

# Some ideas for improving quality of the index tracking based on cointegration

Damián Pastor<sup>1</sup>, Tomáš Sabol<sup>1</sup>, Jozef Glova<sup>1</sup>

<sup>1</sup> Faculty of Economics, Technical University of Košice, B. Nemcovej 32, 040 01 Košice, Slovak Republic

Corresponding author: *Damián Pastor* ([damian.pastor@tuke.sk](mailto:damian.pastor@tuke.sk))

Received 7 December 2015  
Accepted 24 February 2016  
Published 10 November 2016

Academic editor:  
*Živilė Tunčikienė*

## Keywords

Cointegration  
cointegrating vector  
long-run equilibrium relationship  
Minkowski metric  
Engle-Granger methodology  
Johansen's cointegration method

## JEL Classification

G11

## Abstract

Cointegration approach to the passive portfolio management enables to replicate the selected stock index and to construct a portfolio with profitability and risk similar to market. This paper analyzes several options for improving this method. It focuses on one of the key tasks, which is an estimate of long-run equilibrium relationship. Five different methods were proposed and compared. The results confirmed the relevance of using the Engle-Granger methodology in all previous surveys, but it also suggested some interesting properties related to the estimate of regression coefficients based on different variants of the Minkowski metric or to estimate regression equation without intercept.

## Introduction

The traditional construction of a financial portfolio is based on an analysis of the correlation structure among the particular financial assets involved in the portfolio. It was Harry Max Markowitz (1952) 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-variate 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 (Holton 2003). 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.

The concept of cointegration was first introduced by Granger (1981) in the article "Some Properties of Time Series Data and Their Use in Econometric Model Specification". Elaboration of cointegration was published in 1987 in the paper written by Engle and Granger (1987). As reported by these authors, there are time series with values that are very likely not far apart at any point in

time. Economic theory generally assumes forces that keep these time series together. Typical examples are time series of short-term and long-term interest rates. Alexander (2008) provided an overview of the use of cointegration for other time series such as spot and futures prices, stock indices, and exchange rates. Alexander also proposed the use of cointegration for index tracking and hedging (Alexander 1999), and long-short investment strategies (Alexander et al. 2002).

Relevance of using cointegration in portfolio management was supported also by Thomaidis (2013), and Lam, Jamaan (2013), and their part in the personnel management were mentioned by Merkevičius et al. (2015). Thomaidis (2013) extended long-short strategy based on cointegration and his portfolios outperformed benchmarks. Lam, Jamaan (2013) compared goal programming and cointegration approaches and concluded that latter is more appropriate for investors in Malaysia.

The focus of this paper is on cointegration approach to index tracking. Index tracking belongs to the passive portfolio management strategies. Its main aim is to construct a portfolio that will replicate a reference index. In the case of index tracking based on cointegration, the constructed portfolio is cointegrated with tracked index. Values of such portfolio and reference index are tied together and are rarely apart. Extensive research in this field was published by Alexander, Dimitriu (2002). They applied cointegration approach to index tracking on Dow Jones Industrial Average (DJIA) historical prices and reported some positive results. Created portfolio had a similar profitability and volatility as the reference index, returns of the created portfolio and the tracked index were highly positively correlated. Similar results were presented by Dunis, Ho (2005). They have used the same methodology for tracking the Dow Jones EuroStoxx50. Cointegration approach is considered to be appropriate for tracking stock indices also by Maurer (2008), who applied this method for tracking the DJIA, and parts of the Dow Jones Composite Average and FTSE 100.

Acosta-González et al. (2015) proposed a new stock selection procedure based on optimizing the cointegration level of the tracking portfolio and benchmark. They proved that this strategy is able to decrease the number of stocks required to track a stock index successfully.

Alexander, Dimitriu (2005) compared the cointegration and traditional approach to index tracking and concluded that the out-of-sample performance of cointegration-based strategies is similar to that of the traditional tracking error variance-minimizing model. Grobys (2010) made the same comparison using data from Swedish stock market with conclusion that cointegration based models dominate.

Authors of this paper consider as a weakness of the prior research in this area that in all published papers only one method of estimating the cointegration vector – the OLS has appeared. Gonzalo (1994) in his paper compares five methods of estimating the cointegration vector, i.e. estimating the long-run equilibrium relationship,

and mentions other four methods. He further stated that even though these methods are superconsistent, estimates made by different methods may vary significantly. Different estimates of cointegration vector could significantly affect the results of index tracking.

In this paper we will discuss several ways how to improve an estimate of long-run equilibrium relationship (cointegrating vector) that is related to stock weights in a portfolio and could lead to an improvement of the index tracking based on cointegration.

In the first section of this paper we define cointegration and cointegrating vector. The second section is devoted to the estimation of the long-run equilibrium relationship. We introduce five different methods that are later used for the estimation of cointegrating vector and compared. Then we will briefly describe the cointegration approach to index tracking and method of portfolio creation. In the fourth section we list the results.

## 1. Cointegration

Suppose we have the time series  $x_t$  and  $y_t$ , that are  $I(d)$ , i.e. the integrated processes of the same order. Time series is  $I(d)$  – integrated process of order  $d$ , if differencing this time series  $d$  times yields a stationary process. For the mentioned time series that are  $I(d)$  is generally true that their linear combination is also  $I(d)$ . However, if there exists  $a$  such that:

$$z_t = x_t - ay_t, \quad (1)$$

where  $z_t \sim I(d-b)$ ,  $b > 0$ , then the time series  $x_t$  and  $y_t$  are cointegrated. Vector  $(1, -a)$  is called a cointegrating vector. Time series  $z_t$  is a deviation from the long-run equilibrium. Therefore, if  $z_t = 0$ , then we say that the system is in the long-run equilibrium. (Engle and Granger 1987; Alexander 2008)

Cointegration in finance is usually connected with time series that are integrated processes of the first order. The condition of cointegration between two  $I(1)$  time series is that their linear combination is stationary ( $I(0)$ ).

Vector  $a$  expresses the long-run equilibrium relationship between cointegrated time series. Assume that the mean of  $z_t$  is equal to 0. Then, if  $a = 1$ , the difference between  $x_t$  and  $y_t$  will be usually approximately equal to 0. These time series will move apart rarely, on the contrary, their difference will tend to constantly converge to 0 (mean reverting process). If  $a$  has value other than 1, then the time series will diverge. We know that a time series  $x_t$  will have the values about  $ay_t$ , and again, the system will tend to return to the state of the long-run equilibrium. In the case of cointegration we are not able to estimate the specific values of cointegrated time series in the future, but we can define their long-run equilibrium relationship ( $x_t = ay_t$ ). They will move toward this relationship and the large deviation from this equilibrium is exceptional.

Cointegration relationship may exist between more than two time series. Consider a vector  $x_t$  with components that are  $I(d)$ . If there is a vector  $a (\neq 0)$  such that  $z_t = a'x_t \sim I(d-b)$ ,  $b > 0$ , then the components of the vector  $x_t$  are cointegrated (Engle and Granger 1987).

## 2. Estimates of long-run equilibrium relationship

The easiest way to estimate the long-run equilibrium relationship (cointegrating vector), as suggested by Engle and Granger (1987), is to use the Ordinary least squares (OLS). Other possible methods mentioned by Gonzalo (1994) are nonlinear least squares, principal components, canonical correlations, instrumental variables, spectral regression, and maximum likelihood in a fully specified error correction model (also called Johansen’s method). Engle-Granger methodology and Johansen’s cointegration method are the most used methods for estimation of cointegrating vector.

### 2.1. Engle-Granger methodology

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 in the case of index tracking 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  $t$ ,  
 $\alpha$  is the intercept,  
 $\beta_k$  is the regression coefficient of stock  $k$ ,  
 $P_{kt}$  is the price of stock  $k$  in time  $t$ ,  
 $\varepsilon_t$  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.

### 2.2. Engle-Granger methodology without intercept in the regression equation

The methodology is almost the same as described above, except that the regression equation is without intercept so it looks as follows:

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

### 2.3. Engle-Granger methodology with Minkowski metric

Engle and Granger (1987) estimate cointegrating vector by OLS. OLS is the most widely used method for estimating the parameters in a linear regression equation. Hatrák (2007) mentions another possible method – the method of minimizing the absolute values of the deviations. However, this method is computationally demanding and the large errors have the same weights as the small ones. OLS can be computed analytically, large deviations are given higher weights (due to the squaring) and estimate has good statistical properties.

If the minimizing the absolute values of deviations and the squares of deviations can be used for estimating the regression coefficients, it is theoretically possible to use also higher powers and minimize the sum of absolute values of deviations to the third or fourth power. If we neglect the computational demands (we have to use the numerical methods) the use of higher powers gives greater weights to larger deviations, what in many cases may not have a negative impact, but on the contrary, it can be a desired feature of the estimation. Statistical properties mentioned by Hatrák (2007) are not as good as in estimate by OLS, but if the result would be improved quality of index tracking, that is the creation of the portfolio more similar to the reference index, then the replacement of OLS is worth it. Attempt to use method of estimating the regression coefficients other than OLS can be found in the paper by Petras and Podlubny (2007) who criticize this method and propose the method of least circles.

In this article we are going to minimize not only the sum of squared errors (OLS), but also sum of absolute values of errors raised to the other powers. Differences raised to the various powers are commonly used in calculation of distance between objects. Similarly, the most used is Euclidean metric based on squared differences:

$$d_e(x_1, x_2) = \sqrt{\sum_i (x_{i1} - x_{i2})^2} \quad (4)$$

The aforementioned sum of the absolute values of deviations is linked to another metric known as the City-block metric. The super-metric that includes both of these metrics is the Minkowski metric. According to Polovina and Hill (2007) it is calculated as follows:

$$d_m(x_1, x_2) = \sqrt[k]{\sum_i |x_{i1} - x_{i2}|^k} \quad (5)$$

In our work we consider the different values of  $k$  (see Eq. (5)) and we are going to find the values of regression coefficients that minimize the sum of the absolute values of the deviations raised to the power of  $k$ .

### 2.4. Johansen’s cointegration method

Based on the articles by Asteriou, Hall (2011) and Cholotá (2009) we briefly present the steps of Johansen’s approach. The first step is the same as in Engle-Granger

methodology – testing the order of integration of time series. Again, all time series should be integrated of the same order. The second step is to select an appropriate lag length for VAR model. The information criterion (AIC, SBC, or other) are used. The selected model should pass the diagnostic tests, i.e. the tests for the presence of autocorrelation, heteroskedasticity, ARCH effect and normality tests. The third part of the procedure is to create a VECM model, consisting of a long-run model (cointegrating equation) and the short-run model (VAR model). VECM model can be written as follows:

$$\begin{aligned} \Delta Z_t = & \Gamma_1 \Delta Z_{t-1} + \Gamma_2 \Delta Z_{t-2} + \dots + \\ & + \Gamma_{k-1} \Delta Z_{t-k+1} + \Pi Z_{t-1} + u_t \end{aligned} \quad (6)$$

where:

$$\Gamma_i = - \sum_{j=i+1}^k \Pi_j \quad (i = 1, 2, \dots, k-1) \quad ;$$

$$\Pi = - \left( I - \sum_{j=1}^k \Pi_j \right) .$$

The content of the matrix  $\Pi$  is an information about long-run relationships and its rank  $r$  indicates the number of cointegrating vectors. After we create a VECM, we test whether an intercept or/and a trend enter the short-run or the long-run model. There are 5 possible alternatives, in practice we meet with the following three:

- a) Intercept (no trend) in the long-run model, no intercept and no trend in VAR model,
- b) Intercept (no trend) in both models,
- c) Intercept and trend in the long-run model, intercept and no trend in VAR model.

The fourth step is to determine the number of cointegrating vectors. Two methods are used: trace statistics and maximum eigenvalue statistics. For the trace statistics the null hypothesis  $H_0$  is: the number of cointegrating vectors is less than or equal to  $r$ ; alternative hypothesis  $H_1$  is: the number of cointegrating vectors is  $N$ , where  $N$  is the number of endogenous variables for  $r = 0, 1, \dots, N-1$ . The maximum eigenvalue statistics has the null hypothesis  $H_0$ : the number of cointegrating vectors is equal to  $r$ ; and  $H_1$ : the number of cointegrating vectors is equal to  $(r + 1)$ .

If we know the number of cointegrating vectors, we can estimate these vectors by maximum likelihood method. Asteriou, Hall (2011) mention two more steps of Johansen's approach: testing for weak exogeneity and testing for linear restrictions.

## 2.5. Using Engle-Granger methodology and Johansen's cointegration method for index tracking

In terms of the index tracking there are some differences between these two methods. As noted by Alexander (2008), Johansen's cointegration method tries to find the linear combination that is the most stationary. Engle-Granger methodology is looking for a stationary linear combination with minimum variance. While Johansen's method maximizes the stationarity of tracking error, Engle-Granger methodology minimizes its variance. If we measure the risk of the index tracking by variance of tracking error, then Engle-Granger methodology seems to be better.

The second difference is in the choice of explained variable. Returning to equation (1), where  $z_t = x_t - ay_t$ . Then  $x_t = ay_t + z_t$ . If Johansen's cointegration method is applied, then we can replace explained variable and explanatory variable and we get  $y_t = 1/a x_t + z_t$ . For Engle-Granger methodology this is not true and the value of  $a$  is not the same in both equations. As Asteriou, Hall (2011) noted, in case of Engle-Granger methodology the change of explained variable may also cause a different results when testing for cointegration, because the time series  $z_t$  has changed. Although it can be demonstrated that if the samples run to infinity, the results of testing for cointegration are same in both cases, in practice if  $y_t$  is cointegrated with  $x_t$ , then  $x_t$  could be not cointegrated with  $y_t$ . This feature is the great disadvantage of Engle-Granger methodology. If we talk about index tracking the choice of explained variable is clear – it is the price of the followed stock index. Therefore, this disadvantage is irrelevant.

A possible problem for index tracking is existence of more than one cointegrating vector when using the Johansen's approach. Engle-Granger methodology always provides only one cointegrating vector that minimizes the variance of the error term of regression model. But how to choose between several cointegrating vectors? While finding more cointegrating relationships is usually presented as an advantage, for index tracking it is a significant problem.

If we take aforementioned properties into account, Engle-Granger approach seems to be the better choice for index tracking. Nevertheless, we were also interested in the results obtained by Johansen's cointegration method, because we have not heard about its application in this context yet.

## 3. Index tracking

The aim of index tracking is to construct a portfolio that will be as similar in terms of risk and profitability as possible to the followed stock index. Such a portfolio should have approximately the same performance and volatility as the reference index, and minimum tracking error volatility. The Pearson's correlation coefficient between



the returns of the tracking portfolio and the returns of the tracked index should be close to the value of 1.

For index tracking the estimate of the long-run equilibrium relationship is very important. When using Engle and Granger methodology we are interested in estimates of the regression coefficients  $\beta_1, \beta_2, \dots, \beta_k$ . The output of Johansen's cointegration method is cointegrating vector  $(1, -a)$ , where the vector  $a$  is a variant of the vector  $\beta$  (vector of the regression coefficients).

Estimating the cointegrating vector is followed by determining the weights of the stocks in constructed portfolio. Alexander, Dimitriu (2002) calculate the weight of individual stock as a ratio of the estimated regression coefficients to the sum of all regression coefficients:

$$\Pi = - \left( I - \sum_{j=1}^k \Pi_j \right) , \quad (9)$$

Such a portfolio should have the above-mentioned characteristics. Cointegration between the index prices and the stock prices should ensure that the tracking portfolio and tracked index have never drift too far apart and after the deviation from the long-run equilibrium appears, the system tends to come back to the equilibrium.

### 3.1. Portfolio construction

In this article we analyze how the quality of index tracking is affected by estimating the long-run equilibrium relationship. At first we used the daily close prices of the stock index Dow Jones Composite Average (DJA) and the daily close prices of its components adjusted for splits and dividends over the period 29-Dec-00 to 31-Dec-14. The data were downloaded from <http://finance.yahoo.com>. We edited the index in accordance with the procedure used by Alexander, Dimitriu (2002) and created "a reconstructed index". From the components of the DJA on 1-Jan-15 we omitted the shares whose price history was shorter than the specified period. So, we omitted the stocks of American Water Works Company, Inc. (AWK), Delta Air Lines Inc. (DAL), JetBlue Airways Corp. (JBLU), United Continental Holdings (UAL), and Visa Inc. (V).

Based on the paper by Alexander and Dimitriu (2002) we identified four key factors that influence the characteristics of tracking portfolios: method of stock selection, the number of selected stocks, interval of reselection, and calibration period. In this work, we add the method of estimating the long-run equilibrium relationship and we consider the following options of these factors.

#### *Method of stock selection*

We applied only basic strategy of stock selection that was used by Alexander, Dimitriu (2002) and consist of the choice of  $x$  shares with the highest prices at a given time and thus with the highest weights in the index.

#### *The number of selected stocks*

Index tracking usually aims to replicate the benchmark with lower number of stocks. However, low number of stocks are sometimes not able to ensure cointegration. According to Alexander, Dimitriu (2002) 20 stocks are required in a case of DJIA. Therefore, our portfolios were composed of 20, 30, and 40 stocks.

#### *Interval of reselection*

Interval of reselection is the time period after which the portfolio is adjusted. The adjustment of the portfolio includes the stock selection based on above-mentioned method, estimating the cointegrating vector, testing for cointegration, and determining the new weights of the stocks. The length of the interval was 21, 126 or 252 trading days.

#### *Calibration period*

For estimating the long-run equilibrium relationship we used the data over the period of  $t$  years before the date of reselection. This period is called "calibration period". In this paper we used the calibration periods of 5 years.

#### *Method of estimating the long-run equilibrium relationship*

Long-run equilibrium relationship was estimated by: Engle-Granger methodology, Engle-Granger methodology without intercept in the regression equation, Engle-Granger methodology with Minkowski metric (with parameters  $k = 1, 5$ ), and Johansen's cointegration method.

45 combinations of these factors were created in total.

#### *Creation of similar stock indices*

We made a tracking portfolio for each of 45 combinations and we got some surprising results (see Table 1 in Appendix). For example, the final values of some portfolios based on Johansen's cointegration method were very high while some others had final values lower than zero. We wanted to know if these results are just the mistakes or they are repeatable in the future or for other indices. As it is necessary to use a long period for cointegration, we did not consider cutting the specified period into shorter samples as a good solution. Therefore we decided to create similar stock indices as DJA from stocks of S&P 500.

From components of S&P 500 we dropped the companies with history of stock prices shorter than specified period and the share classes other than "A". It yields 421 stocks. We omitted also shares of AIG, because they were the most expensive at the beginning of the year 2001. 420 stocks were then divided to 7 groups alphabetically according to ticker. Price-weighted index was calculated for each group of 60 stocks, so we created 7 indices similar to DJA. We made again 45 tracking portfolios for each index and we compared their performance with performance of portfolios aimed to track original DJA.

### Transaction costs

In accordance with Alexander and Dimitriu (2002), we considered the transaction costs of 0.2% of the trade value (value of purchased and sold shares), and also the zero transaction costs.

### 3.2. Followed portfolio characteristics

We have focused on following characteristics of the tracking portfolios that are important in terms of the index tracking: profitability and volatility of the portfolio, value of Information Ratio, correlation between the returns of the tracking portfolio and the returns of the tracked index, correlation between the returns of the tracked index and the tracking error, and tracking error volatility. Profitability was measured by the final value of the portfolio (value on 31-Dec-14). The starting value (value on 1-Jan-06) was equal to 1. The volatility of the portfolio was determined as the annualized standard deviation of logarithms of daily returns assuming 252 trading days per year. Volatility of the tracking error was determined similarly. Information Ratio is the ratio of the excess return to tracking error volatility. To measure the correlation we used the Pearson's correlation coefficient.

For a successful index tracking is desired to have: similar profitability and volatility of constructed portfolio and tracked index, minimal volatility of tracking error, highly positively correlated returns of the tracking portfolio and the returns of the tracked index, and correlation between the returns of the tracked index and the tracking error close to 0.

## 4. Results

This paper is focused on comparison of portfolios created with an intention of index tracking based on cointegration and with a long-run equilibrium relationship (cointegrating vector) estimated by five different ways. Four of them are based on Engle-Granger methodology. In addition to frequently used estimate of cointegrating vector by OLS, we have considered omitting an intercept in the regression model and estimating based on two different parameters of Minkowski metric ( $k = 1, 5$ ). When estimating the regression coefficients by using Minkowski metric, we tried to minimize a sum of absolute values of the residuals raised to the power of  $k$ . The last method used was Johansen's cointegration method.

When using Johansen's method we applied the procedure described in the second chapter. Lag length for VAR model was chosen based on AIC criterion. Problems appeared after the first VAR model was estimated. Assumptions of no autocorrelation, no ARCH effect, and normality were violated. These assumptions were not fulfilled even with other lag lengths. We have decided to estimate the cointegrating vector anyway. As a final vector from which the weights of stocks were calculated, we selected the vector normalized with respect to the variable DJIA. This variable represents close prices of the tracked index.

Basis of cointegration approach to index tracking is cointegrating relationship between stocks in tracking portfolio and tracked index. Existence of this relationship depends on the estimate of the long-run equilibrium relationship. We tested for cointegration at each portfolio reselection. Hypothesis of cointegration was not rejected when using any of the methods. All methods are able to find a cointegrating relationship between the tracked index and the stocks in constructed portfolio.

Followed portfolio characteristics (they are included in Tables 1 and 2 of Appendix) are very similar in portfolios constructed with OLS and with other parameters of Minkowski metric. We used Wilcoxon signed-rank test and we found out that there are two statistically significant differences between OLS and Minkowski metric with parameter  $k = 5$ . Portfolios based on Minkowski metric ( $k = 5$ ) have the correlation between tracking error and returns of the tracked index closer to zero and lower standard deviation of returns. This fact slightly favors Minkowski metric with  $k = 5$ . Portfolios constructed by this method with 40 stocks and annual reselection tracked the benchmark properly. Portfolios composed of 30 and 20 stocks significantly fall behind in their profitability.

Portfolios created with OLS without intercept have some interesting features. These portfolios have Information ratio values significantly lower compared to portfolios constructed with OLS with intercept, mainly due to the high volatility of the tracking error. Omitting intercept is also reflected in lower correlation between the returns of the tracking portfolio and the returns of the tracked index, and higher transaction costs. On the other hand, lower volatility of daily returns is positive feature. The big difference between these two methods is in correlation between the tracking error and the returns of the tracked index. With intercept the Pearson's correlation coefficient is 0.14 on average, without intercept it is  $-0.54$  on average.

Portfolios constructed based on Johansen's cointegration method do not copy the tracked index in terms of profitability and volatility. We consider uncommon high final values of these portfolios fortunate coincidence and not easily repeatable in the future. Johansen's method is the worst also when looking in correlation between the returns of the tracked portfolio and the returns of the tracked index, or in transaction costs.

## Conclusions

The aim of this paper is to improve cointegration-based index tracking. We focused on the way how to estimate the long-run equilibrium relationship (cointegrating vector), which is connected with determining the weights of stocks in the portfolio. Long-run equilibrium relationship was estimated by Johansen's cointegration method, Engle-Granger methodology based on OLS with intercept and without intercept, and Engle-Granger methodology based on Minkowski metric with parameters  $k = 1, 5$ . Theoretical assumption that Engle-Granger methodology

is better for index tracking than Johansen's method was proved by our results. Portfolios created by Johansen's cointegration methods failed to replicate DJIA in terms of profitability and risk. Significantly better results were achieved by Engle-Granger methodology.

Differences between values of regression coefficients estimated by various parameters of Minkowski metric (including  $k = 2$ , that is OLS) were small, just as differences in followed portfolio characteristics. In terms of risk and minimum correlation between the tracking error and the returns of the tracked index, the best performing portfolios were constructed by use of Minkowski metric with parameter  $k = 5$  with 40 stocks and annual reselection. The relevance of the use of Minkowski metric with parameter  $k = 5$  and its advantages over OLS should be confirmed in future research.

We have identified significant differences in the characteristics of the portfolios between the portfolios constructed using OLS with intercept in the long-run equilibrium equation and without intercept. Portfolios without intercept have lower profitability, higher volatility of tracking error, lower correlation of returns with returns of the tracked index, and higher transaction costs. Thus intercept positively affects the quality of index tracking. On the other hand, estimation of long-run equilibrium equation without intercept would be suitable when decline in the stock market is expected. If stock markets decline, then the negative correlation between the tracking error and the returns of the tracked portfolio lead to higher profitability of portfolio than profitability of tracked index.

## References

- Acosta-González E, Armas-Herrera R, Fernández-Rodríguez F (2015) On the index tracking and the statistical arbitrage choosing the stocks by means of cointegration: the role of stock picking, *Quantitative Finance* 15(6): 1075–1091. <https://doi.org/10.1080/14697688.2014.940604>
- Alexander C (2008) *Market Risk Analysis, Volume II: Practical Financial Econometrics*. Wiley, Hoboken, 426 pp.
- Alexander C (1999) Optimal hedging using cointegration. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 357(1758): 2039–2058. <https://doi.org/10.1098/rsta.1999.0416>
- Alexander C, Dimitriu A (2002) The cointegration alpha: Enhanced index tracking and long-short equity market neutral strategies. ISMA Finance Discussion Paper No. 2002-08. Available at SSRN. <https://doi.org/10.2139/ssrn.315619>
- Alexander C, Dimitriu A (2005) Indexing and statistical arbitrage. *The Journal of Portfolio Management* 31(2): 50–63. <https://doi.org/10.3905/jpm.2005.470578>
- Alexander C, Giblin I, Weddington W (2002) Cointegration and asset allocation: a new active hedge fund strategy. *Research in International Business and Finance* 16: 65–90.
- Asteriou D, Hall SG (2011) *Applied Econometrics*. Palgrave Macmillan, New York, 512 pp.
- Chocholatá M (2009) Aplikácia Johansenovej kointegračnej metódy pri testovaní platnosti parity kúpnej sily na Slovensku. *Ekonomické rozhlady* 38(4): 481–495.
- Dunis CL, Ho R (2005) Cointegration portfolios of European equities for index tracking and market neutral strategies. *Journal of Asset Management* 6(1): 33–52. <https://doi.org/10.1057/palgrave.jam.2240164>
- Engle RF, Granger CW (1987) Co-integration and error correction: representation, estimation, and testing. *Econometrica: Journal of the Econometric Society* 251–276. <https://doi.org/10.2307/1913236>
- Gonzalo J (1994) Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics* 60(1): 203–233. [https://doi.org/10.1016/0304-4076\(94\)90044-2](https://doi.org/10.1016/0304-4076(94)90044-2)
- Granger CW (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16(1): 121–130. [https://doi.org/10.1016/0304-4076\(81\)90079-8](https://doi.org/10.1016/0304-4076(81)90079-8)
- Grobys K (2010) Correlation versus cointegration: Do cointegration based-index-tracking portfolios perform better? Evidence from the Swedish stock-market. *Zeitschrift für Nachwuchswissenschaftler – German Journal for Young Researchers* 2(1): 72–78.
- Hatrák M (2007) *Ekonometria*. Iura Edition, Bratislava, 502 pp.
- Holton G (2003) *Value-at-Risk: Theory and Practice*. Academic Press, Amsterdam and Boston, 405 pp.
- Lam WS, Jamaan SHH (2013) Comparison between goal programming and cointegration approaches in enhanced index tracking. *Proceedings of the 20<sup>th</sup> National Symposium on Mathematical Sciences: Research in Mathematical Sciences: A Catalyst for Creativity and Innovation* 1522(1): 1111–1115. AIP Publishing. <https://doi.org/10.1063/1.4801255>
- Markowitz HM (1952) Portfolio selection. *The Journal of Finance* 7(1): 77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- Maurer TA (2008) Cointegration in Finance: an Application to Index Tracking. Available at SSRN 1586997. <https://doi.org/10.2139/ssrn.1586997>
- Merkevičius J, Davidavičienė V, Raudeliūnienė J, Buleca J (2015) Virtual organization: specifics of creation of personnel management system. *E+M Ekonomikė and Management* 18(4): 200–210. <https://doi.org/10.15240/tul/001/2015-4-014>
- Petrás I, Podlubny I (2007) State space description of national economies: the V4 countries. *Computational Statistics & Data Analysis* 52(2): 1223–1233. <https://doi.org/10.1016/j.csda.2007.05.014>
- Polovina UPS, Hill R (2007) Conceptual structures: knowledge architectures for smart applications. 15<sup>th</sup> International Conference on Conceptual Structures, ICCS 2007, 22–27 July 2007, Sheffield, UK.
- Thomaidis NS (2013) On the application of cointegration analysis in enhanced indexing. *Applied Economics Letters* 20(4): 391–396. <https://doi.org/10.1080/13504851.2012.705420>

## APPENDIX

Table 1. Characteristics of portfolios constructed to track DJIA.

Interval of reselection	OLS			OLS without $\beta_0$			Minkowski k = 1			Minkowski k = 5			Johansen		
	Number of stocks in portfolio														
	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40
<b>Final value of the portfolio / Final value of the tracked index (without transaction costs)</b>															
21	0.873	1.028	1.032	0.656	0.726	0.691	0.877	1.027	1.031	0.854	1.028	1.032	135.34	0	9.930
126	0.924	1.014	1.093	0.823	0.730	0.736	0.917	1.014	1.091	0.913	1.014	1.093	0	0.483	0.085
252	0.925	1.010	1.043	0.680	0.776	0.771	0.919	1.010	1.036	0.933	1.010	1.043	0	0.323	0.040
<b>Final value of the portfolio / Final value of the tracked index (including transaction costs)</b>															
21	0.800	0.967	0.977	0.560	0.643	0.621	0.803	0.966	0.976	0.779	0.967	0.977	9.201	0	0
126	0.899	0.993	1.073	0.787	0.702	0.708	0.891	0.993	1.071	0.887	0.993	1.073	0	0.351	0.004
252	0.909	0.996	1.031	0.662	0.757	0.751	0.903	0.996	1.024	0.917	0.996	1.031	0	0.211	0
<b>Value of Information ratio</b>															
21	-5.142	-1.124	-0.996	-6.447	-5.359	-5.774	-5.025	-1.178	-1.035	-5.757	-1.123	-0.996	0.965	-	-
126	-2.514	-0.227	3.030	-2.576	-4.374	-4.149	-2.699	-0.227	2.962	-2.771	-0.227	3.030	-	-3.980	-3.742
252	-2.175	-0.124	1.320	-4.573	-3.411	-3.339	-2.312	-0.124	1.032	-1.957	-0.124	1.321	-	-1.868	-
<b>Annualized standard deviation of logarithms of daily returns</b>															
21	0.206	0.203	0.203	0.175	0.170	0.165	0.206	0.203	0.203	0.206	0.203	0.203	2.299	4.776	2.143
126	0.205	0.202	0.200	0.172	0.172	0.163	0.205	0.202	0.200	0.206	0.202	0.200	0.494	0.333	1.518
252	0.209	0.200	0.201	0.179	0.176	0.163	0.209	0.200	0.201	0.209	0.200	0.201	1.924	0.841	3.610
<b>Annualized standard deviation of logarithms of tracking error</b>															
21	0.043	0.030	0.023	0.090	0.083	0.083	0.044	0.030	0.023	0.043	0.030	0.023	2.301	4.771	2.133
126	0.042	0.031	0.023	0.093	0.081	0.083	0.043	0.031	0.023	0.043	0.031	0.023	0.512	0.263	1.510
252	0.044	0.031	0.023	0.090	0.082	0.086	0.044	0.031	0.023	0.044	0.031	0.023	1.933	0.832	3.627
<b>Correlation between the returns of the tracking portfolio and the returns of the tracked index</b>															
21	0.977	0.989	0.993	0.895	0.915	0.917	0.977	0.989	0.993	0.978	0.989	0.993	0.037	0.055	0.099
126	0.978	0.988	0.993	0.887	0.917	0.917	0.978	0.988	0.993	0.978	0.988	0.993	-0.041	0.611	0.106
252	0.978	0.988	0.994	0.894	0.914	0.910	0.977	0.988	0.994	0.978	0.988	0.994	0.025	0.160	-0.057
<b>Correlation between the returns of the tracking portfolio and the tracking error</b>															
21	-0.001	-0.003	0.048	-0.491	-0.554	-0.606	-0.002	-0.003	0.048	0.002	-0.003	0.048	-0.051	0.029	0.005
126	-0.003	-0.060	-0.091	-0.523	-0.537	-0.623	-0.002	-0.060	-0.088	0.006	-0.060	-0.091	-0.265	0.009	-0.027
252	0.072	-0.116	-0.068	-0.452	-0.494	-0.615	0.081	-0.116	-0.060	0.071	-0.116	-0.068	-0.099	-0.070	-0.110

Note: Table shows the values of followed characteristics of portfolios created with the intention of DJIA tracking based on cointegration. Their long-run equilibrium relationship was estimated by five different methods: Engle-Granger methodology based on OLS with intercept (OLS) and without intercept (OLS without  $\beta_0$ ), Engle-Granger methodology based on Minkowski metric with parameters k = 1 (Minkowski k = 1) and k = 5 (Minkowski k = 5), and Johansen's cointegration method (Johansen). Portfolios were constructed with 20, 30 or 40 stocks, and interval of reselection of 21, 126 or 252 days. Some values of Information ratio for portfolios based on Johansen's cointegration method are missing, because it is not possible to calculate logarithmic return for portfolios with final value equal to 0.



**Table 2.** Characteristics of portfolios constructed to track created price-weighted indices.

Interval of reselection	OLS			OLS without $\beta_0$			Minkowski k = 1			Minkowski k = 5			Johansen		
	Number of stocks in portfolio									20	30	40	20	30	40
<b>Average final value of the portfolio / Final value of the tracked index (without transaction costs)</b>															
21	0.909	0.882	0.936	0.852	0.866	0.831	0.898	0.880	0.933	0.903	0.908	0.944	1.026	1.915	1.199
	(0.172)	(0.061)	(0.095)	(0.435)	(0.275)	(0.220)	(0.160)	(0.067)	(0.096)	(0.176)	(0.057)	(0.120)	(2.041)	(2.915)	(1.534)
126	0.847	0.872	0.965	0.802	0.846	0.835	0.837	0.874	0.963	0.856	0.874	0.966	3.459	0.288	0.558
	(0.099)	(0.089)	(0.131)	(0.320)	(0.222)	(0.241)	(0.096)	(0.091)	(0.128)	(0.108)	(0.084)	(0.136)	(8.631)	(0.475)	(0.658)
252	0.792	0.889	0.952	0.821	0.892	0.818	0.799	0.889	0.952	0.792	0.897	0.956	0.379	1.145	0.256
	(0.125)	(0.098)	(0.089)	(0.404)	(0.353)	(0.225)	(0.118)	(0.101)	(0.089)	(0.150)	(0.103)	(0.103)	(0.432)	(2.741)	(0.340)
<b>Average final value of the portfolio / Final value of the tracked index (including transaction costs)</b>															
21	0.812	0.802	0.874	0.728	0.756	0.735	0.801	0.800	0.871	0.801	0.822	0.878	0.000	0.000	0.089
	(0.166)	(0.062)	(0.088)	(0.344)	(0.240)	(0.199)	(0.157)	(0.068)	(0.090)	(0.166)	(0.056)	(0.111)	(0.000)	(0.000)	(0.236)
126	0.817	0.847	0.942	0.764	0.808	0.798	0.807	0.849	0.939	0.825	0.849	0.942	0.082	0.000	0.200
	(0.102)	(0.087)	(0.127)	(0.299)	(0.212)	(0.230)	(0.099)	(0.090)	(0.124)	(0.108)	(0.083)	(0.132)	(0.216)	(0.000)	(0.385)
252	0.775	0.873	0.937	0.796	0.867	0.795	0.782	0.873	0.938	0.774	0.880	0.942	0.132	0.875	0.191
	(0.125)	(0.098)	(0.086)	(0.388)	(0.344)	(0.220)	(0.118)	(0.100)	(0.086)	(0.149)	(0.102)	(0.100)	(0.349)	(2.316)	(0.327)
<b>Average value of Information ratio</b>															
21	-2.684	-3.961	-3.232	-3.468	-2.776	-3.089	-2.830	-4.003	-3.304	-2.952	-3.574	-3.182	-	-	-
	(2.282)	(1.188)	(2.229)	(3.732)	(2.659)	(2.362)	(2.130)	(1.287)	(2.223)	(2.381)	(1.199)	(2.768)	-	-	-
126	-2.497	-3.051	-1.756	-2.776	-2.096	-2.430	-2.628	-3.019	-1.802	-2.448	-3.040	-1.777	-	-	-
	(1.251)	(1.743)	(3.290)	(2.949)	(2.173)	(2.591)	(1.183)	(1.793)	(3.229)	(1.463)	(1.664)	(3.459)	-	-	-
252	-3.172	-2.522	-1.799	-2.579	-1.736	-2.323	-3.046	-2.532	-1.779	-3.229	-2.396	-1.741	-	-	-
	(1.963)	(1.968)	(2.202)	(2.952)	(3.102)	(2.513)	(1.830)	(2.010)	(2.192)	(2.370)	(2.042)	(2.395)	-	-	-
<b>Average annualized standard deviation of logarithms of daily returns</b>															
21	0.258	0.247	0.244	0.198	0.191	0.184	0.258	0.247	0.244	0.257	0.247	0.244	1.913	1.793	2.051
	(0.018)	(0.013)	(0.010)	(0.010)	(0.008)	(0.008)	(0.018)	(0.013)	(0.010)	(0.017)	(0.012)	(0.009)	(0.981)	(0.606)	(1.739)
126	0.259	0.245	0.237	0.200	0.192	0.185	0.260	0.244	0.237	0.258	0.244	0.237	1.526	1.324	0.901
	(0.016)	(0.013)	(0.008)	(0.010)	(0.009)	(0.007)	(0.017)	(0.014)	(0.008)	(0.015)	(0.013)	(0.008)	(0.752)	(0.467)	(0.461)
252	0.263	0.243	0.236	0.210	0.193	0.187	0.263	0.243	0.236	0.263	0.243	0.236	1.538	1.624	0.901
	(0.016)	(0.014)	(0.010)	(0.013)	(0.011)	(0.009)	(0.016)	(0.014)	(0.010)	(0.016)	(0.014)	(0.010)	(1.031)	(0.676)	(0.480)
<b>Average annualized standard deviation of logarithms of tracking error</b>															
21	0.080	0.056	0.043	0.121	0.113	0.110	0.080	0.056	0.043	0.079	0.056	0.043	1.672	1.673	1.112
	(0.015)	(0.010)	(0.009)	(0.010)	(0.009)	(0.008)	(0.015)	(0.010)	(0.009)	(0.014)	(0.009)	(0.008)	(0.890)	(0.628)	(0.157)
126	0.081	0.055	0.042	0.122	0.114	0.109	0.081	0.055	0.042	0.079	0.055	0.042	1.414	1.147	0.869
	(0.014)	(0.008)	(0.006)	(0.011)	(0.009)	(0.008)	(0.015)	(0.008)	(0.006)	(0.012)	(0.008)	(0.006)	(0.708)	(0.434)	(0.470)
252	0.084	0.055	0.043	0.122	0.114	0.111	0.084	0.054	0.043	0.084	0.055	0.043	1.467	1.454	0.844
	(0.014)	(0.009)	(0.007)	(0.018)	(0.011)	(0.009)	(0.014)	(0.009)	(0.007)	(0.014)	(0.009)	(0.008)	(1.002)	(0.578)	(0.505)
<b>Average correlation between the returns of the tracking portfolio and the returns of the tracked index</b>															
21	0.954	0.975	0.985	0.851	0.872	0.882	0.955	0.975	0.985	0.955	0.976	0.985	0.173	0.079	0.173
	(0.009)	(0.006)	(0.004)	(0.024)	(0.022)	(0.017)	(0.009)	(0.007)	(0.004)	(0.010)	(0.006)	(0.004)	(0.081)	(0.088)	(0.107)
126	0.954	0.975	0.984	0.848	0.869	0.884	0.954	0.975	0.984	0.954	0.975	0.984	0.174	0.234	0.294
	(0.009)	(0.006)	(0.004)	(0.024)	(0.022)	(0.017)	(0.009)	(0.006)	(0.004)	(0.008)	(0.006)	(0.004)	(0.066)	(0.061)	(0.138)
252	0.952	0.975	0.983	0.851	0.870	0.877	0.952	0.976	0.983	0.952	0.975	0.983	0.222	0.236	0.404
	(0.011)	(0.007)	(0.005)	(0.043)	(0.027)	(0.020)	(0.010)	(0.006)	(0.005)	(0.011)	(0.007)	(0.005)	(0.152)	(0.136)	(0.225)
<b>Average correlation between the returns of the tracking portfolio and the tracking error</b>															
21	0.188	0.194	0.227	-0.509	-0.558	-0.613	0.187	0.191	0.227	0.180	0.194	0.226	0.051	-0.046	0.005
	(0.170)	(0.143)	(0.123)	(0.060)	(0.054)	(0.052)	(0.169)	(0.142)	(0.123)	(0.160)	(0.135)	(0.110)	(0.065)	(0.084)	(0.080)
126	0.203	0.154	0.079	-0.493	-0.550	-0.607	0.206	0.152	0.080	0.191	0.146	0.075	-0.012	0.040	-0.057
	(0.136)	(0.190)	(0.118)	(0.054)	(0.064)	(0.050)	(0.138)	(0.191)	(0.118)	(0.136)	(0.183)	(0.108)	(0.125)	(0.095)	(0.091)
252	0.237	0.121	0.058	-0.422	-0.545	-0.587	0.237	0.121	0.059	0.231	0.114	0.054	0.001	0.075	0.022
	(0.169)	(0.186)	(0.128)	(0.112)	(0.084)	(0.072)	(0.170)	(0.188)	(0.127)	(0.162)	(0.184)	(0.123)	(0.152)	(0.051)	(0.094)

Note: Table shows the average values and standard deviations of followed characteristics of portfolios created with the intention of index tracking based on cointegration. Their long-run equilibrium relationship was estimated by five different methods: Engle-Granger methodology based on OLS with intercept (OLS) and without intercept (OLS without  $\beta_0$ ), Engle-Granger methodology based on Minkowski metric with parameters k = 1 (Minkowski k = 1) and k = 5 (Minkowski k = 5), and Johansen's cointegration method (Johansen). Portfolios were constructed with 20, 30 or 40 stocks, and interval of reselection of 21, 126 or 252 days. Values of Information ratio for portfolios based on Johansen's cointegration method are missing, because it is not possible to calculate logarithmic return for portfolios with final value equal to 0.