough set is an important tool to deal with uncertain or vague where it helps the researchers in identifying the future trends in the markets. Besides, it is a good indicator for the investors or the decision makers to make better investment decisions.

Although the philosophy of the rough set approach and econometric analysis is very different, their comparison on a common set of data is well founded because the two methodologies can be applied to the market prediction problems. The choice of rough set analysis has been made since widely used for classification and prediction in financial problems (see Altman et al., 1981; Altman, 1993; Dimitras et al., 1995). econometric analysis must respect several statistical assumptions such as normality of distributions of the variables, equality of variance and covariance of the variables for the groups of investors. It is hard indeed to meet all these assumptions in practice; this is why their validity has been questioned by several authors (see Eisenbeis, 1977; Ohlson, 1980; Zavgren, 1985). Unlike discriminant analysis, the rough set approach does not need any assumptions about data prior to the analysis (see Krusinska et al. (1992) and Stefanowski (1992) for a critical discussion of the two methods).

It can be observed that better results are obtained for the econometric model derived from the complete information table, however, these results are less satisfactory than those obtained using the rough set approach.

Finally, compared to other existing methods, rough set approach offers the following advantages: It discovers important facts hidden in data and expresses them in the natural language of decision rules. · It accepts both quantitative and qualitative attributes and specifies their relevance for approximation of the classification. It can contribute to the minimization of the time and cost of the decision making process (rough set approach is an information processing system in real time).

It offers transparency of classification decisions, allowing for their argumentation. It takes into account background knowledge of the decision maker. We investigate the use of the rough set model for financial time-series data analysis and forecasting.

The rough set model is an emerging technique for dealing with vagueness and uncertainty in data. It has many advantages over other techniques, such as fuzzy sets and neural networks, including attribute reduction and variable partitioning of data. These characteristics can be very useful for improving the quality of results from data analysis. We demonstrate a rough set data analysis model for the discovery of decision rules from time series data. Rules are generated through reducts and can be used for future prediction. A unique ranking system for the decision rules based both on strength of the rule and stability of the rule is used in this study. The ranking system gives the user confidence regarding their market decisions. Our experiment results indicate that the forecasting of future stock index values using rough sets obtains decision rules with high accuracy and coverage.

REFERENCES

Al-Qaheri, H., Hassanien, A. E. and Abraham, A. (2008) Discovering Stock Price Prediction Rules Using Rough Sets. Neural Network World Journal, (to appear).

Altman, E.I., 1968. Financial ratios discriminate analysis and the prediction of corporate bankruptcy. The Journal of Finance 23, 589-609.

Altman, E.I., 1984. The success of business failure prediction models: An international survey. Journal of Banking and Finance 8 (2), 171-198.

Altman, E.I., 1993. Corporate Financial Distress and Bankruptcy. John Wiley, New York.

Altman, E.I., Marco, G., Varetto, F., 1994. Corporate distress diagnosis: Comparisons using discriminant analysis and neural networks (the Italian experience). Journal of Banking and Finance 18, 505±529.

Barber, Brad M., Lee, Yi-Tsung, Liu, Yiu-Jane, and Oden, Terrance, 2007, Just how much do individual investors lose by trading? Working paper, University of California at Davis.

Beaver, W.H., 1966. Financial ratios as predictors of failure. Empirical Research in Accounting: Selected Studies. Journal of Accounting Research, Supplement to vol. 4, pp. 71± 111.

Cai, Fang, and Zheng, Lu, 2004, Institutional trading and stock indices, Journal of Finance, 61 (2).

Campbell, John Y., Ramadorai, Tarun, and Vuolteenaho, Tomo, 2005, Caught on tape: institutional order flow and stock returns, NBER Working Paper No. 11439.

Chakravarty, S., 2001, Stealth trading: which traders move stock prices? Journal of financial economics 61 (2).

Chen, C. Y., Guoli, Z, Qiao, S. Y. and Wen, S. P. 2003 Study on discertization in rough set based on genetic algorithm, IEEE Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an.

Cohen, R., 1999, Asset allocation decision of individuals and institutions, Harvard University, Working Paper.

Courtis, J.K., 1978. Modeling a financial ratios categorical framework. Journal of Business Finance and Accounting 5 (4), 371±386.

Del Guercio, D., 1996, The distorting effect of the prudent- man laws on institutional equity investment, Journal of Financial Economics 40, (31).

Dennis, Patrick J., Strickland, Deon, 2002, Who blinks in volatile markets, individuals or institutions? Journal of Finance 57 (5).

Dimitras, A.I., 1995. Multicriteria methods for business failure prediction (in Greek). Ph.D. Dissertation, Technical University of Crete, Chania, Greece.

Dimitras, A.I., Zanakis, S.H., Zopounidis, C., 1996. A survey of business failures with an emphasis on prediction methods 278.

Dimitras, A.I., Zopounidis, C., Hurson, Ch., 1995. A multicriteria decision aid method for the assessment of business failure risk. Foundations of Computing and Decision Sciences 20 (2), 99±112.

Eisenbeis, R.A., 1977. Pitfalls in the application of discriminant analysis in business and economics. The Journal of Finance 32, 875±900.

Friedman, H., Altman, E.I., Kao, D.-L., 1985. Introducing recursive partitioning for financial classification: the case of financial distress. The Journal of Finance 40 (1), 269±291.

Golan, R.H. and Ziarko, W. (1995) A methodology for stock market analysis utilizing rough set theory. Proceedings of the IEEE/IAFE Computational Intelligence for Financial Engineering, 9-11 Apr, pp. 32-40.

Gompers, Paul A., Metrick, Andrew, 2001, Institutional investors and equity prices. Quarterly Journal of Economics, 116, 229-259.

Griffin,John, M., Harris, Jeffrey, Topaloglu, Selim, 2003. The Dynamics of institutional and individual trading, Journal of Finance, 58, 2285-2320.

Gupta, Y.P., Rao, R.P., Bagghi, P.K., 1990. Linear goal programming as an alternative to multivariate discriminant analysis: a note. Journal of Business Finance and Accounting 17 (4), 593±598.

Jaaman, S. H., Shamsuddin, S. M., Yusob, B. and Ismail, M. (2009) a predictive for Malaysian stock market returns, International Research Journal of Finance and Economics Euro Journal publishing, Inc. I. 30.

Jones, F.L., 1987. Current techniques in bankruptcy prediction. Journal of Accounting Literature 6, 131±164.

Kaniel, Ron, Saar, Gideon, Titman, Sheridan, 2008, individual investor trading and stock returns, Journal of Finance, 63, 273-310.

Keasey, K., Watson, R., 1991. Financial distress prediction models: a review of their usefulness. British Journal of Management 2, 89±102.

Lawrence R. (1997) Using Neural Networks to Forecast Stock Market Prices, Department of computer Science, University of Manitobo, pp. 5-10. Retrieved September 2006 from: http://www.cs.uiowa.edu/~rlawrenc/research/Papers/nn. pdf.

Luoma, M., Laitinen, E.K., 1991. Survival analysis as a tool for company failure prediction. Omega 19 (6), 673±678.

Messier, W.F., Hansen, J.V., 1988. Including rules for expert system development: an example using default and bankruptcy data. Management Science 34 (12), 1403±1415.

Nguyen H S. and Skowron A. (1995) Quantization of real value attributes. Proceedings of Second Joint Annual Conf. on Information Science, Wrightsville Beach, North Carolina, p34-37

Nofsinger, John R., Sias, Richard W., 1999, Herding and feedback trading by institutional and individual investors, Journal of Finance 54. 2263-2295.

Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research (Spring), 109±131.

Pawlak Z. (1982) Rough Sets, Int. J. of information and computer science., V. 11 p. 341-356.

Pawlak Z., Grzymala, B. J., Slowinski R. and Ziarko, W. (1995) Rough sets. Communications of the ACM, V. 38, No. 11, pp. 89-95.

Pawlak, Z. (1991) Rough Sets. Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishers, The Netherlands.

Polkowski, L. (2003) Rough Sets: Mathematical Foundations. Physica-Verlag.

Scott, J., 1981. The probability of bankruptcy: a comparison of empirical predictions and theoretical models. Journal of Banking and Finance 5,317±344.

Siskos, Y., Zopounidis, C., Pouliezos, A., 1994. An integrated DSS for financing firms by an industrial development bank in Greece. Decision Support Systems 12, 151±168.

Tay, F. E.H. and Shen, L. (2002) "Economic and financial prediction using rough sets model Computing", European Journal of Operational Research 141, p 641–659

Vasanthi, S., Subha, M. V. & Nambi, S. Th. (2011) An empirical study on stock index trend prediction using Markov chain analysis, Sri Krishna International Research & Educational Consortium, JBFSIR V. 1, I. 1.

Vermeulen, E.M., Spronk, J., Van der Wijst, N., 1998. The application of the multi-factor model in the analysis of corporate failure. In: Zopounidis, C. (Ed.), Operational Tools in the Management of Financial Risks. Kluwer Academic Publishers, Dordrecht, pp. 59±73.

Wermers, **Russ**, 1999, Mutual Fund trading and impact on stock prices, Journal of finance, 54, 581-622.

Zavgren, C.V., 1983. The prediction of corporate failure: the state of the art. Journal of Financial Literature 2, 1±37.

Zmijewski, M.E., 1984. Methodological issues related to the estimation of financial distress prediction models. Studies on Current Econometric Issues in Accounting Research, pp. 59±82.

Zopounidis, C., 1987. A multicriteria decision making methodology for the evaluation of the risk of failure and an application. Foundations of Control Engineering 12 (1), 45±67.

Zopounidis, C., Dimitras, A.I., Le Rudulier, L., 1995. A multicriteria approach for the analysis and prediction of business failure in Greece. Cahier du LAMSADE, no. 132, Universit_e de Paris Dauphine.