

Jozef GLOVA, PhD

E-mail: jozef.glova@tuke.sk

Damián PASTOR, PhD Student

E-mail: damian.pastor@tuke.sk

Professor Tomáš SABOL, PhD

E-mail: tomas.sabol@tuke.sk

Faculty of Economics

Technical University of Kosice, Slovak Republic

COINTEGRATION AND ITS SPECIFIC APPLICATIONS IN PORTFOLIO MANAGEMENT

***Abstract:** Cointegration becomes the prevalent statistical tool in financial economics. In passive stock portfolio management enables the replication of stock index and construction of portfolio with better characteristics than index itself. It is a powerful technique for investigating long term dependence in multivariate time series. In our paper we construct cointegration based portfolios that differ in number of selected stocks, interval of reselection, calibration period, and transaction costs. We considered an allocation into portfolio consisting of Dow Jones Industrial Average components and thereafter we compare long term return and risk profile of portfolios focus on cointegration selection process and index DJIA. The cointegration technique enabled us to use long calibration period and provided that portfolio weights do not change too much over time and outperform the index DJIA in post-sample performance measurement.*

***Keywords:** Index Tracking, Cointegration, Cointegration vector, long-run equilibrium relationship, Engle-Granger methodology, Portfolio Risk and Return.*

JEL Classification: C61, G11

1. Introduction

The conventional construction of a financial portfolio is based on the analysis of correlation structure among the particular financial assets involved in the portfolio. It was Harry Max Markowitz in early 1950's who published a revolutionary paper on how does one select an efficient set of risky investment or so called efficient frontier. This theory provides the first quantitative view of portfolios variance, where co-movements in securities returns are considered. So, the variance of portfolios is not a simple product of the particular investment proportion and their variances. Instead of it one has to consider covariance structure implicitly involved in multi-variant distribution of securities returns. Almost three decades ago the general approach RiskMetrics was developed by J.P. Morgan during the late 1980's and has been commonly applied by financial market participants for more than two decades. Unfortunately the concept lacks of

accuracy if the correlation structure varying in time. From this perspective the traditional portfolio needs rebalance repeatedly, what could increase the cost structure of the portfolio dramatically. In general the use of the traditional concept is delimited and depends on the level of change within the portfolio volatility.

While the traditional approach considers historical time series returns of the selected set of financial assets and their replication against the return of a particular index the cointegration analysis uses assets' time series appearing and behaving as random processes or processes of the so-called random walk. In our study we use the second mentioned concept, cointegration. The classical papers on cointegration are by Granger (1986) and Engle and Granger (1987).

The cointegration is based on the long-term relationship between time series. One can consider the cointegration, if there is such linear combination of the non-stationary time series that is stationary. The passive index tracking strategy tries to achieve equal return as well as the underlying index, and concurrently tries to diminish the volatility of the tracking error, thus a difference between the portfolio return and underlying index.

The paper is divided as follows: at the beginning we briefly start with an overview of time series stationarity, a specific assumption that is expected to be fulfilled for applying the cointegration approach. A difference between correlation and cointegration is being explained in a brief form. Further we describe cointegration analysis and the possible fields and forms of its applicability. All this effort is summarized in an overview the theory and the state of the art. Engle-Granger method has been applied as a technical part of our research methodology. We considered an allocation into portfolios consisting of Dow Jones Industrial Average (DJIA) components. At first we describe methodology with a description of data and later the further attributes for asset allocation are specified. Beyond the current research in this field we consider particular modifications of key parameters and them sensibility change in a form of different number of stocks, reselection interval, calibration period and strategy used as well as level of transaction expenses. At the end the discussion is provided.

2. Literature review on equity portfolio management style

Passive and active equity portfolio management style is usually discussed and described in financial literature. The crucial phase in the investment process is allocation what for equity style portfolios means stock picking or stock selection. It was Harry M. Markowitz (1952, 1959) who made the first quantitative and empirical contribution to portfolio selection. According to Reilly and Brown (2012) no middle ground exists between active and passive equity management strategies. They also argue that "hybrid" active/passive equity portfolio management style exists, in a form of enhanced indexing, but such styles are variations of active management philosophies.

Focusing on passive equity portfolio management means a long-term buy-and-hold strategy. Very often some authors like Gibson (2013) or Nofsinger (2013)

referee about indexing strategy, because of the goal of tracking an index. In this context only occasional rebalancing is needed, specifically because dividends and their reinvesting, stocks merge or change in the index construction. In traditional literature one can find three basic techniques for constructing a passive index portfolio – full replication, sampling, and quadratic optimization or programming. Full replication technique helps ensure close tracking, but it may be suboptimal because of transaction cost connecting with purchase of many securities and dividend reinvestments. With sampling technique we need to buy a representative sample of stocks that comprise the benchmark index. The last passive technique is quadratic optimization or quadratic programming based on historical information on price changes and correlations between securities as inputs to a computer program that determines the composition of a portfolio that minimize tracking error with the benchmark. This technique lack of accuracy because it relies on historical price changes and correlation. According to Alexander (2008) correlation reflects co-movements in returns, which are liable to great instabilities over time. Returns have ‘no memory’ of a trend so correlation is intrinsically a short term measure. As she further explains that is why portfolios that have allocations based on a correlation matrix commonly require frequent rebalancing and long-short strategies that are based only on correlations cannot guarantee long term performance because there is no mechanism to ensure the reversion of long and short portfolios. That’s the reason why Alexander (1999), Alexander and Dimitriu (2005) and Dunis and Ho (2005) proposed to use cointegration analysis as a sound statistical methodology for modelling the long term equilibrium. According to Alexander and Dimitriu the phenomenon of equity indexing has attracted considerable interest in the last two decades, whereby the passive investment industry as a whole has witnessed a remarkable growth.

3. Cointegration

Cointegration was first introduced by Granger in 1981, but was extensively discussed later in Engle and Granger (1987). Today’s it is a standard econometrics tool that gives us an opportunity to detect and characterise the comovement and long-run equilibrium relationship of a system of diverse processes integrated of order one or higher that share one or more common stochastic trend.

Cointegration analysis, as well as countless number of other time series models, is based on the stationarity concept. Strictly stationary are considered discrete time stochastic processes only if their probability is invariant in time. We usually work with weakly stationary or covariance stationary process. In our paper under the term of stationarity we understand its weakly form.

Weakly stationarity of a discrete time stochastic process $\{X_t\}_{t=1}^T$ have to fulfil three main conditions:

- Mean value of time series $E(X_t)$ is a finite constant;
- Variance $V(X_t)$ is a finite constant;
- $cov(X_t, X_s)$ depends only on absolute value of $|t - s|$.

The first two conditions make us aware that time series is detrended and a majority of observations is approaching their expected or mean value. This condition, as mentioned by Alexander (2008), depicts equal common dependency of two variables in each moment.

One says a time series $x = (x_t)$ is integrated of order d , denoted as $x \sim I(d)$ if, after differencing d times, it turns out to be stationary. Mathematically written as $x_t(1 - B)^d = z_t$, where B is the backshift operator ($x_t B^k = x_{t-k}$) and z is a stationary series ($z \sim I(0)$).

Given two time series x_t and y_t , where both series are integrated of order d , denoted as $I(d)$. Their linear combination is

$$z_t = x_t - ay_t, \tag{1}$$

and is mostly integrated of order d or $I(d)$, too. There is possible that $z_t \sim I(d - b)$, $b > 0$. In this case time series x_t and y_t are co-integrated and vector $(1, -a)$ is called the cointegration vector. We call the state when $x_t = ay_t$ the equilibrium and z_t expresses a deviation from the equilibrium state.

Let's suppose both x_t and y_t are both integrated as $I(1)$. The condition of cointegration will be fulfilled if z_t is stationary, hence $I(0)$. If $a = 1$, then z_t is the difference between x_t and y_t . This time series tends to be almost persistent and a finite variance exists („mean reversion“).

It is well known that correlation and regression analysis using raw data or using $I(1)$ processes produce misleading results. Therefore the first differences of data have been used as a possible solution to the problem of spurious correlation.

Cointegration measures long term dependency between capital asset prices. This concept differs from correlation and regression analysis, whose severe limitations as a dependency measure. It is also set apart from copulas, which are typically used to construct joint distributions of asset returns. Cointegration models are usually constructed in two phases: the first phase look into a long term equilibrium between prices of a set of assets, and the second phase is a dynamic model of correlation or so called an error correction model (ECM).

In general we can say cointegration and correlation are related but different concepts. High correlation does not automatically imply high cointegration nor vice versa. If there is cointegration or not, high correlation can occur. But to distinguish both terms we need to note that correlation tells us nothing about the long term relationship or behaviour between two assets. So correlation is not adequate measure over long periods of time. Correlation only reflects co-movements in returns, which have no 'memory' of a trend, so is intrinsically a short term measure. In Figure 1 on the left we see time series, which are not cointegrated with relatively high correlation coefficient of 0,975. On the right in the same figure we see cointegrated time series, but the lower correlation is observed, correlation of 0,486. In Figure 2 we see two graphs of cointegrated time series with high correlation values, but with different values of a , where a determines the relationship between time series in equilibrium: $x_t = ay_t$.

In summary, high correlation values can occur when time series are cointegrated, but even when there is no evidence of cointegration. Correlation tells us nothing about the long-term relationship between a pair of financial assets

As we already mentioned in our papers in Glova (2013a, b), the co-movements between stocks can be due to a single or multiple indices. So the correlation or covariance structure of security returns might be obtained by relating the return on a stock to the return on a stock market index or other non-market indices. Unfortunately as mentioned by Alexander (2008) so created portfolios require frequent rebalancing because there is nothing to prevent the tracking error from behaving in the unpredictable manner of random walk.

To conclude, since correlation tells us nothing about long term performance there is a need to augment standard risk-return modelling methodologies to consider long term trends in prices. Therefore as mentioned by Alexander and Dimitriu (2005) portfolio management strategies based on cointegrated financial assets should be more effective in the long term.

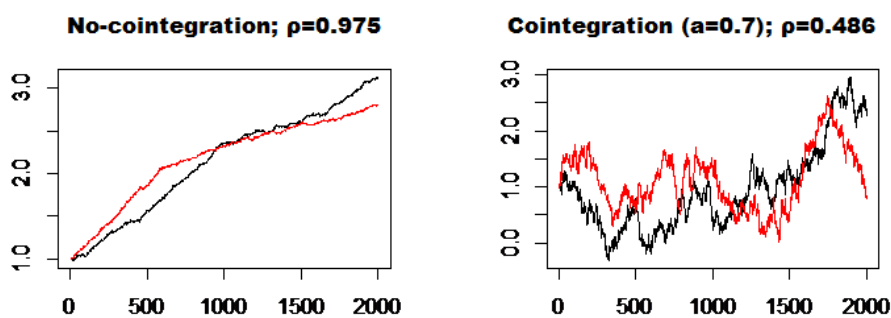


Figure 1. Simulated time series without cointegration, but with correlation of 0,975 (on the left) and cointegrated time series with correlation of 0,486 (on the right)

The presence of cointegration implies there is a statistical causality between returns. Return on an asset commoves with time lagged returns of another asset. We also say that x Granger causes y if lagged values of x help to predict current and future values of y better than just lagged values of y alone.

Financial assets pricing processes are usually figured out as random walk models and one supposes to be $I(1)$ processes. As Bossaerts (1988) states cointegration does not imply market inefficiency. Moreover Dwyer and Wallace (1992) have shown that the presence of cointegration in selected financial markets is compatible with market efficiency. Though, Dwyer and Wallace (1992) state that there are no risk-free returns above opportunity cost available to agents given transaction costs and agents' information, or explained in other form, there is

where is market efficiency, there is the absence of arbitrage. Alexander and Dimitriu (2005a, 2005b, 2005c) discuss the use of cointegration methods for enhanced indexation.

As written by Alexander (2008) no financial systems have higher cointegration than term structure. As extended by Bradley and Lumpkin (1992), Alexander and Johnsons (1992, 1994), Hall et al. (1992), Davidson et al. (1994), Brenner et al. (1996) cointegration and correlation go together in the yield curve, and we often find strongest cointegration where correlations are highest.

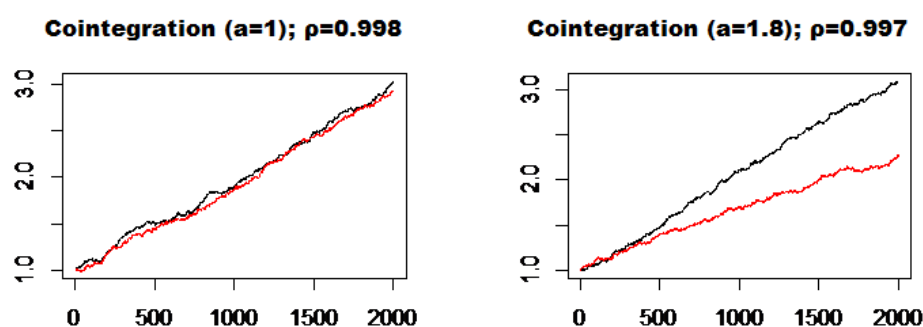


Figure 2. Simulated time series with different values of a coefficient

We may as well find cointegration between a sufficiently diversified stock portfolio and market indices. This is also the field our paper focuses on. Cointegration based portfolios provides that the index weights do not change too much over time as properly described by Alexander (1999), Alexander and Dimitriu (2005 a, b) and Dunis and Ho (2005). Very inspiring is the contribution written by Maurer (2008) that describes very well two different index tracking strategies, one based upon cointegration analysis and the other based on a market equilibrium. Alexander (2008) also summarizes the existence of cointegration in other segments of financial market, e.g. spot and futures prices, commodities, spread options, market integration, and foreign exchange.

According to Maurer (2008) there are several ways a cointegration system may be estimated and tested, among which the Engle-Granger, introduced by Engle and Granger (1987), and the Johansen methods, presented by Johansen (1988, 1991) and Johansen and Juselius (1990), are the most famous and widespread. Since our paper applies the approach of Engle and Granger we shortly look into this method. The above mentioned methodology of Engle and Granger (1987) is a two step estimation approach. Firstly, a cointegration vector is estimated using an optimization set-up as simple as OLS estimation. Secondly, an error correction model or ECM is constructed again applying simple OLS

estimation. The previous estimations may be used in place of the unknown, true co-integrating vector.

The conventional construction of a financial portfolio is based on an analysis of the correlation structure among the particular financial assets involved in the portfolio. Almost three decades ago the general approach RiskMetrics was developed by J.P. Morgan and has been commonly applied by financial market participants for more than two decades. Unfortunately the concept lacks of accuracy if the correlation structure varying in time. From this perspective the traditional portfolio needs rebalance repeatedly, what could increase the cost structure of the portfolio dramatically. In general the use of the traditional concept is delimited and depends on the level of change within the portfolio volatility.

While the traditional approach considers historical time series returns of the selected set of financial assets and their replication against the return of a particular index the cointegration analysis uses assets' time series appearing and behaving as random processes or processes of the so-called random walk. In our study we use the second mentioned concept, cointegration.

The cointegration is based on the long-term relationship between time series. One can consider the cointegration, if there is a such linear combination of the non-stationary time series, that is stationary. The passive index tracking strategy tries to achieve equal return as well as the underlying index, and concurrently tries to diminish the volatility of the tracking error, thus a difference between the portfolio return and underlying index.

Index tracking based on Cointegration

Stock market indices are weighted sums of stock prices. A good diversified portfolio of stocks will be usually cointegrated with the indices. This can also provide a stationarity in portfolio's weights, that don't change so much over time. An introduction to index tracking using cointegration was published by Alexander (1999) and is based on Engle-Granger methodology. There is also further evidence provided by Alexander and Dimitriu (2005a, 2005b, 2005c), Dunis and Ho (2005) and others. The optimization criterion used in index tracking applying Engle-Granger methodology is to minimize the variance of the tracking error whilst also ensuring the tracking error is stationary. As mentioned by Alexander (2008) the Engle-Granger procedure is a better choice than others for the benchmark tracking problem.

Engle-Granger methodology begins by testing the order of integration. All the variables should be integrated of the same order. One of the tests of stationarity can be used, for example ADF test. The null hypothesis of ADF test is the presence of unit root in time series, which means that the time series is non-stationary. If all variables are integrated of the same order, then we choose the explained variable, establish regression equation, and estimate the long-run equilibrium relationship by OLS. The regression equation has the following form:

$$\ln(I_t) = \alpha + \sum_{k=1}^n \beta_k \ln(P_{kt}) + \varepsilon_t \quad (2)$$

where: I_t is the price of stock index in time t ,
 α is the intercept,
 β_k is the regression coefficient of stock k ,
 P_{kt} is the price of stock k in time t ,
 ε_t is the error term.

After estimating the long-run equilibrium relationship, it is necessary to test whether the variables are cointegrated or not. Engle and Granger (1987) proposed to test the stationarity of the residuals and suggest the ADF test. If the error term is stationary, then the variables are cointegrated.

4. Research methodology

In this paper we applied index tracking on Dow Jones Industrial Average (DJIA) and Dow Jones Composite Average (DJCA). We used daily close prices of these indices and daily close prices of their components adjusted for splits and dividends for the period from 29-Dec-2000 to 31-Dec-2013. The data was downloaded from <http://finance.yahoo.com>. To mark selected stocks we used standardized identifiers – the ticker symbols.

Both indices were modified and we created "reconstructed indices". Current components of these indices form the basis of reconstructed indices. We excluded shares with price history shorter than period mentioned above and shares whose log prices are not integrated process, which is one of the conditions for cointegration relationship. Reconstructed DJIA consists of 30 shares and reconstructed DJCA of 60 shares. In the case of the reconstructed Dow Jones Industrial Average we excluded the shares of Cisco, Nike and Visa, and we replaced them by former components of DJIA: Alcoa Incorporated, Bank of America Corporation and Hewlett Packard Company. Reconstructed Dow Jones Composite Average has 60 components, while 30 is the same as the reconstructed DJIA, and from other 35 current components of DJCA we excluded shares of Delta Air Lines, Jet Blue Airways, Public Service Enterprise Group, Southern Company and United Continental Holdings. Components of the indices were unchanged throughout the period.

Since both stock indices are price-weighted, we calculate their prices as the sum of the prices of all components divided by Dow Divisor. In our case, we chose the value of divisor so that the initial value of index (value on 1-Jan-2006) was 1.

The aim of index tracking is to replicate the selected stock market index. As reported by Alexander and Dimitriu (2002), we expect that created portfolio will have the same return and volatility as benchmark and high correlation of the returns with the returns of the benchmark. Index tracking should also minimize the tracking error, i.e. the difference between the returns of the created portfolio and

the returns of the tracked index. Correlation between the tracking error and the returns of the tracked index should be minimal. Of course, positive tracking error (excess return of the portfolio) with minimum volatility may be interesting for potential investors.

We followed the following characteristics of the created portfolios: profitability and volatility of the portfolio, Information ratio, correlation between the returns of the tracked index and the returns of the portfolio, correlation between the returns of the tracked index and the tracking error and volatility of the tracking error. In terms of profitability, we compared the final values of portfolios (values on 31-Dec-2013). We considered the starting value of each portfolio (value on 1-Jan-2006) equals to 1. Volatility of the portfolio corresponds to an annualised standard deviation of daily log returns of the portfolio, while we considered 252 trading days in the year. Similarly, we determined the volatility of the tracking error. Information ratio is defined as the tracking error divided by its volatility. For measuring the correlation we used Pearson's correlation coefficient.

Cointegration approach to index tracking is based on the existence of a long-run relationship between the log prices of the index and the log prices of the stocks in the portfolio. A necessary condition for the existence of this relationship is that all variables in the regression equation are integrated of the same order. The results of the ADF test, which is presented in Appendix A, shows that this assumption is fulfilled for daily log prices of reconstructed indices and their components.

Log prices of the reconstructed index is explained variable in the long-run equilibrium equation; logarithms of closing prices of selected stocks are explanatory variables. It is natural to expect that an index can be closely tied to a particular portfolio consisting of shares that are part of it. For this reason, we are looking for cointegration relationship between the reconstructed index and its components.

$$\ln(rDJA_t) = \alpha + \sum_{k=1}^n \beta_k \ln(P_{kt}) + \varepsilon_t \quad (3)$$

Equation parameters have been estimated based on the method of ordinary least squares (OLS). Then, we tested the cointegration relationship between the log prices of the index and log prices of selected stocks using already mentioned Engle-Granger methodology. Residuals of the regression equation were tested for the presence of unit root (stationarity) using ADF test for the chosen significance level ($\alpha = 0.05$). If the residuals are stationary, then the variables in the regression equation are cointegrated.

Weight of individual stock in the portfolio was calculated as the regression coefficient of the stock divided by the sum of all regression coefficients except the constant-level. Weight could be a negative number, and the sum of the weights is equal to 1.

Based on our calculations and in terms of research by Alexander and Dimitriu (2002) and Dunis and Ho (2005), we identified four basic factors that

affect monitored portfolio characteristics: method of stock selection, number of selected stocks, reselection interval, and calibration period.

Method of stock selection

Stock selection was made from components of the reconstructed indices by two methods. The first method (Method A) consists of selecting x shares with highest price and thus also the highest weight in the index at a given time. The second method (Method B) is the selection of x shares that were most correlated with the index (the highest correlation coefficient between the returns of the stock and returns of the index) over the calibration period.

The number of selected stocks

For the first method we have chosen for tracking DJIA successively 15, 20 and 25 shares, and for tracking DJCA 20, 30, 40 and 50 shares. For the second method we have chosen 10, 15, 20 and 25 shares for tracking DJIA and 10, 20, 30, 40, 50 shares for tracking DJCA. We also constructed the portfolios composed of the full number of shares.

Reselection interval

Portfolios were re-balanced at different time intervals. Intervals considered were 10, 21, 63, 126 and 252 trading days. During rebalancing we selected shares under the above mentioned methods, we estimated parameters of the regression equation, we tested for cointegration and we defined the new weights of the individual components.

Calibration period

For the estimate of regression we have used t years of data prior to the date of re-balancing. This period can be called "calibration period". In our work, we selected three different calibration periods: 3, 4, and 5 years.

We have created all possible combinations of these factors, a total of 270 portfolios.

Transaction costs

Transaction costs can substantially reduce the profitability of the portfolio. Jones (2002) defined the overall transaction costs as the sum of half the quoted spread and fees. For DJIA shares he determined the transaction costs in 2000 of around 0.2%. In determining transaction costs effective spread is used more often than quoted spread. Jain (2003) calculated the effective spread in 2000 for NYSE (0.1%) and Nasdaq (0.51%). Normally, fees are paid for each trade or per share and the amount depends on the chosen broker, the number of purchased/sold shares and their prices.

In this work, we decided to analyse constructed portfolios assuming the absence of transaction costs and assuming the transaction costs of 0.2% and 0.5% of the trade value. We expect the real transaction costs to be between 0.2% and 0.5%.

5. Results and discussion

Alexander and Dimitriu (2002) tested for cointegration between the tracked index and the selected shares after each re-balancing. The results of their work show that the existence of cointegration relationship requires a minimum number of shares and the minimum length of the calibration period, what is confirmed in our results. Problems with rejection of the hypothesis of cointegration emerged in the portfolios consisting of 10 shares and portfolios consisting of 15 shares with a three year calibration period that tracked DJIA. Number of reselections for individual portfolio, in which we reject the hypothesis of stationarity (cointegration) at significance level of $\alpha = 0.05$ is presented in Table 1. In all other portfolios we accepted the hypothesis of cointegration for each reselection.

Table 1. Number of reselections for individual portfolio that have not accepted the hypothesis of cointegration

			Reselection interval													
Tracked index	Selection method	Number of shares	Calibration period: 3 years					4 years					5 years			
			10	21	63	126	252	10	21	63	126	252	10	21	63	
DJIA	1.	15	4	1	1	0	0	0	0	0	0	0	0	0	0	0
DJIA	2.	10	18	10	2	1	1	16	8	3	1	1	7	3	1	1
DJCA	2.	10	5	4	2	1	1	1	0	0	0	0	1	0	0	0

In accordance with the conclusions of Alexander and Dimitriu (2002), we can conclude that increasing number of shares in the portfolio and the extension of the calibration period has a positive effect on existence of the cointegration relationship. Portfolios with high number of shares and long calibration period are characterized by strong cointegration relationship with the reference portfolio and it seems that they are the most appropriate for the index tracking.

Method A: stock selection based on the highest price

Before including transaction costs the returns of the constructed portfolios were higher than the returns of the tracked indices in almost all cases. Exceptions were some portfolios composed of 20 shares that tracked DJCA. When including transaction costs, the highest return between the portfolios tracked DJIA was achieved by the portfolio with 15 shares, interval of reselection of 252 trading days, and the calibration period of 5 years. Between the portfolios tracked DJCA it was a portfolio of 50 shares, reselection interval of 126 trading days, and the calibration period of 5 years. Highest value of Information Ratio (including transaction costs of 0.5%) between the portfolios tracked DJIA has a portfolio composed of 25 shares, length of reselection interval of 252 trading days, and the calibration period of 5 years. Between the portfolios tracked DJCA it is a portfolio composed of the full number of shares, reselection interval of 126 trading days, and the calibration period of 4 years.

In terms of profitability, reselection interval is very important. Intervals of reselection of 10 and 21 trading days have high transaction costs that reduce the profitability. Based on our results, we consider the optimal intervals of reselection of 126 or 252 trading days.

Moderate influence on the level of the transaction costs, and therefore on the profitability of the portfolio, has a length of the calibration period. Longer calibration period is mostly related to the stability of the stock selection and lower transaction costs.

Logical assumption that with increasing number of shares the correlation between the returns of the constructed portfolio and the returns of the benchmark index will increase was confirmed by our results. More shares in the portfolio also resulted in a lower volatility of the tracking error in all cases.

Taking into account all monitored characteristics of constructed portfolios, we consider the portfolios with reselection interval of 126 or 252 trading days, 5 year calibration period, and the number of shares of at least two thirds of the total number of shares of the benchmark index to be the most appropriate for the index tracking. All such portfolios achieved excess return and the value of the correlation coefficient between their returns and the returns of the tracked index is more than 0.99. The annualised standard deviation of these portfolios was higher than the annualised standard deviation of the tracked index by less than 1% and the annualised standard deviation of the tracking error was less than 3%.

Method B: stock selection based on the highest value of the correlation coefficient

Before including transaction costs all created portfolios had higher returns than the returns of the benchmark indices. After including transaction costs the highest final value between the portfolios tracked DJIA had the portfolio consisting of 10 shares, reselection interval of 126 trading days, and the calibration period of 5 years. Between the portfolios tracked DJIA it was also the portfolio with 10 shares, calibration period of 5 years, but with reselection interval of 63 trading days. These portfolios also had the highest values of Information ratio. Their profitability and the values of Information ratio were higher than the portfolios composed by using the first method.

Compared to the first method, the second method seems to be more risky. Portfolios created by the first method and composed of the same number of shares, the same reselection interval and the same calibration period have higher correlation of their returns with the returns of the tracked index, less volatility of the returns and less volatility of the tracking error. The second approach has also higher correlation of tracking error with the returns of tracked index, when the average value of the Pearson's correlation coefficient is about 0.29.

An extension of the calibration period or reselection interval decreases the transaction costs for most of the portfolios, as it was in the case of the first method. It also remains true that an increasing number of shares in the portfolio increases the correlation between the returns of the portfolio and the returns of the index and decreases the volatility of the tracking error.

Based on the criteria of profitability and Information ratio, investors should prefer the portfolios with reselection interval of 63 or 126 trading days, the calibration period of 5 years, and the number of shares from 10 to 15 for DJIA and from 10 to 30 for DJCA.

But the indices are better replicated by portfolios with reselection interval of 126 or 252 trading days, the 3 year calibration period and the number of shares at least 2/3 of the total number of shares of the tracked index. All this created portfolios achieved excess returns. The value of the Pearson's correlation coefficient between their returns and the returns of the tracked index is more than 0.97. The maximum difference between the annual volatility of the portfolio and the annual volatility of the index is 2.7% and the maximum annualised standard deviation of the tracking error is 5%.

Because of the paper size limitation only summarized results are included in Appendix B.

6. Conclusion

The aim of the index tracking, which belongs to the passive investment strategies, is the replication of the selected stock index. Cointegration approach to index tracking is based on the existence of the long-run equilibrium between the prices of shares in the constructed portfolio and the prices of the tracked index. Its main advantage is the minimization of the tracking error variance and that deviations from long-run equilibrium tend to come back to it (mean reverting process). In our work we applied the methodology presented in the work of Alexander and Dimitriu on tracking of DJIA and DJCA. In construction of portfolios, we considered various combinations of four key parameters that influence the properties of created portfolios and quality of the index tracking. It is a method of stock selection, number of shares in the portfolio, reselection interval, and calibration period. For all portfolios, we considered three different levels of transaction costs. Portfolios were analysed over a period of eight years.

A key factor in the construction of the portfolio has been the method of the stock selection. In this work we used the stock selection method based on the highest weight in the tracked index (applied in the works of Alexander and Dimitriu, Dunis and Ho), but also a new method of selecting stocks that most correlated with the tracked index over the calibration period. Portfolios constructed using this new method achieved higher returns, but with higher volatility and lower correlation with the tracked index. The high ratio of return on a risk of these portfolios (assessed by Information ratio) may be interesting for less risk averse investors. We consider the creation of other new criteria for the stock selection to be the possible way to modify the properties of constructed portfolios.

Among other factors, the profitability is greatly affected by reselection interval. Longer intervals of reselection, such as 126 or 252 trading days, is connected with lower transaction costs. Transaction costs may significantly reduce the profitability of the portfolio. For a higher correlation with the tracked index and

lower volatility of the portfolio returns and the tracking error, it is appropriate to choose a higher number of shares in the portfolio - at least two thirds of the total number of shares of the reference index. As the ideal length of calibration period we consider the 5 years for the first method and 3 years for the second method. Portfolios with these combinations of key factors provide: excess return with a similar volatility as a reference index, a high correlation between the returns of the portfolio and returns of the reference index, and a low volatility of the tracking error. Cointegration approach to index tracking therefore appears to be successful.

Acknowledgement

This research was supported by Slovak national basic research funding VEGA No. 1/0795/13.

REFERENCES

- [1] **Alexander, C. (1999), *Optimal Hedging Using Cointegration***; Philosophical Transactions of the Royal Society Series A 357, 2039-2058;
- [2] **Alexander, C. (2008), *Market Risk Analysis Volume II – Practical Financial Econometrics***; John Wiley & Sons, Inc. West Sussex;
- [3] **Alexander, C. and Dimitriu, A. (2002), *The Cointegration Alpha: Enhanced Index Tracking and Long-Short Equity Market Neutral Strategies***; ISMA Discussion Papers in Finance;
- [4] **Alexander, C. and Dimitriu, A. (2004), *A Comparison of Cointegration and Tracking Error Models for Mutual Funds and Hedge Funds***; ISMA Center Discussion Papers in Finance;
- [5] **Alexander, C. and Dimitriu, A. (2005), *Hedge Fund Index Tracking***; In: G.N. Gregoriou, G. Hubner, N. Papageorgiou, and F. Rouah (eds), *Hedge Funds: Insights in Performance Measurement, Risk Analysis, and Portfolio Allocation*, pp. 165-179. John Wiley & Sons, Inc. Hoboken, NJ;
- [6] **Alexander, C. and Dimitriu, A. (2005), *Indexing, Cointegration and Equity Market Regimes***; *International Journal of Finance and Economics* 10, pp. 213-231;
- [7] **Alexander, C. and Johnson, A. (1992), *Are Foreign Exchange Markets Really Efficient?***; *Economics Letters* 40, 449-453;
- [8] **Alexander, C. and Johnson, A. (1994), *Dynamic Links***; *Risk* 7(2), 56-61;
- [9] **Bradley, M., Lumpkin, S. (1992), *The Treasury Yield Curve as a Cointegrated System***; *Journal of Financial and Quantitative Analysis* 27, 449-463.
- [10] **Bossaerts, P. (1988), *Common Nonstationary Components of Asset Prices***; *Journal of Economic Dynamics and Control*, Vol 12;
- [11] **Brenner, R. J., Harjes, R. H. and Kroner, K. F. (1996), *Another Look at Alternative Models of the Short Term Interest Rate***; *Journal of Financial and Quantitative Analysis* 31(1), 85-108;

- [12] Davidson, J., Madonia, G. and Westaway, P. (1994), *Modelling the UK Gilt-Edged Market*; Journal of Applied Econometrics 9(3), 231-253;
- [13] Dunis, C. and Ho, R. (2005), *Cointegration Portfolios of European Equities for Index Tracing and Market Neutral Strategies*; Journal of Asset Management 6, pp. 33-52;
- [14] Dwyer, G. and Wallace, M. (1992), *Cointegration and Market Efficiency*; Journal of International Money and Finance, Vol. 11;
- [15] Engle, R. F. and Granger, C.W. J. (1987), *Co-integration and Error Correction: Representation, Estimation and Testing*; Econometrica 55(2), pp. 251-276;
- [16] Gibson, R. C. and Sidoni, CH. J. (2013), *Asset Allocation. Balancing Financial Risk*; McGraw-Hill;
- [17] Glova, J. (2013a), *Determinacia systematickeho rizika kmenovej akcie v modeli casovo-premenliveho fundamentalneho beta*; E+M Ekonomie a Management, Vol. 16, No. 2, p. 139. ISSN 1212-3609;
- [18] Glova, J. (2013b), *Exponential Smoothing Technique in Correlation Structure Forecasting of Visegrad Country Indices*; Journal of Applied Economic Sciences. Vol. 8, No. 2, pp. 184-190. ISSN 1843-6110;
- [19] Granger, C. W. J. (1986), *Developments in the Study of Cointegrated Economic Variables*; Oxford Bulletin of Economics and Statistics 42(3), 213-227;
- [20] Hall, A. D., Anderson, H. M. and Granger, C. W. J. (1992), *A Cointegration Analysis of Treasury Bill Yields*; Review of Economics and Statistics 74(1), 116-126;
- [21] Jain, P. (2003), *Institutional Design and Liquidity at Stock Exchanges Around the World*; working paper, Indiana University, Bloomington, IN;
- [22] Johansen, S. and Juselius, K. (1990), *Maximum Likelihood Estimation and Inference on Cointegration – with Applications to the Demand for Money*; Oxford Bulletin of Economics and Statistics 52(2), 169-210;
- [23] Jones, C. M., Kaul, G. and Lipton, M. L. (2002,1994), *Transactions, Volume and Volatility*; Review of Financial Studies 7, 631-651;
- [24] Markowitz, H. M. (1952), *Portfolio Selection*; Journal of Finance; pp. 77-91;
- [25] Markowitz, H. M. (1959), *Portfolio Selection. Efficient Diversification of Investments*; John Wiley & Sons, Inc., New York;
- [26] Maurer, T. A. (2008), *Cointegration in Finance: An Application to Index Tracking*; Available at SSRN 1586997;
- [27] Nofsinger, J. (2013), *Psychology of Investing*; Pearson Series in Finance. Prentice Hall;
- [28] Reilly, F. K. and Brown, K. C. (2012), *Investment Analysis & Portfolio Management*; South-Western Cengage Learning.

Appendix A

Table 1. Testing the order of integration of log prices of stock indices and their components, and selected measures of descriptive statistics

	DJIA	DJCA	AA	AEP	AES	ALK	AXP	BA	BAC	CAT	CHRW	CNP	CNW	CSX	CVX	D	DD	DIS	DJK	ED	EIX
adf stat	-2.426	-2.263	-2.467	-2.483	-2.889	-1.955	-1.837	-1.836	-2.222	-2.419	-1.723	-2.580	-2.868	-2.431	-2.906	-2.736	-2.345	-2.489	-2.342	-3.110	-1.992
p-value	0.398	0.467	0.381	0.374	0.202	0.597	0.647	0.648	0.484	0.401	0.696	0.333	0.211	0.396	0.195	0.267	0.432	0.371	0.434	0.108	0.582
diff adf stat	-15.220	-15.547	-15.240	-16.794	-14.840	-15.713	-16.215	-15.690	-17.123	-14.718	-16.956	-15.898	-15.807	-16.377	-16.360	-16.740	-15.221	-16.195	-16.298	-15.361	-17.546
diff p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
mean	7.126	7.673	2.896	3.313	2.498	2.880	3.672	3.956	3.033	3.802	3.508	2.437	3.526	2.294	3.989	3.395	3.530	3.314	3.619	3.505	3.246
median	7.127	7.704	3.109	3.275	2.525	2.716	3.704	4.037	3.134	3.953	3.765	2.480	3.506	2.347	4.063	3.367	3.468	3.283	3.646	3.456	3.435
sd	0.270	0.309	0.506	0.279	0.588	0.537	0.345	0.404	0.586	0.571	0.619	0.448	0.248	0.644	0.501	0.361	0.253	0.378	0.359	0.317	0.552
	EXC	EXPD	FDX	FE	GE	GMT	GS	HD	HPQ	IBM	INTC	JBHT	JNJ	JPM	KEX	KO	KSU	LSTR	LUV	MATX	MCD
adf stat	-0.677	-2.054	-2.105	-1.843	-1.745	-2.965	-2.169	-1.107	-1.960	-2.868	-3.209	-2.829	-2.412	-3.420	-2.571	-2.832	-1.941	-1.881	-2.005	-2.294	-3.410
p-value	0.973	0.555	0.534	0.645	0.686	0.170	0.507	0.921	0.595	0.211	0.086	0.227	0.404	0.050	0.337	0.226	0.603	0.629	0.576	0.454	0.052
diff adf stat	-16.281	-16.792	-16.144	-15.022	-15.855	-14.720	-15.727	-15.987	-14.563	-14.518	-14.293	-15.678	-15.666	-16.326	-15.818	-15.825	-16.696	-16.648	-14.983	-16.263	-15.635
diff p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
mean	3.455	3.343	4.325	3.454	3.043	3.309	4.736	3.461	3.248	4.626	2.931	3.033	3.930	3.487	3.401	3.100	3.338	3.396	2.563	2.698	3.635
median	3.560	3.510	4.420	3.514	3.111	3.359	4.753	3.431	3.232	4.533	2.952	3.167	3.932	3.520	3.538	2.996	3.262	3.631	2.630	2.791	3.711
sd	0.399	0.459	0.313	0.333	0.291	0.356	0.348	0.347	0.397	0.367	0.206	0.819	0.218	0.269	0.608	0.294	0.680	0.586	0.268	0.373	0.625
	MMM	MRK	MSFT	NEE	NI	NSC	PCG	PFE	PG	R	T	TRV	UNH	UNP	UPS	UTX	VZ	WMB	WMT	XOM	
adf stat	-2.150	-2.348	-3.087	-2.437	-1.235	-2.862	-2.486	-0.937	-2.977	-2.850	-2.954	-2.911	-1.939	-2.572	-2.107	-2.958	-2.449	-2.317	-1.978	-2.180	
p-value	0.515	0.431	0.118	0.393	0.901	0.213	0.372	0.949	0.165	0.218	0.174	0.192	0.604	0.336	0.533	0.173	0.388	0.444	0.588	0.502	
diff adf stat	-15.351	-15.553	-15.519	-17.097	-14.672	-16.806	-17.897	-16.495	-15.235	-15.410	-16.164	-15.332	-16.159	-16.955	-14.885	-15.452	-15.249	-15.994	-16.691	-16.120	
diff p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	
mean	4.175	3.453	3.144	3.603	2.709	3.568	3.247	2.966	3.856	3.540	3.054	3.698	3.528	3.856	4.048	3.900	3.237	2.650	3.864	3.972	
median	4.179	3.463	3.137	3.739	2.655	3.733	3.431	2.956	3.920	3.637	3.056	3.633	3.550	3.779	4.059	3.974	3.163	2.772	3.817	4.093	
sd	0.270	0.236	0.173	0.454	0.302	0.561	0.503	0.234	0.295	0.403	0.290	0.326	0.443	0.618	0.205	0.417	0.278	0.714	0.186	0.398	

Note: Table shows ADF test statistics for the log prices (adf stat) with p-value (p-value), ADF test statistics for the log returns (diff adf stat) with p-value (diff p value), mean, median, and the standard deviation (sd) of the log prices

Cointegration and its Specific Applications in Portfolio Management

Appendix B

Table 2. Final values and values of Information ratio of portfolios tracked DJIA (method A of stock selection)

	DJIA				Final value tc=0				Final value tc=0.002				Final value tc=0.005				Inf ratio tc=0				Inf ratio tc=0.002				Inf ratio tc=0.005								
	Number of selected stocks																Number of selected stocks																
	Calibration period: 3 years																Calibration period: 3 years																
	15	20	25	30	15	20	25	30	15	20	25	30	15	20	25	30	15	20	25	30	15	20	25	30	15	20	25	30	15	20	25	30	
Reselection interval	10	2.164	2.133	2.275	2.159	2.012	1.984	2.184	2.112	1.785	1.762	2.049	2.041	10	0.219	0.258	0.696	0.561	0.027	-0.013	0.484	0.408	-0.289	-0.459	0.148	0.172							
	21	2.123	2.222	2.292	2.170	2.017	2.118	2.226	2.132	1.859	1.962	2.126	2.076	21	0.166	0.409	0.764	0.605	0.034	0.230	0.603	0.482	-0.178	-0.055	0.354	0.293							
	63	2.158	2.182	2.287	2.137	2.095	2.127	2.244	2.113	2.002	2.046	2.178	2.076	63	0.207	0.358	0.732	0.541	0.131	0.259	0.629	0.451	0.014	0.105	0.472	0.316							
	126	2.284	2.256	2.226	2.158	2.243	2.220	2.197	2.140	2.181	2.166	2.154	2.112	126	0.372	0.516	0.539	0.623	0.323	0.449	0.476	0.556	0.246	0.346	0.380	0.455							
	252	2.082	2.285	2.226	2.159	2.054	2.263	2.210	2.147	2.011	2.228	2.186	2.129	252	0.122	0.552	0.617	0.617	0.084	0.512	0.576	0.574	0.026	0.450	0.516	0.510							
	Reselection interval	Calibration period: 4 years																Calibration period: 4 years															
			15	20	25	30	15	20	25	30	15	20	25	30		15	20	25	30	15	20	25	30	15	20	25	30	15	20	25	30		
		10	2.068	2.024	2.291	2.139	1.917	1.891	2.206	2.097	1.691	1.691	2.078	2.033	10	0.109	0.067	0.803	0.502	-0.109	-0.213	0.586	0.361	-0.470	-0.670	0.244	0.143						
		21	2.104	2.140	2.342	2.158	2.000	2.045	2.279	2.124	1.845	1.902	2.184	2.072	21	0.160	0.290	0.937	0.575	0.013	0.106	0.779	0.459	-0.223	-0.185	0.533	0.283						
		63	2.080	2.010	2.231	2.108	2.022	1.960	2.193	2.086	1.935	1.886	2.138	2.053	63	0.125	0.040	0.644	0.431	0.044	-0.067	0.548	0.351	-0.083	-0.231	0.401	0.229						
		126	2.214	2.198	2.208	2.107	2.176	2.163	2.182	2.091	2.119	2.112	2.142	2.066	126	0.313	0.430	0.549	0.422	0.261	0.361	0.484	0.364	0.182	0.255	0.385	0.276						
		252	2.074	2.229	2.250	2.099	2.048	2.206	2.235	2.088	2.010	2.172	2.212	2.071	252	0.121	0.472	0.685	0.397	0.084	0.428	0.646	0.356	0.028	0.363	0.587	0.294						
Reselection interval		Calibration period: 5 years																Calibration period: 5 years															
			15	20	25	30	15	20	25	30	15	20	25	30		15	20	25	30	15	20	25	30	15	20	25	30	15	20	25	30		
		10	2.040	2.003	2.252	2.138	1.888	1.873	2.175	2.099	1.660	1.677	2.059	2.040	10	0.067	0.023	0.664	0.488	-0.150	-0.242	0.475	0.362	-0.511	-0.671	0.179	0.167						
		21	2.099	2.144	2.263	2.156	1.996	2.049	2.205	2.124	1.841	1.907	2.118	2.077	21	0.147	0.288	0.702	0.558	0.006	0.112	0.559	0.454	-0.219	-0.168	0.337	0.294						
		63	2.239	2.078	2.241	2.134	2.178	2.029	2.205	2.112	2.086	1.955	2.152	2.081	63	0.320	0.177	0.649	0.514	0.245	0.077	0.560	0.439	0.127	-0.076	0.424	0.325						
	126	2.373	2.249	2.230	2.145	2.336	2.216	2.204	2.129	2.279	2.166	2.167	2.104	126	0.493	0.547	0.605	0.553	0.447	0.480	0.544	0.495	0.378	0.377	0.451	0.409							
	252	2.431	2.387	2.345	2.189	2.404	2.366	2.329	2.177	2.364	2.334	2.305	2.160	252	0.558	0.793	0.922	0.710	0.526	0.754	0.883	0.670	0.478	0.694	0.825	0.609							

Note: white colour – final values greater than final value of DJIA (1.991); light grey colour – values between 95% and 100% of the final value of DJIA; dark grey colour – values between 90% and 95% of the final value of DJIA; black colour – values lower than 90% of the final value of DJIA

Table 3. Final values and values of Information ratio of portfolios tracked DJIA (method B of stock selection)

	DJIA												Inf ratio												
	Final value				tc=0				Final value				tc=0.002				Final value				tc=0.005				
	Number of selected stocks												Number of selected stocks												
Reselection interval	Calibration period: 3 years												Calibration period: 3 years												
		10	15	20	25	10	15	20	25	10	15	20	25	10	15	20	25	10	15	20	25	10	15	20	25
	10	3.210	2.885	2.761	2.259	2.884	2.597	2.538	2.125	2.396	2.166	2.203	1.924	0.771	0.753	0.796	0.452	0.598	0.539	0.590	0.233	0.298	0.170	0.245	-0.124
	21	3.130	2.773	2.910	2.265	2.864	2.535	2.715	2.155	2.465	2.177	2.423	1.989	0.735	0.676	0.936	0.462	0.590	0.492	0.764	0.282	0.345	0.181	0.482	-0.005
	63	3.406	3.065	2.912	2.304	3.221	2.912	2.779	2.234	2.943	2.682	2.579	2.130	0.890	0.905	0.976	0.558	0.796	0.796	0.855	0.440	0.645	0.622	0.662	0.257
	126	2.879	3.248	2.757	2.256	2.781	3.142	2.674	2.207	2.635	2.983	2.548	2.132	0.644	1.012	0.819	0.464	0.583	0.942	0.740	0.381	0.486	0.832	0.617	0.253
	252	2.820	2.601	2.430	2.210	2.765	2.554	2.388	2.182	2.684	2.483	2.324	2.139	0.573	0.545	0.494	0.374	0.541	0.506	0.449	0.328	0.490	0.447	0.380	0.257
	Calibration period: 4 years												Calibration period: 4 years												
		10	15	20	25	10	15	20	25	10	15	20	25	10	15	20	25	10	15	20	25	10	15	20	25
	10	3.343	3.557	2.541	2.132	3.071	3.245	2.348	2.031	2.664	2.777	2.058	1.879	0.897	1.068	0.598	0.239	0.750	0.899	0.404	0.069	0.504	0.612	0.081	-0.203
	21	3.211	3.722	2.496	2.134	2.984	3.452	2.334	2.050	2.644	3.047	2.091	1.925	0.834	1.145	0.549	0.244	0.705	1.006	0.386	0.103	0.493	0.776	0.118	-0.120
	63	3.206	3.134	2.584	2.117	3.057	2.986	2.475	2.062	2.835	2.764	2.312	1.979	0.830	0.850	0.649	0.219	0.746	0.758	0.541	0.124	0.614	0.612	0.370	-0.023
	126	3.135	2.980	2.468	2.110	3.045	2.876	2.395	2.070	2.909	2.720	2.287	2.010	0.792	0.739	0.519	0.199	0.740	0.672	0.446	0.133	0.659	0.568	0.333	0.031
	252	2.371	2.336	2.005	2.159	2.328	2.289	1.971	2.132	2.264	2.218	1.920	2.092	0.306	0.281	0.016	0.272	0.274	0.245	-0.024	0.229	0.223	0.189	-0.085	0.165
	Calibration period: 5 years												Calibration period: 5 years												
		10	15	20	25	10	15	20	25	10	15	20	25	10	15	20	25	10	15	20	25	10	15	20	25
	10	3.347	3.023	2.218	2.027	3.135	2.792	2.074	1.945	2.817	2.446	1.860	1.822	0.846	0.798	0.251	0.062	0.739	0.646	0.095	-0.082	0.564	0.392	-0.159	-0.310
21	3.401	3.055	2.242	2.075	3.213	2.860	2.123	2.004	2.931	2.569	1.944	1.896	0.880	0.822	0.271	0.145	0.786	0.695	0.146	0.021	0.635	0.488	-0.055	-0.173	
63	3.677	2.982	2.347	2.086	3.526	2.857	2.263	2.036	3.299	2.668	2.137	1.961	1.001	0.785	0.391	0.166	0.932	0.700	0.304	0.079	0.822	0.566	0.167	-0.054	
126	3.937	2.918	2.093	2.000	3.832	2.835	2.045	1.967	3.673	2.711	1.973	1.917	1.085	0.728	0.121	0.015	1.041	0.671	0.065	-0.046	0.971	0.584	-0.022	-0.140	
252	2.922	2.359	2.198	2.093	2.867	2.323	2.168	2.071	2.784	2.268	2.121	2.039	0.642	0.346	0.238	0.187	0.608	0.314	0.204	0.148	0.557	0.264	0.151	0.089	

Note: white colour – final values greater than final value of DJIA (1.991); light grey colour – values between 95% and 100% of the final value of DJIA; dark grey colour – values between 90% and 95% of the final value of DJIA; black colour – values lower than 90% of the final value of DJIA