Stock Exchange Market Analysis Using Data Engine and Winrosa

Thomaidis Nikos*, Dounias George*, Tselentis George**.
*Technical University of Crete, Chania, Greece
Department of Production Engineering and Management
Technical University of Crete
University Campus, Kounoupidiana, 73100, Chania, Greece
Phone: +30-821-37323, Fax: +30-821-69410
email: thomaid@dpem.tuc.gr, dounias@dpem.tuc.gr

* MIT – Management Intelligent Technologies GmbH
Promenade 9, D-52076 Aachen Germany
Phone: +49-2408-94580, Fax: +49-2408-94582
email: gt@mitgmbh.de

ABSTRACT: This work attempts to apply advanced data analysis and rule extraction tools and techniques, in order to obtain useful information and possible relations hidden inside files containing daily measures of various international stock exchange market indexes. Such file in fact, consist a common report in daily news worldwide, and is well known that meaningful affections are suspected among the strongest international indexes and other developing or rising stock exchange markets. In the current approach we try to relate the evolution of the Greek Stock Exchange Market Index in comparison with the strongest international markets, i.e. USA, Japan, Germany, England and France. Daily measures were though collected for the above markets concerning a period of three years (1996-1999), while the analysis and interpretation of results was performed with the aid of experts. As the examined domain is strongly chaotic, we do not expect clear and accurate statements on the investor's side, however useful relations should be of additional advise and thus should be taken into account.

KEYWORDS: Stock Exchange Market, Fuzzy Rule Base, Inductive Machine Learning, Information Entropy

APPLICATION DOMAIN: DATA & SINGULARITIES

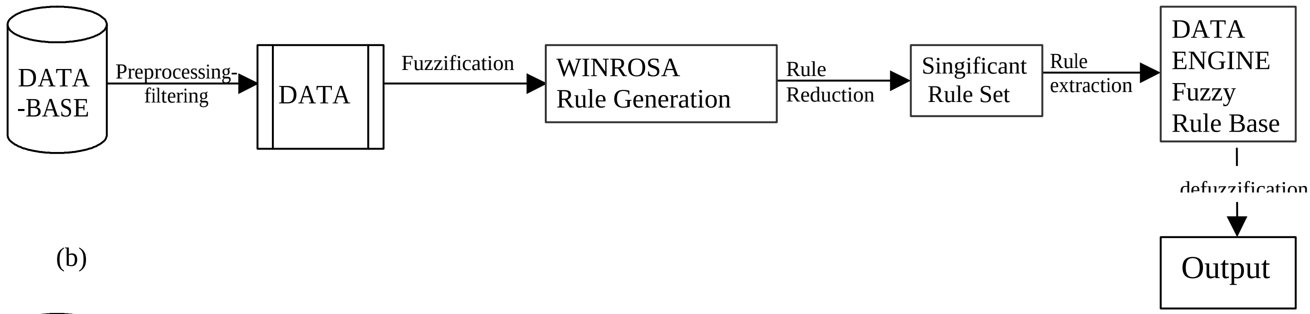
We've examined a somewhat opportune but complicated domain as well, in order to test the efficiency of DataEngine and Winrosa in real world financial decision making. Specifically we collected general stock exchange indexes from the five most powerful economies worldwide, in daily basis for the last three years. We also collected the corresponding Greek Stock Exchange Index values for the above time period, aiming at extracting useful relations among indexes in the form of decision rules. Such information obviously is not an attempt to identify a precise model for describing existing relations and certain affections among indexes of different countries. On the contrary, as the stock exchange domain of application has been proved strongly chaotic in literature [1], the idea behind our analysis, is the extraction of knowledge that we strongly believe it can be of great assistance to stock analysts for the decision making process in financial investments.

The collected data were initially processed in numerical form and accordingly also transformed in linguistic form, with the guidance of related expert staff. A preprocessing strategy has also been adopted including erasing of records corresponding to missing values or dates where no transactions in Greek Stock market took place.

BASIC METHODOLOGICAL STEPS

Below we graphically represent the basic steps of our methodology:

(a)



(b)

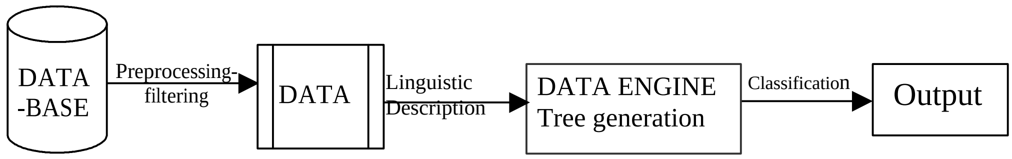


Figure 1: (a)The Fuzzy rule based methodology (b) Entropy Information Based methodology

The first experiment (Figure 1a) concerns the creation of a fuzzy rule base for classifying each case (e.g. each record of the proportional index variations[1]) in a group (cluster). Each group corresponds to a specific index fuzzy set. The methodology is analytically presented in the following steps:

1. *Generation of fuzzy rules*. In this step a set of fuzzy rules is generated with respect to the training file, using the *Fuzzy-R.O.S.A (Rule_Oriented_Statistical_Analysis)* method [3]. The training file consists of six columns, five of which correspond to each international (U.S, English, German, French and Japanese) indexes proportional variation on a specific date. The last column corresponds to the Greek index proportional variation of the following day (one day after). This time delay stimulates the prediction's time interval. This means that each rule of the rule set refers to the prediction of tomorrow's index variation. For the generation of fuzzy rules, the domain is modeled in

a)

b)

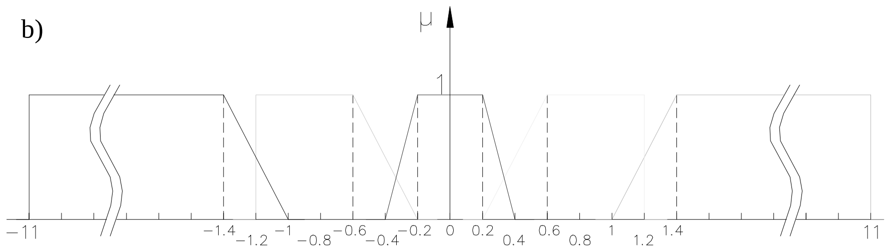


Figure 2: a) Fuzzy sets corresponding to Greek Stock Exchange index b) Fuzzy sets corresponding to other international indexes

[1] From now on whenever we mention "variation" we mean "daily variation"

GREEK INDEX FUZZY SETS					**INTERNATIONAL INDEXES FUZZY SETS**				
Linguistic description	**Characteristic values for the description of trapezoid membership function**				**Linguistic description**	**Characteristic values for the description of trapezoid membership function**			
Steady	-1	-0.5	0.5	1	Steady	-0.4	-0.2	0.2	0.4
Slight positive variation	0.5	1.5	3	3	Slight positive variation	0.2	0.6	1.2	1.2
Slight negative variation	-3	-3	-1.5	-0.5	Slight negative variation	-1.2	-1.2	-0.6	-0.2
Great negative variation	-11	-11	-3.5	-2.5	Great negative variation	-11	-11	-1.4	-1
Great positive variation	2.5	3.5	11	11	Great positive variation	1	1.4	11	11

Table 1: Discriminating values among different fuzzy sets

trapezoid fuzzy sets in agreement with expert's opinion (see Figure 2, Table 1). The reader should have noticed that Greek index fuzzy sets are in general "broader" than the others are. This is to model the observation that Greek Stock Market Index shows greater variations from day to day than the other international Stock Market Indexes, whose performance is more stable, characterized by slight daily deviations. According to experts this observation is quite meaningful, concerning the developing and growing character of Greek economy.

For the purpose of rule space exploration a lot of alternatives are available by the WINROSA [3]:Complete search, Evolutionary search, "Monte Carlo" search e.t.a. Each rule is also characterized by a probabilistic "degree of coverage" with respect to the data set, which results to be the rating index of each rule. In our occasion, *normalized hit rate* is selected as a measure of rule's coverage [3].

2. *Rule reduction.* Once the rules have been produced, tree rule reduction strategies are adopted: Logical Rule Reduction, Reduction of Negative Rules and Reduction based on rule's success rate [3]. Reduction strategies result finally in a significant rule set.

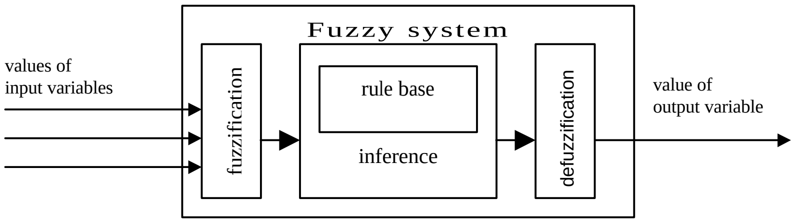


Figure 3**:** The fuzzy system

3. *Rule extraction.* The significant rule set is extracted in a fuzzy rule base [2]. The function of a fuzzy rule base, or a fuzzy system in general, is depicted in Figure 3. In our case the input values (e.g. international indexes proportional variation on a specific date) are fed into the system. The input values "participate" in some degree in the fuzzy sets of the variables of the rules' premise. Hereby, a specific rule set is fired and the *inference* process is completed by the classification of our example in each cluster using a degree of membership. An expert, however, can not easily use in our case the inference process in order to reach a conclusion about the Greek index variation.
4. Hereby, the *defuzzification* process is proposed. It consists of the estimation of the *center of gravity* [3] (see equation 1) for each fuzzy set and the weighted cumulating of centers of gravity, as described in the equation 2 :

$$(1)\quad a_u = \frac{\int_a a m_u(a)da}{\int_a m_u(a)da}$$

$$(2)\quad x_s = \frac{\sum_{u\in U} m_{su} a_u}{\sum_{u\in U} m_{su}}$$

where ì$_{su}$ represents the degree of membership of the example *s* to the cluster (fuzzy set) *u* , ì$_u$(á) represents the degree of membership of value *á* in fuzzy set *u* and á$_u$ represents the center of gravity of fuzzy set *u*.

The weighted cumulating of centers of gravity is such that should not be confused with the trivial mean value. Although both of them are a kind of representative values, the weights that the are used in our case are of different nature. Also, the summation of these weights with respect to the set of clusters is not necessarily equal to one. The estimated value of x_s corresponds to the predicted proportional index variation, e.g. the output value. It is obvious that several outputs can be also obtained by our system such as linguistic values (Steady, Great positive variation e.t.a.), absolute index values and so on.

In the second experiment (Figure 1b), an alternative classification method is applied, using tree induction [4]. The objective is the creation of an inference tree, which can classify each case in a specific cluster. In fact, each case in represented by means of linguistic terms such as Steady, Slight positive variation, Slight negative variation, as concerns the variations of the market indexes (Table 2).

INTERNATIONAL INDEXES CRISP SETS			**GREEK INDEX CRISP SETS**		
Linguistic description	**Low and upper limit**		**Linguistic description**	**Low and upper limit**	
Steady	-0.4	0.4	Steady	-1	1
Slight positive variation	0.4	1.4	Slight positive variation	1	3
Slight negative variation	-1.4	-0.4	Slight negative variation	-3	-1
Great negative variation	-11	-1.4	Great negative variation	-11	-3
Great positive variation	1.4	11	Great positive variation	3	11

Table 2: Discriminating values among different crisp sets for the linguistic transformation phase of the data

The main difference in comparison with the methodology described above is that the classification as well as the description of the domain is totally “crisp”. This means that:

1. Each case can be classified in one and only one cluster (Greek index will be either steady or slight increasing, but not both of them at the same time)
2. The values that stand between low and upper limit have all degree of membership equal to one and the values that are out of set’s limits have all degree of membership equal to zero.

The methodology is analytically presented as follows:

1. *Tree generation:*. In this step, an induction tree is generated using a training file, characterized by linguistic descriptions and time delay (see above). Alternative entropy information criteria are used for the formation of the most representative and accurate tree.
2. *Inference*: Cases are fed into the system. Then, the system forms a decision, according to the cluster in which the case should be classified to.

PRESENTATION AND VALIDATION OF RESULTS

Below we present indicative results of our experiments based on the analysis of specific software tools (Plug-Ins of Data Engine and Winrosa) to our data domain, concerning daily stock exchange index variations worldwide. Initially, in Table 3, we present a sample of seven rules, concerned with the Fuzzy Rule Based Methodology. The referred rule set contains 568 rules in total with rating index varying from 0.2 to 1. The outcomes refer to the affection or not, of the most important international Stock Exchange Market Index variations, to the corresponding Greek General Index, calculated one day later. Briefly we can approximately draw the following prime conclusions on these rules:

1. Greece is positively affected by the increase of the most important Stock Exchange Markets, i.e. England, USA, France and Japan (Rule-1, Rule-2, Rule-3)
2. Slight mixed mode (positive or negative) variations of the most important Stock Exchange Markets do not really affect the Greek Stock Market (Rule-4)
3. Negative variations of the British Stock Market itself (Rule-5) as well as of most of the main international markets (Rule-7), seem to have an obvious equivalent affection to the Greek Stock Exchange Market, while positive variations of the German Stock Market cause similar tends to Greece (Rule-6)

.

RULE-1: if LONDON = Steady *and* NEW YORK = Steady *and* TOKYO = Steady *and* PARIS = *Great positive variation then* GREECE = *Slight positive variation with certainty 1*
RULE-2: if LONDON = *Steady* and NEW YORK = *Great positive variation* and TOKYO = *Slight positive variation* then GREECE = *Slight positive variation* with certainty 1
RULE-3: if LONDON = *Steady* and NEW YORK = *Great positive variation* and FRANKFURT = *Slight positive variation* then GREECE = *Slight positive variation* with certainty 1
RULE-4: if LONDON = *Slight positive variation* and TOKYO = *Slight positive variation* and FRANKFURT = *Slight negative variation* and PARIS = *Slight positive variation* then GREECE = *Steady* with certainty 1
RULE-5: if LONDON = *Great negative variation* and TOKYO = *Steady* then GREECE = *Great negative variation* with certainty 1
RULE-6: if TOKYO = *Steady* and FRANKFURT = *Great positive variation* and PARIS = *Steady* then GREECE= *Great positive variation* with certainty 1
RULE-7: if LONDON = *Slight negative variation* and NEW YORK = *Slight negative variation* and TOKYO = *Steady* and FRANKFURT = *Great negative variation* and PARIS = *Slight negative variation* then GREECE = *Great negative variation* with certainty 0.9996

Table 3*:* Rules showing certainty degree concerning variations between Greek and main International Stock Exchange Market Indexes.

Figure 3 graphically represents the success of our rule base to predict real daily variations of the Greek Stock Exchange Market Index, based on previous day's information on the most important international indexes. In fact our methodology "follows" the actual behavior, which means that our system is easily adapted to real conditions, but faces difficulties when training cases are not recent, or when variations are full of singularities (i.e. ups and downs, sudden changes, etc.). In such a case, a moving window method to select and supply recent training data to our system could be of great assistance.

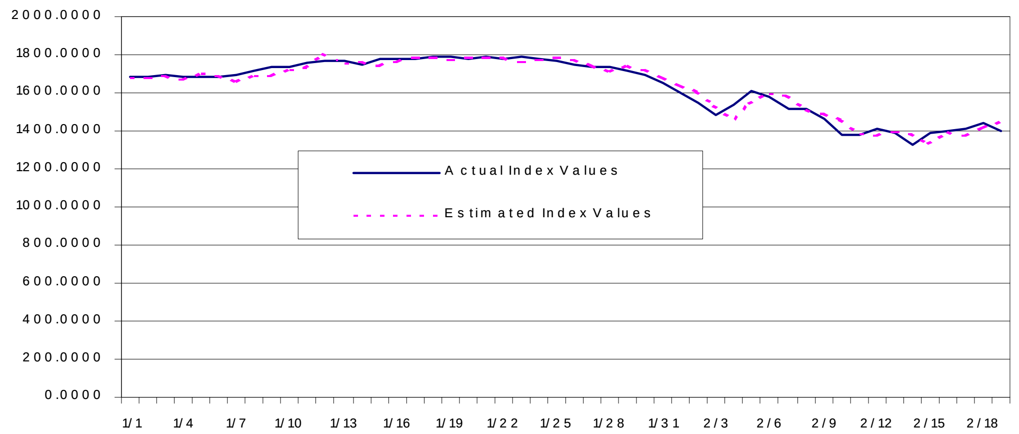


Figure 4: Prediction of Greek Stock Market Index variation within the period 1/1-2/15 1999

The additional rule set sample that follows below was produced by the application of various entropy information based criteria (see Data Engine Tree Plug-In). The finally selected as best decision tree consists of 246 nodes and 441 branches expanded in 6 levels and forming finally 384 decision rules. We see that similar to the previous results can be extracted using the inductive learning approach. We summarize them below:

1. Stability in Greece is in accordance to previously observed market stability internationally (Rule-8).
2. Greece is not affected significantly by small mixed variations of the major international markets (Rule-9, Rule-15).

3. A positive variation of French, English and Japanese Stock Market can affect Greece in the same direction (Rule-11, Rule-12)
4. Negative variation of English Stock Market can cause an opposite effect to the Greek Stock Exchange Market (Rule-13)
5. German Stock Market affects Greece more seriously in negative variations of the Stock Exchange Index (Rule-16)
6. When most international stock markets tend to increase uniformly, Greece is tracking them (Rule-17)

RULE-8: If London is Steady AND Paris is Steady AND Frankfurt is Steady AND Tokyo is Steady and NewYork is Steady THEN Greece is Steady
RULE-9: If London is Slight_Positive_Variation AND Paris is Steady AND Frankfurt is Slight_Negative_Variation AND Tokyo is Great_Positive_Variation and NewYork is Steady THEN Greece is Steady
RULE-10: If London is Slight_Positive_Variation AND Paris is Steady AND Frankfurt is Great_Positive_Variation AND Tokyo is Great_Positive_Variation and NewYork is Steady THEN Greece is Steady
RULE-11: If London is Steady AND Paris is Slight_Positive_Variation AND Frankfurt is Steady AND Tokyo is Great_Positive_Variation and NewYork is Steady THEN Greece is Slight_Positive_Variation
RULE-12: If London is Steady AND Paris is Slight_Positive_Variation AND Frankfurt is Great_Positive_Variation AND Tokyo is Great_Positive_Variation and NewYork is Steady THEN Greece is Slight_Positive_Variation
RULE-13: If London is Great_Negative_Variation THEN Greece is Slight_Positive_Variation
RULE-14: If London is Slight_Negative_Variation AND NewYork is Great_Positive_Variation THEN Greece is Great_Positive_Variation
RULE-15: If London is Slight_Positive_Variation AND Paris is Great_Negative_Variation AND Tokyo is Great_Negative_Variation THEN Greece is Steady
RULE-16: If London is Great_Positive_Variation AND Paris is Great_Positive_Variation AND Frankfurt is Great_Negative_Variation THEN Greece is Great_Negative_Variation
RULE-17: If London is Slight_Positive_Variation AND Paris is Great_Positive_Variation AND Frankfurt is Great_Positive_Variation AND Tokyo is Steady AND NewYork is Great_Positive_Variation THEN Greece is Great_Positive_Variation

Table 4: Inductive Machine Learning Based rules showing relations concerning variation between Greek and the main International Stock Exchange Market Indexes.

One simple way of further validating results is to invent a hypothetical *investment game*. Let’s assume that the investor owns an initial capital of 1000 DRS and is bounded to put his money only in the stocks that are counted for the estimation of the Greek Stock Market Index. Apparently, the investor’s profit is highly correlated with the variations of index values. A great number of strategies exist consisting of investment decisions. There are two alternatives:

1. Investment of the capital
2. Withdrawal of the capital

assuming that each transaction is decided on the remaining capital in daily basis. Supposing we consider four alternatives strategies: Random play, “Dummy” play, Prediction based play and Knowledge based play. Random play means that the investor makes his decision totally by chance (e.g. by flipping a coin). “Dummy” play is equal to the absence of investment strategy. Capital remains bounded in the stock market. Knowledge based play is the most sophisticated strategy as the index value is assumed to be known a priori. It’s clear that a strategy like this is optimal but practically unfeasible at the same time. Nevertheless, it is an objective measure, which can be compared with any alternative strategy with the purpose of estimate its effectiveness. Finally, Prediction based play is the investment strategy indicated by the prediction of index value using the above-mentioned methodology. An objectively profitable strategy is the one that results in percentage of profit similar to that achieved by adopting knowledge based strategy and far from that achieved by adopting the other two strategies.

The results obtained above are of course just indicating what the best decision may be. In other words they are not a definite best solution to investment decisions for analysts, but they can be used as a guide for forming a way of inference on such complex and continuously varying domains. Further assistance can be obtained when experts have examined and expressed their opinion on the above rules in terms of classification accuracy, complexity, correctness, completeness, consistency, etc. Currently our work is in this phase in order to see how the outcomes are interpreted by decision-makers, which consist the candidate set that should use such an analysis. This is a very hard job to do indeed, and is always performed by the knowledge engineer involved in such knowledge-based solutions.

FURTHER RESEARCH

Additional experiments are now performed, using and comparing performance of alternative entropy information criteria, building and testing more accurate fuzzy membership functions, but also attempting different initial time modeling of our overall approach (e.g. weekly variation predictions, season’s behavior e.t.a.).Moreover, as mentioned above, the use of a moving window technique on the learning (training) phase of our approach could prove a very effective way of predicting accurately close variations and affections among international and the Greek Stock Exchange Market Indexes.

ACKNOWLEDGMENTS

The “HNIOXOS A.E.L.B.E.” Company, Chania, Greece is greatly acknowledged for providing us with the Stock Exchange Index Data needed for performing the experiments described in this paper, as well as for validating results and providing us with helpful expert advising. MIT GmbH is also greatly acknowledged for supplying the authors with the software needed for performing the domain analysis.

REFERENCES

1. Lux, Marchesi, 1999 “Scaling and Criticality in a Stochastic Multi-Agent Model of a Financial Market”, *Nature* 397, p.498-500,
2. MIT GmbH, 1997, ‘DataEngine: Overview and User Manual’, Aachen, Germany.
3. MIT GmbH, 1997, ‘WINROSA: Handbook», Aachen, Germany’.
4. Quinlan, 1993,“C4.5: Programs for Machine Learning”, Morgan Kaufmann Publishers, San Mateo, California.