UDC - 336.764.061

PROBLEMS OF THE USE OF COINTEGRATION PAIRS FOR PAIRS TRADING

Rinat V. Faizullin
MIREA - Russian Technological University
78 Vernadsky Avenue, Moscow, 119454, RF
rf85@mail.ru
ORCID iD: 0000-0002-1179-3910
Russian Federation

doi: 10.56243/18294898-2022.2-28

Abstract

This paper describes a pairs trading strategy using cointegration approach. If cointegrated pairs are thought of as such pairs, whose linear combination is a stationary process, that is, a process with stable statistical properties, then any deviation from these characteristics will be transient. If you know that such a deviation has happened, that is, a departure from the long-term equilibrium, you can forecast the direction of stock price movements and execute lucrative trades accordingly. When the difference between stock prices exceeds the prediction, we must sell the overpriced asset and acquire the undervalued one, then close the deals when the price ratio returns to long-term equilibrium. This is one form of statistical arbitrage trading strategy. During the research, it was discovered that cointegration is dependent on a variety of factors, including the time period under consideration, and that this is not the only issue. When more recent data is given more weight, it is suggested that methods for determining cointegration be developed. The importance of setting the conditions for entering and terminating a transaction, as well as the possibility of "disappearing" cointegration are also noted as issues with employing cointegrated pairings for pair trading.

Key words: pair trading, cointegration, cointegrated pairs, time series.

Introduction

Data mining is a modern method of obtaining and identifying ideas about the functioning of an object. In cases of working with dynamic objects, multidimensional time series serve as a source of information. Analysis of the dynamics of time series and forecasting their evolution are of great importance for managing various processes in social (for example, election campaigns), economic (stock, futures and commodity markets) and sociotechnical systems (for example, social networks) [1].

The use of the mathematical apparatus for the transformation of many economic and social processes makes it possible to identify qualitative relationships between various factors and the surrounding space that generates information. Dynamic analysis of digital information makes it possible to develop and implement models to rationalize the control of individual parameters, taking into account the specifics and properties of the object in question. In particular, intelligent algorithmization makes it possible to rationalize the behavior of subjects in specific periods of time by processing a large amount of data and identifying temporary elements, for example, in the context of building investment strategies or mathematical processing of the information environment [2,3].

When predicting the state of some object system, we are faced with time series. One of the directions in the development of forecasting methods is based on the search for interdependencies between time series. If there is a linear combination of time series that has the properties of a stationary process, it means that these time series are cointegrated. The issues of forecasting cointegrated time series are covered in a number of sources both in general form [4,5,6] and in relation directly to economic [7,8,9] or technical parameters [10,11,12,13].

Methods

The derivation of a functional dependence is used to anticipate processes (economic, technical, and biological) defined by time series. Changes in the values of time series occur in real life due to a variety of reasons and variables. It is not always possible to construct a multifactorial model of the classical type due to their multiplicity and the difficulty of recognizing relationships with the resulting variable. It is not always possible to construct a multifactorial model of the classical type due to their multiplicity and the difficulty of recognizing relationships with the resulting variable. As a result, when it comes to time series, it's common to believe that internal patterns in the growth of process create the cumulative influence of factors [14].

The time series' value should then be decided by the current time or the values of the indicators that are known at that time. An autoregressive series is one that is dependent on its own past values:

$$\gamma_t = \phi\gamma_{t-1} + \varepsilon_t \tag{1}$$

where the random component ε_t has zero mathematical expectation and stable variance $\varepsilon_t \sim N(0,\sigma^2)$.

When using statistical methods, consistency of mathematical expectation, variance and autocorrelation is a common criterion. The stationary character of process is defined by these conditions. It ensures that an appropriate mathematical model to represent the dynamics of this process is chosen. Many economic processes, however, are non-stationary; yet, if the values of the time series develop gradually over time, the time series can be made stationary by eliminating the trend line. This transformation is called difference-stationary [15] if it results in a stationary series. The series will usually become stationary if you proceed from the original series to its initial differences

$$\Delta\gamma_t = \gamma_t - \gamma_{t-1}$$

$$\Delta\gamma_t = \gamma_t - \gamma_{t-1} = \varepsilon_t \tag{2}$$

Let us go on to the cointegration description, presuming we've decided what we mean by stationary time series and how to make a non-stationary time series stationary.

They are called cointegrated if a linear combination $I\,(1)$ of time series $Y_t = (y_{1t}, y_{2t} \dots y_{nt})^T$ has the features of a stationary series.

$$\beta^T Y_t = \beta_1 y_{1t} + \beta_2 y_{2t} \dots + \beta_n y_{nt} \sim I(0) \tag{3}$$

If there is at least one such vector β, then the data set is considered to be cointegrated. It is clear that if the original series are non-stationary, then for their linear combination to be stationary, then it must include at least two time series. The use of a large number of time series in the cointegration relation is complicated by interdependencies between non-stationary time series, which causes the effect of multicollinearity.

It is clear that any cointegration vector is not unique, because:

$$k \cdot \beta^T Y_t = (\beta^*)^T Y_t \sim I \tag{4}$$

Usually, the cointegration relation leads to the expression of the dependence of one time series on others, then the cointegration vector can be written as:

$$\beta = (1, -\beta_2, \dots, \beta_n)^T \tag{5}$$

Thus, the cointegration relation can be represented as:

$$\beta^T Y_t = y_{1t} - \beta_2 y_{2t} - \dots - \beta_n y_{nt} \sim I(0) \tag{6}$$

or

$$y_{1t} = \beta_2 y_{2t} + \beta_3 y_{3t} + \dots + \beta_n y_{nt} + u_t \tag{7}$$

where $u_t = \beta^T Y_t \sim I(0)$

It turns out that we assume the existence of some kind of long-term equilibrium equation:

$$y_{1t} = \beta_2 y_{2t} + \beta_3 y_{3t} + \dots + \beta_n y_{nt} \tag{8}$$

The fact that some pairs of financial assets are interrelated or dependent on one set of external factors explains cointegration between financial time series. If the presence of cointegration necessitates the formation of a stationary process by the combining of time series. It goes without saying that if there are many time series, there can be multiple such combinations, hence we refer to a set of cointegrating vectors as follows:

$$B^T Y_t = \begin{pmatrix} \beta_1^T Y_t \\ \vdots \\ \beta_r^T Y_t \end{pmatrix} = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{rt} \end{pmatrix} \sim I \tag{9}$$

The time series dependence model can be described in the form of an error-correction model (ECM) in the situation of cointegration of $I(1)$ series [16]. That is, a functional dependency can be formalized as follows:

$$\Delta y_{1t} = f(y_{1t-1} - (\beta_2 y_{2t-1} + \beta_3 y_{3t-1} + \dots + \beta_n y_{nt-1})) = f(u_{t-1}) \tag{10}$$

The random component u_t has no mathematical expectation, but the value of the random component in the previous step determines the next change in the time series. As a result, if the processes are cointegrated, they gravitate to the values established by the long-term equilibrium (8). Returning to the financial market, if two stock prices are cointegrated, their price ratio may be violated at different times, but it will eventually return (tend) to a recognized level. You can use this to monitor the value of the random component u_t, and if its analysis shows that this value began to behave differently than a stationary process at some point, you can suspect the occurrence of some changes in the processes, that is, fix a violation of the long-term equilibrium. In technology, this could mean a breakdown, an accident, the need to replace equipment, and so on, and in the economy, it could mean some structural changes. While the procedures for evaluating time series for stationarity and cointegration are widely documented while methods for anticipating changes in cointegration relationships are not.

Results

The process of identifying cointegration between time series requires mathematical justification. The presence of visual similarity of the series and strong correlation does not guarantee the existence of causal relationships.

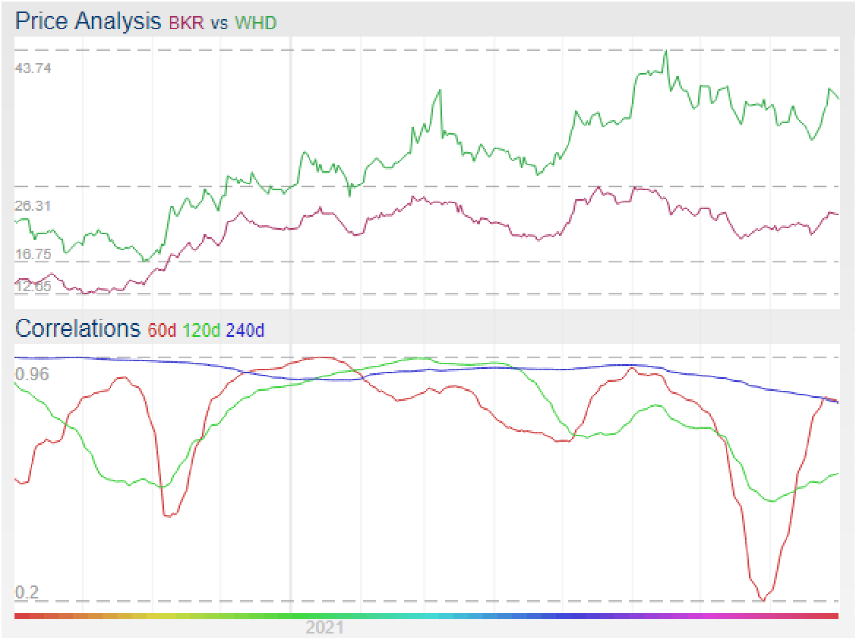

Fig. 1 Stock price dynamics and values 60/120/240 daily Values pairwise correlations for BKR and WHD assets

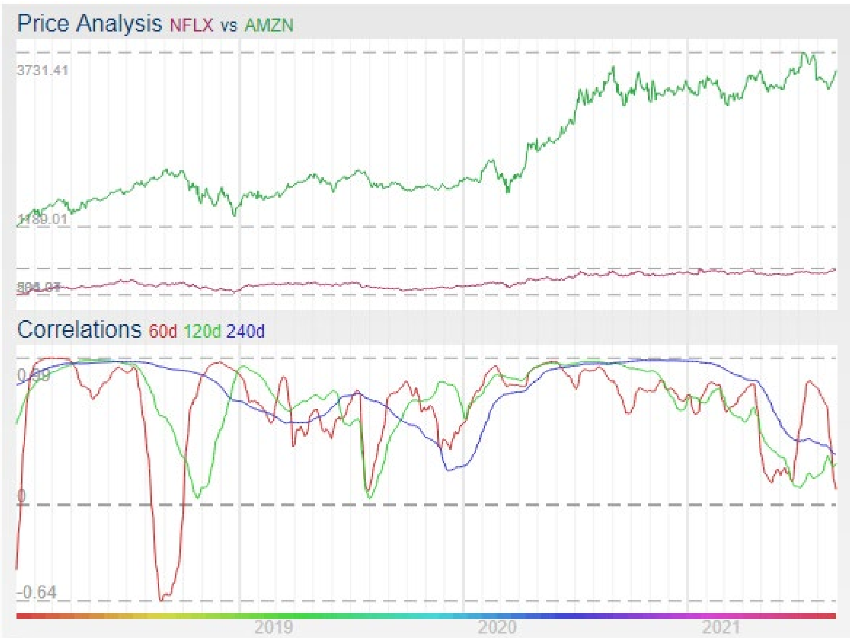

Fig. 2 Stock prices dynamics and values 60/120/240 of daily pairwise correlations for NFLX and AMZN assets from 2018-01-02 to 2021-09-01

There is more similarity between the BKR and WHD stock price charts (Fig. 1) than between the NFLX and AMZN stock price charts (Fig. 2) and the value of the correlation coefficient is higher in the first pair of stocks, however, the Engle-Granger test shows that there is no causal relationship in the first pair. connections (Fig. 3), and in the second pair there is (Fig. 4).

Engle-Granger test: ***BKR(t) = β * WHD(t) + α + r(t)***	
ADF p-value (BKR):	0.3905
H_0 not rejected => BKR is probably I(1) process	
ADF p-value (WHD):	0.6656
H_0 not rejected => WHD is probably I(1) process	
Cointegration p-value:	0.5648
95% confidence: H_0 not rejected => cointegration test not passed	
99% confidence: H_0 not rejected => cointegration test not passed	
Regression coefficient α:	6.626160
Regression coefficient β:	0.472348
Regression R^2:	0.7227
Regression Adjusted R^2:	0.7216

Fig. 3 Engle-Granger test results for BKR and WHD assets

Engle-Granger test: ***AMZN(t) = β * NFLX(t) + α + r(t)***	
ADF p-value (AMZN):	0.8285
H_0 not rejected => AMZN is probably I(1) process	
ADF p-value (NFLX):	0.4908
H_0 not rejected => NFLX is probably I(1) process	
Cointegration p-value:	0.0079
95% confidence: H_0 rejected => cointegration test passed	
99% confidence: H_0 rejected => cointegration test passed	
Regression coefficient α:	-597.308000
Regression coefficient β:	7.273440
Regression R^2:	0.9043
Regression Adjusted R^2:	0.9042

Fig. 4 Engle-Granger test results for assets NFLX and AMZN for the period from: 2018-01-02 to:2021-09-01

Let us look at several cointegration-based trading strategies. Pair trading, or cointegration-based trading, is when trades are conducted with each of the assets in a pair [17]. Private cointegration was utilized by Clegg, M., and Krauss, C. (2018) to select trading pairings and create buy and sell signals. They note that such a trading strategy has good performance indicators. “We compared private cointegration with several classic pair trading options from 1990 to 2015 on a data set for S&P 500 participants. The annual return was more than 12 per cent after operating costs.

These results can only be partly explained by common sources of systematic risk, and are significantly superior to classical distance or cointegration-based pair trading options in our dataset [18].

When a temporal divergence happens, the higher performing stocks (stocks that have increased in value) must be sold, while the lower performing stocks (stocks that have decreased in value) must be bought at the same time. At the same time, it's important to remember that the spread between the two equities will eventually equalize (8). This will occur either as a consequence of the higher performing stocks falling back, or as a result of the lower performing stocks increasing, or both, resulting in a profit on your transaction. There is no profit or loss if both equities move up or down at the same time without changing the spread between them [19].

The spread value is the $u_t = y_{1t} - (\beta_2 y_{2t} + \beta_3 y_{3t} + \cdots + \beta_n y_{nt})$ from formula (7) in general and in particular in pair trading

$$u_t = y_{1t} - (b_1 + b_2 y_{2t}), \tag{11}$$

where y_{1t}, y_{2t} are share prices of the cointegrated pair, b_1, b_2 are coefficients of the regression equation that determine the relationship between them, then the spread has the properties of a stationary process.

Therefore, trading in pairs can be considered as a market-neutral trading strategy that allows traders to profit in different market conditions: rising and falling markets or sideways. This strategy belongs to the category of statistical arbitrage trading strategies.

If the strategy tracks the relationship between the returns of a pair of market assets, then it remains to consider how stable such a relationship is over time. The essence of the strategy is that they can deviate from some of their long-term equilibrium, but return to it. Therefore, trading in pairs can be viewed as a market-neutral trading strategy that allows traders to profit in different market conditions: rising A divergence within a pair can be caused by temporary changes in supply/demand, large buy/sell orders for one security, a reaction to important news about one of the companies, etc.

Due to market neutrality, this trading strategy can be very safe (provided it is diversified) and immune to a global market crisis, even when the entire market or sector collapses. If you trade enough pairs at the same time, your pairs trading portfolio can perform well in difficult market situations as well.

Indeed, in order for the strategy of trading in concatenated pairs to be profitable, no matter in what direction the market or stock prices within the pair move, the main thing is that the linear combination of a stock pair In cases where the difference between stock prices turns out to be greater than the forecast, we need to sell the overvalued asset and buy the undervalued one, and when the price ratio returns to long-term equilibrium, close the deals. known to us can be described as a stationary process with a stable mathematical expectation.

Consider the situation of losing trades. Losses can occur when the cointegration ratio has changed or the cointegration between rows has disappeared. However, rapid market changes can break the relationship (namely a structural gap), which will further lead to huge losses in intraday trading. [20]. That is, if transactions are made for a pair of assets in some ratio, in the hope that the current deviation (10) from this ratio (9) will eventually become 0 (for example), and ratio (9) has changed under the influence of some factors and is no longer a long-term equilibrium, then these transactions will incur losses. It must be understood that cointegration can appear and disappear. Let's consider specific examples. For example, Fig. 2 and 4 show a pair of stocks of Netflix, Inc. and Amazon.com, Inc. and when analyzing the entire period From:2018-01-02 To:2021-09-01, cointegration is observed. Engle-Granger test indicates the presence of a causal relationship. The same test for the period From: 2018-01-02 To: 2018-12-31 indicates a lack of communication (Fig. 5, 6).

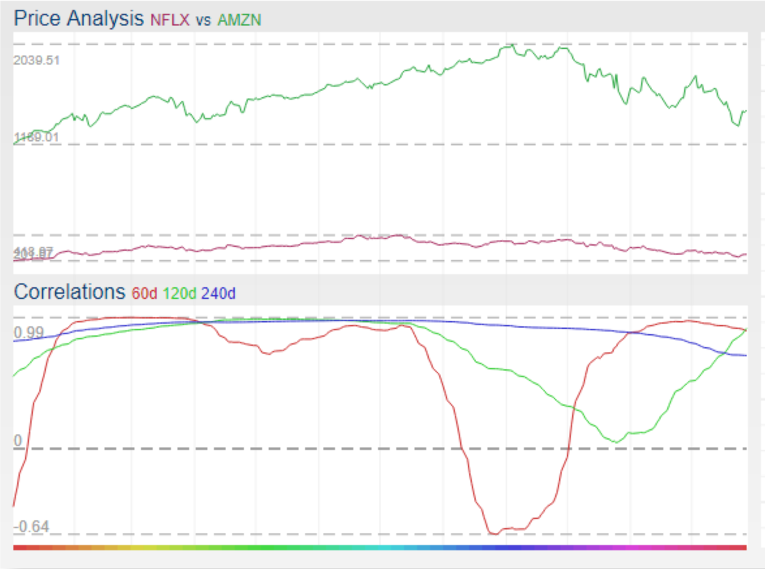

Fig. 5 Stock price dynamics and 60/120/240 daily pair correlations for BKR and WHD assets over the period from 2018-01-02 to 2018-12-31

Engle-Granger test: *AMZN(t) = β * NFLX(t) + α + r(t)*	
ADF p-value (AMZN):	0.2089
☑ H_0 not rejected => AMZN is probably I(1) process	
ADF p-value (NFLX):	0.2243
☑ H_0 not rejected => NFLX is probably I(1) process	
Cointegration p-value:	0.6790
⊗ 95% confidence: H_0 not rejected => cointegration test not passed	
⊗ 99% confidence: H_0 not rejected => cointegration test not passed	
Regression coefficient α:	638.845000
Regression coefficient β:	3.139750
Regression R^2:	0.6089
Regression Adjusted R^2:	0.6074

Fig. 6 Engle-Granger test results for assets NFLX and AMZN for the period from 2018-01-02 to 2018-12-31

The regression equation and the coefficient of determination have also changed.

Conclusion

An analysis of the possibilities of using cointegrated pairs for pair trading showed a number of problems that need to be solved for their successful use.

The first problem associated with determining which period to take for analysis is more general than the problem of analyzing cointegrated pairs. This aspect is important for the analysis of any kind of time series. It seems promising to give more recent data more importance in the analysis, as, for example, is done with exponential smoothing.

Even if we assume that the cointegration ratio is stable, and we believe that it will not change in the future, then the question remains of determining the values of the entry and exit criteria for transactions and whether these criteria will change.

The next objective of the researchers should be to address the fact that the cointegration ratio or the presence of cointegration as a whole can alter (cointegration can “disappear”), and that this process must be anticipated (forecast). As a result, the trader must be given the conditions by which he can stop considering the pair to be cointegrated and stop using it for pair trading. If we assume that the cointegration ratio is constant and will not change in the future, the question of calculating the values of transaction entry and exit criteria as well as whether these criteria will change or remains.

References

1. Andrianova E.G., Golovin S.A., Zykov S.V., Lesko S.A., Chukalina E.R. Review of modern models and methods of analysis of time series of dynamics of processes in social, economic and socio-technical systems (2020) //Russian Technological Journal, 2020, 8(4).- p. 7-45 (In Russ.) https://doi.org/10.32362/2500-316X-2020-8-4-7-45
2. Zaytsev A., Dmitriev N., Fayzullin R., Mihel E. Formation of Investment Behavior Strategy using the Game-theoretic Method (2021) //TEM Journal, 2021, № 2.- p. 673-681.
3. https://www.scopus.com/record/display.uri?eid=2-s2.0-85107796524&origin=resultslist;
4. Rodionov D., Zaytsev A., Konnikov E., Dmitriev N., Dubolazova Y. Modeling Changes in the Enterprise Information Capital in the Digital Economy (2021) //Journal of Open Innovation: Technology, Market, and Complexity, 2021, № 7.- p. 166.
5. https://www.scopus.com/record/display.uri?eid=2-s2.0-85109579005&origin=resultslist
6. Pfaff B. Analysis of integrated and cointegrated time series with R., Springer Science & Business Media, 2008.
7. Granger C.W.J. Time series analysis, cointegration, and applications (2004) //American Economic Review, 2004, T. 94, №. 3.- p. 421-425.
8. Hargreaves C. Non-stationary time series analysis and cointegration (1994) //Oxford University Press.
9. Kunst R.M., Franses P.H. The impact of seasonal constants on forecasting seasonally cointegrated time series (1998) //Journal of Forecasting, 1998, vol. 17, №. 2.- p. 109-124.
10. Engle R., Granger C. Long-run economic relationships (1991) //Readings in cointegration, Oxford University Press.
11. Bahmani-Oskooee M. Export growth and economic growth: An application of cointegration and error-correction modeling (1993) //The Journal of Developing Areas, 1993, vol. 27, №. 4.- p. 535-542.
12. Shi H., Worden K., Cross E.J. A cointegration approach for heteroscedastic data based on a time series decomposition: an application to structural health monitoring (2019) //Mechanical Systems and Signal Processing, 2019, vol. 120.- p. 16-31.
13. de Souza Goncalves E., Rosa P.F.F. Sensor fusion with cointegration analysis for IMU in a simulated fixed-wing UAV //2017 International Conference on Military Technologies (ICMT), IEEE, 2017.- p. 493-499.
14. Li W., Kong D., Wu J. A new hybrid model FPA-SVM considering cointegration for particular matter concentration forecasting: a case study of kunming and yuxi, China (2017) //Computational intelligence and neuroscience, vol. 2017.
15. Faizullin R.V., and Hering S. “Cointegration analysis method for fault detection based on sensor data” (2020) //IOP Conference Series: Materials Science and Engineering. Vol. 971, No. 4, IOP Publishing.

ՀԱՄԱԻՆՏԵԳՐՎԱԾ ԶՈՒՅԳԵՐԻ ՕԳՏԱԳՈՐԾՄԱՆ ԽՆԴԻՐՆԵՐԸ ԶՈՒՅԳ ԹՐԵՅԴԻՆԳԻ ՀԱՄԱՐ

Ֆայզուլլին Ռ.Վ.
МИРЭА - ՌԴ տեխնոլոգիական համալսարան

Քննարկվում է զույգ թրեյդինգի համար համաինտեգրված զույգերի օգտագործման հնարավորությունը: Եթե համաինտեգրված զույգերի տակ հասկացվում են այնպիսի զույգեր, որոնց գծային համադրությունը ստացիոնար գործընթաց է, ապա ենթադրվում է, որ այդ բնութագրերից հնարավոր շեղումը ժամանակավոր կլինի: Երբ տեղի է ունենում նման շեղում, հնարավոր է դառնում կանխատեսել բաժնետոմսերի գների փոփոխությունների ուղղությունը և կատարել համապատասխան շահութաբեր գործարքներ: Ուսումնասիրության ընթացքում հաշվի է առնվել, որ համաինտեգրումը կախված է մի շարք պարամետրերից, այդ թվում՝ դիտարկվող ժամանակահատվածից, և դա միակ խնդիրը չէ: Առաջարկվում է մշակել համաինտեգրման հաշվարկման մեթոդներ, երբ առավել թարմ տվյալներին մեծ կշիռ է տրվում: Հետաքրքրություն են ներկայացնում նաև զույգ թրեյդինգի համար համաինտեգրված զույգերի օգտագործման խնդիրները: Կարևորվում է գործարքից մուտքի և ելքի չափանիշների որոշումը:

Բանալի բառեր. զույգ թրեյդինգ, կոինտեգրացիա, կոինտեգրված զույգեր, ժամանակավոր շարքեր:

ПРОБЛЕМЫ ИСПОЛЬЗОВАНИЯ КОИНТЕГРИРОВАННЫХ ПАР ДЛЯ ПАРНОГО ТРЕЙДИНГА

Файзуллин Р.В.
МИРЭА - Российский технологический университет

В статье рассматривается возможность использования коинтегрированных пар для парного трейдинга. Если под коинтегрированными парами понимаются такие пары, линейная комбинация которых представляют собой стационарный процесс, то есть как процесс со стабильными статистическими характеристиками, то подразумевается что возможное отклонение от этих характеристик будет временными. Если знать что произошло такое отклонение, то есть отклонение от долгосрочного равновесия, то возникает возможность спрогнозировать направление изменений цен акций и сделать соответствующие выгодные сделки. В тех случаях, когда разница между ценами акций оказывается больше прогнозной, нам необходимо продать переоцененный актив и купить недооцененный, а когда соотношение цен вернется к долгосрочному равновесию, закрыть сделки. Это один из видов статистических арбитражных торговых стратегий. В ходе исследования рассмотрено что коинтеграция зависит от ряда параметров, в том числе от рассматриваемого периода и это является не единственной проблемой. Возникает рекомендация разработки методов расчета коинтеграции когда более свежим данным придается больший вес. Также в качестве проблем использования коинтегрированных пар для парного трейдинга выделены важность определения критериев входы и выхода из сделки, а также возможность «исчезновения» коинтеграции.

Ключевые слова: парный трейдинг, коинтеграция, коинтегрированные пары, временные ряды.

Submitted on 01.04.2022.
Sent for review on 01.04.2022.
Guaranteed for printing on 14.06.2022.