"Proposal of creation of a portfolio with minimal risk"

| AUTHORS | Vincent Šoltés https://orcid.org/0000-0002-2656-5582 |
| | Jakub Danko https://orcid.org/0000-0002-1790-3324 |

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
12	1	5

Vincent Šoltés, CSc., Doctor of
National Sciences, Professor,
Department of Finance, Faculty of
Economics, Technical University of
Košice, Slovak Republic.

Jakub Danko, Department of Finance,
Faculty of Economics, Technical
University of Košice, Slovak Republic.

Vincent Šoltés (Slovak Republic), Jakub Danko (Slovak Republic)

PROPOSAL OF CREATION OF A PORTFOLIO WITH MINIMAL RISK

Abstract

The aim of this work is to propose a method for creating portfolios with a minimal expected risk. The proposed method consists of two steps. In the first step, the authors use a method for finding a minimum spanning tree. It is a graph theory tool, which is the field of discrete mathematics. Graph is defined as a set of vertices and edges. By this method the authors distribute assets, for example a stock index, into several subgroups. From each group it is then chosen an asset, from which most of the edges come out. These selected assets will be used to create a portfolio. In the second step, the authors will use a method of minimizing the standard deviation of the portfolio to calculate the weight of its assets. By this method, first it is found the weight of each asset so that the resulting portfolio would have the lowest possible expected risk. Then the authors find the portfolio with the lowest possible expected risk at required yield and create investment strategies. These strategies are compared during the time and between each other based on the variation coefficient. The article can be a practical guide for an individual investor during the minimal risk portfolio creation and shows him, which assets (and which asset weights) of the selected index to purchase.

INTRODUCTION

Yield and *risk* are two categories that are inseparable in financial markets. If you only invest in one of the underlying assets, the expected yield is proportional to the expected risk measured, for example using standard deviation. If we divide invested capital into several underlying assets, then we talk about the process of *diversification*, resulting in a *portfolio*. The aim of diversification is to create a portfolio that has a high expected rate of yield and low level of risk. It is obvious that not every diversification meets this objective.

Should diversification be effective, first it is necessary to answer two questions. The first is the choice of the underlying assets in the portfolio. By different sets of underlying assets is it possible to achieve different rate of reduction of expected risk of portfolio?

The aim of the investor is to select a set of underlying assets so that the expected portfolio risk decreased significantly, that it should be lower than the expected risk of investment into any single asset.

The second question that occurs after solving the first one is what should be the amount of resources invested in each selected underlying asset, so that the resulting portfolio would have the lowest expected risk? In other words, if we change the weights of assets, will the resulting portfolio always have lower expected risk?

Pioneering work in modern portfolio theory was done especially by authors Harry Markowitz (1952, 1959) and William Forsyth Sharpe (1992). Authors have studied the effect of asset risk, return, correlation and diversification on expected investment portfolio returns. Asset selection process (the investment process) is in most cases realized by two main approaches. The first is by using technical analysis and the second by fundamental analysis, however both of the approaches simultaneously are rarely utilized by the investor. The popular technical analysis method is the Japanese candlestick method, which was used by authors Yeongija Goo, Darhsin Chen and Yiwei Chang (2007) for analysis of the Taiwan Top 50 Tracker Fund and Taiwan Mid-Cap 100 Tracker Fund. The methodology is used to predict future price trends based on the relationships among opening, high, low and closing prices of the analyzed financial instruments. If we invest in the larger number of high-risk assets, we can use the reverse stock split method, which allows us to effectively merge these assets to form a smaller number of proportionally more valuable shares. This method was used by Manuela Raisová, Martin Užik and Christian M. Hoffmeister (2016) for analysis of V4 (Vysegrad group) countries' assets. The quality of theoretical introduction to quantify return and risk in case of two and three asset portfolio is written by Michal Šoltés (2003, 2012). New approaches for portfolio selection using, for example, neural networks are introduced by authors Alberto Fernández and Sergio Gómez (2007) and Ali S. Hadi, Azza A. El Naggar and Mona N. Abdel Bary (2016).

In this work we propose to create a portfolio with minimal expected risk divided into two steps. In the first step by using the *Minimum Spanning Tree* method we select the most appropriate underlying assets in the portfolio and in the second step by using *Generalized Reduced Gradient* method we calculate their weights in the proposed portfolio.

1. DESCRIPTION OF THE DATABASE

We used the historical prices of stocks comprising the index *Standard and Poor's 500* (hereinafter referred to as the *S&P 500*). Shares included in the index represent about 70% of the total capitalization of the US stock market. According to many experts, the index is considered to be the standard measure of the performance of the US stock market. It has been compiled since 1957: index creators have since determined its value retroactively since 1926.

This index represents the basis for selection of shares for our portfolio. An analysis is performed using data from *finance.yahoo.com*, which are available from 2. 1. 1962. Sufficient time horizon of analysis allows us to compare composition and performance of the created portfolio, taking different historical stock prices constituting the portfolio into account. Since the creation of the index is dynamic and its shares are constantly amended and supplemented, as relevant company shares we consider shares with at least five year price history. Considering that the average month has 20 market trading days (after taking into account weekends and holidays, during which no business is done), we consider that the average year has about 240 business days. Depending on how big a share of a particular company forming index of price history is, we qualify it for an analysis in the following table.

Table 1. Criterion for classification of companies into different price histories

Price history	Since year	Minimal number of business days	Number of companies
5-year history	2011	at least 1 200 business days	478 companies
10-year history	2006	at least 2 400 business days	454 companies
15-year history	2001	at least 3 600 business days	425 companies
20-year history	1996	at least 4 800 business days	371 companies
25-year history	1991	at least 6 000 business days	305 companies
30-year history	1986	at least 7 200 business days	197 companies

Shares, which do not have a price history for at least 1200 days have not been included in the analysis. Survivor bias, in this case, is not relevant because we would like to create a portfolio, so we need only assets of the existing companies. It is also clear that with the growth of the reporting period, the number of companies that meet a given price history decreases. Shares, which are found in 30-year price history are certainly also found for example in 5-year price history, the opposite is not necessarily true. It is therefore logical to watch the effect time period has on identification of shares in portfolio we create. In case we would create portfolios for different historical prices, whose compound somewhat differs, it would mean that the selected time period would have had a significant impact for the method of choosing shares into our portfolio. We calculate the price history to March 2016, when we launched the analysis.

In each share constituting the index we follow the company name, ticker, the sector in which the company operates and subsequently routine quantitative characteristics for each day of the reporting period:

- date;
- open price;
- highest price;
- lowest price;
- close price;
- trade volume;
- adjusted close price.

The *S&P 500* index divides companies into 10 sectors. This distribution has also been used in the process of designing our portfolio:

- consumer discretionary;
- consumer staples;
- energy;
- financials;
- health care;
- industrials;
- information technology;
- materials;
- telecommunication service;
- utilities.

2. STOCK SELECTION IN THE PORTFOLIO

After identification and creation of six groups according to their share price history (5, 10, 15, 20, 25 or 30 years) we calculate the classic daily return for each asset as:

$$r_n = \frac{P_n - P_{n-1}}{P_{n-1}} \cdot 100\%, \qquad (1)$$

where P_n is adjusted close price in time n, P_{n-1} is adjusted close price in previous business day (in time $n-1$) and r_n represents daily return in time n.

A *correlation matrix* has been calculated for each of the six groups of stock daily yield using standard Pearson correlation coefficient. The correlation matrix is a square symmetric matrix of size $n \times n$ (n is the number of shares analyzed) with ones on the main diagonal, in which the value of

the line i and column j represents the Pearson correlation coefficient of the daily yield between shares i and j given by:

$$\rho_{i,j} = \frac{cov_{i,j}}{\sigma_i \cdot \sigma_j}, \qquad (2)$$

The correlation matrices serve as a basis to calculate the distance matrices we get by transformation in accordance with the methodology of Professor Mantegna (1999) according to the following equation:

$$d_{(i,j)} = \sqrt{2 \cdot \left(1 - \rho_{i,j}\right)}, \qquad (3)$$

where $d_{(i,j)}$ represents the distance between shares i and j in index and $\rho_{i,j}$ is the correlation coefficient between these shares.

New mathematical objects, *complete graphs*, which are given by *vertices* and *edges*, are received from a distance matrix with the application of discrete mathematics tools. Vertices of these objects represent shares forming the index, taking
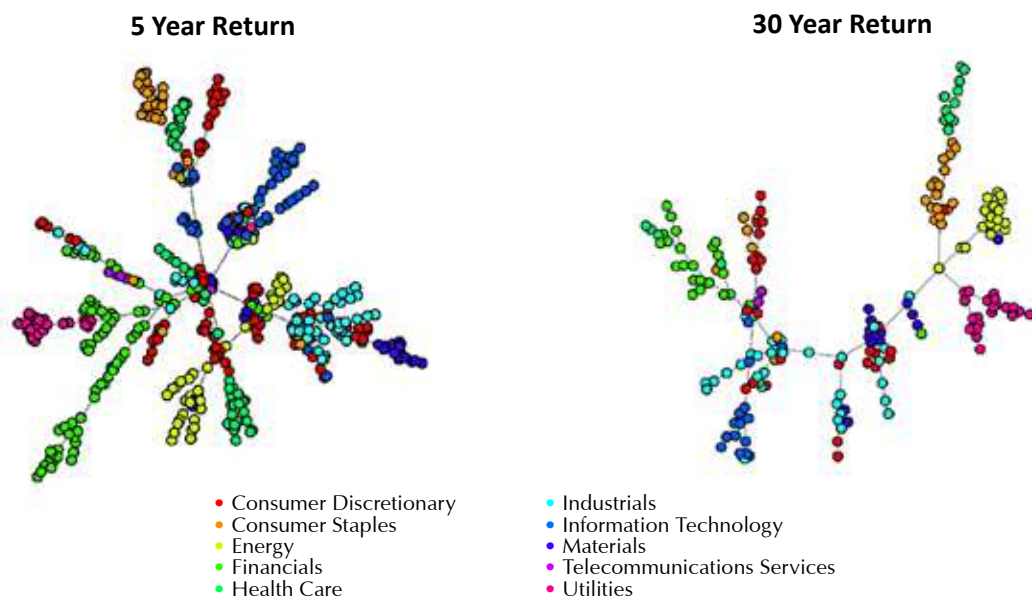
into account different price history. The edges between the vertices represent distances between calculated shares of a distance matrix. On the basis of the relationship for the distance matrices calculation from the correlation matrix, an inverse relationship between these two variables is applied. With correlation growth of two shares their mutual distance decreases, and conversely a fall in the correlation of two shares their mutual distance increases. Given transformation is carried out for the distance to reach non-negative values only, when in fact the correlation coefficient did not meet such criteria.

When applying the method of *minimum spanning tree*, which is finding such a *subgraph* of original graph, that is continuous, does not contain cycles and has minimal edge evaluation considering that there is a path between every pair of vertices, we get the minimum spanning tree for each of six complete graphs. This minimum spanning tree represents such a structure of shares, in which shares that are the closest (they have the greatest possible cross-correlation) are mutually linked. These minimum spanning trees and certain graph characteristics are the basis for selection of shares for our portfolio.

By this, for each price history we obtain one minimum spanning tree that represents mutual link be-

tween shares in this history. At this stage of analysis, we need to choose a certain amount of shares from this structure systematically, so that we can create a portfolio with minimal risk. We select one share from each sector, in order to diversify risks. Our portfolio therefore consists of ten shares. Ten shares are considered to be appropriate for an individual investor. The following figure shows the calculated minimum spanning tree also with color differentiation of individual sectors.

In this part of the analysis, we show only some partial results from our analysis (not all the price histories, only 5 and 30-year historical return). In Figure 1 we have the comparison between 5 and 30-year historical return, in which we graphically illustrate the use of this method, we can also see that the method is very useful in identifying the market structures, according to fact that the company shares are located in the common sector so closely together. Our results confirm the findings of Professor Mantegna, who applied this method when analyzing the structure of the US stock exchange index of Dow Jones Industrial Average. As we can see in the picture below, graph on the left side is significantly thicker than the right one, it is because of different price history. As it is mentioned in Table 1, 5-year history have 478 assets compared with 197 assets with 30-year history. The

5 Year Return   **30 Year Return**

- Consumer Discretionary
- Consumer Staples
- Energy
- Financials
- Health Care
- Industrials
- Information Technology
- Materials
- Telecommunications Services
- Utilities

Figure 1. Comparison between minimum spanning trees of 5 and 30-year historical yield with color-coded sectors

Table 2. The representatives of each sector according to time period

No.	Sector / History (number of years)	5	10	15	20	25	30
1.	Consumer discretionary	VFC	FOX	FOX	HD	HD	HD
2.	Consumer staples	CL	PG	PG	CL	CPB	KO
3.	Energy	MRO	OXY	NBL	APA	APA	HES
4.	Financials	AMP	BEN	BEN	BEN	BEN	AXP
5.	Health care	SYK	JNJ	SYK	PFE	PFE	JNJ
6.	Industrials	HON	HON	ITW	DOV	EMR	EMR
7.	Information technology	ADP	LLTC	LLTC	CSCO	CSCO	INTC
8.	Materials	PPG	PPG	PPG	PPG	PPG	PPG
9.	Telecommunications services	T	T	T	T	T	VZ
10.	Utilities	XEL	DTE	DTE	PEG	AEP	AEP

most important part of the analysis is to identify shares as representatives of particular sector, we select for the portfolio. In this section, two graph characteristics with a degree of vertex and eccentricity of a vertex are considered as decision criterion.

The *degree of vertex* represents the number of edges incident to the vertex. The higher the degree of vertex is, the more it can be considered as central, respectively, it may represent the vertices in close area and describe the graph itself. The vertex with degree one is located at the periphery of the graph, graph with a large number of degree with two vertices has a shape, in terminology of graph theory, called a *Path*. A graph with one high-degree vertex and with other degree vertices is called a *Star*.

Eccentricity of the graph's vertex represents the distance from this vertex to its most distant vertex. The lower the eccentricity of the vertex, the more it can be considered as central. The vertex with the lowest eccentricity is called the *centre* of the graph. The minimum from the eccentricities represents the *radius* of the graph and the maximum from eccentricities represents the *diameter* of the graph. According to previous allegations, it is obvious that graphs in the shape of path have high diameter, on the contrary, stars have very low radius and diameter.

When identifying and compiling the portfolio, we try to ensure that the shares we choose would represent the sector, in which they are located, in the best and most accurate way. In each sector, therefore we are looking for a central vertex: a share that best represents the sector as a whole. Considering that every share represents a specific vertex, we know what is the degree of the vertex, what is the eccentricity and what is the sector of the company the share represents. For each sector, therefore we are looking for such share that has the highest degree of vertex. If in any of the sectors there are several shares with the same highest degree, the decision-making criterion is the minimum from eccentricities. In this way, we clearly identify the representatives of each sector for each price history and we create a portfolio comprising of ten shares.

The results of this part of analysis are presented in the following table (see Table 2), which indicates that despite the change of the time interval the method identifies the sector's representatives relatively reliably, since many companies in the table occur several times. This points to the suitability of using such an identification method that is not very sensitive to the selected time period of performed analysis and the investor does not need a large number of historical data to compile a portfolio.

The aim of this part is to compile the portfolio in a way that share equities of the portfolio (weights) are optimized in order to minimize risk.

3. MINIMIZING PORTFOLIO RISK

After the identification of portfolio shares, based on historical data on prices and calculated profitability, we strive to optimize these portfolios: finding the weights of shares that create the portfolio so that risk is minimized as to such an investor is exposed. The risk in this case is calculated by the standard deviation of expected returns. The higher the standard deviation, the higher the portfolio risk for the investor. The standard deviation of the portfolio is calculated as the square root of the product of the row vector of individual stocks equities (weights) with covariance matrix profitability of these shares and the column vector of these equities:

$$\sigma = \sqrt{w \cdot \Sigma \cdot w^T}, \qquad (4)$$

where w represents row weight vector, Σ covariance matrix of profitability and w^T transposed row weight vector (column weight vector).

The aim is to minimize this value for the portfolio with an individual price history. Optimizing conditions are as follows:

- It is impossible to enter into short positions (the weight cannot be negative). In other words, we offer guidance for an individual investor to create a portfolio by purchasing shares.

- The portfolio can be comprised of any combination of 10 representatives of the sectors identified above while the portfolio can contain from 1 to 10 shares (the weight may also be zero).

- The size of invested funds is given by budgetary constraints of the investor (sum of equities is equal to one).

- Other financial instruments than previously identified shares are not included in the portfolio, the risk-free rate is also excluded.

The covariance matrix is computed for each portfolio and the weights of shares, which minimize the standard deviation of the portfolio determined by *Generalized Reduced Gradient Method (GRG)*. The

fulfilment of this objective allows us to find a combination of individual components equities of a portfolio that has the least possible expected risk, meaning with any other combination of weights the expected risk is higher. However, in addition to risk the investor also takes into account the expected rate of return, while annual investment of returns, specific share respectively, appear as the best indicator for the purpose of comparison. Therefore, the next part of the analysis deals with average annual returns.

For each of the identified shares the average annual rate of return is calculated as the average annual rate of return given by the formula:

$$r_{n_i} = \frac{P_{n_i} - P_{n_{i-1}}}{P_{n_{i-1}}} \cdot 100\%, \qquad (5)$$

where P_{n_i} is adjusted close price on the day n in year i, $P_{n_{i-1}}$ is to this day corresponding adjusted close price in the previous year (on day n in year $i-1$) and r_{n_i} is the annual profitability during day n in year i.

Based on the average annual return on shares in the portfolio, it is clear that the entire portfolio may acquire profitable return in the range from the yield of the lowest return on the share that is composed by to the yield of the highest return on share which it is composed by. However, if the investor would invest all the funds in only one asset, the risk of the investment would be very high. By weighting annual return on shares constituting a separate portfolio weights that minimize the risk of these portfolios, minimum returns that an investor would have required (the yield achieved with minimal risk) are brought for each portfolio. On the basis of the relationship of the risk and size of individual yields on shares in the portfolios, 4 basic investor strategies are formed. The first strategy is a strategy of minimal risk, as described above. The profitability of this strategy is the minimum required yield rate of an investor. If profitability of the portfolio with minimal risk is deducted from the yield on the share with the highest profitability in the portfolio, we get the range of profitability of previously assembled strategies. If this range is divided appropriately, a strategy of gradually widening regular returns can be drawn up. Again, we solve the optimization problem, in

which in comparison with the previous case we have additional condition that the profitability of the portfolio should be equal to a specific value. Again, however, we minimize the standard deviation of the portfolio under the conditions mentioned above.

Thus for each of the six portfolios we get four strategies appropriate for investors with varying degree of risk perception. The first strategy is the already mentioned *minimal risk strategy*. The second strategy is called a *conservative strategy*, the third is called a *balanced strategy* and the last is an *aggressive strategy*. Each additional strategy provides greater expected return and of course higher risk. For each strategy, we calculate the size of the eq-

uities of various financial instruments comprising the portfolio and calculate the expected yield and risk of the strategy.

Data on prices of individual stocks come from freely available database *finance.yahoo.com*. Analysis was performed using the statistical programming language *R*. When working with graphs the tool library *igraph* was used. For optimization problems the *Solver* add-in *MS Office Excel* was used, while making use of the nonlinear algorithm *Generalized Reduced Gradient Method*.

In this part we again present results only for price history for 5 and 30 years. Results are presented in Tables 3 and 4.

Table 3. Results analysis for the 5-year historical yield

HISTORY 5	Minimal risk	Conservative strategy	Balanced strategy	Aggressive strategy
Standard deviation	0.00803266	0.008574246	0.009879435	0.011996454
Average annual return	13.04%	17.57%	22.09%	26.61%
w1 =	0.010915	0.087619	0.150148	0.21905
w2 =	0.225393	0.128771	0.006994	0
w3 =	0	0	0	0
w4 =	0	0	0	0.040009
w5 =	0.028655	0.055977	0.058938	0.000173
w6 =	0	0	0	0
w7 =	0.078404	0.189369	0.258686	0.215034
w8 =	0	0.107177	0.269254	0.489257
w9 =	0.338238	0.133961	0	0
w10 =	0.318395	0.297126	0.25598	0.036477

Table 4. Analysis results for the 30-year historical yield

HISTORY 30	Minimal risk	Conservative strategy	Balanced strategy	Aggressive strategy
Standard deviation	0.01053565	0.01120884	0.013116116	0.01614639
Average annual return	13.20%	16.77%	20.34%	23.91%
w1 =	0.017083	0.189769	0.375219	0.615732
w2 =	0.147868	0.167418	0.166141	0.064059
w3 =	0.044927	0.025534	0	0
w4 =	0	0	0	0
w5 =	0.242356	0.268722	0.275574	0.175763
w6 =	0.043359	0	0	0
w7 =	0.008828	0.05885	0.103246	0.144446
w8 =	0.048777	0.022929	0	0
w9 =	0.149895	0.055303	0	0
w10 =	0.296907	0.211476	0.07982	0

So that we can compare the riskiness of individual portfolios proposed by this research, we calculated their coefficients of variation. The results are shown in Table 5 and riskiness of individual portfolios is distinguished by colors (light shades: the lowest risk level, dark hues: the highest risk level).

Table 5. Riskiness of individual portfolios measured by the variation coefficient

History	Minimal risk	Conservative strategy	Balanced strategy	Aggressive strategy
5	6.16 %	4.88 %	4.47 %	4.51 %
10	11.33 %	9.01 %	8.25 %	8.29 %
15	10.69 %	9.28 %	8.86 %	9.02 %
20	9.76 %	8.67 %	8.31 %	8.85 %
25	9.20 %	7.48 %	7.55 %	8.18 %
30	7.98 %	6.68 %	6.45 %	6.75 %

CONCLUSION

The present method provides practical guidance for identification and selection of shares into investor's portfolio. After the implementation of this method, we are looking for portfolios that offer the lowest risk at predefined performance conditions. A positive relationship between the portfolio's profitability and risk was confirmed. Table 5 shows, however, that the least risky strategy is the balanced one.

In five portfolios it has the lowest coefficient of variation (only in history 25 the conservative strategy is less risky). Paradoxically, the portfolio created by minimal risk strategy is not the portfolio with minimal risk in either case. On contrary, minimal risk strategy is the most risky, because portfolios created by this strategy show the highest coefficient of variation in all periods. The absolutely lowest risk shows the balanced portfolio in the five-year history and on the other hand the absolutely most risky portfolio is the minimal risk one during the ten-year history.

The main benefit of the methodology used in this article is that it combines well-known approach of portfolio optimizing with graph theory, which is not often used in portfolio creation process.

REFERENCES

1. Bonanno, G., Caldarelli, G., Lillo, F., Micciché, S., Vandewalle, N., and Mantegna, R. N. (2004). Networks of Equities in Financial Markets. *The European Physical Journal B, 38*(2), 363-371. Retrieved from https://arxiv.org/pdf/cond-mat/0401300.pdf

2. Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *International Journal, Complex Systems 1695.* Retrieved from http://igraph.org (accessed: June 10, 2016).

3. Fernández, A. and Gómez, S. (2007). Portfolio selection using neural networks, *Computers and Operations Research, 34*(4), 1177-1191. Retrieved from http://www.sciencedirect.com/science/article/pii/S0305054805002042

4. Goo, Y., Chen, D. and Chang, Y. (2007). The application of Japanese candlestick trading strategies in Taiwan. *Investment Management and Financial Innovations, 4*(4), 49-79. Retrieved from http://www.businessperspectives.org/journals/investment-management-and-financial-innovations/issue-67/the-application-of-japanese-candlestick-trading-strategies-in-taiwan

5. Hadi, A. S., El Naggar, A. A., and Abdel Bary, M. N. (2016). New model and method for portfolios selection. *Applied Mathematical Sciences, 10*(6), 263-288. Retrieved from http://www.m-hikari.com/ams/ams-2016-ams-5-8-2016/p/abdelbaryAMS5-8-2016.pdf

6. Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B, 11,* 193-197. Retrieved from https://link.springer.com/article/10.1007/s100510050929

7. Markowitz, H. M. (1952). Portfolio Selection. *The Journal of Finance, 7*(1), 77-91. Retrieved

from https://www.math.ust.hk/~maykwok/courses/ma362/07F/markowitz_JF.pdf

8. Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments*. London: Yale University Press, 344 p.

9. Raisová, M., Užik, M., and Hoffmeister, C. M. (2016). Normal and reverse stock splits in the V4 countries, *Investment Management and Financial Innovations, 13*(4), 94-105. Retrieved from http://www.businessperspectives.org/

journals/investment-management-and-financial-innovations/issue-102/normal-and-reverse-stock-splits-in-the-v4-countries-monetary-stimulus

10. Sharpe, W. F. (1992). Asset Allocation: Management Style and Performance Measurement, *The Journal of Portfolio Management, 18*(2), 7-19. Retrieved from https://www.researchgate.net/publication/247905617_Asset_allocation_Management_style_and_performance_measurement, https://web.stanford.edu/~wfsharpe/art/sa/sa.htm

11. Šoltés, V., Šoltés, M. (2003). Analysis of two-asset portfolio. *E&M Ekonomie a Management, 6,* 63-65.

12. Šoltés, M. (2012). Theoretical Aspects of Three-Asset Portfolio Management. *CurentulJuridic, 15*(4), 130-136. Retrieved from http://revcurentjur.ro/old/arhiva/attachments_201204/recjurid124_13F.pdf