INDUCTIVE LOGIC PROGRAMMING FOR DISCOVERING FINANCIAL REGULARITIES

Boris Kovalerchuk
Department of Computer Science, Central Washington University, Ellensburg, WA 98926-7520, USA E-mail: borisk@tahoma.cwu.edu
Phone: (509) 963-1438, Fax: (509) 963-1449

Evgenii Vityaev
Institute of Mathematics, Russian Academy of Science, Novosibirsk 630090, Russia
E-mail: vityaev@math.nsc.ru Phone: 7 (3832) 35-44-62

Version August 28, 1998

Abstract

The purpose of this work is discovering regularities in financial time series using Inductive Logic Programming (ILP) and related "Discovery" software system [Vityaev et al., 1992,1993] in data mining. Discovered regularities were used for forecasting the target variable, representing the relative difference in percent between today's closing price and the price five days ahead. We describe the method, types of regularities found and analyzed, statistical characteristics of these regularities on the training and test data and the percentage of true and false predictions on the test data. There are more than 130 discovered regularities on 10 year (1985-1994) data. The best of these regularities had shown about 75 % of correct forecasts on test data (1995-1996). The target variable was predicted using separately SP500 (close) and own history of the target variable. Active trading strategy based on discovered rules outperformed buy-and-hold strategy and strategies based on several ARIMA models in simulated trading for 1995-1996. An ARIMA model constructed using discovered rules had shown the best performance among tested ARIMA models. The performance of this model is similar to performance based on discovered rules.

INDUCTIVE LOGIC PROGRAMMING FOR DISCOVERING FINANCIAL REGULARITIES

Boris Kovalerchuk
Department of Computer Science, Central Washington University, Ellensburg, WA, 98926-7520, USA

Evgenii Vityaev
Institute of Mathematics, Russian Academy of Science, Novosibirsk 630090, Russia

Version August 28, 1998

1.Introduction

A machine learning type of method, called Machine Methods for Discovering Regularities (MMDR) is applied for forecasting financial time series in this paper. The method expresses patterns in first order logic and assigns probabilities to rules generated by composing patterns. Currently the majority of learning systems for financial applications concentrate on neural networks, genetic algorithms, and related techniques. In practice, learning systems based on first-order representations have been successfully applied to many problems in chemistry, physics, medicine and other fields [Mitchell, 1997, Kovalerchuk et al, 1997]. Often these methods are called **Inductive Logic Programming** (ILP) methods [Mitchell, 1997].

How can we motivate the choice for MMDR in financial applications? As any technique based on first order logic, this technique allows one to get **human-readable forecasting rules** [Mitchell, 1997, chapter 10], i.e. **interpretable** in ordinary financial language in addition to the forecast. A financial specialist can evaluate the performance of the forecast as well as a forecasting rule. We present examples of human-readable forecasting rules in section 2.2.

Also, as any technique based on probabilistic estimates, this technique delivers rules tested on their **statistical significance.** Statistically significant rules have advantage in comparison with rules tested only for their performance on training and test data [Mitchell, 1997, chapter 5]. Training and testing data can be too limited and/or not representative. If rules rely only on them then there are more chances that these rules will not deliver a right forecast on other data.

What is the motivation to use MMDR in particular? MMDR uses hypothesis/rule generation and selection process, based on fundamental representative measurement theory [Krantz, Luce, Suppes and Tversky, 1971, 1989, 1990.] The original challenge for MMDR was the simulation of discovering **scientific laws** from empirical data in chemistry and physics. There is a well-know difference between "black box" models and fundamental models (laws) in modern physics. The last ones have much longer life, wider scope and a solid background. We have a reason to believe that MMDR caught some important features of discovering these regularities ("laws"). This is an area of extensive computer science research during the last 25 years [Pattern Recognition and Artificial Intelligence, 1989, Zagoruiko, 1979]. MMDR ideas and the "Discovery" system have been successfully used in several Russian and international research projects, e.g., [Kovalerchuk et al, 1997]. There are reasons to believe that usage of MMDR for financial forecast can be beneficial in financial area as well as for further advances of MMDR.

In the paper we study several types of hypotheses/rules presented in first-order logic. They are simple relational assertions with variables (section 2.2). Mitchell [1997] noted the importance that relational assertions "can be **conveniently expressed** using first-order representations, while they are **very difficult** to describe using propositional representations" (pp.275, 283-284). Many well-known rule learners such as AQ, CN2 are propositional [Mitchell, 1997, p.279, 283]. Note that decision tree methods represent a particular type of propositional representation [Mitchell, 1997, p.275]. Therefore decision tree methods as ID3 and its successor C4.5 [Quinlan, 1993] fit better to tasks without relational assertions. Mitchell argues and gives examples that propositional representations

offer no general way to describe the essential *relations* among the values of the attributes (pp. 283-284). Below we follow his example. In contrast with propositional rules, a program using **first-order representations** could learn the following general rule: IF *Father(x,y)* & *Female(y)*, THEN *Daugher(x,y),* where x and y are variables that can be bound to any person. For the target concept $Daughter_{1,2}$ **propositional rule learner** such as CN2 or C4.5, the result would be a collection of very specific rules such as

IF ($Father_1=Bob$)&($Name_2=Bob$)&($Female_1=True$) THEN $Daughter_{1,2}=True$.

Although it is correct, this rule is so specific that it will rarely, if ever, be useful in classifying future pairs of people [Mitchell, 1977, pp.283-284]. In section 5 we show that the close problem exists for ARIMA and Neural Networks methods. First-order logic rules have an advantage in discovering relational assertions because they capture relations directly, e.g., *Father(x,y)* in the example above.

Also, first order rules allow one to express naturally other more general hypotheses not only the relation between pairs of attributes [Krantz et all, 1971, 1989, 1990]. These more general rules can be as for classification problems as for an interval forecast of continuous variable. Moreover these rules are able to catch Markov chain type of models used for financial time series forecast. In section 2.2 we describe first-order rules with relational assertions which we used. We share Mitchell's opinion about the importance of algorithms designed to learn sets of first-order rules that contain variables. "This is significant because first-order rules are much more expressive than propositional rules" [Mitchell, 1997, p.274].

What is the difference of MMDR from other Machine Learning methods dealing with first-order logic [Mitchell, 1997, Russel and Norvig, 1995]? From our viewpoint the main accent in western first-order methods [Mitchell, 1997, Russel and Norvig, 1995] is on two computational complexity issues: how wide is the class of hypotheses tested by the particular machine learning algorithms and how to construct a learning algorithm to find deterministic rules. The emphasis of MMDR is on probabilistic first-order rules and measurement issues, i.e., how we can move from a real measurement to first-order logic representation. This is a non-trivial task [Krantz et all, 1971, 1989, 1990]. For example, how to represent temperature measurement in terms of first-order logic without losing the essence of the attribute (temperature in this case) and without inputting unnecessary conventional properties? For instance, Fahrenheit and Celsius zeros of temperature are our conventions in contrast with Kelvin scale where the zero is a real physical zero. There are no temperatures less than this zero. Therefore incorporating properties of the Fahrenheit zero into first-order rules may force us to discover/learn properties of this convention along with more significant scale invariant forecasting rules. Learning algorithms in the space with those kind of accidental properties may be very time consuming and may produce inappropriate rules. This study is closely related to "learning from hints", which incorporates any invariance hint about unknown function *f* [Abu-Mostafa, 1989] in contrast to learning only from examples. "A hint may take the form of a global constraint on *f*, such as symmetry property or an invariance. It may also be partial information about the implementation of *f*." [Abu-Mostafa, 1989]. Our main source for hints in first-order logic rules is [Krantz et all, 1971, 1989, 1990].

It is well know that the general problem of rule generating and testing is NP-complete. Therefore the discussion above is closely related to the following questions. What determines the number of rules and when to stop generating rules? What is the justification for specifying particular expressions instead of any other expressions? Using [Krantz et all, 1971, 1989, 1990] approach we select rules which are simplest and consistent with measurement scales for a particular task. Section 2.2 presents a set of rules applicable to stock market. The algorithm stops generating new rules when they become too complex (i.e., statistically insignificant for the data) in spite of possible high accuracy on training data. The obvious other stop criterion is time limitation. Detailed discussion about a mechanism of initial rule selection from measurement theory [Krantz et all, 1971, 1989, 1990] viewpoint is out of the scope of this paper. A special study may result in a catalogue of initial rules/hypotheses to be tested (learned) for stock market forecast. In this way any financial analyst can choose rules to be tested without generating them. This paper delivers a preliminary list of rules for that catalogue (see section 2.2).

The critical issue in applying data-driven forecasting systems is generalization. The "Discovery" system generalizes data through "lawlike" logical probabilistic rules. Discovered rules have similar statistical estimate and significance on training and test sets of studied time series. Theoretical advantages of MMDR generalization are presented in [Vityaev,1992] and [Kovalerchuk et al,1996]. This approach has some similarity with mentioned hint approach [Abu-Mostafa, 1990]. We use mathematical formalisms of first order logic rules described in [Russel and Norvig, 1995], [Halpern, 1990] and [Krantz, Luce, Suppes and Tversky, 1971, 1989, 1990].

The relative difference in percent between today's closing price and the price five days ahead was used as a target variable. We also used Standard and Poor 500 close (SP500C), Dow-Jones Industrial Average (DJIA), and

additional generated features:

(i) weekday (indication of a particular day of the week, i.e., Monday, Tuesday, Wednesday, Thursday, Friday) for each value of studied variables and

(ii) the first and second differences for variables (prices and SP500C and DJIA indexes) for various weekdays, which are similar to their first and second derivatives.

All this information is expressed using the first order logic and probability theory through logical probabilistic laws.

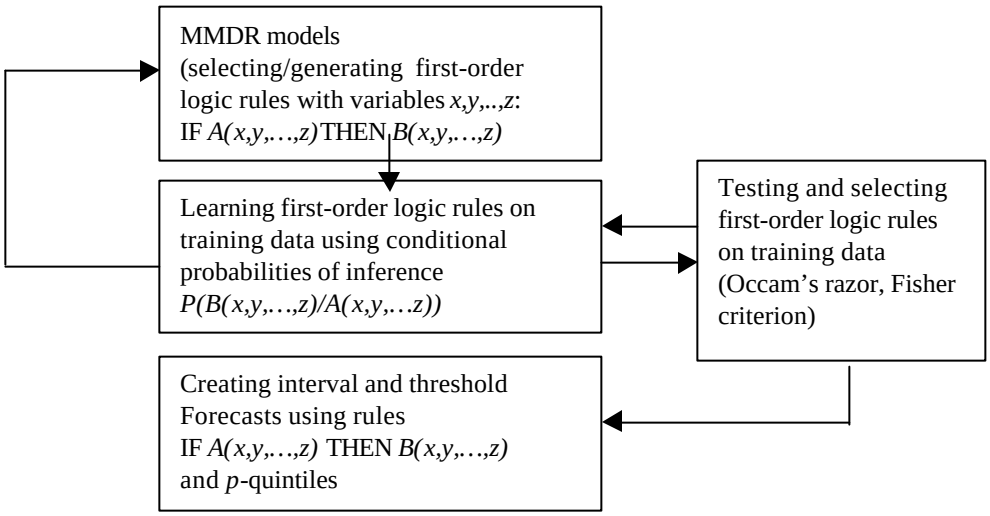


Figure 1. Flow diagram for MMDR: steps and technique applied.

Figure 1 describes the steps of MMDR. On the first step we select and/or generate a class first–order logic rules suitable for a particular task (see section 2.1). The next step is learning the particular first-order logic rules using available training data. Then we test first-order logic rules on training using Fisher statistical criterion. After that we select statistically significant rules and apply Occam's razor principle: prefer the simplest hypothesis (rules) that fits the data [Mitchell, 1977, p. 65]. Simultaneously we use the rules' performance on training for their selection. We may iterate back and forth among these three steps several times to find the best rules. The last step is creating interval and threshold forecasts using selected first-order logic rules: IF *A(x,y,…,z)* THEN *B(x,y,…,z)*.

The paper has the following structure: introduction, method, forecasting results, simulated trading performance, comparison with ARIMA, and concluding remarks.

2. Method

2.1.Variables. Two time series TR (training set) and CT (control/test set) of the target variable were used to train and evaluate a forecasting algorithm, where TR = $\{a_1, \dots, a_{tr}\}$ is ten year data (1985-1994, tr =2528 days) and CT = $\{a_1, \dots, a_{ct}\}$ is two year data (1995-1996, ct = 506 days).

Five sequential days are used as the main forecast unit (an **object**):

$$\mathbf{a_t} = (\mathbf{a^1_t},\mathbf{a^2_t},\mathbf{a^3_t},\mathbf{a^4_t},\mathbf{a^5_t}),$$

where $\mathbf{a^j_t}$ is j-th day of five-day object $\mathbf{a_t}$. We will also use another notation: $\mathbf{a_t}=(a_t,a_{t+1},a_{t+2},a_{t+3},a_{t+4})$, with the following correspondence between notations: $a_{(t-1)+j} = \mathbf{a^j_t}$ for all five days of $\mathbf{a_t}$ (j=1,...,5). A variable Weekday(a_t) has five values:1,2,3,4 and 5 for weekdays, where Weekday(a_t)=1 means Monday and Weekday(a_t)=5 means Friday. Example. Let a_t ="March 3, 1998", Weekday(a_t)=2. Tuesday's code is 2. We do not consider Saturdays and Sundays in this study, because the stock market is closed these days.

Several new variables were generated from SP500C:

(2.1.1) **relative differences** $)_{ij}(\mathbf{a_t}) = (SP500C(\mathbf{a^j_t}) - SP500C(\mathbf{a^i_t}))/SP500C(\mathbf{a^i_t})$, i<j, i,j= 1,...,5;

Example. Let i=1, j=2, t="March 3, 1998" then

$\mathbf{a_t}$ = <March 3, 1998; March 4, 1998; March 5, 1998; March 6, 1998;March 9, 1998> , where $\mathbf{a^1_t} = a_t$ = "March 3, 1998", $\mathbf{a^2_t}$

$=a_{t+1}$ = "March 4, 1998", $\mathbf{a}^3_t = a_{t+2}$ = "March 5, 1998",
$\mathbf{a}^4_t = a_{t+3}$ = "March 6, 1998", $\mathbf{a}^5_t = a_{t+4}$ = "March 9, 1998". Therefore, $\Delta_{12}(\mathbf{a}_t)$ = (SP500C($\mathbf{a}^2_t$) - SP500C($\mathbf{a}^1_t$))/SP500C($\mathbf{a}^1_t$) =(SP500C(March 4, 1998) - SP500C(March 3, 1998))/SP500C(March 3, 1998).

(2.1.2) **differences** $\Delta_{ijk}(\mathbf{a}_t) = \Delta_{jk}(\mathbf{a}_t) - \Delta_{ij}(\mathbf{a}_t)$ **between two relative differences**.

Example. Let k=3, then $\Delta_{ijk}(\mathbf{a}_t) = \Delta_{jk}(\mathbf{a}_t) - \Delta_{ij}(\mathbf{a}_t)$ can be written as

$\Delta_{123}(\mathbf{a}_t)$ = (SP500C(March 5, 1998) - SP500C(March 4, 1998))/SP500C(March 4, 1998) -

-(SP500C(March 4, 1998) - SP500C(March 3, 1998))/SP500C(March 3, 1998).

(2.1.3) **cyclic permutation** B (cycle of length 5) for object **a** and function wd(**a**). wd(**a**) = <1,2,3,4,5> means that **a** represents normal five weekdays from Monday to Friday, but wd(**b**) = < d_1,...,d_5> =<2,3,4,5,1> means the five days from Tuesday to the next Monday, i.e., <Tue, Wed, Thu, Fri, Mon>. Let B(Mon, Tue, Wed, Thu, Fri)=(Tue, Wed, Thu, Fri, Mon). Thus B is a cyclic permutation, which changes the set of five days under consideration for rule extraction from **a** to **b:** < d_1,...,d_5> = B(<Mon, Tus, Wen, Thu, Fri >).

Formally wd(**b**)=< d_1,...,d_5> is equivalent to (Weekday($\mathbf{b}^1$)= d_1) &...& (Weekday($\mathbf{b}^5$)=d_5), and

< d_1,...,d_5> =B(Mon, Tue, Wed, Thu, Fri), where B is a cyclic permutation of weekdays. The Weekday function is defined in this section above.

Note that it is possible to generate hundreds of similar variables in addition to (2.1.1)-(2.1.3). In the time frame of current study we considered only (2.1.1)-(2.1.3) for SP500C and their analogues for the target and DJIA. The first two features (2.1.1) and (2.1.2) catch properties similar to the first and second derivatives of the original time series. Analysis of more variables requires much more runtime. Therefore current results show mostly applicability of the method and its capability as a knowledge acquisition tool for financial time series.

2.2. Hypotheses and probabilistic "laws". Let's introduce notation: $\mathbf{a} = \mathbf{a}_{t1}$, $\mathbf{b} = \mathbf{a}_{t2}$ are arbitrary objects; $(\Delta(\mathbf{a}) \# \Delta(\mathbf{b}))^{\gamma}$ is **any of inequalities** as $(\Delta_{ij}(\mathbf{a}) \# \Delta_{ij}(\mathbf{b}))^{\gamma}$, i < j; i,j = 1,...,5; or $(\Delta_{ijk}(\mathbf{a}) \# \Delta_{ijk}(\mathbf{b}))$, i < j < k; i,j,k = 1,..,5; $\gamma,\gamma 0,\gamma 1,\gamma 3 \in \{0,1\}$, (0) means that there is no negation for the expression $\Delta_{ij}(\mathbf{a}) \# \Delta_{ij}(\mathbf{b})$ and (1) means negation of this expression.

To find probabilistic regularities ("laws") we tested the following hypotheses 2.2.1-2.2.4:

2.2.1. $(wd(\mathbf{a}) = wd(\mathbf{b}) = <d_1,...,d_5>) \& (\Delta(\mathbf{a}) \# \Delta(\mathbf{b}))^{\gamma 1} \to ((target(\mathbf{a}^5) \# target(\mathbf{b}^5))^{\gamma 0}$;

Example: Let **a** = <March 3, 1998; March 4, 1998; March 5, 1998; March 6, 1998;March 9, 1998>, **b** = <March 10, 1998; March 11, 1998; March 12, 1998; March 13, 1998; March 16, 1998>, $\Delta(\mathbf{a}) = \Delta_{12}(\mathbf{a}_t)$, $\Delta(\mathbf{b}) = \Delta_{12}(\mathbf{b}_t)$, $\gamma 1 = 0$ (no negation of relation $\Delta(\mathbf{a}) \# \Delta(\mathbf{b})$), $\gamma 0=1$(negation of relation (target($\mathbf{a}^5$) # target($\mathbf{b}^5$)). The last one means that we test the rule with relation (target($\mathbf{a}^5$) > target($\mathbf{b}^5$)). Let also < d_1,...,d_5>=< Tue,Wed, Thu, Fri, Mon >. Therefore, the tested rule/hypothesis is ((wd(March 3, 1998; March 4, 1998; March 5, 1998; March 6, 1998; March 9, 1998) = wd(March 10, 1998; March 11, 1998; March 12, 1998; March 13, 1998; March 16, 1998) = < Tue,Wed, Thu, Fri, Mon >) &($\Delta(\mathbf{a}) \# \Delta(\mathbf{b})$) → (target($\mathbf{a}^5$) > target($\mathbf{b}^5$)). This means that we test all five-day objects, which begin with Tuesday. The tested statement is: "IF for any five-day objects **a** and **b** the SP500C difference $\Delta_{12}(\mathbf{a}_t)$ is smaller than $\Delta_{12}(\mathbf{b}_t)$ THEN the target stock for the last day of **a** will be greater than for the last day of **b".**

2.2.2. $(wd(\mathbf{a}) = wd(\mathbf{b}) = < d_1,...,d_5>) \& (\Delta(\mathbf{a}) \# \Delta(\mathbf{b}))^{\gamma 1} \& (\Delta(\mathbf{a}) \# \Delta(\mathbf{b}))^{\gamma 2} \to (target(\mathbf{a}^5) \# target(\mathbf{b}^5))^{\gamma 0}$;

These hypotheses have similar interpretation. The only difference from 2.2.1 is that now we consider two differences in the rules. For example one of the tested statements is: "IF for any five-day objects **a** and **b** the SP500C difference $\Delta_{12}(\mathbf{a}_t)$ is smaller than $\Delta_{12}(\mathbf{b}_t)$ AND the SP500C difference $\Delta_{23}(\mathbf{a}_t)$ is greater than $\Delta_{23}(\mathbf{b}_t)$ THEN the target stock for the last day of **a** will be greater than for the last day of **b".**

2.2.3. $(wd(\mathbf{a}) = wd(\mathbf{b}) = < d_1,...,d_5>) \& (\Delta(\mathbf{a}) \# \Delta(\mathbf{b}))^{\gamma 1} \& (\Delta(\mathbf{a}) \# \Delta(\mathbf{b}))^{\gamma 2} \&$

$(\Delta(\mathbf{a}) \# \Delta(\mathbf{b}))^{\gamma 3} \to (target(\mathbf{a}^5) \# target(\mathbf{b}^5))^{\gamma 0}$;

These hypotheses have similar interpretation. The only difference from 2.2.2 is that now we consider three differences in the rules. For example one of the tested statements is: "IF for any five-day objects **a** and **b** the SP500C difference $\Delta_{12}(\mathbf{a}_t)$ is smaller than $\Delta_{12}(\mathbf{b}_t)$ AND the SP500C difference $\Delta_{23}(\mathbf{a}_t)$ is greater than $\Delta_{23}(\mathbf{b}_t)$ AND the SP500C difference $\Delta_{123}(\mathbf{a}_t)$ is greater than $\Delta_{123}(\mathbf{b}_t)$ THEN the target stock for the last day of **a** will be greater then for the last day of **b".**

2.2.4. $(wd(\mathbf{a}) = wd(\mathbf{b}) =< d_1,...,d_5>) \& (\Delta(\mathbf{a}) \# \Delta(\mathbf{b}))^{\gamma 1} \& ... \& (\Delta(\mathbf{a}) \# \Delta(\mathbf{b}))^{\gamma k} \to (target(\mathbf{a}^5) \# target(\mathbf{b}^5))^{\gamma 0}$.

These hypotheses allow us generate more than three relations including $\Delta_{ijk}(\mathbf{a}_t)$. For example, one of the tested statements is: "IF for any five-day objects **a** and **b** the SP500C difference $\Delta_{12}(\mathbf{a}_t)$ is smaller than $\Delta_{12}(\mathbf{b}_t)$ AND the SP500C difference $\Delta_{23}(\mathbf{a}_t)$ is greater than $\Delta_{23}(\mathbf{b}_t)$ AND the SP500C difference $\Delta_{123}(\mathbf{a}_t)$ is greater than $\Delta_{123}(\mathbf{b}_t)$ AND …..

AND …. THEN the target stock for the last day of **a** will be greater than for the last day of **b".**

Figure 2, tables 1 and 2 present examples of hypotheses 2.2.1-2.2.4 in usual financial terms.

IF

current 5 days end on Monday and there are some other ("old") five days (from the history of years 1984-1996) that end on Monday too

AND

the relative SP500C difference between Tuesday and Thursday for the old five days is no greater than between Tuesday and Thursday for the current five days

AND

the relative SP500C difference between Tuesday and Monday for the old five days is greater than between Tuesday and Monday for the current five days

AND

the relative difference between SP500C differences for Tuesday, Wednesday and Wednesday, Thursday) for the old five days is no greater than for the pairs of days for the current five days

AND <we omit linguistic description of $(\ )_{245}(\mathbf{a}) > (\ )_{245}(\mathbf{b})$, which is similar to previous one>

THEN

the target value for Monday from the current 5 days should be no greater than target value for the Monday from the old five days, i.e., we forecast that a target stock five days ahead from current Monday will grow no greater than it was five days ahead from the old Monday.

Figure 2. Diagram for rule 2.4.1.

2.3. Transition probabilities. Many well-known prediction methods used for stock market study can be written in terms similar to 2.2.1-2.2.4, e.g., Markov chains and other methods exploiting conditional/transition probabilities. Two simple Markov chains are presented in [Hiller, Lieberman, 1995, p. 632].

Chain 1 has four states with conditional probabilities presented in table 1.

Table 1. Transition matrix for chain 1

	The stock increases today	The stock decreases today
The stock increased yesterday	0.7	0.3
The stock decreased yesterday	0.5	0.5

In particular, table 1 gives a rule:

IF the stock increased yesterday THEN the stock increases today with probability 0.7.

Chain 2 has more states and presented in table 2.

This chain has similar interpretation. For instance, table 2 gives a rule:

IF the stock increases today and decreases yesterday THAN the stock will increase tomorrow with probability 0.6. This rule is a combination of bold cells in table 2. If part is taken from the first column. Similarly the bold cell on the first row is used as a THEN part of the rule. Probability 0.6 can be found in the intersection of respective row and column.

Below we show how this Markov chain type of models can be embedded into first-order logic rules and discovered using this technique. We evaluate expressions 2.2.1-2.2.4 on training data using conditional probabilities. Let us consider six-day objects instead of five-day objects:

$<d_1,...,d_5,d_6>=<$Mon,Tues,Wed, Thur, Fri, Mon$>$, $(wd(\mathbf{a}) = wd(\mathbf{b}) = < d_1,...,d_5,d_6>$, $\mathbf{a}=\mathbf{a_t}$ and $\mathbf{a}^6_t = \mathbf{a}^1_{t+1}=\mathbf{b}^1_t$,i.e., **a** is some six days and **b** is the next six days excluding Saturday and Sunday with overlapping the end of **a** and the beginning of **b.** Also we generate relative difference (2.1.1) for the same target stock price (S): $delta_{ij}(\mathbf{a_t}) = (S(\mathbf{a}^j_t)-S(\mathbf{a}^i_t))/S(\mathbf{a}^i_t)$. This variable is equal to $target(\mathbf{a_t})$ five days back. The target represents five-day forecast in contrast with $delta_{ij}(\mathbf{a_t})$.

Let us suppose that we have found the following conditional probabilities on training set TR sore some fixed days i and j from $\mathbf{a_t}$ and $\mathbf{a_{t+1}}$:

0.31 for *Rule1*: $(delta_{ij}(\mathbf{a_t}) < delta_{ij}(\mathbf{a_{t+1}}) \Rightarrow (target(\mathbf{a}^6_t) < target(\mathbf{a}^6_{t+1}))$

0.69 for *Rule 2*: $(delta_{ij}(\mathbf{a_t}) < delta_{ij}(\mathbf{a_{t+1}}) \Rightarrow NOT(target(\mathbf{a}^5_t) < target(\mathbf{a}^5_{t+1}))$

0.65 for *Rule 3*: NOT (delta$_{ij}$($\mathbf{a}_t$) < delta$_{ij}$($\mathbf{a}_{t+1}$) => (target($\mathbf{a}^5_t$) < target($\mathbf{a}^5_{t+1}$))

0.35 for *Rule 4*: NOT(delta$_{ij}$($\mathbf{a}_t$)<delta$_{ij}$($\mathbf{a}_{t+1}$) => NOT(target($\mathbf{a}^5_t$) < target($\mathbf{a}^5_{t+1}$))

Table 2. Transition matrix for chain 2

	The stock increases both tomorrow and today	The stock increases tomorrow and decreases today	The stock decreases tomorrow and increases today	The stock decreases both tomorrow and today
The stock increased both today and yesterday	0.9	0	0.1	0
The stock increased today and decreased yesterday	**0.6**	0	0.4	0
The stock decreased today and increased yesterday	0	0.5	0	0.5
The stock decreased both today and yesterday	0	0.3	0	0.7

It can be represented with a matrix of transition probabilities, used in Markov chains for forecasting:

		Target	
		0	1
Delta	0	0.31	0.69
	1	0.65	0.35

We use 0 for "up", i.e., if delta$_{ij}$($\mathbf{a}_t$) < delta$_{ij}$($\mathbf{a}_{t+1}$) and 1 for "down", i.e., if delta$_{ij}$($\mathbf{a}_t$) ≥ delta$_{ij}$($\mathbf{a}_{t+1}$). The same notation is used for the target. Rule 2 can be described using usual financial language: "If delta goes UP then target goes DOWN with probability 0.69". Several of these expressions were used to study a forecast horizon for consecutive days and weeks by changing < d$_1$,...,d$_k$> and i,j day, where < d$_1$,...,d$_k$> was extended from 5 days up to 12 weeks.

For each probabilistic law C=(A$_1$(x,y,…z)&…&A$_k$(x,y,…z)Y A$_0$(x,y,…,z)) we obtain an estimate of conditional probability (relative frequency) P(A$_0$/A$_1$&…&A$_k$) using training data. Remind that all expressions A$_i$ in first-order rules depend on variables x,y,…,z, i.,e. A$_i$(x,y,…z) in contrast with propositional logic expressions, which do not have variables x,y,…z.. These probabilities are used as evaluation functions combined with a test for statistical significance. This is a relatively common way designing an evaluation function. The relative frequency is used in AQ method and statistical significance in evaluated in CN2 method, but for entropy. We exploited an original search mechanism to select appropriate expressions [Vityaev, 1992; Vityaev, Moskvitin 1993] using mentioned evaluation criteria (conditional probability P(A$_0$/A$_1$&…&A$_k$) and Fisher statistical significance test). The search is in the line with "current-best-hypothesis" search and "least-commitment" search [Russel and Norvig, 1995, pp.546-552] but it is applied for probabilistic hypothesis, which is more complex. The used search was arranged in accordance with a definition of semantic probabilistic inference [Vityaev, 1992]. After finding several probabilistic laws E$_1$ ¤ E$_2$ ¤ ... ¤ E$_{k-1}$ the search for a new one was done by adding to the IF part of the rule (antecedent) a new atomic logical expression ( ) (**a**) #) (**b**))[9]. This adding is also known as specialization [Russel and Norvig, 1995, p.546]. We find mentioned new logical expression using the full search in 2.21-2.2.4 logical expressions.

The Fisher criterion was used on each step to test statistically each generated hypothesis 2.2.1-2.2.4 to be a probabilistic "law" [Vityaev, 1992]. We delete one of the atoms ( ) (**a**) #) (**b**)) and test if the rest of the expression has less conditional probability. If the conditional probability decreases statistically significantly (with a certain level of confidence) then we accept a tested hypothesis as a probabilistic "law" E$_k$. This idea is close to idea of finding a

generalization of a hypothesis [Russel and Norvig, 1995, p.546].

2.4. Learning. We tested hypotheses 2.2.1-2.2.4 using training set TR = $\{a_1, ... ,a_{tr}\}$ and randomly chosen pairs of objects **a**, **b** from TR by the "Discovery" system. To test hypotheses from section 2.3 we used all sequential pairs of objects from TR. The result of learning is a set **Law** of possible probabilistic "laws" found on TR. Each of these probabilistic "laws" was described with conditional probabilities on TR. To test if a "law" is a stable "law" we evaluated also its conditional probability on CT. However we did not use these conditional probabilities to choose preferred "laws" for forecast to guarantee independence of the forecast test on the control set (CT).

Let us give three examples of laws with relatively high conditional probabilities for both training and test/control sets TR and CT:

2.4.1. $(wd(\mathbf{a}) = wd(\mathbf{b}) = <2,3,4,5,1>)\&(\ )_{13}(\mathbf{a}) \# \ )_{13}(\mathbf{b}))\&(\ )_{15}(\mathbf{a}) > \ )_{15}(\mathbf{b}))\&$
$(\ )_{234}(\mathbf{a}) \# \ )_{234}(\mathbf{b}))\&(\ )_{245}(\mathbf{a}) > \ )_{245}(\mathbf{b})) \rightarrow (target(\mathbf{a}^5) \# target(\mathbf{b}^5))$;

Frequency on TR is 0.6385 and frequency on CT is 0.7609. This "law" can be translated to the normal financial language (see figure 2 above). That statement is true only statistically as it reflects by frequencies: the frequency on TR equals to 0.64 and the frequency on CT equals to 0.76. It means that for about 70% of those cases we have found an upper limit for the target value, which is the target value for the old Monday.

Let us suppose that the last target value is -3%, i.e. closing price for the old five days decreased. It means that we also will have a decrease from the current Tuesday to current Monday and amount of decrease will be greater than -3% with probability 0.7, for example, it can be -5%.

We present the next two examples without a linguistic description.

2.4.2. $(wd(\mathbf{a}) = wd(\mathbf{b}) = <2,3,4,5,1>)\&(\ )_{24}(\mathbf{a}) \# \ )_{24}(\mathbf{b}))\&(\ )_{145}(\mathbf{a}) \# \ )_{145}(\mathbf{b}))\&(\ )_{234}(\mathbf{a}) > \ )_{234}(\mathbf{b}))\&(\ )_{235}(\mathbf{a}) \# \ )_{235}(\mathbf{b}))$
$\rightarrow (target(\mathbf{a}^5) > target(\mathbf{b}^5))$;

Frequency on TR is 0 .63 and frequency on CT is 0.66.

2.4.3. $(wd(\mathbf{a}) = wd(\mathbf{b}) = <2,3,4,5,1>)\&(\ )_{25}(\mathbf{a}) \# \ )_{25}(\mathbf{b}))\&(\ )_{45}(\mathbf{a}) > \ )_{45}(\mathbf{b}))\&\ (\ )_{124}(\mathbf{a}) > \ )_{124}(\mathbf{b}))$
$\rightarrow (target(\mathbf{a}^5) > target(\mathbf{b}^5))$;

We have found **134 regularities ("laws")** of that kind connecting SP500C and the target.

The process of generating new rules is finished when there are no more rules with higher conditional probability and still statistically significant. Note this stop criterion does not require itself to restrict the set of tested rules a priory. The restriction can be based on the volume of available data, acceptable levels of conditional probabilities and significance. For practical computations often we stop computations earlier, reaching some runtime limit or/and acceptable level of conditional probabilities. The average of conditional probabilities of these regularities on training data TR is 0.5813 and the average of conditional probabilities on test data CT is 0.5759. All conditional probabilities are evaluated as relative frequencies on TR and CT respectively as it common in Machine Learning [Mitchell, 1997, p.282]. Thus performance (conditional probability) is sufficiently stable when we are moving from training to test data. The difference is 0.0054=0.5813-0.5759, i.e. 0.54%. Nevertheless this difference has a variation. Typical difference is no greater than K3% (53 regularities, i.e., 40%). There are also regularities with significantly higher differences. It tells us that some regularities became stronger and some weaker in financial time series for the last two years. Sometimes frequencies dropped down by 50%. It can mean changing market conditions, business strategy of the target company, stockholders' behavior and even that regularities have become known and people used them (market efficiency). Thus there are three types of regularities:

(1) Regularities/rules with similar performance on training and test data. Frequency difference range is K3% (53 regularities, 40%) with only 0.14% of the average decrease of frequencies;
(2) Regularities/rules with increasing performance on test data. Frequency increased on 38 regularities (28%) with 5.8% of the average increase of frequency;
(3) Regularities/rules with decreasing performance on test data. Frequency decreases on 43 regularities (32%) with 6.6% of average decrease of frequency.

Noise issue. It is possible that rules may not work out of sample due to noise. This is a common problem of all forecast methods. Probably MMDR suffers less from noise than other methods. If MMDR captured a "critical mass" of noise this noise would be a part of statistically significant rule (MMDR selects only statistically significant rules). So it is questionable should it be called noise. We interpret this situation as discovering different laws on different data as it is common in scientific discovery. For example, some laws of physics identified using data from Earth do not work on Moon or Mars with other gravitational levels.

Often the reason that rules may not work out of the sample is that the method is very sensitive to initial assumptions. In Neural Networks initial assumptions include such parameters as weight functions, number of layers,

and so on. MMDR is relatively robust in this sense, because MMDR pays special attention to minimize the set of assumptions.

2.5. Forecast. We can use regularities from **Law** set for forecasting only if we know right-side (target($\mathbf{a}^5$)) or left-side (target($\mathbf{b}^5$)) value of inequality (target($\mathbf{a}^5$) # target($\mathbf{b}^5$))$^{\gamma 0}$, which is a part of a found regularity. If we take both objects **a** and **b** from CT, then a forecast is impossible, because both target values are unknown. Therefore five days from TR with a known target value were used. If we take, for example, object **a** from TR and object **b** from CT then we will have a lower bound for unknown target($\mathbf{b}^5$) if $\gamma 0 = 1$ and an upper bound if $\gamma 0 = 0$, because the value target($\mathbf{a}^5$) is known. If we take object **a** from CT and object **b** from TR, we will have an upper bound if $\gamma 0 = 1$ and a lower bound if $\gamma 0 = 0$ for unknown value of target($\mathbf{a}^5$). In the "If part" of the rule 5.2.4 ( ) (**a**) #) (**b**))$^{\gamma 1}$& ... &( ) (**a**) #) (**b**))$^{\gamma k}$ values of all inequalities for objects **a,b** are defined in TRcCT and this part of the rule is an expression, which relates training and test/control objects. This expression shows similarity of objects **a** and **b**.

We forecast a target value for object **a** from CT by applying all regularities from the **Law** set to two sets of pairs of objects {<**a,b**>***b** 0 TR} and {<**b,a**>***b** 0 TR}. The first of these sets for each regularity gives a set of upper bounds Up1($\mathbf{a}^5$) = {target($\mathbf{b}^5$)}, if $\gamma 0 = 1$ and a set of lower bounds Low1($\mathbf{a}^5$) = {target($\mathbf{b}^5$)} if $\gamma 0 = 0$ for unknown value of target($\mathbf{a}^5$). Similarly the second of these sets gives lower bounds Low2($\mathbf{a}^5$) = {target($\mathbf{b}^5$)} if $\gamma 0 = 1$ and a set of upper bounds Up2($\mathbf{a}^5$) = {target($\mathbf{b}^5$)}, if $\gamma 0 = 0$ for an unknown value of target($\mathbf{a}^5$). The whole sets of upper and lower bounds Up1($\mathbf{a}^5$), Up2($\mathbf{a}^5$), Low1($\mathbf{a}^5$), Low2($\mathbf{a}^5$) for target($\mathbf{a}^5$) are obtained by joining these bounds for all individual regularities.

Our regularities allow us to forecast only for the last days of a five-day cycle (not necessarily Friday). It means that if there is a holiday within these five days we do not have enough data for forecast. We made a forecast for 442 days from 506 on CT.

This is not a principal restriction of the method. Regularities could be discovered with missing days, but it would take more runtime. Note that regularities without identification of a particular day of the week have significantly less prediction power.

Next we use order statistics with a confidence level. We compute **p-quintile** (p = 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90) for the **upper bound** of target($\mathbf{a}^5$) and **(1-p)-quintile** for the **lower bound** of target($\mathbf{a}^5$). For each value of p-quintile (p = 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90) we have the upper bound $Up_p(\mathbf{a}^5)$ for target($\mathbf{a}^5$), taken from Up1($\mathbf{a}^5$cUp2($\mathbf{a}^5$) and we have the lower bound $Low_p(\mathbf{a}^5)$ for target($\mathbf{a}^5$), taken from Low1($\mathbf{a}^5$)cLow2($\mathbf{a}^5$).

We assign $Low_p(\mathbf{a}^5) = -4$ for large p values (e.g. 0.80, 0.90, 0.95) if (1-p)-quintile is less than the least value of the lower bound for target($\mathbf{a}^5$). Also we assign $Up_p(\mathbf{a}^5) = +4$ for large p values (e.g. 0.80, 0.90, 0.95), if p-quintile is greater than the largest value of the respective upper bound.

There is no the forecast if the lower bound $Low_p(\mathbf{a}^5)$ is greater than the upper bound $Up_p(\mathbf{a}^5)$. It took place sometimes for small p (e.g., 0.55, 0.60, 0.65). We also refuse the forecast if we get p-interval [-4,+4]. Note that the p-intervals [$Low_p(\mathbf{a}^5)$, $Up_p(\mathbf{a}^5)$] for an unknown value of target($\mathbf{a}^5$) are nested for growing p values, i.e., $Low_{p1}(\mathbf{a}^5)$ # $Low_{p2}(\mathbf{a}^5)$ 4 $Up_{p1}(\mathbf{a}^5)$ $ $Up_{p2}(\mathbf{a}^5)$, if p1 > p2. All these intervals are results of forecast.

3.Results

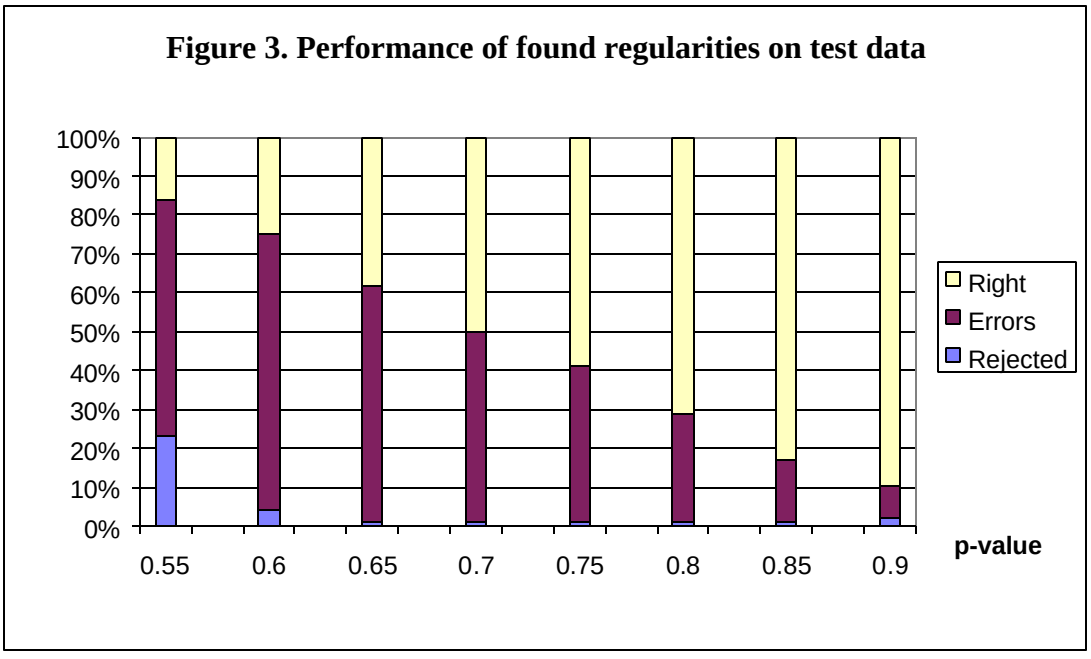
3.1.Results for set of expressions from section 2.2. We have evaluated the performance of the forecast for each p-quintile and for all objects from CT using six parameters:

(1) percentage of **Rejections**,
(2) percentage of **Errors**,
(3) percentage of **Right** forecasts,
(4) mean length of the p-intervals for all (right and wrong) forecasts (**ML)**
(5) mean length of the p-intervals for all right forecasts (**MLR**) and
(6) bound forecast mean square error (**BF MSE**), i.e. mean square difference between the forecast and **the nearest p-interval bound** for forecasts which are out of p-interval.

For cases when one of the bounds is not defined (we did not find "good" regularities for that bound) we took a doubled distance from target($\mathbf{a}^5$) obtained by forecast and a known bound, i.e., 2*(target($\mathbf{a}^5$) – $Low_p(\mathbf{a}^5)$), if we found the lower bound. If we have the upper bound 2*($Up_p(\mathbf{a}^5)$ - target($\mathbf{a}^5$)) is used. Table 3 and Figure 3 show performance metrics for test set CT of listed parameters. Figure 3 graphically represents first four columns of table 3. It reflects that with growth of *p* percent of correct forecast is growing too.

Table 3. Performance metrics for a set of regularities

p-value	Rejections	Errors	Right Forecast	ML -Mean length of the p-intervals for all (right and wrong) forecasts	MLR- Mean length of the p-intervals for all right forecasts	BF MSE Bound Forecast Mean Square Error
0.55	102 (23%)	268 (61%)	72 (16%)	0.5432	1.2148	2.0113
0.60	17 (4%)	315 (71%)	110 (25%)	0.8236	1.3325	1.5903
0.65	4 (0.9%)	279 (61%)	168 (38%)	1.2404	1.5755	1.7527
0.70	4 (0.9%)	215 (49%)	223 (50%)	1.7630	2.0140	1.9939
0.75	3 (0.7%)	176 (40%)	263 (59%)	2.3310	2.5875	1.7256
0.80	3 (0.7%)	125 (28%)	314 (71%)	3.0392	3.2493	1.3803
0.85	3 (0.7%)	71 (16%)	368 (83%)	3.9483	4.0957	1.2254
0.90	10 (2.2%)	35 (7.9%)	397 (90%)	5.1923	5.2540	1.1093

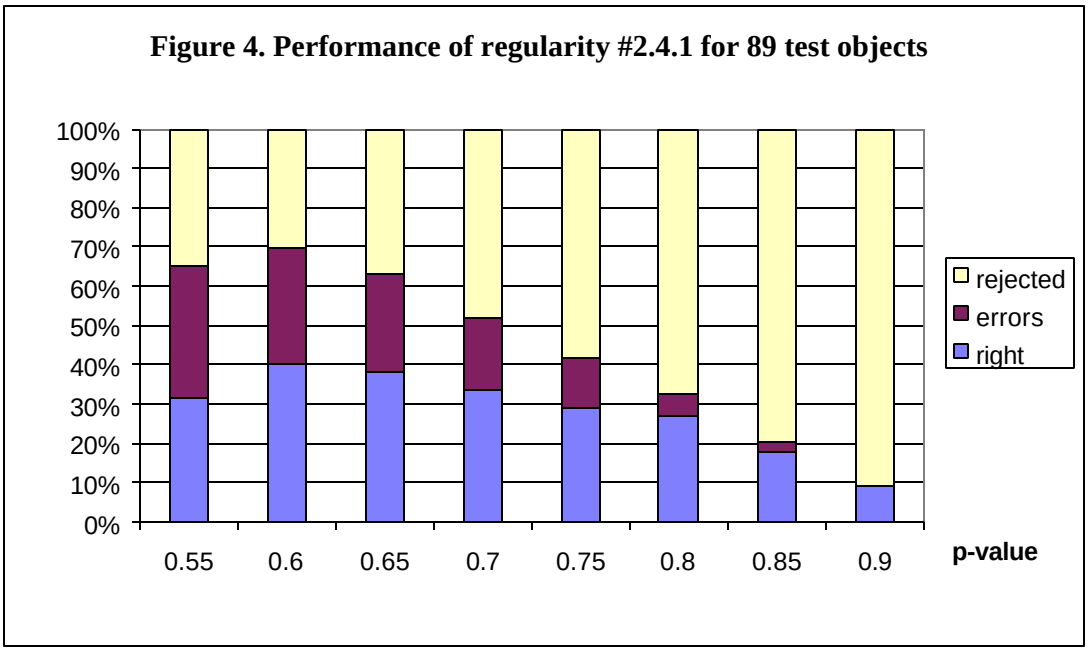


3.2.Results for regularity (2.4.1). Regularity (1) from section 2.4 was identified with 440 objects from training set TR. There are also 89 five-day sequences available in test set CT to test regularity (1). We considered different p-values and found the number of objects from those 89 objects, which are related to a particular p-value. For example p=0.55

brings us 58 objects and 28 of them were predicted correctly (in relatively narrow forecast interval). See table 4. Increasing p allowed us to get up to 100% correct forecast, but with a wider forecast interval and less number of objects (see figure 4 and table 4). It means that for practical forecast we need to choose some intermediate acceptable level of p. Figure 4 shows approximately equal number of right forecasts, wrong forecasts and rejections for p=0.55 and growth of rejections and decreasing the number of wrong forecasts.

Table 4. Performance for regularity #2.4.1

p-value	Right Forecast	ML- Mean length of the p-intervals for all (right and wrong) forecasts	MLR- Mean length of the p-intervals for all right forecasts	BF MSE Bound Forecast Mean Square Error
0.55	28 from 58 (48,3%)	2.806	0.269	2.640
0.60	36 from 62 (58.1%)	3.111	0.925	3.347
0.65	34 from 56 (60.7%)	3.471	1.386	2.146
0.70	30 from 46 (65.2%)	4.081	2.119	1.989
0.75	26 from 37(70.3%)	5.059	3.172	0.604
0.80	24 from 29 (82.8%)	4.962	4.013	0.114
0.85	16 from 18 (88.9%)	6.129	5.411	0.029
0.90	8 from 8 (100%)	6.221	6.221	0.000



This choice depends on investor's individual purposes, acceptable risk level and environment. Therefore it should be a part of a trading strategy, which requires a special study probably similar to portfolio selection with risky securities [Hiller, Lieberman, 1995, pp.561-563]. We leave systematic study of this issue out of the paper. Without that analysis we assume that reasonable level of p-value for data presented in table 3 would be [0.65, 0.75]. **A**lso the interval stock forecast can be viewed as an integral part of portfolio selection with risky securities. One of the known non-linear optimization models for portfolio selection is based on study done by Markowitz and Sharpe that helped them win the 1991 Nobel Prize in Economics [Hiller and Lieberman, 1995, pp.561-563]. Let us present basic elements of that model.

There are n stocks considered for inclusion in the portfolio, x_j is the number of shares of stock j, m_{ij} and s_{ij} are estimated mean and variance, respectively, of the return on each share of stock j, where s_{jj} measures the risk of the stock, s_{ij} is a covariance of the return on one share each of stock i and stock j. Two functions $R(x)=\sum_{j=1}^{n} m_j x_j$ and $V(x)=\sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij} x_j x_i$ are introduced. The first one represents the total **return** and the second one the **risk** associated with the portfolio. The objective function to be maximized is $f(x)=R(x)-ßV(x)$, where the parameter $ß$ is a nonnegative constant that reflects the investor's desired trade-off between expected return and risk. Choosing $ß=0$ implies that risk should be ignored completely, whereas choosing a large value for $ß$ places a heavy weight on minimizing risk. This way investor's expected utility can be maximized if the model captured investor's utility function (relative value to the investor of different total returns) [Bazaraa et al, 1993].

There is a bottleneck in this model to identify $ß$, $\{F_j\}$ and $\{s_{ij}\}$.The bottleneck related to $\{F_j\}$ and $\{s_{ij}\}$ is that they are only mean and variance over a period of time. Different periods may give different mean and variance. Also it is not clear which of them will be similar to them for portfolio selection term. Therefore the usage of forecast values of mean and variance appears a natural extension of that model. We think that our forecasted intervals for a stock j can be natural substitution for the stock variance and the middle of that interval could substitute mean F_j. Also having probabilities and respective forecast for discovered rules ("laws"), it is possible to incorporate them into decision analysis models to maximize payoff over possible alternatives [Hiller, Lieberman, 1995, pp.832-901] in stochastic programming setting.

Let us comment an advantage to predict the target using a particular regularity as (2.4.1). If we exploit all 134 found regularities **the target can be predicted practically for all possible objects**, but for some of them forecast interval can be very large and useless. Using a particular regularity from (2.4.1) we often can **predict the target only for few specific objects** but much more accurately. It means that we refuse to make any stock market decision for objects where we found that there is insufficient information for an accurate forecast. This approach looks more reasonable than other approaches delivering forecasts for all objects and always using only one formula (rule). In section 3.3 we illustrate this with some discovered rules. Those rules are applicable only for 55 days of 1995-1996, but they outperformed buy-and-hold strategy in simulated trading (see section 4).

3.3. Results for expressions from section 2.3. We found transition probabilities for several expressions on training data and confirmed them on test data. These expressions and probabilities can be used for forecast similarly to expressions described in 3.1 and 3.2. In 3.1 and 3.2 we used SP500C to predict the target. The main difference is that now we use previous values of the target itself to predict the target similarly as it is done in Markov chains models (see section 2.3). In 3.1 and 3.2 we used SP500C to predict the target. Similarly, we show that first-order logic can help to discover Markov chains type of models automatically.

Below we present examples of studied prediction statements:

(i) If the target decreases from previous Monday to current Monday then the target will grow for the next Tuesday with probability P1.

(ii) If the target increases from previous Monday to current Monday then the target will decrease for the next Tuesday with probability P2.

How did we select these types of rules? There are four reasons. The first one is based on Occam's razor principle: prefer the simplest hypothesis that fits the data [Mitchell, 1977, p.65]. What does it mean fit the data? We use a specific version of Occam's razor: prefer the simplest rules with maximum expected forecasting stability when moving from sample to a real forecast. This expected stability is evaluated using statistical criteria. The second one is

that given simple rules (based on relational assertions) well fit to the first-order logic. They can be naturally discovered with first-order logic in contrast with rules in propositional logic (see introduction to this paper). The third reason is based on fundamental results [Krantz et al, 1971, 1989,1990] showing that any fundamental measurement can be presented using first-order logic assertions. Krantz et al [1971, 1989,1990] developed these presentations for the most popular scales as order, interval, relational and absolute scales. The forth reason is that conditional probabilities found in Markov chain type of models have direct relationship with performance of the forecast, i.e., how these rules are confirmed on control/test data (CT). For testing rules we only need to find these probabilities on CT and compare them with probabilities on training data (TR). If probabilities on CT are similar or higher than on TR then rules are confirmed.

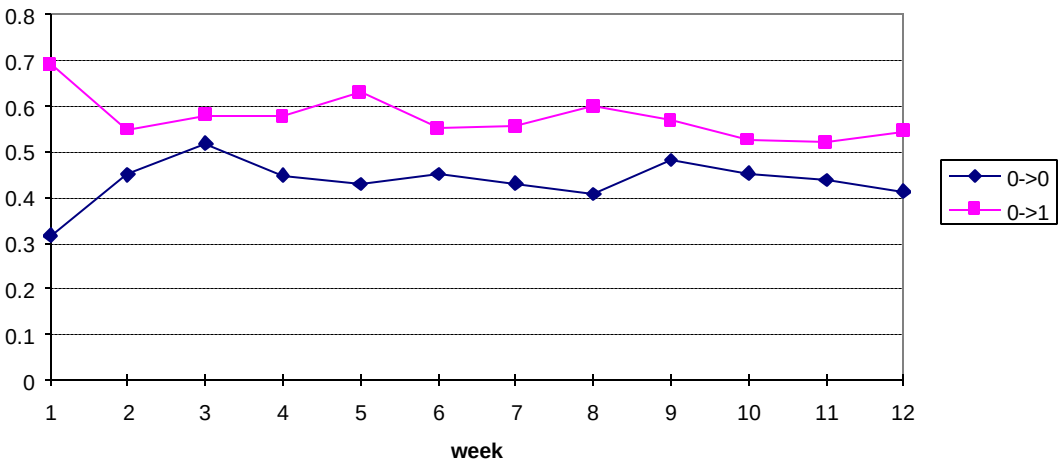
Regularities with different days in (i) and (ii) were tested. One of the found pairs of these regularities has relatively good prediction power (**probabilities 0.66 and 0.75 on test set CT**) and can be tested for practical forecast of the target. Actually our conditional probabilities are very close to a *confusion matrix* used in [Swanson, White, 1995] as a measure of forecast performance that is calculated how well a given forecasting procedure identifies the direction of change in the spot rate.

The natural buy/sell prediction strategy is based on this directional forecast [Cheng and Wagner, 1996]

$$\text{Prediction} = \begin{cases} Buy, \text{ if } UP \\ Sell, \text{ if } DOWN \end{cases}$$

We also studied the horizon of that forecast (see figure 5). Regularities (i) and (ii) are restricted. They are

Figure 5. Transition probabilities for 12 weeks (training data)



not applicable for all days, e.g., (i) and (ii) are not applicable for Wednesday and Thursday forecast. The most promising is forecast for one week among found regularities. There are also some lower chances for success for 5 and 8 weeks (see figure 5). But the majority of probabilities beyond one-week horizon is too close to 50:50. The upper line shows transition probabilities for $0 \rightarrow 1$ and $0 \rightarrow 0$, i.e., from up to down and from up to up (see section 2.3 for more details about notation).

4. Simulated trading performance

In this section we discuss testing the discovered regularities on the test data (1995-1996) using simulated trading performance. Direct testing of regularities described in sections 2.3 and 3.3 on the test data are given in table 6 (one-week horizon). This testing confirms the discovered regularities. In fact they are even more confirmed in 1995-1996 data than in the training data (1985-1994). In table 6 we use notation from section 2.3. Transition probability for training data is 0.69 for transition from 0 to 1 (0? 1) and 0.65 is for transition from 1 to 0 (1? 0). As we used above codes 0 and 1 for target and delta represent up and down, respectively (section 2.3.) Table 6 shows 0.84 and 0.7 for

them, respectively on the test data (1995-1996).

Table 5. Transition probabilities for test data

	Target	
Delta	0 (up)	1(down)
0 (up)	0.3	0.7
1 (down)	0.84	0.16

The comparison with another methods is more complicated. Regularities 2.2.1-2.2.3 are used to deliver interval forecasts. Regularities (2.2.4) are used to deliver so called "threshold forecasts", e.g., stock price *c* will be no less than the threshold *C* (*c=C*). See section 2.2. There are also "point" forecasts, where a particular value of the stock is forecasted. So there is a problem to compare threshold, interval and "point" forecasts directly, i.e., to find out which one is closer to the actual value of a stock.

Fortunately, different forecasts can be compared using trading performance. A forecast giving the best performance obviously has an advantage. We simulated trade over 1995-1996 years and compared results of simulated trading with stock prices and gain/loss in these years using our regularities in form of 2.2.4 (sections. 2.2 and 2.3) for forecast.

This testing of forecast of time series requires using a trading strategy. So we test a forecast together with a trading strategy. Each of them can be wrong/ineffective and can corrupt/hide a real advantage of another one. Therefore, this comparison can not be a final comparison of forecast methods, but it gives a useful output about the practical value of a forecast method in trade.

A simulated trading performance for the target (*T*) was evaluated on the test data (1995-1996). The target was scaled using the formula *T=10*(t+5)* to get more convenient larger numbers. The scaling does not change the performance. An active trade strategy was compared with buy-and-hold strategy for entire 1995-1996 years (table 6 and figure 6).

Buy-and-hold strategy means in our simulation "buying" *n* shares at the first trading day of 1995 and "selling" them at the last trading day of 1996. This way we "bought" 48 shares for 55.6 each (total investment 2668.7) on January 3, 1995 and "sold" for 60.36 on December 31, 1996 with gain 228.58 (8.56% of the initial buy-and-hold investment).

Table 6. Simulated trading performance

	Active trading	Buy-and-hold
Average investment for 1995-1996	994.53	2668.7
Final number of shares	48	48
Gain for 1995-1996	1059.87	228.37
Gain (% to the final capital)	52.92%	7.88%
Gain (% to the average active trading investment)	106.57%	Not applicable
Gain (% to the initial buy-and-hold investment)	Not applicable	8.56%

The active trading delivered simulated gain of 1059.87 (for 48 shares) in contrast with 228.37 in buy-and-hold strategy for the same 48 shares (table 6). To simplify consideration we ignore all taxes in this consideration. The initial investment in our active strategy is much smaller (169.68) with average investment over two years equal to 994.53 in contrast with 2668.7 in buy-and-hold strategy. It means that the active strategy does not require "freezing" 2668.7 in shares for two years. The gain is 52.92% of the final capital for the active strategy and 7.88% is a gain for to the final capital for buy-and-hold strategy (table 6. Therefore the active strategy outperformed buy-and-hold strategy. All taxes were ignored as we mentioned before.

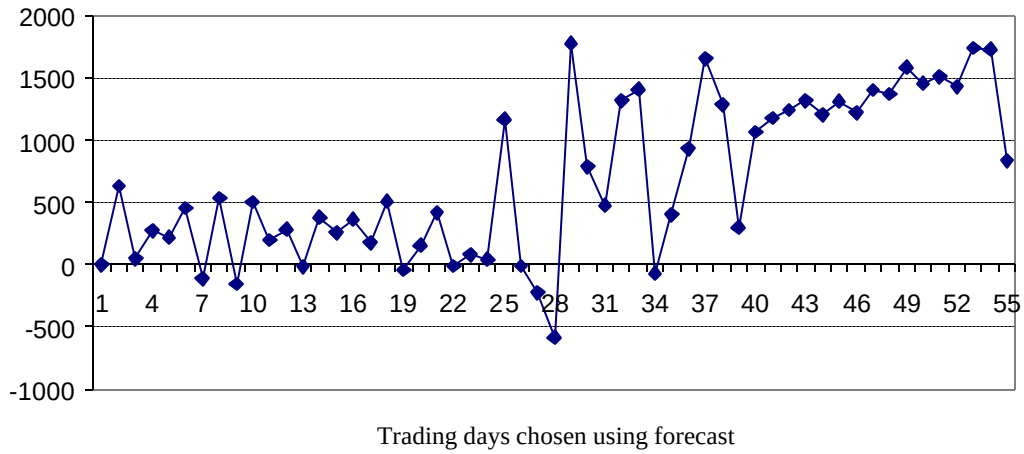
Figure 6 shows how the active strategy outperformed buy-and-hold strategy in gain/loss dynamics. Trading days are numbered on these figures from 1 to 55. These days were chosen over 1995-1996 using our rules and forecast bases on these rules. The rules used are applicable only for these days of 1995-1996.

To illustrate how we reached this result and have chosen trading dates we present simulated trade for January 1995 in more detail. The real trading interval in our active trading strategy is shorter. It begins on January 16,1995. How did we choose January 16 instead of the first trading day on 1995? January 16 is the first day with enough data to use tested regularities. They are applicable only for trading days with particular properties. Therefore

the scope of these regularities can be relatively narrow. Nevertheless they can deliver more exact forecast than a regularity applicable for all trading days, because they capture deeper and more specific properties. An extended set of rules can be discovered the same way and they can be applicable for a wider set of trading days. Meanwhile we show that an active strategy with discovered regularities already outperforms buy-and-hold strategy.

There is a mechanism within an active trading to chose whether we should trade on a particular day, e.g., January 3. In buy-and-hold strategy this date is chosen formally as a beginning of the year. In active strategy we used rules discovered on TC, i.e., data from 1985-1994. Only data of 1995-1996 were used to make trading decisions using these rules. Only on January 16 we get enough data for our rules and can check which rule is applicable (IF part of the rule is equal to 1). The applicable rules forecast on January 16 that the price will go up on January 23. Therefore, January 16 is time to buy shares and January 23 is time to sell them, if the forecast is correct. We buy shares on January 16. On January 23 rules forecast that the price will go down on January 30 with high probability. Thus, January 23 is time to sell, but there is also a relatively high probability that the price will continue to grow so we can sell only half of our shares. This can give us some extra gain, but force us to use a more complex active strategy. In this paper we follow a simpler active strategy--to sell all shares bought just before. This way we get a gain on January 23. As it was mentioned the forecast for January 30 is that shares will go down. Thus they should be sold out, but we already sold out all our shares. Therefore there is no trade for us on January 30. Having more shares on that day we could get an extra gain, because share's value is higher on January 30,1995. Next forecast on February 6 shows us that shares will go up on February 13, so February 6 is a time to buy. The further dynamics of gain/losses is shown in figure 6. They show that the active strategy, using discovered regularities and forecast outperformed buy-and-hold strategy for test years (1995-1996).

Figure 6. Performance difference between active and buy-and-hold trading (1995-1996)



5. Comparison with ARIMA

We made comparison of our results with ARIMA time series model [Pankratz, 1983, Montgomery at al, 1990]. Studied ARIMA models are presented in table 8, where *s* periodic parameter (*s*=5 days), *t* is a day and *T(t)* is the Target stock for day *t*, *a*,*b*,*c* and *q* are model coefficients. These coefficients were evaluated for models 3-5 using non-linear optimization methods (Solver Tool, MS Excess'97) and test data (1995-1996 years). The sign of *T(t+1)-T(t)* was forecast with high accuracy, but not an absolute value. Correct forecast of the sign is sufficient to form a successful buy/sell strategy. The sign forecast is simpler than absolute value forecast and first-order logic methods fit to discover sign forecast rules.

Model (1) called a random walk is considered "…a good ARIMA model for many stock-price series"

[Pancratz, 1983, p.410]. Nevertheless this model is not applicable to generate a trading strategy. This model does not create ups and downs needed for a trading strategy. We can not say that models presented in table 8 utilize all the power of ARIMA. ARIMA model requires adjusting parameters. Therefore it is possible to find a better ARIMA model. There are many ways how ARIMA parameters can be adjusted. We experimented with an adjustment mechanism within licensed ARIMA software ("Forecast Expert", SBS Inc.). The SBS Inc. offers this software for investors. After adjustment done by this software we have got 58% of correct sign forecast of Dow-Jones Industrial Averag1e (DJIA). We forecasted for 25 days ahead during 1994. This result shows us that there is a room to improve

Table 7. Comparison with ARIMA

	Model	Data	Simulated trading performance (correct buy/ sell signal)	Comment
1	$T(t+1)=T(t)$	1995-1996 all days	Not applicable	Buy/sell strategy is based on non-zero difference between $T(t+1)$ and $T(t)$. This model has zero difference for all days. Only random advice is possible here with 50% of success
2	$T(t+1)=aT(t)+b$	1995-1996 all days	62.58%	This result is less precise than the one for a subset selected by our method (table 5).
3	$T(t+4)=(1-q)*$ $*(T(t+3)+qT(t++2)$ $++q*qT(t)+c$	1995-1996 all days	58.84%	This result is less precise than the one for subset selected by our method (table 5)
4	$T(t+s)=aT(t)+b$	1995-1996 s=5, t is a fixed weekday	79.6%	This simple Markov process is fully consistent with our approach. We tested all weekdays t (Mondays, Tuesdays, Wednesdays, Thursdays and Fridays) using our approach to discover appropriate s and t
5	$T(t+2s)=aT(t+s)+$ $+bT(t)+c$	1995-1996 s=5, t is a fixed weekday	75.92%	This model exploits exactly the same data as we used to discover regularities. The model gives 75.92%. We got 84-70% (see table 5). This result confirms our approach.

an adjustment mechanism in ARIMA. MMDR and other first-order logic methods can be effective for these purposes. Model 4 in table 8 shows that when we found parameters s and t with MMDR we were able to adjust ARIMA parameters and to get 79% of correct sign forecast.

General problems of ARIMA model are presented in Montgomery et al [1990, p.288-290]. "There is no at present a convenient way to modify or update the estimates of the model parameters as each new observations become available, such as there is in direct smoothing. Future evolution of the time series will be identical to the past, that is the form of the model will not change over time."

The most significant advantage of the first order methods and MMDR, in particular, is that they can **forecast directly the sign of the difference instead of the value** as ARIMA does. ARIMA can generate a sign forecast using a forecasted value. The forecast of a value is more complex and available data may not fit for value forecast. The value forecast can be inaccurate and statistically insignificant but the forecast of the sign can be accurate and statistically significant for the same data.

The similar problem exists for neural networks. Chang and Wagner [1996] noted that their individual neural networks were capable of sign forecast, but not for the forecast of the absolute magnitude of the price.

6. Conclusion

Computational experiments presented in this paper show that first-order logic Machine Learning methods and Machine Methods for Discovering Regularities (MMDR), in particular, are able to discover useful regularities in financial time series for stock market prediction. In the time frames of the current study we obtained positive results using separately SP500C and history of target itself for target forecast. The best of these regularities had shown about 75 % of correct forecasts on test data (1995-1996). The target variable was predicted using separately SP500 (close) and own history of the target variable. Active trading strategy based on discovered rules outperformed buy-and-hold strategy and strategies based on several ARIMA models in simulated trading for 1995-1996. An ARIMA model constructed using discovered rules had shown the best performance among tested ARIMA models.

Combined usage of SP500C, target history, DJIA and other indicators can give more powerful regularities and a better forecast in future study.

References

1. Abu-Mostafa [1990]"Learning from hints in Neural Networks", *Journal of complexity*, 6, pp.192-198.
2. Bazaraa M., Sherali H., Shetty C. [1993] *Nonlinear Programming: theory and Algorithms*, Wiley, NY.
3 Chang. W., Wagner L[1996] Forecasting the 30-year U.S. Treasure Bond with a system of neural networks. (*http://ourworld.compuserve.com.homepages/FTPPub/cheng.txt*)
4 Halpern J.Y.[1990] "An analysis of first-order logic of probability", *Artificial Intelligence*, v.46, pp.311-350.
5 Henriksson R., Merton R [1981], On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills, *Journal of Business*,54, pp.513-533.
6 Hiller F., Lieberman [1995] *Introduction to Operations Research*, McGraw-Hill, NY.
7 Int. J. of Pattern Recognition and Artificial Intelligence [1989], v.3, N.1.
8 Kendall M.G., Stuart A. [1966] *The advanced theory of statistics*, v.2, Charles Griffin & Co LTD, London.
9 Kovalerchuk B., Vityaev E., Ruiz J. F. [1997] "Design of consistent system for radiologists to support breast cancer diagnosis", *Joint Conf. of Information Sciences*, Duke University, NC, USA, v.2, pp. 118-121.
10 Kovalerchuk, B., Triantaphyllou, E., Despande, A., Vityaev, E. [1996] "Interactive Learning of Monotone Boolean Function". *Information Sciences*, Vol. 94, issue 1-4, pp. 87-118.
11 Krantz D.H., Luce R.D., Suppes P., Tversky A.[1971, 1989, 1990] *Foundations of measurement*. Vol.1-3, Acad. Press, NY, London
12 Mitchell T. [1997] *Machine Learning*, McGraw-Hill, NY.
13 Montgomery D., Johnson L., Gardiner J. [1990] *Forecasting & time series analysis*, NY, McGrow-Hill.
14 Quinlan J.R.[1993] *C4.5:Programms for Machine Learning*. San Mateo, CA; Morgan Kaufmann.
15 Pankratz A. [1983] *Forecasting with univariate Box-Jenkins models*, NY, Willey
16. Russel S. and Norvig P. [1995] *Artificial Intelligence*. *A modern approach*, Prentice Hall
17 Swanson N., White H.[1995] A model-selection Approach to assessing the information in the term structure using linear models and artificial neural networks, *Journal of Business & Economic Statistics*, July, Vol. 13, n.3, pp.265-275.
18. Vityaev E., Moskvitin A.[1993] "Introduction to Discovery theory: Discovery system", *Computational Systems*, v.148, Novosibirsk, p.117-163 (in Russian).
19 Vityaev E. [1992] "Semantic approach to knowledge base development: Semantic probabilistic inference", *Computer Systems*, v. 146, Novosibirsk, p.19-49 (in Russian).
20 Zagoruiko N.G. [1979] *Empirical prediction*, Nauka, Novosibirsk (in Russian).