EFFECTIVE PORTFOLIOS — ECONOMETRICS AND STATISTICS IN SEARCH OF PROFITABLE INVESTMENTS*

Urszula Skórnik-Pokarowska

Department of Econometrics and Computer Science, SGGW
Nowoursynowska 159, 02-787 Warsaw, Poland

e-mail: skornik@alpha.sggw.waw.pl

(Received April 19, 2005)

Methods of constructing effective portfolios based on mathematical statistics are presented and compared with methods based on econometrics, multi-criteria optimization and on-line investments.

PACS numbers: 89.75.–k

1. Introduction

The construction of a profitable investment portfolio is a difficult task and can by tackled from the point of view of various disciplines. There are many methods that allow constructing a portfolio, such as *e.g.*, Sharpe's model or Markowitz model. Most of them are viewed as multi-criteria optimization problems, (see [1,2]), or econometric problems as far as forecasting of future prices is concerned (*e.g.*, see [3,4]). The multi-criteria optimization approach as well as econometric models *e.g.*, GARCH (see [1, 2, 4, 5]) take usually into account historic share prices and through time series analysis evaluate risk and returns. Models built for forecasting allow short term investments only. In our search for profitable long term investment portfolio we will concentrate not only on the share prices but also on financial condition of the traded firm. We are convinced that a long term investment portfolio should consist of the best companies (see [2, 6, 7]). Therefore, the selection of companies that constitute portfolio will be based on analysis of their financial reports. This leads to statistical analysis that results in selecting the best companies from the economic view point. One cannot argue which methods are best. They can be used on different stages of portfolio construction.

* Presented at the First Polish Symposium on Econo- and Sociophysics, Warsaw, Poland, November 19–20, 2004.

2. Selection of companies

Warsaw Stock Exchange Index, WIG, comprises many companies, but some of them are not traded often enough for our needs; some are traded for a certain time and disappear from the Index. Therefore, we have chosen 15 companies within WIG-20 and have added to this number further 9 companies from WIG. The set includes Boryszew, Świecie, Żywiec, Orbis, Agora, Jutrzenka, PKNOrlen, Dębica, Kęty, Telmax, Sokołów, KGHM, Kable, TPSA, Comarch, Budimex, Prokom, Irena, PGF, Krosno, Vistula, Compland, Elektrim, Swarzędz.

3. Cluster analysis

For the chosen set of companies the clusters analysis was performed. Its aim was to distinguish groups of companies that are similar. The taxonomy based on share prices can be found in [8, 9]. Here the analysis was based on similarity measure between stocks characterized by 8 various economic indicators suggested in [10]: return on equity, return on assets, amount due turnover, inventory turnover, liabilities turnover, assets turnover, debt margin, profitability ratio (see Table I).

The first 7 ratios are connected with fundamental analysis of the investigated company, only one indicator, the ratio of profitability, was based on the stock price. The economic ratios were calculated based on financial reports for the year 2003. The profitability ratio was calculated based on daily stock prices in 2003. The ratio values build an observation matrix. Before any further analysis can be done, the observation matrix has to be normalized. Many different methods of normalization have been discussed in *e.g.*, [11–14]. First of all the ratios are stated in different scales. Moreover, some of them have disstimulating values, which means that high values are not desired from the point of view of a general characterization of the company. The normalization methods play an essential role when one intends to aggregate the variables, *i.e.*, instead of using 8 variables describing a company uses one variable, being a linear combination of the initial set of variables. The methods of ranking companies by aggregating ratios have been discussed *e.g.*, in [10]. The coefficients of linear combination of n ratios are calculated as

$$\frac{V_i}{\sum_{j=1}^{n} V_j}\,, \tag{1}$$

where V_i is the coefficient of variation of the i-th indicator. The method of aggregating ratios allows ranking the companies. Our aim is, however, to find groups of companies that are similar and to distinguish those coefficients that are responsible for similarities and dissimilarities between the companies.

TABLE I

Matrix of indicators for the year 2003.

	ROE	ROA	Amount due turnover	Inventory turnover	Liabilities turnover	Assets turnover	Debt margin	Profit-ability ratio
Agora	0.00	0.00	51.9	6.6	35.6	574.6	0.217	0.000
Boryszew	0.18	0.11	71.8	55.8	126.0	333.2	0.39	0.007
Budimex	0.09	0.03	86.5	14.9	85.6	226.5	0.599	0.000
Compland	0.05	0.02	114.4	33.0	94.5	278.7	0.587	0.000
Comarch	0.090	0.035	124.7	17.5	96.5	376.1	0.536	0.002
Dębica	0.166	0.105	84.6	45.8	89.1	290.3	0.365	0.003
Elektrim	0.000	−.057	77.9	8	231.3	972.8	0.889	0.002
Irena	0.062	0.041	152.4	59.5	96.4	430	0.343	0.001
Jutrzenka	0.057	0.046	95.8	39.1	45.4	260.7	0.193	0.002
Kable	0.049	0.020	81.3	35	259.7	442.3	0.586	0.004
Kęty	0.127	0.079	56.9	50.4	99.4	335.8	0.380	0.003
KGHM	0.103	0.047	41.1	64.9	148.5	660.3	0.539	0.001
Krosno	0.079	0.037	63.1	106.7	120.5	331.1	0.540	0.002
Orbis	0.037	0.025	22.1	7	124.4	1112.7	0.328	0.001
PGF	0.112	0.068	26.3	39.3	48.9	200.2	0.389	0.001
PKNOrlen	0.112	0.068	26.3	39.3	48.9	200.2	0.289	0.001
Prokom	0.006	0.003	129.9	4.6	181	540.1	0.455	0.001
Sokołów	0.029	0.016	38.6	18.3	66.3	197.3	0.443	0.001
Świecie	0.217	0.136	69.2	23.7	90.5	304.7	0.372	0.001
TPSA	0.053	0.022	63.7	2.7	211.7	868.9	0.583	0.000
Żywiec	0.185	0.122	33.3	15.9	118.3	486.4	0.338	0.000
Vistula	−0.193	−0.069	54.8	72.8	159.2	271.3	0.643	0.004
Swarzędz	1.317	0.111	336.1	20.8	2925.3	3474.1	0.916	0.002
Telmax	−0.111	−0.084	112.6	7.4	78.8	346.5	0.237	0.004

Before calculating the distances between the companies we have normalized the ratios according to the formulas owed to Kukuła (see [11]). All ratios with stimulating values (return on equity (ROE), return on assets (ROA), assets turnover, profitability ratio) were calculated according to the formula:

$$z_{ij} = \frac{x_{ij} - \min_j x_{ij}}{\max_j x_{ij} - \min_j x_{ij}}\,, \tag{2}$$

while the disstimulating ratios (amount due turnover, inventory turnover, liabilities turnover and debt margin) values were recalculated by

$$z_{ij} = \frac{\max_j x_{ij} - x_{ij}}{\max_j x_{ij} - \min_j x_{ij}}\,. \tag{3}$$

The obtained matrix is a normalized matrix which can be used for further analysis. Based on the normalized matrix the squares of Euclidean distance between the companies were calculated and then the clusters of similar companies were obtained by Ward's method [15], (see Fig.1).

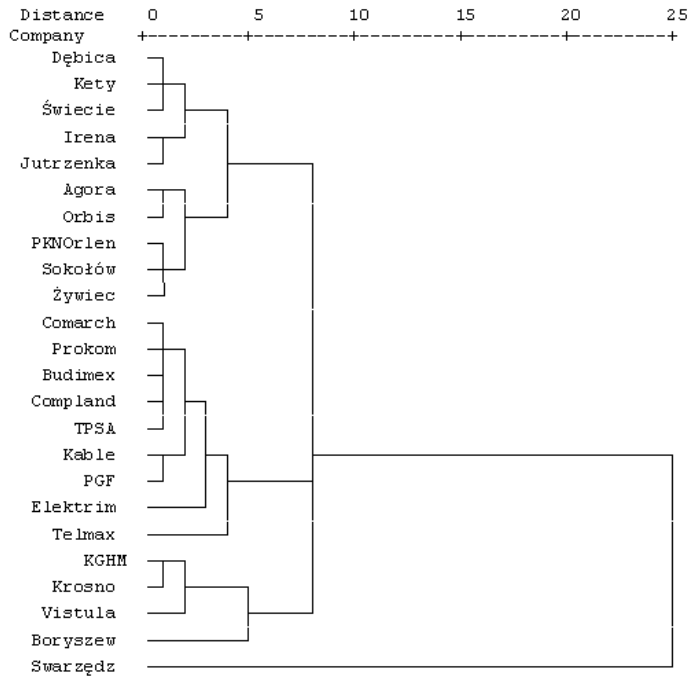

Fig. 1. Dendrogram obtained using Ward's method.

We have distinguished 7 clusters — see Table II. In order to examine more precisely the similarities within companies that belong to the same cluster we use the Principal Component Method (PCM).

TABLE II

Clusters of companies obtained using Ward's method.

Cluster		
1	Agora, Orbis PKNOrlen, Sokołów, Żywiec	5
2	Boryszew	1
3	Budimex, Compland, Comarch, PGF Prokom, TPSA, Kable, Elektrim	8
4	Dębica, Świecie, Kęty Irena, Jutrzenka	5
5	KGHM, Vistula, Krosno	3
6	Swarzędz	1
7	Telmax	1

4. Principal Component Method

The Principal Component Method (PCM) is frequently used in sociological, psychological and marketing research, (see [12, 16]). It can be used to rank the companies but its main application is distinguishing ratios that determine division into clusters as well as examining the diversity of the considered stocks. The PCM was applied to normalized matrix with the normalization formulas (2) and (3). The normalized matrix has been additionally standardized according to the formula $(z_{ij} - \bar{x}_j)/s_j$. The correlation matrix was used to distinguish principal components and the Varimax rotation for ease of interpretation, (see [16,17]). The eigenvalues and correlation coefficients are shown in Table III.

The obtained values are given by

$$r_{\bar{z}_i x_j} = q_{ij}\bar{\lambda}_i\,, \tag{4}$$

where: q_{ij} is the component of the i-th eigenvector of the variable x_j, $j = 1, \ldots, 8$, are eigenvalues of the correlation matrix. The values are the correlation coefficients between a given component and original variables. The absolute values of the correlation coefficients indicate the relative importance of original variable j in the new derived component i. The indicators with the largest absolute values have the strongest discrimination power and influence in the most significant level the division of companies into clusters as far as two or three principle factors are concerned. We have obtained

TABLE III

Rotated factor matrix with three largest eigenvalues. Source: our own calculations.

	Factor		
	Z_1	Z_2	Z_3
Indicator	$r_{\bar{z}_1 x_j}$	$r_{\bar{z}_2 x_j}$	$r_{\bar{z}_3 x_j}$
Return on equity (ROE)	-0.892^*	0.026	0.432
Return on assets (ROA)	-0.169	-0.077	0.939^*
Amount due turnover	0.874^*	-0.003	-0.031
Inventory turnover	-0.122	0.809^*	-0.142
Liabilities turnover	0.975^*	-0.011	-0.128
Assets turnover	-0.925^*	0.145	0.068
Debt margin	0.663	0.065	0.421
Profitability ratio	-0.053	-0.817^*	-0.073
Eigenvalue	3.94	1.40	1.15
% of variance	49.27	17.53	14.42
Cumulative %	49.27	66.80	81.22

* Indicate largest correlation coefficients.

the following rotated matrix (see Table III) with three components with eigenvalues greater then 1 that account for over 81% of total variance of the original variables. This means that instead of 8 ratios we can concentrate on 3 extracted components. The first component explains almost 50% of the total variance. It is highly positively correlated with liabilities turnover and amount due turnover, and negatively with assets turnover, and return on equity. Over 17% of total variance is accounted for by the second component. The second component is highly correlated with inventory turnover and profitability ratio. The third component is correlated positively with return on asset. Debt margin has little discrimination power. The mean values of the ratios with largest discrimination power for 7 distinguished groups of clusters are presented in Table IV.

The companies in cluster 6 are characterized by large return on equity and very large liabilities and assets turnover. The companies in cluster 2 and 4 are characterized by large return on equity and moderate inventory, liabilities and assets turnover. What distinguishes the clusters 2 and 4 is very high value of profitability for cluster 2 that consists of one stock only,

TABLE IV

Mean values of the ratios with largest discrimination power. Source: our own calculations.

Cluster	ROE	ROA	Amount due turnover	Inventory turnover	Liabilities turnover	Assets turnover	Profitabil. ratio
1	0.07	0.05	34.44	17.42	78.70	514.24	0.001
2	0.18	0.11	71.80	55.80	126.00	333.20	0.007
3	0.07	0.02	94.66	17.50	157.45	481.40	0.001
4	0.13	0.08	91.78	43.70	84.18	324.30	0.002
5	0.00	0.00	53.00	81.47	142.73	420.90	0.002
6	1.32	0.11	336.10	20.80	2925.31	3474.10	0.002
7	−0.11	−0.08	112.60	7.40	78.80	346.50	0.004

namely Boryszew. Cluster 7 has negative value of return of equity and large value of profitability. Clusters 1 and 3 have moderate values of return on equity and small values of profitability. Cluster 5 has small values of ROE and ROA. The analysis indicates that the best companies from the long term investment point of view are in cluster 2. The PCM has some other applications. It enables ranking the companies. Multiplying the first component scores by the scores of the standardized matrix of observation we obtain distinct values describing companies, (see [12, 16]). The values can be used to rank the companies. The results of such ranking and comparison with linear ordering obtained with use of the synthetic variable, aggregated with formula (1), are shown in Table V. One can see that the best three companies are the same for each method. We verify the portfolio by buying one share of each stock: Boryszew, Świecie, Żywiec, Orbis, Agora, in January 2004. If we sell the shares at the beginning of November 2004 the profit obtained is 9%.

5. On-line investment strategies

The above portfolio construction allowed choosing a group of companies that might perform well on the WSE. The method, however, does not answer the question how to divide the capital between the stocks. This can be determined by optimization techniques (see [1, 2]) or by on-line investment strategies applied to a chosen set of companies. We have performed the Universal Portfolio (see [8, 18]) strategy on Warsaw Stock Exchange in the period when Agora trading price was rapidly falling down. The port-

TABLE V

Comparison of rankings obtained by synthetic variable and by PCM. Source: our own calculations.

Linear ordering by PCM		Values of the synthetic variable	
Boryszew	11.48	Boryszew	0.636
Świecie	10.52	Świecie	0.630
Żywiec	10.45	Żywiec	0.630
Orbis	10.41	Dębica	0.586
Agora	9.81	Orbis	0.567
Jutrzenka	9.29	Kęty	0.564
PKNOrlen	9.28	Swarzędz	0.554
Dębica	9.10	PKNOrlen	0.548
Kęty	9.09	Jutrzenka	0.537
Telmax	8.38	PGF	0.525
Sokołów	8.11	Comarch	0.525
KGHM	6.40	Agora	0.523
Kable	6.24	TPSA	0.523
TPSA	6.24	KGHM	0.518
Comarch	5.78	Kable	0.512
Budimex	5.49	Budimex	0.507
Prokom	5.26	Sokołów	0.507
Irena	4.59	Prokom	0.489
PGF	4.52	Irena	0.485
Krosno	4.27	Compland	0.471
Vistula	4.21	Telmax	0.465
Compland	3.76	Krosno	0.463
Elektrim	2.73	Elektrim	0.462
Swarzędz	−12.05	Vistula	0.396

folio consisted on two stocks, Agora and Dębica. Without any analysis of distribution of returns, without any care about the financial standing of the considered firms, the method gradually eliminated Agora from the portfolio (see Fig. 2). In view of that result the Universal Portfolio seems to be a good

method to be applied for Warsaw Stock Exchange. The only disadvantage is computational complexity and the need to track on-line daily trading prices of a vast amount of stocks.

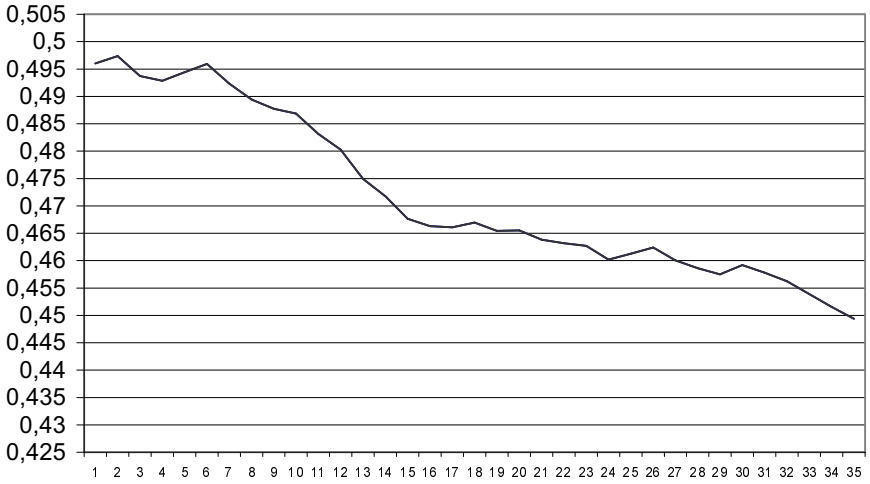

Fig. 2. The percentage of capital invested in Agora, January 2003.

We have also made calculations for Universal Portfolio consisting of two assets, PGF and Dębica. The results are shown in Fig. 3 and Fig. 4.

Fig. 3. The changes of wealth invested in Agora and Dębica for the universal portfolio strategy in January 2003 *i.e.*, in the period of Agora trading price rapid decrease.

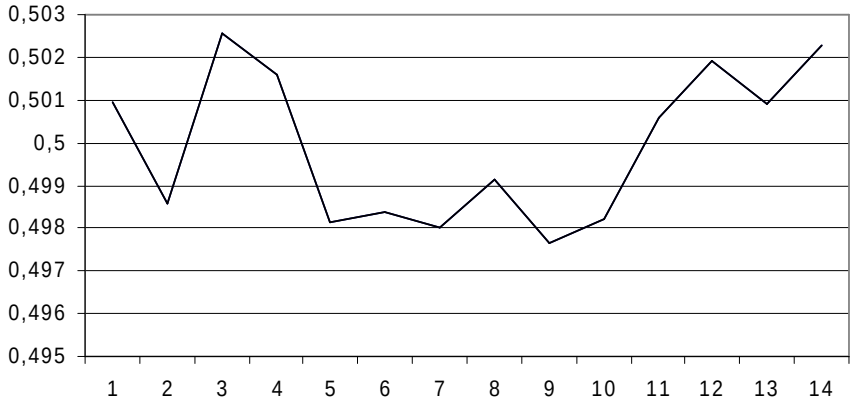

Fig. 4. The percentage of wealth invested in PGF for the universal portfolio strategy for two assets PGF and Dębica in January 2004.

6. Conclusions

The multivariate classification methods allowed choosing companies that can be regarded as good from the point of view of an investor. The PCM enabled investigating the structure of similarities between companies in the same clusters. It also allowed distinguishing ratios that have little discrimination power as well as those that influence the diversity within companies. It is worth noticing that both, the method of aggregating coefficients and the PCM indicated Boryszew as the best company. Fig. 5 shows the stock prices of Boryszew from the beginning of 2004 until November 2004.

The computations we have made for WGPW proved that standard taxonomy methods can be effective when they are supported by principle component analysis (or discriminant analysis). It seems that good results in

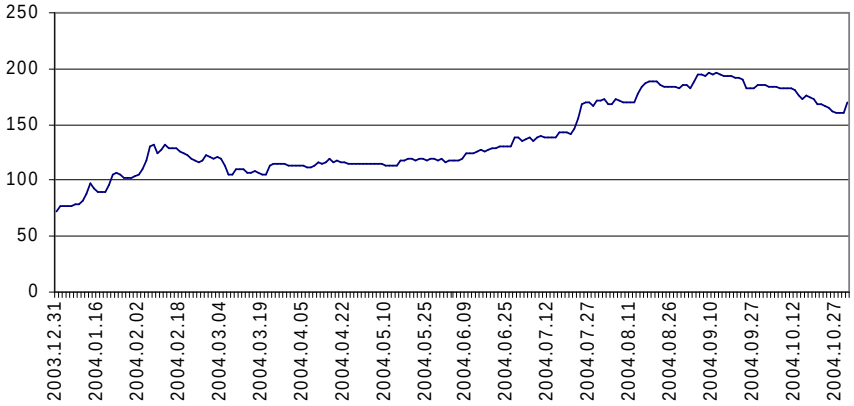

Fig. 5. The stock price of Boryszew from PCM.

portfolio management can be obtained by Universal Portfolio method. In contrast to standard methods of portfolio diversification, the above mentioned methods are not complicated and do not require advanced mathematical calculus. The computations were made with support of SPSS and Excel.

REFERENCES

[1] K. Jajuga, T. Jajuga, *Inwestycje*, Wydawnictwa Naukowe PWN, Warszawa 2002, in Polish.

[2] W. Tarczyński, *Rynki Kapitałowe*, Agencja Wydawnicza Placet, Warszawa 1997, in Polish.

[3] T. Pisula, J. Pisula, *Rynek Terminowy* **17**, 113 (2002), in Polish.

[4] K. Piontek, *Modelowanie Preferencji a Ryzyko '03*, Ed. T. Trzaskalik, Wydawnictwo AE, Katowice 2003, p. 491–504, in Polish.

[5] R.N. Mantegna, H.E. Stanley, *An Introduction to Econophysics. Correlations and Complexity in Finance*, Cambridge University Press, Cambridge 2000, [in Polish: *Ekonofizyka. Wprowadzenie*, Wydawnictwa Naukowe PWN, Warszawa 2001].

[6] B. Batóg, K. Wawrzyniak, *Rynek Kapitałowy. Skuteczne Inwestowanie*, Szczecin 2004, p. 275–284, in Polish.

[7] M. Gruszczyński, K. Wrona, *Rynek kapitałowy. Skuteczne Inwestowanie*, Szczecin 2004, p. 119–131, in Polish.

[8] U. Skórnik-Pokarowska, A. Orłowski, *Physica A* **344**, 81 (2004).

[9] U. Skórnik-Pokarowska, J. Różański, A. Orłowski, *Metody Ilościowe w Badaniach Ekonomicznych IV*, Warszawa 2004, p. 267–279, in Polish.

[10] W. Tarczyński, M. Łuniewska, *Dywersyfikacja Ryzyka na Polskim Rynku Kapitałowym*, Agencja Wydawnicza Placet, Warszawa 2004, in Polish.

[11] K. Kukuła, *Zero Unitarization Method*, Wydawnictwa Naukowe PWN, Warszawa 2000.

[12] W. Ostasiewicz, *Statistical Methods of Data Analysis*, Wydawnictwo Akademii Ekonomicznej, Wrocław 1999.

[13] M. Walesiak, *Metody Analizy Danych Marketingowych*, Wydawnictwa Naukowe PWN, Warszawa 2000, in Polish.

[14] A. Zeliaś, *Taksonomiczna Analiza Przestrzennego Zróżnicowania Poziomu Życia w Polsce w Ujęciu Dynamicznym*, Wydawnictwo Akademii Ekonomicznej, Kraków 2000, in Polish.

[15] D.F. Morrison, *Multivariate Statistical Methods*, McGraw-Hill Book Comp., New York 1976.

[16] J.H. Ward, Jr., *J. Am. Statist. Assoc.* **58**, 236 (1963).

[17] H.F. Kaiser, *Psychometrika* **23**, 187 (1958).

[18] T. Cover, *Math. Finance* **1**, 1 (1991).