IFA IRAN FINANCE ASSOCIATION

Iranian Journal of Finance

Print ISSN 2676-6337
Online ISSN 2676-6345

Pairs Trading Based on Empirical Mode Decomposition (EMD)

Bahareh Zarintaj

PhD Candidate of Industrial Management, Department of Management, Dehaghan Branch, Islamic Azad University, Dehaghan, Iran. (Email: khoshouei@khuisf.ac.ir)

Saeed Aghasi *

*Corresponding Author, Assistant Professor, Department of Management, Dehaghan Branch, Islamic Azad University, Dehaghan, Iran. (Email: esmail.kazemi@khuisf.ac.ir)

Forozan Baktash

Assistant Professor. Department of Economy, Dehaghan Branch, Islamic Azad University, Dehaghan, Iran. (Email: f.baktash@gmail.com)

Iranian Journal of Finance, 2023, Vol. 7, No.3, pp. 95-119.
Publisher: Iran Finance Association
doi: https://doi.org/ 10.30699/IJF.2023.369001.1380
Article Type: Original Article

Received: November 08, 2022
Received in revised form: May 12, 2023
Accepted: June 17, 2023
Published online: September 01, 2023

Abstract

As a trading strategy, pairs trading is performed based on the arbitrage opportunities extracted from statistical models. It is an outcome of the distance between an asset pair and the equilibrium state. Consequently, selecting a pair with the potential to form long-term relationships and reverting to the mean is the main challenge associated with pair trading. Cointegration is one of the most famous statistical tests for selecting a pair's trading. The present study uses Empirical Mode Decomposition (EMD) to decompose the time series of an asset pair price into its constituent elements (intrinsic mode functions). This study examined the property of cointegration across different levels and the corresponding levels of 2- time series to find the cointegration pairs in different decomposition levels and finally examine the resulting profitability. To this end, the profitability of the pairs trading system related to 14 stocks of the Tehran Stock Exchange throughout 2012-2021 was investigated based on EMD. The results showed that the outputs are pretty noticeable for the first level of decomposition (the first intrinsic mode function), and the number of trading opportunities increased by more than two times compared to the normal pair trading with cointegration; the daily returns increased by four times; and the Sharpe ratio increased by about two times compared to the normal pairs trading. The system formed based on the second mode function also outperformed the normal cointegration, and the performance of the third intrinsic mode function is almost on par with that of cointegration. Moreover, the mean transaction duration decreased remarkably in the first and second mode functions.

Keywords: Statistical Arbitrage, Pairs Trading, Empirical Mode Decomposition, Sharpe Ratio.

Introduction

Stock market investors are looking for trading strategies with acceptable returns and a low level of risk. Generally, stocks and securities always involve uncertainty and risk. However, many investors seek the potential to make profits in the stock market using information and various trading strategies and systems. Indeed, they seek to make profits based on a proper understanding of the stock market and its behavior. Every trading system manifests an algorithm to identify mispriced financial assets using sciences such as mathematics, statistics, and finance. Today, the financial markets are witnessing a growing use of computer programs that run a trading system online and in a real-time manner without the involvement of emotions. Therefore, the profitability of these programs depends on the quality of the trading algorithm implemented. Statistical arbitrage is a subclass of trading algorithms that identify assets mispriced using statistical approaches (statistical models).

Statistical arbitrage includes using arbitrage opportunities calculated and obtained based on statistical models. Pairs trading is one of the most common approaches taken in statistical arbitrage. As its name speaks for itself, this trading strategy is used for an asset pair. Pairs trading is one of the most famous and oldest algorithmic trading systems, the efficiency and profitability of which has been demonstrated in many studies conducted on different financial markets (Clegg & Krauss, 2018). Pairs trading is a statistical arbitrage strategy that exploits short-term price divergences between a pair of assets that have historically moved together. In the stock market, pairs traders open a short position in an overvalued stock and a long position in an undervalued stock. When the prices of the two stocks converge, the opened positions are closed by taking the opposite positions (Kim et al.,2022).

The pairs trading approach is based on buying and selling situations on an asset pair (stocks, commodities, options, or money) that form long-term equilibrium relationships. As an investment strategy, pairs trading is neutral to market changes and trends, which enables the trader to earn profits in any market conditions, including downtrend, uptrend, or asymptotic trends, or even in times of high or low volatility (Brahimipour and Davoudi, 1400). This strategy falls in the statistical arbitrage category, given that it considers the buying and selling of equivalent assets simultaneously and makes profits from the price difference between them.

To date, pairs trading has been implemented based on the price or returns behaviors of the assets making up the asset pair. The present study investigates the equilibrium relationship associated with an asset pair in different frequencies based on the decomposition of time series price. Therefore, this study aims to design and examine the profitability of the pairs trading system in different time scales through the time-scale decomposition using the empirical mode decomposition (EMD) approach. Accordingly, the time series of the price of two assets making up a trading pair is decomposed in different time scales, with the property of cointegration examined in each time scale to assess the equilibrium relationship.

Considering different time scales, one can divide the price time series into short-term and long-term trends. Long-term trends have less volatility, while short-term trends involve high volatility. Consequently, studying the pairs trading strategy in the context of time scale enables us to study the effects of short-term and long-term trends on the profitability of the pairs trading. A given pair may lack cointegration property for a time scale. However, it may have this property for another time scale or yield remarkable profitability at a certain time horizon.

Literature Review

Pairs trading is one of the mainstream statistical arbitrage strategies in current financial market practice that can seize possible arbitrage opportunities caused by mispricing between similar assets (Lin & Tan,2023). A concept related to pairs trading is relative pricing. If the statistical model indicates the relative mispricing of two assets, the relatively undervalued asset is bought because it is expected to increase in price. In contrast, an asset valued relatively higher is sold because its price is expected to decrease (Ramos-Requena et al., 2021). Pairs trading identifies arbitrage opportunities using the statistical analysis of two assets simultaneously. What matters most in pairs trading is that it is neutral to the market trend (downtrend or uptrend) because stocks are always sold short, which yields profit in a downtrend market, while the shares are bought that, yield profit in an uptrend market. Therefore, pairs trading can be implemented in both downtrend and uptrend markets (Elliott & Bradrania, 2018)

Pairs trading is implemented using distance, cointegration, Copula, and Kalman filters. The distance method uses the normalized price difference of two stocks as an indicator for relative mispricing, and the trading time is determined by reaching the difference level already set. The cointegration method is applied to the stock pairs which have cointegration properties. The price series of two stocks with the cointegration property is such that if they move away from each other, they are expected to converge in the future. Using the Copula method in the case of the series of the returns of two shares involves calculating the joint frequency function and then the conditional density function, yielding a criterion for mispricing through the conditional density function. In the case of the Kalman filter approach, the equilibrium distance of assets is considered an unobservable random variable estimated through the forecasting and correction process. Other machine learning-based approaches have also been used to implement the pairs trading system. Below, some studies conducted on pairs trading and the relevant approaches are discussed.

Keshavarz Haddad& Talebi(2023) stated that Market practitioners and speculators attempt to benefit from market price gaps and profit opportunities through arbitrage strategies. Although some investors trade stocks based on a particular share's available financial and fundamental information, others profit from risk hedging and swing trading opportunities. One of these strategies is pairs trading, a statistical arbitrage sub-category. Pairs trading can assure reasonably a risk-free profit gaining. This paper aims to make a hypothetical portfolio composed of pairs of stocks by exploring a significant association between their prices in the Toronto Stock Exchange, TSX. This paper compares the profitability of distance, cointegration, and copula functions as the pair's selection and trading strategy devices in TSX from January 2017 to June 2020. The results showed that the highest profitability comes from trading by the copula method. The time frame includes two heterogeneous pre and post-COVID-19 periods. Although the financial markets have been struggling with a complex situation over the COVID-19 days, the performance of the methodologies is not affected by the crisis.

da Silva et al. (2023 proposed an alternative pairs trading strategy based on computing a mispricing index in a novel way via a mixed copula model, or more specifically, via an optimal linear combination of copulas. This paper evaluates our proposed approach's statistical and economic performances by analyzing S&P 500 daily stock returns between 1990 and 2015. Empirical results are obtained not only from the complete sample analysis but also from subperiod analyses. These subperiods are chosen in two ways: i) fixed time length and ii) bull/bear market dependent. Empirical results suggested that overall the mixed copula strategy has superior performance than the distance approach in terms of average returns and Sharpe ratio, considering or not the cost transaction. The superiority is more evident during crisis periods.

Kim et al. (2022) proposed a hybrid deep reinforcement learning method for pairs trading called HDRL-Trader, which employs two independent reinforcement learning networks; one for determining trading actions and the other for determining stop-loss boundaries. Furthermore, HDRL-Trader incorporates novel techniques, such as dimensionality reduction, clustering, regression, behavior cloning, prioritized experience replay, and dynamic delay, into its architecture. The performance of HDRL-Trader is compared with the state-of-the-art reinforcement learning methods for pairs trading (P-DDQN, PTDQN, and P-Trader). The experimental results for twenty stock pairs in the Standard & Poor's 500 index show that HDRL-Trader achieves an average return rate of 82.4%, 25.7%P higher than the second-best method, and yields significantly positive return rates for all stock pairs.

Davallou & Yazdi (2022) stated that the primary purpose of the present research was to compare the performance of pairs trading based on the Vine Copula, Student's t Copula, and Distance approaches. This was done on the Tehran Stock Exchange (TSE) for the first time. The main advantage of the Copula approach over other common approaches in pairs trading, such as the Distance approach, is that in this approach, the stock return distribution does not need to follow the Gaussian distribution. The results showed that both Copula approach strategies produced higher returns and Sharp ratios than the Distance approach and the Tehran Stock Exchange Index (TEDPIX). However, the Distance approach had a lower performance than the TEDPIX. The annual returns of the Vine Copula strategy, Student's t Copula strategy, and Distance approach stood at 194, 171, and 20 percent, respectively. The Sharp ratios of these three strategies were 0.79, 0.70, and -0.04, respectively. Also, the performance of the pairs trading strategy in non-downtrend market conditions was better than in downtrend market conditions.

Dastori & Moradpour(2021) stated that In this study, the main problem is to improve the performance of the high-frequency pair trading algorithm by using a combination of genetic algorithm and fuzzy statistical quality control. For this purpose, two hypotheses have been developed. The statistical population is companies listed on the Tehran Stock Exchange, the sample was limited to the top 50 companies due to the need for high-volume transactions, and 33 shares in 9 industries were selected. After implementing three basic methods, fuzzy statistical quality control and the combined genetic algorithm-fuzzy statistical quality control method, the methods' performance results were compared. The results showed that in the primary method, 43.10% return; in the fuzzy statistical quality control method, 55.58% return; and in the combined genetic algorithm-fuzzy statistical quality control method, the average return was 63.59%. The t-test shows a statistically significant difference between the specific performance of the basic methods and fuzzy statistical quality control, as well as the basic methods and the combined genetic-fuzzy statistical algorithm quality control. Based on the results of the fuzzy statistical quality control model and genetic algorithm, which has a significant increase compared to previous models in increasing the average return.

Brahimipour and Dawoodi (2021) investigated the profitability of the pairs trading strategy based on the stock market's linear state-space system and the Kalman filter. The study uses a pairs trading strategy based on the description of the observable process, i.e., the residuals of the cointegration model in terms of an unobservable process with the property of mean reversion and a state-space model. The profitability of the pairs trading strategy on 21 stocks from a subset of petroleum products stocks and basic metals industries in the Tehran Stock Exchange shows that the pairs trading model used in the study is more profitable in terms of Sharpe ratio compared to pairs trading in terms of cointegration and market performance.

Ramos et al. (2021) selected pairs with highly stable series. Eventually, they developed a mean reversion strategy for these pairs, showing that adopting such an approach to pairs trading yields positive results in uptrend markets and periods of low volatility. Lu et al. (2021) designed a two-phase pairs trading strategy optimization framework, i.e., a pair's strategy that warns of the break mode using machine learning techniques. The first phase is a hybrid model that extracts frequency and time domain features to detect structural breaks, and the second phase implements pairs trading by sensing significant risks, including structural break and close market risks, using a new reinforcement learning model. Han et al. (2021) developed a pairs trading strategy through unsupervised learning. They identified the pairs by combining company features (basic information) and price information. The firm features contribute essential information which can be used to identify pairs and to improve the performance of the pairs trading strategy significantly.

Requena et al. (2020) developed a pairs trading strategy using a Hurst exponent approach. A pair with a Hurst exponent less than 0/5 can be selected as a paired trading pair and become open and close trading situations based on the normalized series distance of the two stocks. The results of the research on the stocks of the S&P group from 2000 to 2015 indicate the better performance of pairs trading using the Hurst exponent approach compared to distance and self-integration methods.

Fallahpour and Hakimian (2019) selected trading thresholds and appropriate time windows using reinforcement learning to maximize returns and minimize harmful risks in pairs trading. The analysis results of intraday data of selected stock pairs show that using reinforcement learning methods for designing the trading system in pairs trading is more advantageous than those found in the previous works. Dastoori et al. (2017) investigated one of the limitations of the pairs trading algorithm, i.e., limits and fixed rules, which cannot account for some system dynamism. The results showed that the modified algorithm increased the returns by 57/95% in the same period of investment, while the basic model yielded a 46/17 % return for the investor.

Clegg and Krauss (2018) examined the profitability of using the concept of partial cointegration in the pairs trading system. They measured the parameters related to partial cointegration by representing the linear state space and obtaining the approximate coefficients using maximum likelihood. The results show that the profitability of the mentioned strategy is around 12% per year, including transaction costs. This return is higher than the distance approach and market performance. Elliott and Bardrania (2018) developed a dynamic model for discrete pairs trading. For this purpose, they considered several hidden states based on the Markov switching approach for the distance process. The results show that the selection of states number affects the estimation of model parameters.

Tadi et al. (2017) applied a pair strategy to stock pairs from the metal ore mining industry and price data over 2015. They compared the performance of the pairs trading strategy to the buy-and-hold strategy by applying the pair's strategy as a post-test. The results indicated that assuming the existence of a short selling system with the range of the optimal threshold, the pairs trading yields higher returns than that of the buy and hold strategy. Fallahpour and Hakimian (2016) examined the performance of the pairs trading system in the Tehran Stock Exchange using a cointegration system by calculating and examining the returns and the Sortino ratio. The analysis results of selected stock pairs in the Tehran Stock Exchange show that using the pairs trading system as a neutral trading system to market changes and trends yields higher returns than the normal return on stocks in the same period.

Stübinger and Bredthauer (2017) examined the profitability of various pairs trading strategies in the New York Stock Exchange from 1998 to 2015. The data used had a high frequency and was analyzed minute by minute. The results showed that the best pair's strategy has an annual return of 0.5050 with a Sharpe ratio of 8.14 with transaction costs included. Research shows that although the profitability of pair's strategies has decreased over time, it can still be a good option for entering trading situations. Moura et al. (2016) sought to explain a pairs trading strategy by representing the cointegration and distance in a linear space state system. In this model, the random process of the cointegration and distance is modeled based on reversion to the mean, which acts as an unobservable process in the state space system. Applying this system to the American and Brazilian stock markets yielded higher returns than the markets.

Stock prices can be studied as a time series variable. The time series is usually studied in a time context, and it can also be studied in other spaces, such as frequency space, to show the hidden information in the time series. From this perspective, the time series or signal is considered a linear combination of a series of basic signals with a certain frequency. Here, high frequencies correspond to short-term behaviors, and low frequencies represent the dominant trend and the direction of long-term movement.

High frequencies can also be deemed as noise occurring in the short term. For example, short-term fluctuations lasting just a few minutes as noise can influence the main price trends in the stock market. The behavior of a series or signal in the frequency space can be investigated using a mathematical theory of Fourier analysis. Also, the time-frequency space can be examined using a wavelet. In the Fourier transform, the time series is written in terms of sine and cosine waves, and in the case of the wavelet transform, a series of basic signals are used, called wavelets, whose spatial and temporal scales make up the original series. Wavelets are a set of mathematical functions used to decompose a continuous signal into its frequency components, with the resolution of each component being equal to its scale. Wavelet transform of a function decomposition is based on wavelet functions. Wavelets (daughter wavelets) are transferred and scaled copies of a function (mother wavelet) with a finite length and highly damped oscillation.

Wavelet transform has superior localization features compared to the Fourier transform. Generally, frequency or spectral analysis involves transforming a time series into components with different amplitudes and frequencies using an orthogonal transformation. Each of these components is a new time series with features other than preserving the energy of the original time series. They have other additional features (Chacon et al., 2020). Traditional spectral analysis methods, such as Fourier analysis, have limitations, including their unsuitability for analyzing sudden changes, discontinuities, or other local changes in time series. In short, it can be said that the traditional spectral analysis tools are inefficient in the case of non-stationary time series (Zhifeng & Huan Zhu, 2020)

As another commonly used method, Empirical Mode Decomposition decomposes the time series into time-frequency space, taking advantage of the time series and its changes. It creates the series making up the primary signal and can show short-term and long-term behaviors in the main series. Indeed, the term "empirical" refers to the fact that there is no predetermined basic signal with a certain frequency to measure the similarity of the series under study, and the series itself is used to calculate its frequencies (Noemi et al., 2018). The empirical mode decomposition (EMD) is a very effective tool for analyzing nonlinear and non-stationary time series, among which the financial time series is the most attractive. The EMD method was proposed by Norden Huang, aiming at tackling the non-stationary of many real-life datasets) Yang et al. 2019). Generally, Empirical Mode Decomposition decomposes a series into several sub-series at different frequencies and some residuals so that the information hidden in the series can be shown as constituent sub-series. Empirical mode decomposition mainly decomposes natural, non-stationary, and nonlinear signals into their constituent components (modes). Natural signals can be created from the combination of different sources (such as music); therefore, each source can show itself in a time interval. Hence, natural mode decomposition is superior to Fourier series and wavelet. Empirical mode decomposition has many applications, including biology, neuroscience, chemistry, physics, finance, and image processing. (Yujun et al., 2020).

In finance, empirical mode decomposition is often used to predict the returns of financial assets. For example, Yang et al. (2023) presented a method answer to the experimental situation analysis and modal synthesis method. Jin et al. (2022) stated that This article aims to propose a deep learning model that autonomously mines the statistical rules of data and guides the financial market transactions based on empirical mode decomposition (EMD) with back-propagation neural networks (BPNN). The intrinsic wave pattern was obtained and then decomposed through the characteristic time scale of the data. Financial market transaction data were analyzed, optimized using PSO, and predicted. Combining the nonlinear and non-stationary financial time series can improve prediction accuracy. Based on the analysis of massive financial trading data, the predictive model of deep learning can forecast the future trend of financial market price, forming a trading signal when particular confidence is satisfied. The empirical results show that the EMD-based deep learning model has an excellent predicting performance. Yujun et al. (2020) sought to predict stock prices using the combined method of empirical mode decomposition and short-term and long-term memory networks (as a tool in deep learning). The primary signal, i.e., the price data series, is decomposed into a set of sub-series using empirical mode decomposition. The short-term and long-term memory model predicts these sub-series, which are finally cointegrated to form the final signal. The prediction results of 5 indices from S&P showed that the combined model outperforms other modes such as support vector and ordinal regression. Chakan et al. (2020) used the combined method of empirical mode decomposition and recurrent neural networks to predict stock prices. In this model, the empirical decomposition involves using the entropy sampling method to remove noise.

Zhifeng & Huan Zhu (2020) used the combined method of the sum of components and mode decomposition of cointegration to predict stock returns. This method makes it possible to determine returns based on three linearly combined components. The components are cash profit forecast, growth rate, and revenue growth. Awajan et al. (2019) used a combined method, including empirical mode decomposition and the Holt-Winters forecasting method, to predict the stock market index of Brazil, Indonesia, and India. Noemi et al. (2018) used a combined support vector regression approach and empirical mode decomposition to forecast stock prices. In this approach, the analysis series obtained from the empirical mode decomposition is factored into the support vector regression. The results showed that this method is efficient for companies with high price frequency, which has good accuracy in short-term time horizons of 20 to 30 minutes, while it yields higher accuracy for stocks with low frequency in longer horizons.

The literature review on pairs trading and empirical mode decomposition reveals that previous studies have yet to examine the use of empirical mode decomposition in implementing the pairs trading system. The current study uses empirical mode decomposition to decompose the time series of an asset pair into intrinsic mode functions (the series obtained from the decomposition) to figure out the short-term and long-term behaviors of the time series. The rationale for using empirical mode decomposition in pairs trading strategy is based on two notions. One is that the two-time series may have a different cointegration property at the level. However, they have this property at the corresponding levels derived from the decomposition, or they may have excellent and appropriate performance at a specific level of the decomposition demonstrated by this methodology. For example, at the levels related to high frequencies, there is expected to be a reduction in the average transaction time. The following section gives a detailed description of the implementation of the models.

Research Model

In this study, the pairs trading system is implemented based on empirical mode decomposition using the cointegration property to examine the existence of a long-term equilibrium relationship. Indeed, two series with cointegration properties are considered suitable pairs for pairs trading. This property and its contribution to pairs trading system performance are initially discussed. As a broad concept, cointegration involves a stationary relationship between the non-stationary variables. The cointegration analysis is mainly based on the presumption that although many economic time series are non-stationary and have a random rising or falling trend, a linear combination of these variables may always be stationary with no random trend in the long term.

From an economic perspective, cointegration is concerned with relating two or more time series underpinned by a theoretical foundation aimed at froing a long-term equilibrium relationship. Although these time series may be random (non-stationary), they follow each other over time, so the differences between them are stationary. Therefore, the concept of "cointegration" invokes a long-term equilibrium relationship. For example, take $y_t = (\mathrm{y}_{1t}, \mathrm{y}_{2t}, \ldots, \mathrm{y}_{nt})'$ represents a vector of n first-order cointegrated time series, i.e., the series itself is not stationary, but its difference series is stationary. In other words, $\forall i : \mathrm{y}_{it} \approx I(1)$ Vector y_t has a cointegrating property when there is a real vector $\beta = (\beta_1, \beta_2, \ldots, \beta_n)'$ so that

$$\beta' y_t = \beta_1 y_{1t} + \beta_2 y_{2t} + ... + \beta_n y_{nt} \approx I(0) \qquad (1)$$

In Equation (1), the vector β is called a cointegrating coefficient vector and $I(0)$ denotes "stationary." In a sense, several non-stationary time series form a linear stationary combination where it β is normalized. A vector with normalized coefficients is represented as $\beta = (1, -\beta_2, ..., -\beta_n)'$. Thus, the cointegrating relationship is shown as follows:

$$y_{1t} = \beta_2 y_{2t} + ... + \beta_n y_{nt} + u_t \qquad u_t \approx I(0) \qquad (2)$$

Where u_t represents the non-equilibrium error. Thus, if the pair $y_t = (y_{1t}, y_{2t})'$ is cointegrated

$$\beta y_t = y_{1t} - \beta_2 y_{2t} \approx I(0) \qquad (3)$$.

Engle and Granger (1987) showed that a cointegrating relationship as an error correction model can be represented as follows:

$$\begin{aligned} \Delta y_{1t} &= c_1 + \alpha_1 (y_{1t-1} - \beta_2 y_{2t-1}) + \sum_j \psi_{11}^j \Delta y_{1t-j} + \sum_j \psi_{12}^j \Delta y_{2t-j} + \varepsilon_{1t} \\ \Delta y_{2t} &= c_2 + \alpha_2 (y_{1t-1} - \beta_2 y_{2t-1}) + \sum_j \psi_{21}^j \Delta y_{1t-j} + \sum_j \psi_{22}^j \Delta y_{2t-j} + \varepsilon_{2t} \end{aligned} \qquad (4)$$

Which accounts for the dynamic behaviors of y_{1t} y_{1t} and. The Engle-Granger method adopts a two-step approach to investigating cointegration. In the first step, $y_{1t} = c + \beta y_{2t} + u_t$ regression is approximated using the least squared errors. Then, the regression residuals are examined using the stationarity tests. If the stationarity is confirmed, the time series y_{1t} y_{1t} are cointegrated with the vector of cointegration coefficients $(1, -\beta)'$.

Following the introduction of the cointegration property, the pairs trading mechanism based on this property is discussed. Cointegration-based pairs trading involves considering all possible candidate financial asset pairs. Then, the property of cointegration for each pair is investigated using the Engle-Granger test. The output of this test includes the p-Value and the cointegration coefficient. For example, if the p-value of the test is below 0.05, two series have the property of cointegration at the confidence level of 0.95; therefore, they will form an equilibrium in the long term, following each other. According to this, if the two price series p_t^Y p_t^X have cointegration properties, the deviation of the two series from their equilibrium state is calculated as $d_t = p_t^X - \beta p_t^Y$.

Then, the average and standard deviation of the series d_t in the in-sample data denoted by μ σ and respectively are calculated, and the normalized series $s_t = \frac{d_t - \mu}{\sigma}$ is used as a criterion for relative pricing. In the case of out-of-sample data, when the normalized distance is more than two, the x share is sold (short selling), and the β number of shares y is bought. If the distance is less than minus two, share Y is sold (short selling), and the β number of shares X is bought. The trading situation is closed when this distance reaches zero or a predetermined value. In the present study, since short selling does not occur in Tehran Stock Exchange, implementing the pair's strategy involves only the purchase transaction.

What follows is a description of how the EMD algorithm is used to decompose the price series of an asset. This decomposition yields series or functions forming a somewhat orthogonal basis. These functions are called Intrinsic Mode Functions (IMF). Decomposition decomposes a time series or signals into several IMF, each one of which has the two following characteristics:

1-There is only one extremum between two consecutive segments of zero (two consecutive zeros).

2- Their mean is zero.

Any other fluctuation fluctuating around a point shares the two last characteristics. The process or the algorithm used to find IMFs is called the sifting algorithm that is comprised of three steps:

1-Finding the local extrema of the signal

2- Joining local maxima using a cubic spline. The resulting diagram is called an upper envelope.

3- Joining the local minima using a cubic spline. The resulting diagram is called a lower envelope.

The mean of two series of the upper and lower envelope is called m_1. The first series, h1, is formed by subtracting m_1 from the main signal.

$$X(t) - m_1 = h_1. \tag{5}$$

The expressed algorithm recurs h_1, and the upper and lower envelopes are calculated h_1. As equation (6) shows, it h_{11} is obtained by subtracting the upper mean series m_{11} h_1.

$$h_1 - m_{11} = h_{11}. \tag{6}$$

The continuation of this process for k iterations yields the first IMF, i.e., c_1

$$h_{1(k-1)} - m_{1k} = h_{1k}, \quad c_1 = h_{1k}. \tag{7}$$

K is often selected in such a way that the criterion

$$SD_k = \sum_{t=0}^{T} \frac{| h_{k-1}(t) - h_k(t) |^2}{h_{k-1}^2(t)}. \tag{8}$$

It is less than a predetermined value. Having the first IMF makes it possible to calculate the residual r_1 using the following equation.

$$X(t) - c_1 = r_1 \tag{9}$$

$X(t)$ It is replaced with, r_1 and the sifting process continues r_1 to calculate the second IMF, and…..n IMFs will be obtained after calculations.

$$X(t) = \sum_{j=1}^{n} c_j + r_n. \tag{10}$$

Cauchy criterion (10) was used to examine whether or not a subseries can be used as an intrinsic state function. The research model, i.e., EMD-based pairs trading, applies the property of cointegration to the series extracted from the decomposition of the price series of an asset pair. Accordingly, two price time series in an asset pair were first decomposed using empirical mode decomposition to calculate the intrinsic mode functions and residuals. Then, the cointegrating relationship between the intrinsic modes functions located at the same level of decomposition was investigated. For example, if two series (x_t, y_t) represent the price series of two assets in a pair, and the empirical mode decomposition decomposes them as follows:

$$x_t = IMF_{1t}^{x} + IMF_{2t}^{x} + ... + IMF_{nt}^{x} + \mathrm{Re}\, s_t^{x} \tag{11}$$
$$y_t = IMF_{1t}^{y} + IMF_{2t}^{y} + ... + IMF_{nt}^{y} + \mathrm{Re}\, s_t^{y}$$

Then, the cointegration property is also examined in the following pairs.

$$(IMF_{1t}^{x}, IMF_{1t}^{y}), (IMF_{2t}^{x}, IMF_{2t}^{y}), ..., (IMF_{nt}^{x}, IMF_{nt}^{y}), (\mathrm{Re}\, s_t^{x}, \mathrm{Re}\, s_t^{y}) \tag{12}$$

If the cointegration property is confirmed, the pairs trading system is implemented on them using the historical data. This would determine their performance in terms of the criteria, such as mean return, risk, and Sharpe Ratio. The following sections describe the actual implementation of the proposed model.

Data Analyses

This section presents an overview of the statistical data used in this study. The data includes 2431 daily returns from 14 industries or indexes obtained from the Tehran Stock Exchange throughout 2012-2021. Table (1) shows these statistical data.

Table1. Descriptive statistics of daily returns on assets

Statistical index industry	mean	median	maximum	minimum	SD	Sharpe Ratio	p-value Jarque-Bera Statistic
1-mineral ore	0.001941	-0.00025	0.204	-0.122	0.051	0.038	0
2-cement	0.001739	-0.00046	0.183	-0.131	0.042	0.041	0
3-tile	0.002093	0	0.344	-0.161	0.049	0.042	0
4-vehicle	0.001601	0	0.245	-0.152	0.064	0.025	0
5-chemicals	0.00191	0.000111	0.275	-0.117	0.041	0.046	0
5-technical	0.001699	0	0.252	-0.173	0.066	0.025	0
6-metals	0.001891	-0.00011	0.23	-0.155	0.047	0.040	0
7 *pharmaceutical*	0.001901	0	0.188	-0.145	0.041	0.046	0
8-food	0.001924	0	0.278	-0.127	0.045	0.042	0
9-sugar	0.001924	0	0.278	-0.127	0.045	0.042	0
10-machinery	0.001985	0.000439	0.221	-0.129	0.043	0.046	0
11-oil products	0.001716	0.000202	0.218	-0.108	0.04	0.042	0
12-investment	0.001432	-0.00021	0.261	-0.123	0.046	0.031	0
13-bank	0.00193	0	0.256	-0.18	0.046	0.041	0

As Table (1) shows, the average daily return of the indices falls within the range of 0.0014 to 0.002. In the finance context, the standard deviation is one of the risk assessment tools that indicate the dispersion of data around the average. For example, the pharmaceutical industry has an average daily return of 0.0019 and a risk of 0.041. The Sharpe Ratio is also obtained by dividing return by risk, showing how much return is obtained per unit of risk, which falls in the range of 0.02 to 0.05. The p-value corresponding to the Jarque-Bera statistic shows that the daily return distribution does not belong to the normal distribution at the confidence level of 0.095.

Based on the pairs trading system, the property of cointegration in the combinations of different pairs from 14 industries was first investigated using the Engel-Grancher test at a confidence level of 0.95. Based on the obtained results, 28 stock pairs were found to have the property of cointegration. For example, the pair (1 and 7) are cointegrated, and their normalized residuals are presented in graph (1).

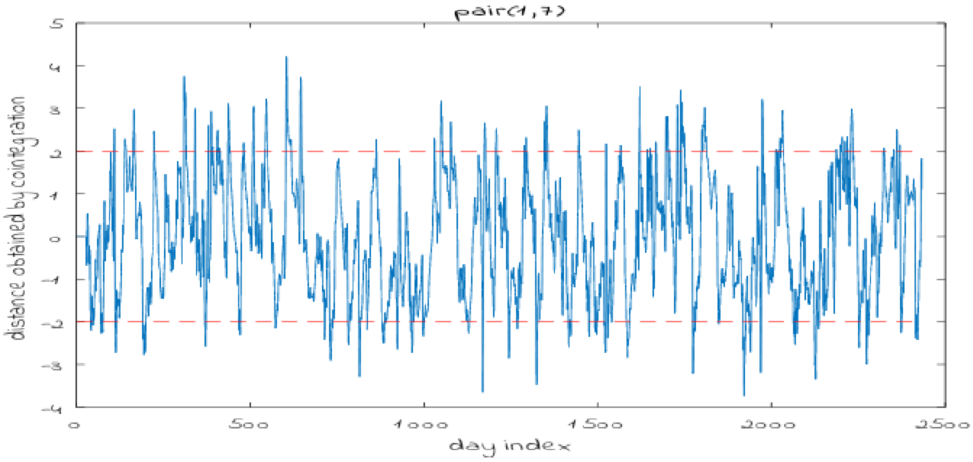

Fig. 1. The normalized cointegrating residuals for the pair(1,7)

In this study, the purchase thresholds were set at two numbers, namely, 2 and -2. Thus, when the normalized returns difference series reaches these two numbers, the asset is purchased at a lower price. In diagram (1), the buying and selling thresholds are marked in red. Sixty trading situations were calculated by implementing the cointegration-based pairs trading strategy based on the data. Each situation has a return, trading duration, and daily return equivalent (Table 2). Given the different durations of the extracted trading situations, the daily return equivalent assesses the returns obtained per one-day unit.

Table 2. Descriptive statistics of pairs trading situations with cointegration

Performance Statistical index	Situation return	Daily return equivalent	Duration(day)
Mean	0.032573	0.002158	24.46667
median	0.013175	0.000684	22.00000
maximum	0.479803	0.025342	72.00000
minimum	-0.227036	-0.006754	3.000000
Standard deviation	0.116348	0.005412	15.86219
skewness	1.244785	1.588981	1.086151
kurtosis	5.992224	7.253955	3.874528
Jarque-Bera Statistic	0.129508	37.87841	13.70923
p-Value	0.001728	0.000000	0.001055

Diagram (2) displays the daily returns for 60 situations.

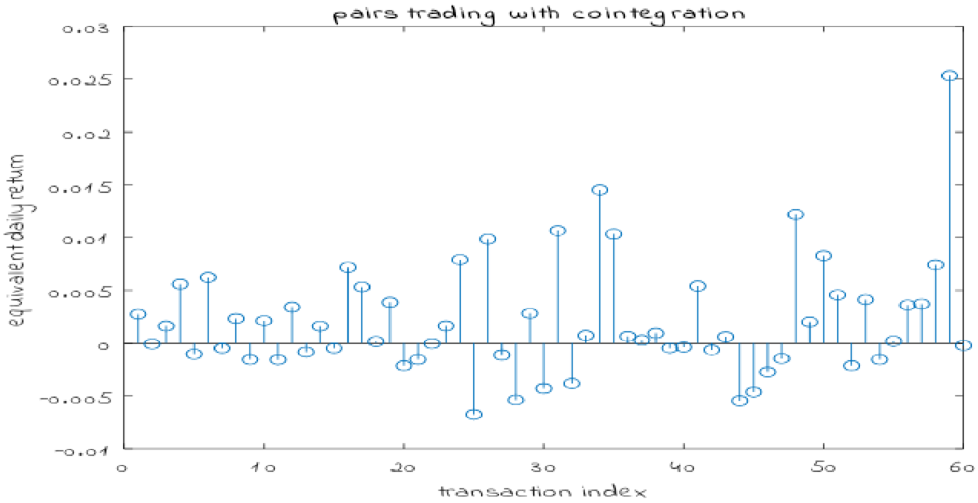

Fig. 2. daily returns equivalents for the pairs trading situations with cointegration.

Table (3) shows the profitability of the pairs trading system based on cointegration following the extracted situations.

Table 3. The profitability of pairs trading system based on cointegration

Performance criterion	value
Number of transactions	60
Mean daily returns	0.002158
Standard deviation	0.005412
Sharpe Ratio	0.398832
The mean duration of the transaction	24.46667
VaR (0.95)	0.004997

Table (3) shows that this trading system yields a mean daily return of 0.002158, which requires bearing a risk (standard deviation) of 0.0054. The Sharpe Ratio for the last trading system is equal to 0.398832, and the value at risk(VaR) shows that the amount of loss does not exceed 0.004997 at the confidence level of 0.95 (if the initial capital is equal to one unit).

After examining the pairs trading, which adopts a general cointegration approach, pairs trading based on empirical mode decomposition (EMD) is discussed. Empirical mode decomposition (EMD) decomposes a time series or signal into several time series called intrinsic mode functions or IMF, almost forming an orthogonal basis. EMD is implemented in MATLAB software using the EMD command. For example, Table (3) shows the output of this command for the second index based on 3 intrinsic mode functions or IMF.

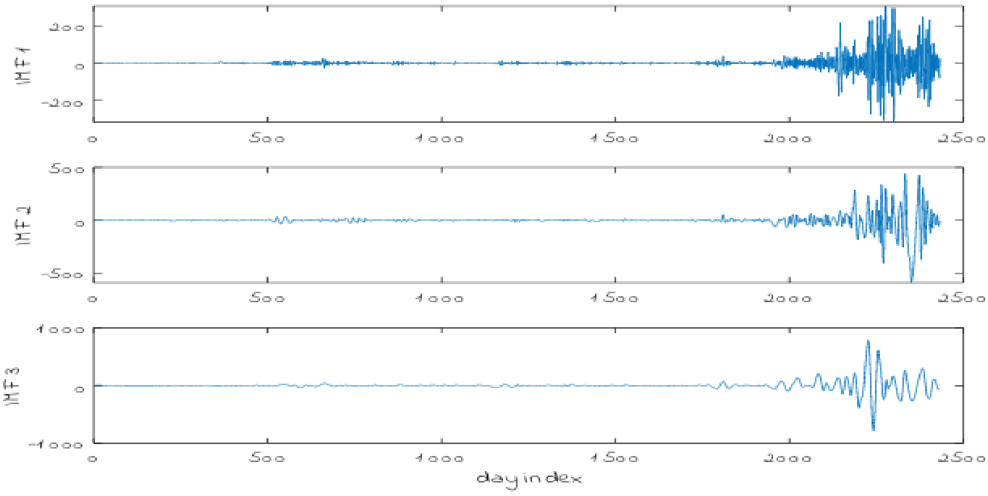

Fig. 2. EMD at three levels for the second index

The EMD was performed at three levels for all pairs, with the property of cointegration examined at the three corresponding levels in each pair. In the case of the confirmation of cointegration property, the resulting trading situations were calculated. Diagram (4) shows the daily returns for all obtained situations.

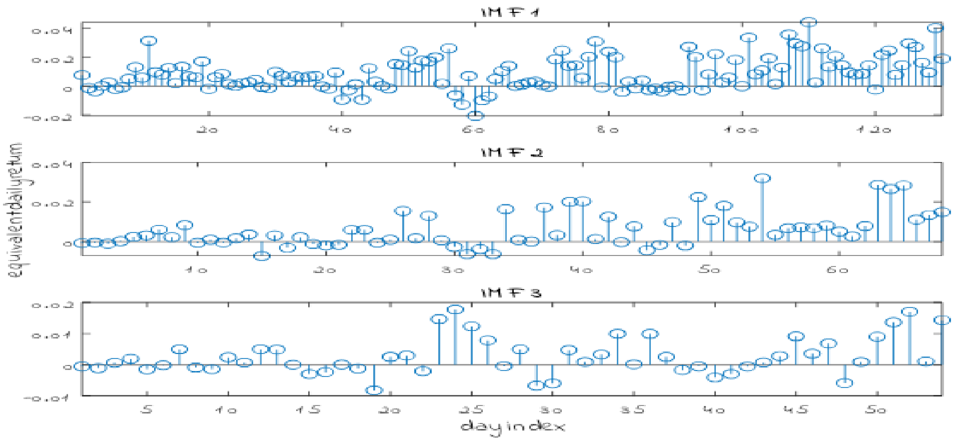

Fig. 4. The daily returns equivalents for the pairs trading situations based on EMD

Table (4) shows the descriptive statistics of the pairs trading system based on empirical mode decomposition in terms of the daily returns equivalents for the extracted situations.

Table 4. Statistical performance of pairs trading system based on EMD

Level of decomposition Statistical index	IMF1	IMF2	IMF3
mean	0.008824	0.006075	0.002660
median	0.006775	0.003250	0.000917
maximum	0.044390	0.031977	0.017735
minimum	-0.020753	-0.007212	-0.008167
SD	0.011467	0.008923	0.005938
skewness	0.597298	1.059608	0.784024
kurtosis	3.259992	3.595497	3.171180
Jarque-Bera Statistic	8.096054	13.72947	13.72947
p-Value	0.017457	0.001044	0.001044

Table 5 shows the profitability of the pairs trading system based on EMD regarding the daily returns equivalents for the extracted situations.

Table 5. Profitability of pairs trading system based on EMD

Decomposition Level Performance index	IMF1	IMF2	IMF3
Number of transactions	130	68	54
Mean daily return	0.008824	0.006075	0.00266
SD	0.011467	0.008923	0.005938
Sharpe Ratio	0.769458	0.680858	0.447927
Mean Duration of Transaction	3.484615	7.205882	20.59259
VaR (0.95)	0.006324	0.004567	0.005913

Given the results presented in Table (5) and Table (3), the pairs trading based on the first intrinsic mode function yield twice the number of trading situations compared to normal cointegration, and the average transaction duration decreases to one-sixth. The average daily returns and Sharpe ratio increased four times and two times, respectively. As a measure of adverse risk, the VaR also grew by 1.26 times. The pairs trading based on the second intrinsic mode function was found to have almost as many trading situations as the normal cointegration has. However, the average duration of a transaction decreased to about a third. The mean daily returns and Sharpe ratio increased by about three times and 1.7 times, respectively. There was no significant change in risk which is shown as VaR. Pairs trading based on the third intrinsic mode function yielded similar results as the normal cointegration in terms of all the mentioned criteria.

Conclusion

The main aim of the present study was to investigate the property of cointegration in an asset pair at different levels through decomposing the price signal using the empirical mode decomposition. Thus, the number of trading situations is expected to increase at higher frequencies, and the duration of each trading situation will decrease. The results of this study confirmed this hypothesis so that the first and second levels of decomposition yielded an increase in trading situations and a reduction in the time of each transaction. Often, there is a positive correlation between risk and return in financial markets, so an increase in risk leads to an increase in return. The return growth in the first and second decomposition levels increased by four and three times, respectively, which can be highly attractive for risk-seeking investors. For risk-averse investors, however, the calculated risk is of great importance to receive returns; thus, the measure of risk-adjusted return, or the Sharpe ratio, is of great importance.

The Sharpe ratio at the first and second levels grew by 2 and 1.7, respectively, which is remarkable, and it can encourage risk-averse investors to use this trading system. In the third level of decomposition, which corresponds to long-term behaviors, the performance criteria are on par with those related to normal cointegration. Comparing Table (5) and the descriptive statistics table also shows that the paired system of this study outperformed the buy and hold approach (selecting a random day and carrying out buying and selling operations) in terms of mean daily return and Sharpe ratio. Based on these findings, investors are advised to decompose the cointegration behavior of the set of asset pairs at different decomposition levels for profitability. Accordingly, given their risk-taking power, they can use the proper level of decomposition to implement a pairs trading system.

Research Limitations

The lack of short selling in the Tehran Stock Exchange is the main limitation of the research. Therefore, implementing the pairs trading system involves only buying the assets with a lower price when the trading signal is formed based on the pairs trading approach.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest concerning the research, authorship and, or publication of this article.

Funding

The authors received no financial support for the research, authorship and, or publication of this article.

References

Awajan, A., Ismail, M., & Alwadi, S. (2019). Stock market forecasting using empirical mode decomposition with Holt-winter. *AIP Conference Proceedings*. 218050006.10.1063/1.5136394.

Brahimipour, Mohammad Mehdi; Davodi, Seyyed Mohammadreza. (2021). Investigating the profitability of the pairs trading strategy based on the linear state-space system and the Kalman filter in the stock exchange. Investment Knowledge, 10(37), 57-75

Chacón, h., Kesici. E., & Najafirad.P. (2020). Improving Financial Time Series Prediction Accuracy Using Ensemble Empirical Mode Decomposition and Recurrent Neural Networks. *in IEEE Access*, 8, 117133-117145,

Clegg, M., & Krauss .C. (2018). Pairs trading with partial cointegration. *Quantitative Finance*, 18(1), 121-138.

da Silva, F. A. S., Ziegelmann, F. A., & Caldeira, J. F. (2023). A pairs trading strategy based on mixed copulas. The Quarterly Review of Economics and Finance, 87, 16-34.

Dastori, M., & Moradpour, S. (2021). Optimization of High-frequency Pair Trading Algorithm Using a Combination of Genetic Algorithm and Fuzzy Statistical Quality Control. *Journal of Investment Knowledge*, *10*(40), 471-484.

Davallou, M., & Yazdi, A. (2022). Pairs Trading; A Comparison between Student-t and Vine Copulas. *Financial Research Journal*, *24*(1), 104-143.

Fallahpour, Saeed; Hakimian, Hassan. (2016). Investigating the performance of pairs trading system in Tehran Stock Exchange: Cointegration approach and Sortino ratio investigation. Financial Engineering and Securities Management, 8(30), 1-17.

Fallahpour, Saeed; Hakimian, Hassan. (2019). Optimizing the pairs trading strategy using the reinforcement learning method and intraday data in the Tehran Stock Exchange. Financial Research, 21(1), 19-34.

Haddad, G. K., & Talebi, H. (2023). The profitability of pair trading strategy in stock markets: Evidence from Toronto stock exchange. *International Journal of Finance & Economics*, *28*(1), 193-207.

Han, C., He, Z., & Toh, A. J. W. (2021). Pairs Trading via Unsupervised Learning. *Available at SSRN 3835692*.

Kim, S. H., Park, D. Y., & Lee, K. H. (2022). Hybrid deep reinforcement learning for pairs trading. Applied Sciences, 12(3), 944.

Lin, B., & Tan, Z. (2023). Exploring arbitrage opportunities between China's carbon markets based on statistical arbitrage pairs trading strategy. *Environmental Impact Assessment Review*, *99*, 107041.

Lu, J. Y., Lai, H. C., Shih, W. Y., Chen, Y. F., Huang, S. H., Chang, H. H., ... & Dai, T. S. (2021). Structural break-aware pairs trading strategy using deep reinforcement learning. *The Journal of Supercomputing*, 1-40.

Moura. C., Pizzinga, Z., & Jorge Zubelli. (2016). A pairs trading strategy based on linear state space models and the Kalman filter. *Quantitative Finance*, 16(10), 1559-1573.

Noemi. N., Matteo, D., & Tomaso. A. (2018). Financial Time Series Forecasting Using Empirical Mode Decomposition and Support Vector Regression. *Risks*, 6(1),1-21.

-Ramos-Requena, J. P., López-García, M. N., Sánchez-Granero, M. A., & Trinidad-Segovia, J. E. (2021). A Cooperative Dynamic Approach to Pairs Trading. *Complexity*, *2021*.

Requena. R., Pedro. J., & Juan. T. (2020). Some Notes on the Formation of a Pair in Pairs Trading, *Mathematics*, 8 (3),1-17.

Robert J. E., & Bradrania. R. (2018). Estimating a regime-switching pairs trading model. *Quantitative Finance*, 18(5), 877-883.

Satoory, Mojtaba; Fallahpour, Saeed; Tehrani, Reza; Mehrgan, Mohammadreza. (2017). Algorithm of high-frequency pairs trading using fuzzy statistical quality control. Financial Engineering and Securities Management, 9(37), 23-41.

Stübinger. J., & Bredthauer. J. (2017). Statistical Arbitrage Pairs Trading with High-frequency Data. *International Journal of Economics and Financial*, 7(4), 650-662.

Tadi, Masoud; Abkar, Majid; Motaharinia, Vahid. (2017). Evaluation of pairs trading strategy using a distance approach in Tehran Stock Exchange. Investment Knowledge, 7(26), 99-112.

Yang, L., Zhao, L., & Wang, C. (2019). Portfolio optimization based on empirical mode decomposition. *Physica A: Statistical Mechanics and its Applications*, *531*, 121813.

Yang, J., Fu, Z., Zou, Y., He, X., Wei, X., & Wang, T. (2023). A response reconstruction method based on empirical mode decomposition and modal synthesis method. *Mechanical Systems and Signal Processing*, *184*, 109716.

Yujun, Y., Yimei, Y., & Jianhua, X. (2020). A hybrid prediction method for stock price using LSTM and ensemble EMD. *Complexity*, *2020*.

Zhifeng, D., & Huan Zhu, H. (2020). Forecasting stock market returns by combining sum-of-the-parts and ensemble empirical mode decomposition. *Applied Economics*, 52(21), 2309-2323.

Jin, Z., Jin, Y., & Chen, Z. (2022). Empirical mode decomposition using deep learning model for financial market forecasting. *PeerJ Computer Science*, *8*, e1076.

Bibliographic information of this paper for citing:

Zarintaj, Bahareh; Aghasi, Saeed & Baktash, Forozan (2023). Pairs Trading Based on Empirical Mode Decomposition (EMD). *Iranian Journal of Finance*, 7(3), 95-119.