INTELLECTUAL SUPPORT OF INVESTMENT DECISIONS BASED ON A CLUSTERING OF THE CORRELATION GRAPH OF SECURITIES

Izabella V. Lokshina
Division of Economics and Business
State University of New York
Ravine Parkway
Oneonta, NY 13820
USA

Abstract: The paper presents a report on implementation of expert system that to be used to form the investment policy based on cluster analysis (including creation of correlation graphs, cluster forming and diversified portfolio synthesis). The paper examines a complex of algorithms developed for synthesis of investors' diversified portfolios, which provide high reliability and stability of decisions under possible trend deviations and with the existence of 'noisy' components. Diversified portfolio synthesis software was developed based on the principle of optimum decision-making in financial policy management.

Expert system is working perfectly in condition of unstable Russian securities market, but could be used for unstable sectors of American market, such as Real Estates, or Bonds.

Key-Words: Diversify Portfolio, Investor, Expert System, Mathematical Model, Discrete Space, Correlation Structure, Dynamic Trajectory, Non-equilibrium Stochastic System, Cluster Analysis.

1 Introduction

The rapid evolution and specific nature of the Russian security market stimulated the need to quickly develop suitable operational support technology (in a broad sense). An important aspect of the specific nature of the Russian security market is seen in the fuzzy information background of stock operations. That is, (1) it is extremely difficult to get information about stocks as well as about transactions; (2) there are no stable criteria of participants' reliability; and (3) there is no formal information infrastructure.

As a result of these problems, information flows in the Russian stock market contained a high 'noisy' component. This was a reason for the market's reduced efficiency and the inability to synthesize the experience of countries with a long history of securities markets onto Russian conditions.

These conditions, as well as the technological incompatibility of Western instruments with Russian reality in the present time, led to the wide use of a subjective expert approach, which is good perhaps only for management of a small investment portfolio. Moreover, this approach did not produce high quality results, because of a shortage of highly qualified experts in the fledgling Russian market. [1, 2]

2 Problem Formulation

The nature of the Russian security market showed a need for creating a new approach to support of investment decisions. It was concluded that modeling of an investor's diversified portfolio could be based on a concept using discrete space of correlation structure of dynamic trajectories of the non-equilibrium system, which is formed by stock market transactions.

2.1 ES, Supporting Investment Decisions

The analysis of MIS functions was made. In modern conditions of investment market in Russia, the important roles belong to informational-analytical functions in addition to banking and organizational-administrative functions. [1, 2]

An expert system for investment decisions support was constructed using the above-mentioned principles.

The Knowledge Base (KB) was transformed into a decision tree. Such an approach enabled showing each relationship more thoroughly and made the fullness and high correlation of the model much simpler.

The importance of such expert system development was proved by the observation that investors acting in the financial market were driven by the principle of maximum efficiency of their assets. As a result, they kept close track of the balance between the profitability of investment operations and the level of commercial risk connected with deal being made. The use of such an expert system in conditions of the Russian stock market could provide a significant increase of investment efficiency.

The system of active financial management in conditions of Russian stock market required a special approach, in order to resolve the problem of optimum investor's portfolio and to improve related software products.

2.2 Application Of Theory Of Financial Management

The main characteristics of a stock portfolio formed on the basis of financial analysis are:

Efficiency of portfolio *Rp*,

(1) $Rp = \sum Rj Xj$, $j=[1,n]$;
$\sum Xj=1$, $j=[1,n]$.

where n is a quantity of security types,
Xj is a share of j-type security,
Rj is an efficiency of j-type security.

Risk of portfolio σp,

(2) $\sigma p = (\sum Xj^2\, \sigma j^2)^{½}$, $j=[1,n]$;

Where $\sigma j^2 = E\,[(Rj - mj)^2]$;

$mj = E\,[Rj]$.

The influence of correlation on portfolio efficiency is estimated by correlation coefficients:

$\rho ij = Vij/\sigma i \sigma j$;

$Vij = E\,[(Ri - mi)\,(Rj - mj)]$;

Where Vij is co-variation of ij-type security efficiency.

Dispersion of effect Vp is,

$Vp = \sum\sum (\sigma i\, Xi)(\sigma j\, Xj)\, \rho ij$, $i=[1,n]$, $j=[1,n]$.

The expected portfolio efficiency, co-variation and dispersion will be dependent on its structure.

A formalized statement of the optimization task is next: To find the values Xj, which provide a minimization of variation of portfolio efficiency.

(3) $Vp = \sum\sum Vij\, Xi\, Xj$, $i=[1,n]$, $j=[1,n]$;
under conditions $\sum mj\, Xj = mp$, $j=[1,n]$;
$\sum Xj = 1$, $j=[1,n]$.

The decision of task (3) gets new peculiarities if to consider that, in addition to risked securities, there are no-risk (or almost no-risk) state securities in the market. So, in theory and in practice the main task of investment policy is correct capital distribution between risk and no-risk investments.

The characteristics of combined investment of joint portfolio (1) can be transformed into

(4) $mp - ro = (mp - ro)\, \sigma p / \sigma r$;

where ro is the characteristic of a no-risk investment;
mr, σr are expected efficiency and dispersion of risked securities.

Margin for risk of any security included into optimal portfolio is proportional to margin for risk of whole portfolio.

(5) $mj - ro = (\beta^*)j\ (mp - ro)$;

$$(\beta^*)j = Cov\ [Rj, (R^*)p] / (V^*)p;$$

where $(\beta^*)j$ is the beta-coefficient defining investment efficiency related to optimal portfolio;
$(R^*)p$, $(V^*)p$ are efficiency and variation of optimal portfolio.

The beta-coefficient is an index of market sensitivity. It defines relative variation of j-type of stock to average stock, or "market".

The beta-coefficient of a portfolio is

$$Bp = \sum \beta j\ Xj,\ j=[1,n].$$

The desire not to have risk makes investors prefer creating portfolios with smaller level of risk when they choose between securities with equivalent norm of margin.

For correlation analysis and choosing between risk and margin, the next norms are used:

K - required norm of margin, or minimum expected margin which push investors buy the security;

Km – expected norm of margin, or norm of margin, which the investor expects to receive in future.

$$vr = K - Rf,$$

where Rf = no-risk norm of margin.

Prices and norm of portfolio margin deviate by absolutely the same way as the security market does.

Any required norm of security margin is satisfied to (5) and is defined as a model of active capital cost creation:

(6) $K = Rf + vr = Rf + \beta\ (Km - Rf)$

An equation (6) is a fundamental equation for financing.

An investor can change his own risk in the market by using a different financial policy (strategy). The beta-coefficient of a company also can be varied as a result of strong competition in the branch of industry, changes in management, and so on. If such variation takes place, required norms of margin also will be changed, as well as prices of stocks of company.

2.3 Application Of Theory Of Non-Equilibrium Processes

Traditional methods of financial management are based, more or less, on relatively stable economic development and the absence of critical situations.

For condition of the Russian market, trends of stocks are presented as realizations of non-equilibrium processes x(t),

$$x(t) = a(t) + u(t),$$

where u(t) is the result of equilibrium processes;
a(t) is a given function, which is repeated for each application;
x(t) is estimated on the basis of correlation analysis.

Let us consider the correlation structure of non-equilibrium processes [x(t)] and [y(t)]. At any moment in time t=T, the average values are defined as:

(7) $\mu x(T) = E[x(t)]$,

$$\mu y(T) = E[y(t)].$$

Co-variation functions for any fixed meanings of time t1, t2 are defined as:

$$Rxx(t1,t2) = E[x(t1), x(t2)];$$

$$Ryy(t1,t2) = E[y(t1), y(t2)]$$

$$Rxy(t1,t2) = E[x(t1), y(t2)]$$

where Rxx(t1,t2), Ryy(t1,t2) are non-equilibrium co-variation functions;
Rxy(t1,t2) is the non-equilibrium inter-co-variation function.

For the average assembly estimation we have:

(8) $\check{R}xx(t1,t2) = \{\sum xi(t1)xi(t2)\}/N,\ i=[1,n].$

Under fixed t1 = t, t2 = t- ŧ

(9) $\check{R}xx(t,t\text{-}ŧ) = \{\sum xi(t)xi(t\text{-}ŧ)\}/N,\ i=[1,n];$

where ŧ is fixed time interval.

An analogous method is applicable for measuring of non- equilibrium co-variation functions:

(10) $\check{R}xy(t1,t2) = \{\sum xi(t1)yi(t2)\}/N,\ i=[1,n];$

$\check{R}xy(t,t\text{-}ŧ) = \{\sum xi(t)yi(t\text{-}ŧ)\}/N,\ i=[1,n].$

2.4 Portfolio Selection for Investment

On the basis of estimations (6)-(10), the statistical analysis of traded security prices of 29 the largest Russian companies, listed on the Russian stock market, was made with the purposes of forecasting cost deviation and forming of an investor's portfolio.

Expert methods were applied only as an addition to statistical methods caused by absence of complete and representative statistics of object characteristics and the fuzzy environment of object functioning.

The use of different methods of non-formalized information development has a result in the acquisition of new knowledge, enlarging of the KB, and increasing forecast quality.

The main method of knowledge acquisition was by interviewing financial analysts and experts. Feedback was provided using a procedure based on the Delphi method.

3 Problem Solution

The investor's portfolio was formed on the basis of cluster analysis. The decision is presented as a separation, which is satisfied to some optimal criteria Q, and is received as resolving of optimization task:

(11) Extr Q(S) , S⊂ A;

where A is set of all acceptable separations.

The choice of a metric, or a measure, for definition of the closeness of the relationship between the objects, which are represented by values of a multidimensional indicator, is central to the research. This metric defines the final version of the classification on the basis of any algorithm of separation. This problem was resolved using the method of correlation analysis.

As a basic method of automated classification, the method of dynamic concentrations was applied to the cluster analysis procedure.

The main definitions and common scheme of a method developed in research are following:

Let us assume, that X = {Xi,…,Xn} – a set of objects for research, each of them is characterized by a p-dimensional vector of indicators.

Xe = (Xei,…,Xep).

Space of k-coverage Sk is a set, each element of which S = (S1,…,Sk) presents a system of subsets of elements X, or classes, which are satisfied by given class structure.

Space of representative L is a set, each element of which can be considered as a representative of class of elements X. The measure of closeness D(X,l) of the relationship between object X and representative l is to be chosen.

Space k representatives Lk for space of k-coverage Sk is a set of subsets l = (li,…,lk), l⊂ L.

For creation of representative L for coverage S:

1. Space of representative L and measure of close relationship D(X,l) are chosen;
2. Representative function g is chosen, which is related to class Si of representative li, e.g., g(Si) = li;

For creation of coverage S related to representative l:

1. Space of coverage Sk is chosen;
2. Appointment function f is chosen, which makes an appointment of object X to special class, e.g., f(l) = S.

The method of dynamic concentration provides the next steps:

- Choice of the coverage space;
- Choice of the representative space and the measure of closeness of the relationship;
- Choice of optimization criteria W(S,l), which allow level of adequacy between every coverage and every representative to be estimated;
- Minimization of criteria W; choice of g- and f- functions;
- Creation of algorithms of dynamic concentrations, which presents a number of iterative g- and f- applications starting from some coverage $So \subset Sk$ or representative $lo \subset Lk$;
- Study of agreement of algorithm of dynamic concentrations.

For resolving of optimization task (11), the derivative from the constructed graph on event dG/dS was used. It was based on the definition of the derivative in Discrete Mathematics and on the definition of the relationship matrix. [3]

Each event defines some incidence matrix:

(12) $Q(\psi) = [qij]$,
where qij = 1, if j-condition belongs to i-realization of event;
qij = 0, if j-condition does not belong to i-realization of event.

After a correlation graph of securities was constructed, a nominalization was made on the basis of the expert method. Each boundary weight was equal to correlation coefficient between securities, and the procedure of nominalization concluded in the estimation of the primary information quality threshold. If the value of the relationship was less than value of the primary information quality threshold kij, the link was not considered at all, because of weak correlation.

On the basis of the boundary correlation matrix, all complete sub graphs (clicks) were formed. All complete sub graphs were built as a tree, which root is correspondent to graph G, and suspension nodes are correspondent to complete sub graphs.

Obviously, non-correlated securities correspond to nodes, which form the layer. Correlated securities form a path on the graph.

A node in the layer is chosen as a given birth, which has a maximum level of coverage Smax(vi) = ai.

δ, the inverse of the derivative obtained from the correlation graph, was used as a measure for closeness of the relationship,

$$\delta = (dG/dS)^{-1} = fij / (fi + fj - 2fij),$$

where S is a set of duo-correlated securities (ai, aj).

For calculation of the measure of closeness of the relationship, an incidence frequency matrix was created on the basis of the transposing of incidence matrix.

$$F = Q^T \times Q$$

After a clustering was produced using the criteria of maximizing the measure of the closeness relationship inside each cluster and minimizing the measure of the closeness relationship outside, we have received the synthesized optimal investor’s portfolio.

4 Conclusion

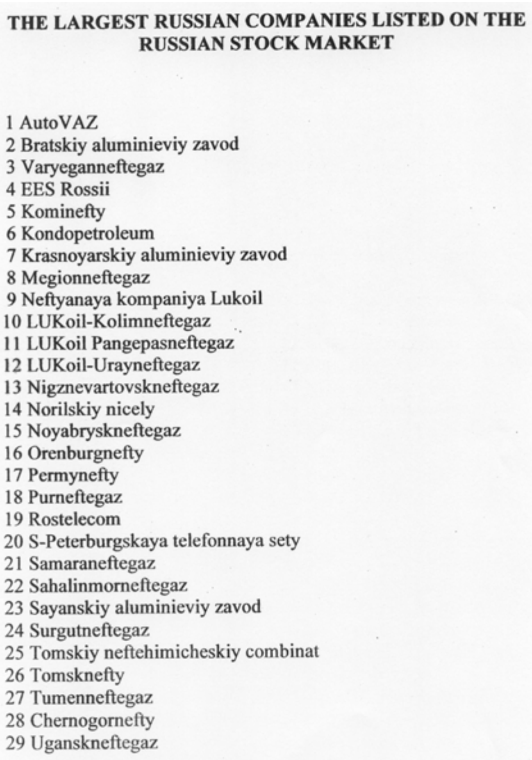
Research with real objects of the Russian financial market confirmed that an expert system, realized on the basis of the principles discussed above, supports the making of reliable decisions, and it was recommended for practical application in the field of financial management. [1, 2]

The general module of the system, called CREATE, which supports decision-making in the field of investment policy, has been implemented and used successfully at a number of Joint Stock Companies in Russia.

The approach based on a clustering of the correlation graph of securities could be used for intellectual support of investment decisions at unstable sectors of American market (such as Real Estates, Bonds, etc).

Appendix

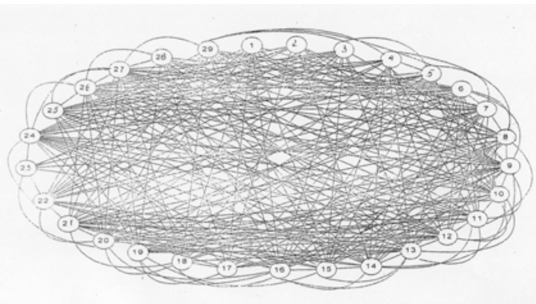
1. List of securities from Russian stock market



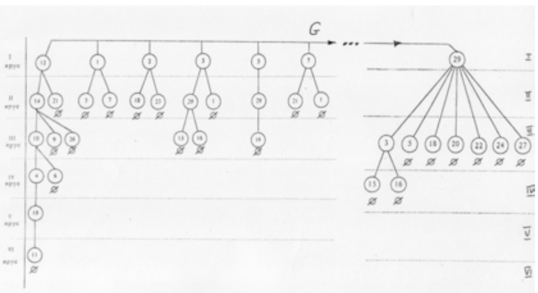
THE LARGEST RUSSIAN COMPANIES LISTED ON THE RUSSIAN STOCK MARKET

1 AutoVAZ
2 Bratskiy aluminieviy zavod
3 Varyeganneftegaz
4 EES Rossii
5 Kominefty
6 Kondopetroleum
7 Krasnoyarskiy aluminieviy zavod
8 Megionneftegaz
9 Neftyanaya kompaniya Lukoil
10 LUKoil-Kolimneftegaz
11 LUKoil Pangepasneftegaz
12 LUKoil-Urayneftegaz
13 Nigznevartovskneftegaz
14 Norilskiy nicely
15 Noyabryskneftegaz
16 Orenburgnefty
17 Permynefty
18 Purneftegaz
19 Rostelecom
20 S-Peterburgskaya telefonnaya sety
21 Samaraneftegaz
22 Sahalinmorneftegaz
23 Sayanskiy aluminieviy zavod
24 Surgutneftegaz
25 Tomskiy neftehimicheskiy combinat
26 Tomsknefty
27 Tumenneftegaz
28 Chernogornefty
29 Uganskneftegaz

2. Correlation graph of securities



3. Decision Tree (some clicks of graph G)



4. Synthesized optimal investor's portfolio

K1 = {12,14,10,4, 6, 25, 19, 11, 9, 7}
K2 = {2, 18, 23, 22}
K3 = {3, 29, 15, 16, 1, 5, 18, 27, 20, 24, 21}
K4 = {13, 9, 24, 8}
/π/ = 4

References:

[1] Lokshina, Izabella V. *Expert system of security market estimation and forecasting in Russia*, - in volume *Work In The Information Society* – Helsinki, Finland, 1996.

[2] Lokshina, Izabella V. *Development of ISFT supporting decision-making for company investment policy creation,* - in volume Mathematics and Simulation with Biological, Economical and Musicoacoustical Applications, WSES Press, 2001 p.p. 225-231

[3] Gorbatov, V.A. , 1968