Associate Professor Cristinca FULGA, PhD
Lecturer Silvia DEDU, PhD Candidate
Lecturer Florentin ŞERBAN, PhD
Department of Mathematics
The Bucharest Academy of Economic Studies

PORTFOLIO OPTIMIZATION WITH PRIOR STOCK SELECTION

Abstract. *We consider the problem of a decision maker, who is concerned with the management of a single-period portfolio that consists of holdings in n risky assets and is adjusted at the beginning of the time-period. The portfolio optimization problem consists in choosing the optimal rebalancing decisions in response to new information on market future prices (returns) of the risky assets in the portfolio in order to maximize the expected value of the end of period wealth in the presence of transaction costs, while satisfying a set of constraints. Rebalancing decisions are manifested in the revision of holdings through sales and purchases of assets. We assume that the assets are sufficiently liquid that market impacts can be neglected.*

We propose to solve the portfolio optimization problem in two steps: first, the phase of stock selection and second, the asset allocation phase. For the stock selection, we use principal component analysis to reduce the number of characteristics that will be taken into account. After that, applying clustering techniques, we find the similarities between the assets and we obtain a partition of the set of assets in clusters. Taking one element from each class, we get the set of assets that will be used to build the optimal portfolio. Once the stock selection completed, the optimal portfolio is obtained using an algorithm that combines the specific features of the convex programming with approximation techniques.

The unique nature of our work is the combination of the classification theory with the portfolio optimization techniques and the study of this approach.

To illustrate the behaviour of the proposed method, we consider the case of a portfolio of assets from Bucharest Stock Exchange.

Keywords: *portfolio optimization, stock selection, clustering.*

JEL Classification: C02, C61, G11.

1. INTRODUCTION

Portfolio management is a topic of particular interest to multinational firms, financial intermediaries, institutional investors and individuals. Recent years have seen a growing interest in portfolio optimization problem and therefore, a rich literature on this field. The paper of Best and Hlouskova (2003) deals with the portfolio selection problem of risky assets with a diagonal covariance matrix, upper bounds on all assets and transactions costs. Blog et al. (1983) consider the specific optimal selection problem of small portfolios. Kellerer et al. (2000) introduce mixed-integer linear programming models dealing with fixed costs and minimum lots and propose heuristic procedures based on the construction and optimal solution of mixed integer sub-problems. Konno and Wijayanayake (2001) propose a branch and bound algorithm for calculating a globally optimal solution of a portfolio construction/rebalancing problem under concave transaction costs and minimal transaction unit constraints. Schattman (2000) develops an iterative heuristic for finding a suboptimal solution for the portfolio problem. Meyer (1974) proved the convergence of a class of algorithms that includes the heuristic in this paper. If the portfolio optimization problem is nonlinear, the algorithm presented in Fulga (2006), that combines penalty concepts and sequential quadratic programming techniques can be used. In recent years, a large amount of work has been devoted to the dynamic portfolio optimization problem, see for example Birge (2008), Topaloglu et al. (2008), Fulga (2008, 2009a, 2009b).

In Section 2 we present the mathematical model of the portfolio problem. Section 3 focuses on deriving a method for solving the portfolio optimization problem. A preliminary condition required is building the initial portfolio. The originality of our approach consists in providing a method that takes into consideration both aspects: the prior stock selection and the asset allocation phase. In this paper we develop a method for choosing the risky assets from a large data set. For stock selection, in Subsection 3.1 we use principal component analysis to reduce the number of characteristics that will be taken into account. After that, applying classification techniques, we obtain a partition of the set of assets in clusters. Taking one element from each class, we get the set of assets that will be used to build the optimal portfolio.

The idea of obtaining clusters that characterize a set of assets can be found also in Kaski et al. (2009), Ştefănescu et al. (2008) and Mantegna (1999). The methodology based on clustering techniques is an useful tool for understanding and detecting the global structure, taxonomy and hierarchy in financial data. These methods were successfully applied to analyze stock and exchange markets. Brida and Risso (2007a, 2007b) have applied clustering techniques in order to classify the assets from Milano Stock Exchange and Frankfurt Stock Exchange using Pearson correlation. Once completed stock selection, in Subsection 3.2 the optimal portfolio is obtained using an algorithm that combines approximation techniques with the specific features of the convex programming.

The rest of the paper goes as follows. In Section 4 we present and solve a case study using, besides the proposed methods, fundamental analysis and technical analysis, two techniques belonging to financial analysis.

Section 5 summarizes the conclusions concerning this original approach and reveals the advantages of using the proposed technique.

2. MATHEMATICAL MODEL OF THE PORTFOLIO PROBLEM

We are concerned with the single-period portfolio, that consists of holdings in n risky assets. The portfolio is adjusted at the beginning of the time-period. The goal is to choose the optimal portfolio in order to maximize the expected value of the end of period wealth in the presence of transaction costs, while satisfying a set of constraints on the portfolio.

The current holdings in each asset are the components of the wealth vector $w = (w_1,...,w_n)^T$. The total current wealth is then $\sum_{i=1}^{n} w_i$.

The amount of money transacted in each asset $i = \overline{1,n}$, is denoted by x_i, with $x_i > 0$ for buying, $x_i < 0$ for selling and $x = (x_1,...,x_n)^T \in R^n$ is the vector of transactions. After transactions, the adjusted portfolio is $w + x$. Representing the sum of all transaction costs associated with x by $f(x)$, the budget or self-financing constraint is $\sum_{i=1}^{n} x_i + f(x) = 0$. The adjusted portfolio $w + x$ is then held for a fixed period of time. At the end of that period, the return on asset i is the random variable $\tilde{r}_i$, $i = \overline{1,n}$. All random variables are on a given probability space. We assume knowledge of the first and second moments of the joint distribution of $\tilde{r} = (\tilde{r}_1,...,\tilde{r}_n)^T$, $E(\tilde{r}) = r$, $r = (r_1,...,r_n)^T \in R^n$, $E\left((\tilde{r} - r)(\tilde{r} - r)^T\right) = C$. A riskless asset can be included, in which case the corresponding r_i is equal to its (certain) return, and the i^{th} row and column of C are zero.

The end of period wealth is a random variable, $\tilde{w} = \tilde{r}^T (w + x)$, with expected value and variance given by $E(\tilde{w}) = r^T (w + x)$, respectively, $E\left(\left(\tilde{w} - E(\tilde{w})\right)^2\right) = (w + x)^T C (w + x)$. The budget constraint can also be written as an inequality, $\sum_{i=1}^{n} x_i + f(x) \le 0$. With some obvious assumptions $\left(f \ge 0,\ r_i > 0,\ i = \overline{1,n}\right)$, solving an expected wealth maximization problem with either form of the budget constraint yields the same result. The inequality form is more appropriate for use with numerical optimization methods. For example, if f is convex, the inequality constraint defines a convex set, while the equality constraint does not. We summarize the portfolio selection problem as

$$(PP)\ \begin{cases} \max r^T (w + x) \\ s.t.\ \sum_{i=1}^{n} x_i + f(x) \le 0 \\ w + x \in X, \end{cases}$$

where $r = (r_1,...,r_n)^T \in R^n$ is the vector of expected returns, $w = (w_1,...,w_n)^T \in R^n$ is the vector of current holdings, $x = (x_1,...,x_n)^T \in R^n$ is the vector of amounts transacted, $f : R^n \to R$ is the transaction cost function and $X \subset R^n$ is the set of feasible portfolios.

3. THE METHOD

We propose to solve the portfolio optimization problem in two steps: first, the phase of stock selection and second, the asset allocation phase. For the stock selection, we use principal component analysis to reduce the number of characteristics that will be taken into account. After that, applying classification techniques, we find the similarities between the assets and we obtain a partition of the set of assets in clusters. Taking one element from each class, we get the set of assets that will be used to build the optimal portfolio. Once the stock selection completed, the optimal portfolio is obtained using an algorithm that combines approximation techniques with the specific features of the convex programming.

3.1. THE STOCK SELECTION PHASE

Nowadays huge amount of financial data are available. Consequently, it is very difficult to make use of such an amount of information and to find basic patterns, relations or trends in the data. We will apply data analysis techniques in order to discover relevant information in the field of financial data, that will be useful in stock selection phase and decision making process.

Let consider we have collected information concerning a number of N assets, each of them with P characteristics, that represents financial indices and will be called variables. We denote by y_i^j the value of the j variable for the asset i. The multivariate dataset will be represented using a matrix $Y = \left(y_i^j\right)_{\substack{i=\overline{1,N} \\ j=1,P}}$ and can be visualized as a set of N points in a P - dimensional data space.

Principal Component Analysis (PCA) is an useful data analysis technique for finding patterns in high dimension data. PCA involves a mathematical procedure that transforms the P variables, usually correlated, into a smaller number $p \leq P$ of uncorrelated variables, called principal components. PCA involves the calculation of the eigenvalue decomposition of a data covariance matrix, usually after mean centering the data for each attribute.

After applying PCA, each asset i will be characterized by p variables, represented by an array of the parameters $y_i^1, y_i^2,..., y_i^p$, therefore it is possible to form p - dimensional vectors $Y_i = \left(y_i^1, y_i^2,..., y_i^p\right)$, $i = \overline{1,N}$, which correspond to the set of N assets.

Let suppose that we have obtained a set of analyzed data consisting of N objects: $Y_1, Y_2,..., Y_N$, where $Y_i = \left(y_i^1, y_i^2,..., y_i^p\right)$, $i = \overline{1,N}$. We will use clustering techniques in order to find the similarities and differences between the assets. The clustering idea consists in assigning the vectors $Y_1, Y_2, ..., Y_N$ to one of n classes $C_1, C_2, ..., C_n$. The goal of clustering analysis is to separate data into groups based on individual cases.

After applying clustering techniques, we will build the initial portfolio by taking one asset from each class. We will obtain a portfolio consisting of n assets: $Y_1, Y_2, ..., Y_n$. Our approach has an important advantage, as guarantees the diversity of the portfolio and consequently improve the optimizing process, because it starts with an initial portfolio composed by a wide range of assets.

When the stock selection phase will be accomplished, we will have chosen n representative assets from the N available ones.

3.2. THE ASSET ALLOCATION PHASE

In this section we focus on solving the portfolio problem (PP). We note that the *n* risky assets in the composition of the portfolio were chosen during the prior stock selection phase.

The difficulty about solving (PP) problem consists in the presence of the transaction costs. Transaction costs can be used to model a number of costs, such as brokerage fees, bid-ask spreads, taxes, or even fund loads. In this paper, we assume the transaction costs to be separable, i.e., the sum of the transaction costs associated with each trade is $f(x) = \sum_{i=1}^{n} f_i(x_i)$, where f_i is the transaction cost function for asset $i = \overline{1,n}$.

The simplest model for transaction costs is that there are none, i.e., $f(x) = 0$. In this case the original portfolio is irrelevant, except for its total value. We can make whatever transactions are necessary to get the optimal portfolio.

A better model of realistic transactions costs is a linear one, with the costs for each transaction proportional to the amount traded $f_i(x_i) = \begin{cases} \xi_i(x_i)|x_i|, \ x_i \neq 0 \\ 0, \ x_i = 0 \end{cases}$, $i = \overline{1,n}$, where $\xi_i(x_i) = \begin{cases} \xi_i^{buy}, \ x_i > 0 \\ -\xi_i^{sell}, \ x_i < 0. \end{cases}$

Here $\xi_i^{buy} > 0$ and $\xi_i^{sell} > 0$ are the cost rates associated with buying and selling asset $i = \overline{1,n}$. We will consider a model that includes fixed plus linear costs, but our method is readily extended to handle more complex transaction cost functions. In this case, the transaction cost function is given by

$f_i(x_i) = \begin{cases} \xi_i(x_i)|x_i| + \psi_i(x_i), \ x_i \neq 0 \\ 0, \ x_i = 0 \end{cases}$, $i = \overline{1,n}$, where $\psi_i(x_i) = \begin{cases} \psi_i^{buy}, \ x_i > 0 \\ \psi_i^{sell}, \ x_i < 0 \end{cases}$

and $\psi_i^{buy} > 0$ and $\psi_i^{sell} > 0$ are the fixed costs associated with buying and selling asset $i=\overline{1,n}.$ Further comments on the transaction costs will be made in the next section.

Evidently the function f_i is not convex, unless the fixed costs are zero.

We assume from now on equal costs for buying and selling, the extension for non-symmetric costs being straightforward. The transaction cost function is then

$$f(x)=\sum_{i=1}^{n} f_i(x_i),\ f_i(x_i)=\begin{cases}\xi_i|x_i|+\psi_i,\ x_i\neq 0\\ 0,\ x_i=0\end{cases},\ i=\overline{1,n}.$$

In the general case, costs of this form lead to a hard combinatorial problem.

The simplest way to obtain an approximate solution is to ignore the fixed costs, and solve with $f_i(x_i)=\xi_i|x_i|$. If ψ_i are very small, this may lead to an acceptable approximation. In general, however, it will generate inefficient solutions with too many transactions. Note that if this approach is taken and the solution is computed disregarding the fixed costs, some margin must be added to the budget constraint to allow for the payment of the fixed costs.

On the other hand, by considering the fixed costs, we discourage trading small amounts of a large number of assets. Thus, we obtain a sparse vector of trades; i.e., one that has many zero entries. This means most of the trading will be concentrated in a few assets, which is a desirable property.

We assume that lower and upper bounds for x_i are known i.e., there exist m_i and M_i such that $m_i\le x_i\le M_i.$ We denote by g_i the convex envelope of $f_i,$ which is the largest convex function which is lower or equal to f_i in the interval $[m_i,M_i]$. For $m_i\neq 0$ and $M_i\neq 0,$ the function g_i is given by $g_i(x_i)=\begin{cases}\left(\frac{b_i}{m(x_i)}+a_i\right)|x_i|,\ x_i\neq 0\\ 0,\ x_i=0\end{cases},\ i=\overline{1,n},$ where $m(x_i)=\begin{cases}M_i,\ x_i>0\\ m_i,\ x_i<0\end{cases}.$ Using g_i for f_i relaxes the budget constraint in the sense that it enlarges the search set. Following the approach in Lobo et al. (2007) we consider the portfolio selection problem with g_i replaced for f_i,

$$(PP')\ \begin{cases}\max r^T(w+x)\\ s.t.\ \sum_{i=1}^{n} x_i+g(x)\le 0\\ w+x\in X,\end{cases}$$

where $g(x)=\sum_{i=1}^{n} g_i(x_i).$ This corresponds to optimizing the same objective, the expected end of period wealth, subject to the same portfolio constraints, but with a looser budget constraint. Therefore, the optimal value of (PP') is an upper bound on the optimal value of the unmodified problem $(PP).$ Since the problem (PP')

is convex, we can compute its optimal solution, and hence the upper bound on the optimal value of the original problem (PP), very efficiently.

The algorithm

Step 1. Building the initial portfolio.

Apply clustering techniques, in order to find the composition of the initial portfolio.

Step 2. Initialization.

Set $k = 0$.

Initialize θ and δ.

Step 3. Solving the problem (PP')

Solve the convex relaxed problem (PP') and let $x^1 = \left(x_1^1, ..., x_n^1\right)$ be the optimal solution of this problem.

Step 4. Solving the modified portfolio selection problem (PP'_{mod})

Set $k=k+1$.

Solve the modified portfolio selection problem (PP'_{mod})

$$(PP'_{\text{mod}}) \begin{cases} \max r^T(w+x) \\ s.t. \sum_{i=1}^{n} \left(x_i + g_i^k(x_i)\right) \leq 0 \\ w + x \in X, \end{cases}$$

where $g^k(x) = \sum_{i=1}^{n} g_i^k(x_i)$ and $g_i^k(x_i) = \left(\dfrac{\psi_i}{\left|x_i^{k-1}\right| + \theta} + \xi_i\right)|x_i|, i = \overline{1,n}.$

The optimal solution of this problem is denoted by $x^k = \left(x_1^k, ..., x_n^k\right)$.

Step 5. Checking stopping condition.

If $\left\|x^k - x^{k-1}\right\|_\infty < \delta$, set $x^* = x^k$ and the algorithm stops. Otherwise, go to Step 4.

Remark. In Meyer 1974 the convergence of a class of algorithms that includes our proposed algorithm is established.

In the next section we consider an example of applying the stock selection phase for determining the initial composition of the portfolio in the particular case of Bucharest Stock Exchange.

4. FUNDAMENTAL ANALYSIS AND TECHNICAL ANALYSIS FOR BUCHAREST STOCK EXCHANGE. CASE STUDY

Applying clustering techniques in the stock selection phase must be preceded by two methods specific to financial analysis: fundamental analysis and technical analysis.

Fundamental analysis tries to find a more realistic value for the assets, based on the information regarding the financial situation of the company, the domain of its investments and the goods owned. The goal is to select the assets whose market price is smaller than the price obtained applying fundamental analysis. The result is creating the opportunity that market will recognize the smaller value of these assets in the future and consequently their price will rise.

In other words, fundamental analysis tries to predict the way of evolution of the asset's price during medium and long term, starting from the past and present achievements of the company and from the estimations about their future.

Technical analysis studies the evolution of the transaction price. Technical analysis starts from the idea that all the relevant information regarding the market is included in the price, excepting the events like natural disasters, wars. Technical analysis captures very well the psychology of the investors.

Depending on the access to the necessary information, on the time dedicated to the analysis and to the investment strategy chosen, each investor chooses the type of analysis which is more suitable for him. Thus, speculators use technical analysis and long term investors use fundamental analysis. It is important that the two techniques are used together, in order to confirm the decision of buying or selling based on this kind of analysis.

We have collected information concerning a number of 40 assets, representing the assets from Bucharest Stock Exchange which had profit during the last two years. The goal is to find the similarities and differences between the assets and to build a diversified portfolio. Since Bucharest Stock Exchange is not mature enough, we can not afford to use only one financial index, such as, for example, the closing price.

As we will take into account more characteristics for each asset, we will use data analysis. We consider the values of seven financial indices for each asset. Four of these indices are specific to fundamental analysis: market capitalization, the ratio between the current price and the net profit net per share in the last year, the turnover evolution between 2006 and 2007, the turnover evolution between 2005 and 2006 and the profit evolution between 2006 and 2007. The other three are characteristic to technical analysis: the ratio between minimum price during the last year and current price, the ratio between maximum price during the last year and current price and the ratio between transaction value and capitalization.

Table 1 contains, for each of the 40 assets considered, the values of seven characteristics: market capitalization $\left(P/Bv\right)$, the ratio between the current price and the net profit net per share in the last year (*PER*), the turnover evolution between 2006 and 2007 $\left(T_{2007}/T_{2006}\right)$, the turnover evolution between 2005 and 2006 $\left(T_{2006}/T_{2005}\right)$, the profit evolution between 2006 and 2007, the ratio between minimum price during the last year and current price $\left(P_{\min}/P\right)$, the ratio between maximum price during the last year and current price $\left(P_{\max}/P\right)$ and the ratio between the transaction value and the capitalization $\left(TrV/Cap\right)$. We have used data available on Bucharest Stock Exchange site [18] from January 9, 2008.

Table 1. The values of seven characteristics considered for each asset

Nr.	Symbol	*P/Bv*	*PER*	Turnover evolution		Price evolution		*TrV /Cap*
				$\left(\frac{T_{2007}}{T_{2006}}\right)$	$\left(\frac{T_{2006}}{T_{2005}}\right)$	$\left(\frac{P_{\min}}{P}\right)$	$\left(\frac{P_{\max}}{P}\right)$	
1	ALR	0.55	2.51	0.93	1.36	0.97	6.29	0.016
2	ALT	0.14	1.86	1	1.38	0.86	5.76	0.61
3	ALU	0.44	2.03	1.14	1.21	0.83	8.09	0.38
4	APC	0.61	8.73	1.54	0.98	0.93	2.21	0.08
5	ARS	0.88	5.19	0.99	1.37	0.86	28	0.0345
6	ATB	0.73	6.84	1.17	1.19	0.87	4.83	0.3
7	BCC	1.79	107.26	1.61	1.69	0.912	7.03	0.53
8	BIO	0.6	6.35	1.08	1.14	0.81	6.31	0.92
9	BRD	1.66	4.51	1.8	1.3	0.79	2.81	0.12
10	BRK	0.21	3.64	1.92	1.44	0.81	20.5	5.22
11	BRM	0.34	10.8	0.92	0.8	0.912	7.03	0.27
12	CBC	0.53	21.87	0.96	1.06	0.875	1.97	0.03
13	CMF	3.54	18.9	1.08	1.16	0.97	1.31	0.03
14	CMP	0.09	7.96	1.25	0.7	0.79	12.39	0.52
15	COMI	0.71	5.4	2.05	0.88	0.73	3.06	0.55
16	COTR	0.73	2.84	1.97	1.25	0.86	13.8	0.15
17	DAFR	0.27	2.75	1.94	1.12	1	10	1.2
18	EFO	0.37	12.2	1.07	1.05	0.8	2.9	0.04
19	ENP	0.19	13.66	1.16	1.35	1	6.35	0.37
20	IMP	0.19	6.74	0.93	0.91	0.66	12.4	1.67
21	MECF	0.49	8.36	1.23	0.03	0.91	5.37	0.4
22	OIL	0.63	61.98	1.02	0.94	0.767	4.64	0.07
23	PPL	1.71	26.34	1.01	1.06	0.89	1.18	0.17
24	PTR	1.25	4.19	1.4	1.14	0.51	4.75	0.37
25	SCD	0.49	7.97	0.79	1.26	0.88	3.44	0.2
26	SIF1	0.75	2.67	1.18	1.46	0.66	3.16	1.19
27	SIF2	0.77	3.55	0.8	1.6	0.71	5.14	2.3
28	SIF3	0.53	4.33	1.47	1.57	0.82	6.4	1.35
29	SIF4	0.25	3.73	1.06	1.39	0.31	1.51	0.99
30	SIF5	0.64	4.61	1.37	1.69	0.77	5.86	2.16
31	SNO	0.68	5.08	1.11	1.2	0.9	3.37	0.06
32	SNP	0.82	4.32	0.92	1.3	0.67	2.91	0.05
33	SOCP	0.94	6.83	1.12	0.92	0.97	2.65	1.11
34	SRT	0.23	3.76	0.76	0.84	0.83	2.09	0.92
35	TEL	0.38	5.84	0.95	1.41	0.9	2.83	1.68
36	TGN	0.94	6.35	1.14	1.18	0.77	2.24	0.11
37	TUFE	0.55	10.23	1.16	1.05	0.86	4.45	0.08
38	UAM	0.31	36.8	0.74	1.06	0.7	2.23	0.14
39	VESY	0.3	11.92	0.98	0.98	0.77	2.67	0.67
40	VNC	0.49	5.98	1.32	1.28	0.91	3.22	0.095

We will apply data analysis techniques to find the similarities and differences between the shares from Bucharest Stock Exchange, using StatistiXL 1.8. We will give the interpretation of the results obtained using Principal Component Analysis.

Table 2 contains: the Principal Components (PCs), the eigenvalues of the correlation matrix between the variables, the percentage of the variance explained by each PC and the cumulated variance. The eigenvalue corresponding to the first principal component is 1.824, which represents 26.063% of the sum of the eigenvalues, hence PC 1 explains 26.063% from the total variance. The second principal component corresponds to an eigenvalue of 1.441, which is 20.582% of the total. Cumulatively, PC 1 and PC 2 explain 46.646% of the total variance. The third principal component corresponds to an eigenvalue of 1.167, which is 16.677% of the total. Cumulatively, the first three PC explain 63.323% of the total variance. The succeeding PCs explain the remaining 36.677% of the variance.

Table 2. The eigenvalues corresponding to PC, the percentage of the variance explained by each PC and their cumulative variation percentage

Explained Variance (Eigenvalues)							
Value	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
Eigenvalue	1.824	1.441	1.167	0.804	0.723	0.562	0.479
% of Variance	26.063	20.582	16.677	11.481	10.328	8.024	6.844
Cumulative %	26.063	46.646	63.323	74.804	85.131	93.156	100.000

The graph represented in Figure 1 reveals the contributions of the original variables to both PC 1 (Principal Component 1) and PC 2. The graph shows the relationship between PC 1 (which explains 26.063% of the total variance) and PC 2 (which explains 20.582% of the total variance). For example, TrV / Cap has a small contribution to PC 2, as seen from its small positive component on the PC 2 axis, but a large contribution to PC 1, as seen from its large positive component on the PC 1 axis. These graphical representations of the contributions of the original variables to the PC axes are normalized to a convenient scale and reflect the relative values of the component score coefficients.

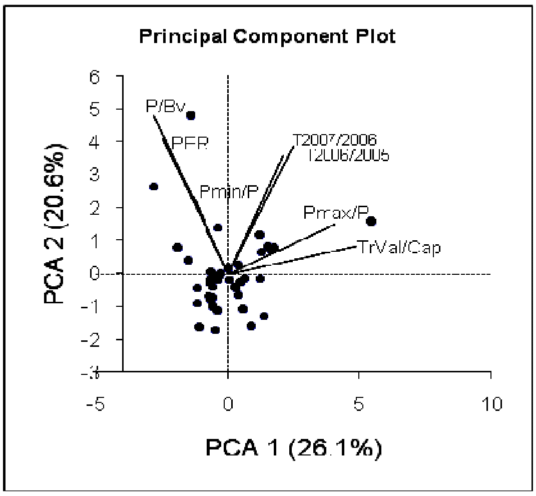

Figure 1. The contributions of each variable to the principal components

Table 3 contains the principal component score coefficients, which represent the correlations between initial variables and principal components. These numbers provide the coefficients by which the principal component axes are defined and from which the actual principal component scores can be computed for each case.

Table 3. Component Loadings (correlations between initial variables and principal components)

Variable	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
P/Bv	-0.471	0.658	-0.142	-0.277	0.197	0.291	0.353
PER	-0.416	0.563	-0.134	0.393	-0.558	-0.162	-0.012
*T*2007/2006	0.409	0.532	0.301	-0.584	-0.207	-0.140	-0.234
*T*2006/2005	0.343	0.497	-0.581	0.215	0.391	-0.021	-0.313
Pmin/P	-0.224	0.326	0.734	0.283	0.397	-0.258	0.008
Pmax/P	0.666	0.207	0.370	0.316	-0.137	0.508	0.007
TrVal/Cap	0.807	0.113	-0.159	0.064	-0.011	-0.325	0.448

The plot represented in Figure 2 shows the successive eigenvalues. It can be used to establish how many principal component axes should be considered as useful. As the first five eigenvalues decline quickly and the remaining eigenvalues form a relatively flat curve, the first five eigenvalues and their corresponding principal components should be retained.

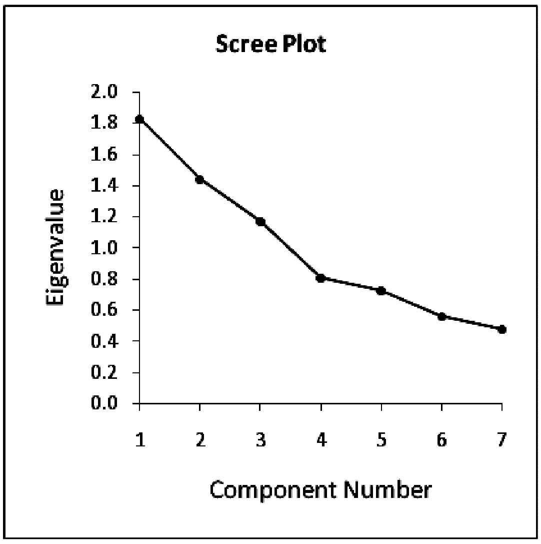

Figure 2. The eigenvalues plot

Figure 3 contains the tree obtained applying the clustering technique, which models the classification process. The tree graph (dendrogram) is a common means of graphically summarizing the clustering pattern. The dendrogram usually starts with all assets as separate clusters and shows the combination of fusions back to a single "root". The order of individuals shown in the dendrogram is that order in which the groups enter the clustering. The assets belonging to the same cluster are similar regarding the characteristics considered.

In order to build a diversified portfolio, we will first chose the number n of clusters that will be taken into account. We will take one asset from each cluster and we will obtain the initial portfolio.

As a consequence of applying clustering techniques, we have selected a portfolio consisting of $n = 10$ assets, each of them representing different classes: PPL, SCD, BRD, SIF1, CMP, BRK, ARS, UAM, OIL and BCC. We will use this initial portfolio to solve the optimal portfolio problem.

5. CONCLUSIONS

Many mathematical methods for the portfolio optimization problem for the real world use a certain composition of the initial portfolio without specifying how the risky assets were chosen.

In this paper we have developed an original algorithm that take into consideration both aspects: the prior stock selection and the asset allocation phase.

We apply the proposed techniques to solve a case study using, besides data analysis methods, tools belonging to financial analysis, such as fundamental analysis and technical analysis.

Computational results show that the approach using classification provides useful results for the portfolio optimization problem. The proposed procedure allows us to build a diversified portfolio and improve the optimization algorithm.

Acknowledgment

This research was supported by the Romanian National University Research Council (CNCSIS), PN II, Program IDEI, Grant No. 1778.

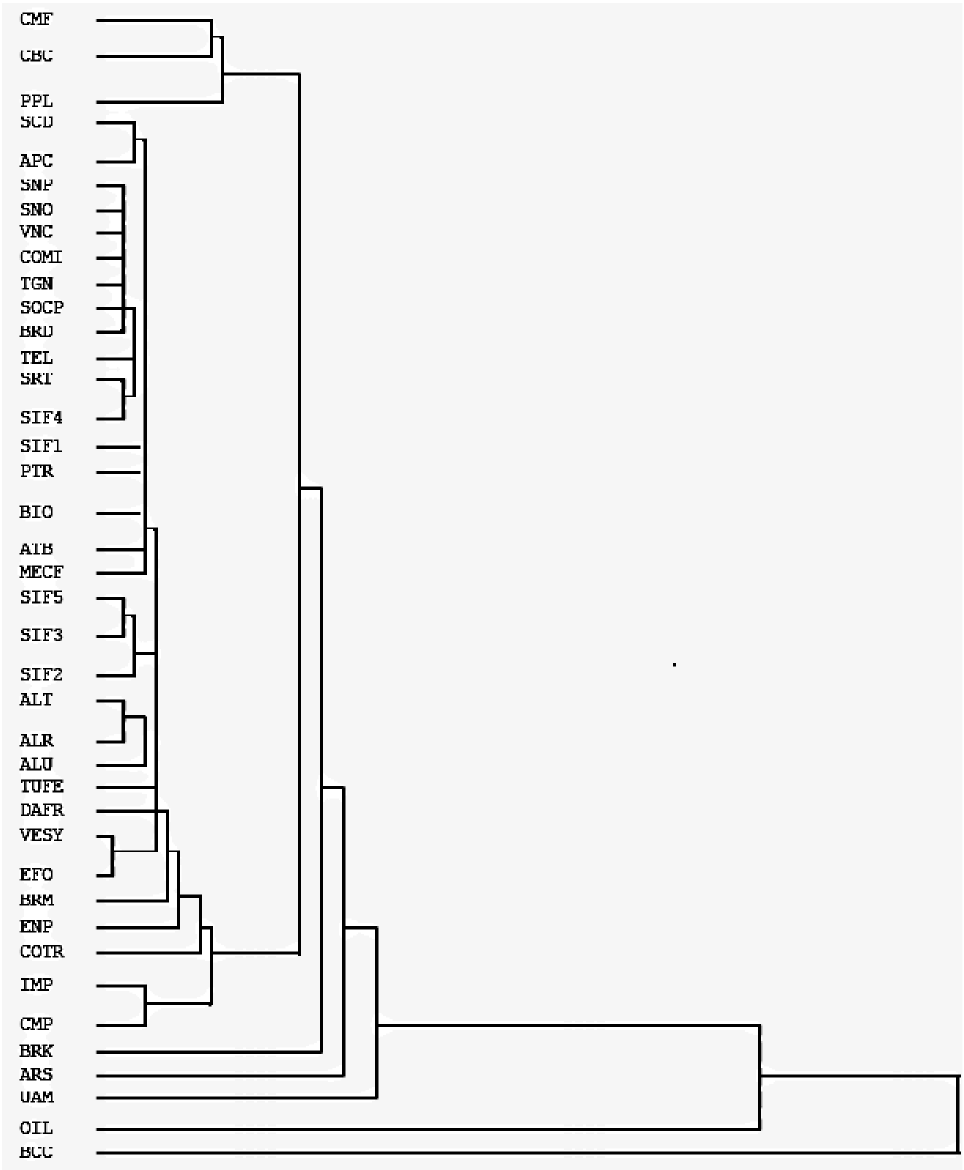

Figure 3. The dendogram obtained applying clustering techniques

REFERENCES

[1] **Agapie, A., Bădin, L. (2009),** ***Confidence Intervals for the Guesstimation Algorithm: A Bootstrap Approach.*** Economic Computation and Economic Cybernetics Studies and Research, 3, 45, 167-178, ASE Publishing House, Bucharest;
[2] **Armeanu, D., Balu, F.O. (2008),** ***Testing the Efficiency of Markowitz Model on Bucharest Stock Exchange.*** Economic Computation and Economic Cybernetics Studies and Research, 1-2, 41, 201-218, ASE Publishing House, Bucharest;
[3] **Best, M.J., Hlouskova, J. (2003),** ***Portfolio Selection and Transaction Costs.*** Computational Optimization and Applications, 24, 95-116;
[4] **Birge, J.R. (2008),** ***Optimization Methods in Dynamic Portfolio Management.*** Handbooks in OR & MS, J.R. Birge and V. Linetsky (Eds.), 15, 845-866;
[5] **Blog, B., Van Der Hoek, G., Rinnooy, A.H.G., Timmer, G.T. (1983)** ***The Optimal Selection of Small Portfolios.*** Management Science, 29(7), 792-798;
[6] **Brida, J.G., Risso, W. (2007),** ***Dynamics and Structure of the Main Italian Companies.*** SSRN: http://ssrn.com/.
[7] **Brida, J.G., Risso, W. (2007),** ***Hierarchical Structure of the German Stock Market***. SSRN: http://ssrn.com/.
[8] **Fulga, C. (2009),** ***Dynamic Model for Portfolio Optimization.*** Lecture Notes in Engineering and Computer Science, Eds.: Ao, S.I., Castillo, O., Douglas, C., Dagan Feng, D., Lee, J.A., Newswood Ltd. I.A.Eng., 2, 2081-2086;
[9] **Fulga, C. (2009),** ***Dynamic Portfolio Optimization for Utility-Based Models.*** Proceedings of the 2009 International Conference on Information and Financial Engineering ICIFE 2009, IEEE Computer Society Press, 117-121;
[10] **Fulga, C., Pop, B. (2008),** ***Single Period Portfolio Optimization with Fuzzy Transaction Costs.*** Proceedings of EURO Conference Continuous Optimization and Knowledge-Based Technologies EurOPT 2008, 125-131;
[11] **Fulga, C. (2006),** ***A Sequential Quadratic Programming Techniques with Two-Parameter Penalty Function.*** Bulletin Mathematique de la Societe des Sciences Mathematiques de Roumanie, 49(97), 4, 335-343;
[12] **Kaski, K., Onnela, J.P., Chakraborti, A. (2009),** ***Dynamics of Market Correlations: Taxonomy and Portfolio Analysis.*** Physical Review E, 68;
[13] **Kellerer, H., Mansini, R., Speranza, M.G. (2000),** ***Selecting Portfolios with Fixed Costs and Minimum Transaction Lots.*** Annals of Operations Research, 99, 287-304;
[14] **Konno, H., Wijayanayake, A. (2001),** ***Portfolio Optimization Problem under Concave Transaction Costs and Minimal Transaction Unit Constraints.*** Mathematical Programming, 89, 233-350;
[15] **Mantegna, R.N. (1999),** ***Hierarchical Structure in Financial Markets.*** The European Physical Journal B, Vol. 11, 193-197;

[16] **Marinescu, D., Marin, D. (2009).** ***Risk and Uncertainty:Analyzing the Income and Substitution Effects.*** Economic Computation and Economic Cybernetics Studies and Research, 1, 43, 41-54; ASE Publishing House;
[17] **Meyer, R.R. (1974),** ***Sufficient Conditions for the Convergence of Monotonic Mathematical Programming*.** Algorithms, Journal of Computer and System Sciences, 12, 108-121;
[18] **Schattman, J.B. (2000),** ***Portfolio Selection under Nonconvex Transaction Costs and Capital Gain Taxes*.** Ph.D. Thesis, Rutgers University;
[19] **Ştefănescu, M.V, Ferrara, M., Dedu, S. (2008).** ***Algorithms for Hierarchical Classification with Applications in Portfolio Management*.** Economic Computation and Economic Cybernetics Studies and Research, 3-4, 42, 109-122, ASE Publishing House, Bucharest;
[20] **Topaloglou, N., Vladimirou, H., Zenios, S. A. (2008),** ***A Dynamic Stochastic Programming Model for International Portfolio Management*.** European Journal of Operational Research 185 (2008) 1501–1524.
[21] www.bvb.ro