USAGE OF THE MAIN COMPONENTS ANALYSIS IN THE MANAGEMENT OF THE INVESTMENT PORTFOLIO

Dan ARMEANU, Ph.D.Professor,
E-mail: darmeanu@yahoo.com
Academy of Economic Studies
Bucharest

Andreea NEGRU, Ph.D.Student,
Academy of Economic Studies
Bucharest

Abstract:

When managing investment portfolios on integrated capital markets, beyond the models put forth by the modern portfolio theory, (the Markowitz model, the CML model, the CAPM model, the Treynor-Black model and more), one can successfully resort to the statistical and mathematical tools made available by the multidimensional data analysis. The reason why we shall use those tools in our analysis is simple: they make it possible to reduce the number of variables in the analysis while preserving much of the information included in the initial data (analysis of the main components); they also outline the extent to which common, latent factors as well as the uncommon factors act upon the respective variables (factorial analysis) and they make it possible to create relevant categories for all observations (shape recognition methods and techniques). In this article we shall deploy informational synthesis instruments (the analysis of main components and factorial analysis) as well as methods and techniques of recognizing shapes (cluster analysis and discriminating analysis) to calculate the financial strength of 20 companies whose shares are describes as "blue chips" on the German, Polish and Romanian stock markets (being included into the DAX, WIG 20 and BET indexes). Namely, they are: ADIDAS AG, BMW AG, COMMERZBANK AG, DAIMLER AG, LUFTHANSA AG, METRO AG, SIEMENS AG, ASSECO POLAND SA, CEZ GROUP, GLOBAL TRADE CENTER SA, LOTOS GROUP, PKO BANK POLSKI, TVN SA, BRD SA, BANCA TRANSILVANIA SA, ALRO SLATINA SA, PETROM SA, ROMPETROL SA, CNTEE TRANSELECTRICA SA and SNTGN TRANSGAZ SA. When managing the investment portfolios on integrated capital markets, the usage of those methods generate relevant classifications of companies that are listed on the stock market and also make it possible to develop very useful prediction instruments.

Numerous economy and financial issues require the detection of functional relationships that occur between several predictor variables which make up *the initial causal zone*, as it is called in the field of data analysis. The higher the amount of variables, the more difficult it becomes to analyze the initial zone, and the more difficult it is to conduct the statistical data analysis. When there are numerous initial variables, it is difficult to gather raw data for the analysis and the calculation becomes much more complex, while it is more likely that the original variables become inter-correlated, which can seriously impair the analysis.

The analysis of the main components is a multidimensional analysis technique with a high degree of applicability in portfolio management on integrated capital markets. The major upside of this model is the decrease of the initial causal zone to an equivalent zone of considerably lower size – the core zone. Switching between the two zones can be done by maximizing the information retained from the first zone, that is, by minimizing the otherwise inevitable loss of information.

For the following data analysis, we have considered seven financial indicators which describe the mentioned companies: the total assets, (AT), the net turnover (CA), operational profit (EBIT), the capital flow (CF), the net result (PN), total debt (DAT TOT) and average stock capitalization (CBM), all indicators applying to the 2010 financial year. The data originates from the companies financial reports, according to the International Financial Reporting Standards, (IFRS) passed by the European Union.

To begin with, we shall present descriptive statistics (the average and the square average margin error) for indicators included in the analysis.

Indicator	Average	Sqr. Avg. Error
AT	64,018.11	167,438.90
CA	19,766.26	30,212.07
EBIT	1,626.20	2,214.13
CF	1,329.14	2,914.43
PN	994.50	1,427.61
DAT TOT	55,785.05	160,542.94
CBM	11,404.46	18,333.43

Table 1. ***Average and square average error for indicators taken into account during the analysis of the main components***
Source: own calculations.

Of course, the higher the square average error, the more the analyzed companies will differentiate themselves from the others, according to the respective indicator. As the previous table indicates, the standard error is high for all seven variables (especially when relating to the average), which means that the market is a complex system, with rich informational contents, where the structure of causal interconnections is very complex, while being hard to identify and quantify in the original seven-dimensions zone.

The correlation matrix of the seven original variables is included in the next table:

Indicator	AT	CA	EBIT	CF	PN	DAT TOT	CBM
AT	1.00	0.26	0.46	-0.12	0.31	1.00	0.25
CA	0.26	1.00	0.90	0.73	0.89	0.21	0.89
EBIT	0.46	0.90	1.00	0.66	0.98	0.41	0.90
CF	-0.12	0.73	0.66	1.00	0.74	-0.17	0.82
PN	0.31	0.89	0.98	0.74	1.00	0.26	0.91
DAT TOT	1.00	0.21	0.41	-0.17	0.26	1.00	0.20
CBM	0.25	0.89	0.90	0.82	0.91	0.20	1.00

Table 2. ***The correlation matrix of the seven variables in the initial causal zone.***
Source: own calculation.

As we expected, the original variables are strongly interrelated, which brings a substantial contribution to decreasing the meaning of those variables, on the one hand, and outlines the presence of some informational redundancies, on the other hand: a significant amount of information is dispersed among the links between variables. Our analytic discourse that uses the method of main components concerns both the decrease of the dimensionality of the original zone, and the elimination of the informational redundancies.

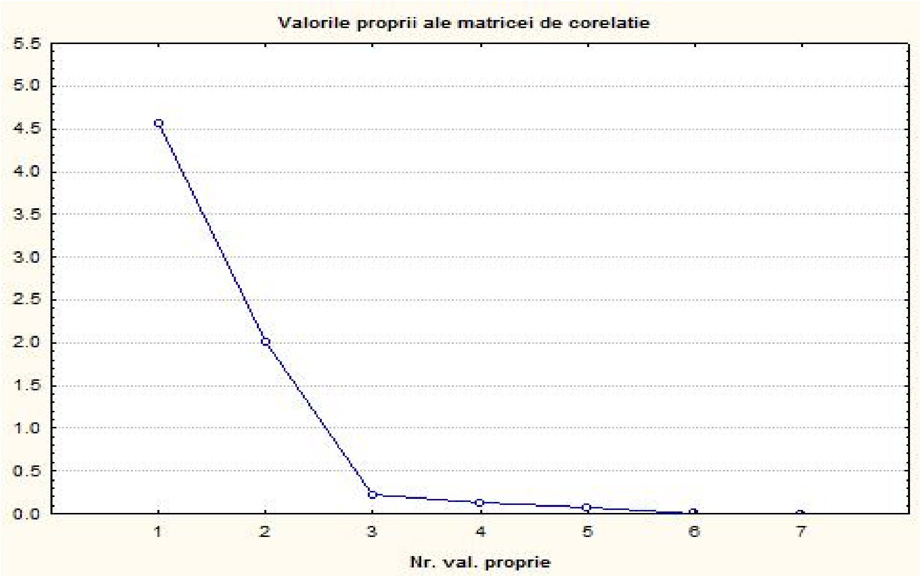
We shall calculate the own values of the correlation matrix [1]. We are interested only in the own values that are higher than one[2], because only those main components are relevant which have a higher informational content than the initial variables. The table below presents the two values of the Σ matrix:

No.	Own value	% Explained variant	Own cumulated value	% Total variance
1	4.5681	65.2587	4.5681	65.2587
2	2.0092	28.7024	6.5773	93.9610

Table 3. ***The own higher than one values of the Σ correlation matrix.***
Source: own calculations.

The graph below depicts all seven own values of the correlation matrix, among which only the first two are representative (that is, they are higher than one), while the rest tend to reach the value of zero.

[1]The original variables will be taken into account only as standard variables, which makes the co-variant matrix to coincide with the correlation matrix.
[2]The Kaiser – Guttman criteria.



Graph 1. ***Graphic representation of the correlation matrix's own values***
Source: own calculation.

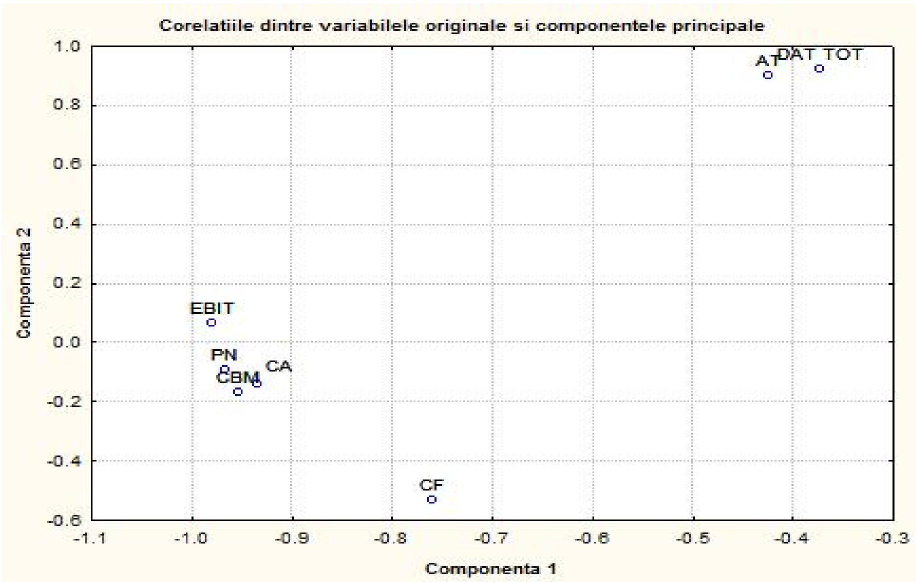
Information presented earlier indicates that there are only own values of the correlation matrix that are higher than the unit an therefore we shall have just as many main components. They explain up to almost 94 percent the variability inside the initial causal zone, which means that the reduction of the analyzed zone from seven to only two variables was possible with an informational loss of about 6 percent, a percentage that can be considered as very good. We also notice that the first main component individually recovers over 65% of the initial causal zone's information, and thus can be used for listing the issuers that are being analyzed.

We then calculated the factor matrix for the resulting main two components. The factor matrix, whose elements - the intensities of the factors – represent the correlation indicators between the initial variables and the main components, is presented in the table below:

Indicator	Comp. 1	Comp. 2
AT	(0.4234)	0.9009
CA	(0.9335)	(0.1406)
EBIT	(0.9795)	0.0660
CF	(0.7592)	(0.5339)
PN	(0.9671)	(0.0938)
DAT TOT	(0.3723)	0.9223
CBM	(0.9527)	(0.1700)
Expl. Var.	4.5681	2.0092
% Total	0.6526	0.2870

Table 4. ***Components of the factor matrix.***
Source: own calculations.

As one can notice, the first component recovers 65,26% of the initial zone's information. It is strongly and negatively correlated with the net turnover indicators, operational profit, capital flow, net profit and average capitalization, thus providing information about the volume, profitability and the ability of companies listed at exchanges in Germania, Poland and Romania to generate cash through their activity, and also about their market value. The second main component carries 28,70% of the variability of the original zone, being strongly correlated, in a positive sense, with the total assets and total debt balance sheet indicators. Therefore, the second component provides information on the size of the analyzed companies and the structure of their financing (through their degree of indebtedness). The next graph illustrates the correlations between the original variable and the main components:



Graph 2. ***Correlations between the original variable and the main components***
Source: own calculation.

Coefficients of linear combinations pertaining to the two main components (own vectors of the correlation matrix) are presented in the table below.

Indicator	Comp. 1	Comp. 2
AT	(0.0927)	0.4484
CA	(0.2044)	(0.0700)
EBIT	(0.2144)	0.0328
CF	(0.1662)	(0.2658)
PN	(0.2117)	(0.0467)
DAT TOT	(0.0815)	0.4590
CBM	(0.2086)	(0.0846)

Table 5. ***Coefficients of linear combinations that define the main components.***
Source: own calculations.

Those coefficients are important because they will be used to calculate the scores (coordinates) of objects in the new zone, the core zone.

The following table presents the coordinates of objects in the core zone:

Issuer	Comp. 1	Comp. 2
ADIDAS	0.2798	(0.2191)
BMW	(1.0169)	0.2258
COMMERZBANK	(0.8701)	4.0903
DAIMLER	(2.5773)	(0.7591)
LUFTHANSA	(0.0632)	(0.3603)
METRO	(0.4774)	(0.3913)
SIEMENS	(2.3814)	(1.0104)
ASSECO	0.6655	(0.1131)
CEZ	(0.1390)	(0.3549)
GTC	0.6704	(0.1035)
LOTOS	0.6022	(0.1228)
PKO	0.2873	0.0544
TVN	0.6680	(0.1154)
BRD	0.5395	(0.0554)
BT	0.6824	(0.0937)
ALRO	0.6895	(0.1077)
PETROM	0.3977	(0.2135)
ROMPETROL	0.6944	(0.1184)
TEL	0.6837	(0.1136)
TGN	0.6651	(0.1184)

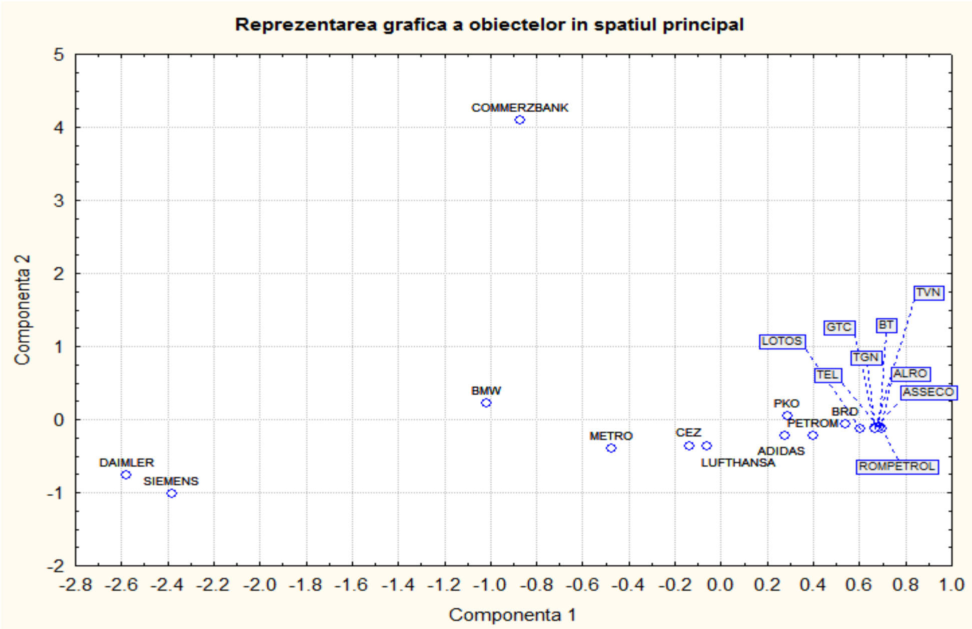
Table 6. ***Coordinates of observations in the new core zone***

Source: own calculations.

As we indicated earlier, the first main component is strongly and negatively correlated with net turnover, operational profit, net profit, operational cash flow and average stock market capitalization indicators. That occurrence was expected because, given the structure of the correlation matrix in table 2, those variables are strongly correlated. In fact, also at an intuitive level it is natural for a company with a large business volume (as indicated by turnover) should have a large operational profit higher net results, which is, of course, beneficial for the company's market value.

Classifying companies that are analyzed according to this first main component is relevant because it makes it possible to simultaneously evaluate the company's business volume (as indicated by the net turnover), the company's ability to generate cash flow from its operational activities (by correlation with the operational cash flow indicator) and the company's profitability (worth on the whole, according to the net profit, as well as at the operational level, through the results of the operation). The fact that the first main component is oppositely correlated with all those indicators makes companies that have the lowest scores for this component should have the highest performance according to indicators mentioned before – in other words, the lower the score of the first main component, the higher the likelihood of dealing with a company that has a large business s volume, high profitability and a great ability to generate cash flow through operational activities. That category will include “blue chips” share listed on the stock exchange.

The second main component is strongly correlated in a positive sense with the total assets, and total debt indicators; as such, the items that register high scores for this component are large companies, with high indebtedness and very likely high leverage. The next graph indicates the distribution of objects subjected to analysis in the main area, where the horizontal axis represents the first component, while the vertical axis represents the second main component.



Graph 3. ***Graphical representation of items in the main zone***

Source: own calculation.

Following the classification of companies according to scores registered for the two main components, one can notice that there are two companies that clearly stand out from the other analyzed companies. They are Daimler and Siemens. Those companies are among the largest at a European level (both in terms of total assets and turnover), they are described by excellent levels of profitability, in terms of both economic and financial profitability (ROA at the level of 3-4% and ROE at the level of 12-13%, high values for developed western markets – in this case, Germany), and their market values are very high, representing 150-200% of the net book value of those companies. Also the quality of profits reported by the two companies is remarkable, the net profit being supported by the net operational cash flow reported by those companies. Meanwhile, the two companies are described by high levels of indebtedness, of over 70%, which means that they benefitted from a significant leveraging effect on their own capital assets, which is reflected in the considerably high rates of financial profitability reported by the respective companies.

The second identifiable class includes Commerzbank, a company described by a high business volume, but with low profitability (ROE of about 5%, ROA below 0.20%), with profits that are not sustained by net cash reserves and with a high level of indebtedness (over 96%) and with a relatively low market value, amounting to the level of 50 percent of the net book value of the company.

The third category of companies that can be reconstructed based on the analysis of main components includes BMW, Metro, CEZ and Lufthansa. They are all listed on western markets, they are all very large (BMW), very profitable (ROE of 14-21%, ROA of 3-9%) and with healthy operational income (except for BMW). They are also relatively deeply indebted – between 60-80% - and have a market value that is usually much higher than the worth of their own funds, except for CEZ, which however make up for the difference through its excellent profitability. By using the method of main components we can outline for instance the fact that CEZ shares should be purchased by investors, because of its outstanding financial health, while the market price of its shares is very low, with excellent potential to grow.

Afterwards, we can make up a group of companies including Adidas, PKO, Petrom, BRD, large companies listed on emerging markets, with high business volumes on those markets (Adidas), described by excellent levels of profitability (ROA 12-18%), high quality, levels of indebtedness ranging from 46 to 88% (high values of the indicators in the cases of PKO and BRD lead to a relatively low rate of ROA economic profitability) and with market capitalizations ranging between 140 and 230% of the companies' net book value. The group includes companies with interesting growth prospects, which investors need to monitor carefully.

And finally, the last category, the fifth, includes small to medium companies on the emerging capital markets, described by profitability levels ranging from good (Rompetrol, GTC, TNV) to very good (ALRO, Transgaz, Asseco Poland), generally moderate levels of indebtedness (below 60%) and good market capitalization (except for Lotos, a company which deserves the "buy" recommendation). Those companies too are worth being monitored, with the mention that in their case, investors need to take into account a higher risk premium than in the case of other categories of companies previously analyzed, against the background of relatively high volatility of the capital markets where companies in this fifth category operate.

Returning to the analysis, we shall present the communalities (calculated as a sum of the square elements on the factor matrix lines) and their specificities (calculated as 1–communality) for the seven descriptive variables considered in the analysis:

Indicator	Communality	Specificity
AT	0.9909	0.0091
CA	0.8912	0.1088
EBIT	0.9638	0.0362
CF	0.8615	0.1385
PN	0.9442	0.0558
DAT TOT	0.9891	0.0109
CBM	0.9366	0.0634

Table 7. ***Communalities and specificities of the indicator variables***
Source: own calculations.

The date presented earlier indicates that common latent factors, which exert a general influence on the analyzed variables, explain the variation of indicators to a degree of over 86%. In the case of some of those variables (total assets, operational results, net worth and average market capitalization) the communality exceeds the level of 90%, which indicates that common factors are decisive, and the influence of specific factors is low. They explain less than 10% of the indicators' variation. Predictor variables mentioned earlier are influenced to a larger extent by the degree of economic development and the situation of the capital market, to give only two examples, than by the individual, company-specific factors. The only descriptive variables that are influenced to a larger extent by specific, directional factors are the net turnover and cash-flow generated by the operational activities, which is normal after all, given the fact that they are more sensitive to the company's field of operations and with the vision of the company's management (which results, among others, into specific business policy and company cash management)

As we have seen so far, the analysis of the main components turns out to be a very useful instrument for the informational synthesis and in eliminating informational redundancies. However, it also serves as a tool for classifying the items subjected to analysis. Thus, given the correlations established between the main components and the variables in the initial causal zone, we are capable to classify analyzed companies according to the main scores. The sense of the existing correlations leads to a primary important result of the main components analysis: companies that score low in the first main component well be described by a high business volume (sales), very good financial performances, an excellent ability to generate cash through their operational activities and by high market value and stock liquidity. In the same way, companies with high scores in the second main component will be the largest on the market and at the same time, they will have the highest level of indebtedness, with an important leverage effect on their own capital assets.

BIBLIOGRAPHY:

1. Benninga, Simon – **Financial modelling**, 3rd edition, *MIT Press*, 2008;
2. Bodie, Zvi; Kane, Alex; Marcus, Alan – ***Essentials of investments***, *The McGraw-Hill Companies,* 2003;
3. Damodaran, Aswath – **Estimating risk parameters** (2008), http://pages.stern.nyu.edu/~adamodar/;
4. Fama, E.F.; K.R. French, **The Cross-Section of Expected Stocks Returns**, *Journal of Finance, 67 (2)*, 1992, pp. 427–465;
5. Friedlob, George; Schleifer, Lydia – **Essentials of financial analysis**, *Wiley,* 2003;
6. Hitchner, James – **Financial valuation: applications and models**, *Wiley,* 2006;
7. Meucci, A.- **Risk and Asset Allocation,** *Springer*, 2005;
8. Ross, Stephen; Westerfield, Randolph; Jordan, Bradford – **Fundamentals of corporate finance**, 8th edition, *McGraw-Hill/Irwin*, 2007;
9. Rubinstein, Mark – **A history of the theory of investments**, *Wiley,* 2006;
10. Beste, A.; Leventhal, D.; Williams, J.; Qin Lu – "*The Markowitz Model. Selecting an Efficient Investment Portfolio*", Lafayette College, Mathematics REU Program, 2002.