+ MODEL
ARTICLE IN PRESS
+ MODEL
Available online at www.sciencedirect.com
Borsa İstanbul Review
Borsa İstanbul Review xxx (xxxx) xxx
http://www.elsevier.com/journals/borsa-istanbul-review/2214-8450

Full Length Article

Pairs trading: is it applicable to exchange-traded funds?

Ekin Tokat [a,*], Ahmet Cevdet Hayrullahoğlu [b]

[a] *TOBB University of Economics and Technology Department of Business Administration Söğütözü Cad, No: 43, Söğütözü, Ankara, 06560, Turkey*
[b] *Roketsan A.Ş., Elmadağ, Ankara, 06780, Turkey*

Received 21 February 2021; revised 16 August 2021; accepted 18 August 2021
Available online ■ ■ ■

Abstract

Among the various statistical trading strategies, pairs trading has been widely employed as a market neutral strategy owing to its simple approach and ease of application. In this context, we develop a cointegration-based pairs trading framework with a set of pre-conditions for pair eligibility and apply it to different asset classes. The performance analysis of a portfolio of 45 pairs is considered for the period of January 2007 to January 2021, which covers the period of a full market cycle of adjacent bull and bear periods; it is studied and benchmarked against the S&P500 index, which is considered as a proxy for the general market. We find an average annual return of 15% with an average Sharpe ratio of 1.43 after considering the transaction costs; we observe that this performance does not vary significantly with a change in the transaction cost levels and does not pass below the risk-free return levels with changing market conditions. Further, the strategy is observed to perform better during bear market conditions. Considering the highly liquid trading environment of the strategy, our findings raise a call for a discussion on the semi-strong form market efficiency.

Keywords: Pairs trading; Cointegration; Quantitative strategies; Exchange traded fund market

1. Introduction

Market practitioners have long been interested in quantitative trading models. Owing to the remarkable development in computer technology in the 1980s, the use of statistical arbitrage strategies has increased and become popular particularly among the hedge fund strategists and investment bank proprietary traders (Gatev et al., 2006). Specifically, pairs trading, a strategy developed by a group of mathematicians, physicists, and computer engineers, received particular attention in the early 1980s (Vidyamurthy, 2004). The idea behind the pairs trading strategy is to take advantage of market inefficiencies; its trading rule is quite straightforward: look for two securities whose prices have been moving together, watch the price spread widen, and then buy the security with a relatively lower price and sell the security with a relatively higher price. If the securities converge to their historical spread pattern, trading will result in profit.

Although hedge funds and investment banks have been extensively using this strategy since the early 1980s, it has been recently gaining increasing attention from academicians. Previous studies related to arbitrage have primarily examined risk-free arbitrage strategies for futures traded on various markets to test the market efficiencies (Fung et al., 2010; Dunis et al., 2010). Further, risk arbitrage has been relatively less discussed while transaction costs have been rarely considered in the existing literature (Chan, 2008, 2013). Analyses of risk-arbitrage, particularly pairs trading, was first introduced by Gatev et al. (1999), followed by many others (Vidyamurthy, 2004; Clegg & Krauss, 2018; Liew & Wu, 2013; Puspaningrum et al., 2009; Rad et al., 2016). However, the focus of prior studies has been mostly restricted to the stock market. Although recent studies have tested the strategy

* Corresponding author.
E-mail addresses: etokat@etu.edu.tr (E. Tokat), ahayrullahoglu@etu.edu.tr (A.C. Hayrullahoğlu).
Peer review under responsibility of Borsa İstanbul Anonim Şirketi.

https://doi.org/10.1016/j.bir.2021.08.001
2214-8450/

Please cite this article as: E. Tokat, A.C. Hayrullahoğlu, Pairs trading: is it applicable to exchange-traded funds?, Borsa İstanbul Review, https://doi.org/10.1016/j.bir.2021.08.001

on exchange traded funds (ETFs), prior studies initially worked with a very limited number of pairs, and later mostly focused on international country ETFs (see Clegg & Krauss, 2018; Schizas et al., 2011; Sipilä, 2013). Hence, the main objective of this study is to examine the profitability of the pairs trading strategy for a variety of asset classes and pair types, such as stock only pairs, ETF only pairs, and stock−ETF pairs. Exploring the differences among the performance outcomes of different asset types may provide a better understanding of market efficiency based on asset class or market type. Although market efficiency has long been studied in finance literature, statistical arbitrage models have been developed by traders to take advantage of the profit opportunities in the markets with different forms of efficiencies. Therefore, this study focuses on developing a pairs trading strategy to examine the performance of a portfolio of 46 paired securities from different asset classes, which are traded on the New York Stock Exchange (NYSE) and National Association of Securities Dealers Automated Quotation (NASDAQ). The developed strategy involves two steps: first, the cointegration method is applied to eligible pairs whose selection is based on a set of pre-conditions, and second, the trading rule for opening and closing positions is set. The pairs are chosen from the equities and ETFs representing different sectors, commodities, and countries. Finally, the strategy performance is evaluated and compared to a benchmark model, which is the buy and hold strategy in the general market, as per Standard and Poor's (S&P) 500 index.

The remainder of the paper is organized as follows. Section 2 discusses risk-free arbitrage, statistical arbitrage, and pair trading strategies based on the existing literature. Section 3 describes the data and strategy design. The results and performance are evaluated in Section 4. We conclude and discuss the critical paths and requirements for a successful pairs trading strategy in Section 5.

2. Pairs trading

Statistical arbitrage strategies in general involve the use of statistical models to analyze price patterns; therefore, variants of statistical arbitrage strategies can be grouped as mean-reverting, momentum, regime shifting, seasonal trending, and high-frequency trading, among others (Chan, 2008). Pairs trading within this statistical arbitrage framework is considered as a basis strategy and one of the simplest approaches. Prior studies have mostly focused on risk arbitrage opportunities in the context of commodity futures spread. Commodity and commodity product futures are found to be good trading pairs for testing market efficiencies as well as risk arbitrage opportunities (Girma & Paulson, 1999; Johnson et al., 1991). The pairs trading strategy was first introduced to the literature of financial econometrics by Gatev et al. (1999). The study shows that using a simple pairs trading rule, which is called the distance method, it is possible to generate profits over a long period of investment time. Considering transaction costs, Do and Faff (2012) show that the algorithm developed by Gatev et al. (1999) is largely unprofitable and therefore inapplicable after 2002. Another method that can be applied to a pairs trading framework is cointegration (Puspaningrum et al., 2009; Vidyamurthy, 2004). Vidyamurthy (2004) emphasizes on the fact that security pairing is a critical step to achieve significant trading performances. As a pairing method, it is suggested to define pair combinations based on statistical significance, that is, cointegration. The use of copulas in pairs trading is another sophisticated approach relative to the distance method. Liew and Wu (2013) propose an application of copulas to pairs trading, while Xie et al. (2016) evaluate the copula-based pairs strategy performance by using utility stocks from the US market. Further, Clegg and Krauss (2018) suggest the use of partial cointegration for pair formation and generating trading rules; their model is benchmarked against distance and cointegration-based trading models and performs well with at least 12% annual return after transaction costs.

Most of the previous studies have analyzed the implementation of the pairs trading strategy on stock pairs. In this study, we focus on comparing the performance of the pairs trading strategy for different asset types, including stocks and ETFs. In contrast to previous studies, which mostly focus on country ETFs, we use a wider range of ETF type in our portfolio and consider sector index, commodity and country ETFs. The algorithm is developed based on the cointegration of pairs, while transaction costs are included to acquire more realistic performance outcomes. Further, a period analysis is conducted for the period of January 1, 2007 to January 1, 2021. Conventionally, a sub-period analysis of the varying market conditions is conducted in studies related to pairs trading; however, this study focuses on a sequential trading schedule without any filtering and splitting of data. Our data covers a full market cycle for the US stock market[1] and provides an effective performance measurement of the strategy concerning all possible market conditions without controlling for market disturbances.

3. Model design

3.1. Cointegration

The idea of the pairs trading strategy comes from the identification of stationary price series. Stock price series are found to be non-stationary due to their stochastic behavior. Following Engle and Granger (1987), if two non-stationary price series are integrated of order one, that is I(1), considering that the first difference of the price series is stationary, that is I(0), then there exists a linear combination of the price series, which forms a stationary process, such that the price series are said to be cointegrated of order one:

[1] The recent full market cycle is usually considered to have occurred between 2007 and 2013 for the US equity market; this is defined as bull, bear, and bull periods, which are adjacent to each other and generally last from as short as 4−5 years to as long as 20 years (see Manning & Napier, 2014; Asymmetry Observations, n.d.).

$$y_t - \beta x_t = e_t,$$

where y_t and x_t are the cointegrated price series, β is the cointegration coefficient, and e_t is the stationary cointegration error. In this framework, the cointegrated price series will show a long-run equilibrium relationship; any deviation from this equilibrium will be corrected in the short-run, which can be shown through an error correction model (ECM) of the following form:

$$y_t - y_{t-1} = \alpha_y(y_{t-1} - \beta x_{t-1}) + \varepsilon_{y_t}$$

$$x_t - x_{t-1} = \alpha_x(y_{t-1} - \beta x_{t-1}) + \varepsilon_{x_t},$$

where $\varepsilon_{y_t}$is the stationary disturbance term, and $\alpha_y(y_{t-1} - \beta x_{t-1})$ is the error correction term for y_t In this framework, the estimated coefficients of the long-run error terms, that is α_y and α_x, reflect the process by which both the price series adjust in the short-run according to their long-run equilibrium paths or as per their speed of adjustment. This mean-reverting property of cointegration is in line with the idea of pairs trading, which was incorporated into the pairs trading strategy by Vidyamurthy (2004). In this study, we employ the cointegration framework formulated in Vidyamurthy (2004) and use ECM to estimate the long-run equilibrium relationship among the paired securities; further, we detect deviations in their long-term relationship, which is used as a sign for taking either a long or short position in paired securities. The estimated eigenvectors from the Johansen cointegration test (Johansen, 1995) are used as hedge ratios to determine the portfolio weights:

$$eigenvalue = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$eigen\text{vector} = \begin{bmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{bmatrix}$$

$$Normalized\ Hedge\ Ratio = h = h_{12}/h_{11}.$$

Spread, which is expected to be stationary, is obtained as follows:

$$Spread = y - hx.$$

To estimate the deviations from the spread, the Z-score, which has a standard normal distribution and helps determine normalized deviations from the long-run relation, is employed as follows:

$$\text{Z} - \text{score} = (Spread - ma_w(Spread))/std_w(Spread).$$

The pairs trading strategy relies on a mean-reverting portfolio, and it involves taking long or short positions with respect to Bollinger Bands. Deviation from pre-defined threshold levels around the moving average forms the baseline for trading decisions. The Bollinger Band approach requires the optimization of threshold Z-score levels and the window size for calculating the moving average. For the window size, weekly (five days), monthly (20 days), semi-annual (120 days), and annual (250 days) values are used and tested. For the threshold levels, the tested standard deviations ranged between 0.1 and 3.1 with 0.2 increments, amounting to a total of 16 threshold levels. The criterion for closing the position requires zero standard deviation of spread from the moving average.

3.2. Backtesting

Backtesting is a crucial component in the development of a trading strategy; it provides the input for optimization and performance enrichment and involves testing the algorithm via historical data. In backtesting the performance of a trading strategy, it is vital to use only the data that would have been available at the time of trade. Otherwise, it is more likely to introduce a look-ahead bias into the system. For example, if a trade position is simulated based on the minimum and maximum daily prices observed on the same trading day, it will diminish the accuracy of the trade's true performance as it is impossible to observe the minimum and maximum daily prices before the trading day ends. A look-ahead bias can be avoided by dividing the data frame into two sub-sets: in-sample and out-of-sample data sets. Accordingly, the coding for hedge ratio, spread, and other parameter calculations will be based on different time periods. In a typical backtesting framework, one year of formation period, during which the hedge ratio is calculated, is followed by one year of trading period (Figure S1, available online).[2]

The use of software technology in finance, particularly for strategy development, has created tremendous trading opportunities. Once the rules of the developed trading system are coded, it is very likely to backtest several trading options and analyze all the combinations of the pre-set parameters to find the best performing rules. With an adequate number of combinations, several rules can be formulated to ensure a good performance. However, an extensive search for variable combinations with different parameters is likely to result in data snooping bias, which is another backtesting bias. The likelihood that a performance result obtained from pure luck will increase with the number of combinations tested. Any trading rule that perfectly fits to its sample data through backtesting may not generate the same performance when it is run against another data set, which will result in the loss of performance persistence. To minimize the probability of snooping bias, in-sample and out-of-sample data sets are used for the estimation of two parameters, which are the entry level of Z-score and window size of the moving average of spread. In order to minimize the probability of snooping bias, in-sample and out-of-sample data sets are used for estimation of two parameters, which are entry level of Z-score and window size of moving average of spread. Sensitivity tests for

[2] We tested the backtesting frameworks with formation and trading periods of longer than one year, however, it reduced the strategy performance (see Table 3).

such parameters are conducted to observe the key component of the developed strategy.

3.3. Data

The criteria for determining the assets that are potentially suitable for pairs trading are defined. Rather than setting various unit root tests as the first condition for an eligible pair (Huck, 2015; Krauss & Herrmann, 2017, Clegg & Krauss, 2018), a set of pre-conditions are used to study the potential pairs: same sector, sub-sector for stocks, weight of stock in holdings of select sector ETF, similar investment grade for country ETFs, geographical proximity for country ETFs, and same commodity class such as precious metals or energy for commodity ETFs. Previous studies (Gatev et al., 2006; Do & Faff, 2012; Clegg & Krauss, 2018) have performed pairs trading on a collection of stock pairs with the same sector restriction, which is the first eligibility criterion of unit root tests. We use the same sector restriction along with other qualitative pre-conditions as the first criterion for pair eligibility and examine stocks and ETFs. The ETFs on selected sectors, commodities, and country indices traded in NYSE and NASDAQ are examined in this framework. An algorithm developed for the analysis extracted 13 years of daily price data from the Thomson Reuters Datastream from January 2007 to January 2021.

An issue in the development of a trading strategy lies in the short-sale constraint. We observe that it could be difficult or impossible to sell the stock because of a small trading volume or the restrictions imposed by the market regulator. To replicate a practical trading environment, stocks with high market caps and high trading volumes are selected. It is most likely for market turmoil and crisis periods to cause structural breaks in price series. Such structural breaks can generate jumps on price series because of a high volatility (Fung et al., 2010). However, the data are not filtered to avoid complications and price jumps such as those observed repeatedly during and in aftermath of the 2007 to 2009 global financial crisis with the aim of imitating a realistic trading environment under uncertainty. Therefore, this study is conducted on overlapping backtesting frameworks, resulting in a sequential trading period analysis with a view of the strategy performance through a full market cycle. Accordingly, we consider the first formation period to start from January 2007, such that the following formation periods overlap with the trading periods, which commence from January 2008. Prior studies on trading show that performance varies over time (Gatev et al., 2006; Do & Faff, 2010; Clegg & Krauss, 2018), although the studies are usually conducted on a basis of sub-period analysis; however, our motive in this study is to test the performance of the pairs trading strategy within a fluctuating market atmosphere for consecutive years. Table 1 provides a summary of the data set used.

3.4. Performance evaluation

The pairs trading strategy performance is measured by cumulative compound returns and Sharpe ratios:

Table 1
Summary of the data set.

Pairs group	Number of stocks	Number of ETFs	Number of STCK−STCK pairs	Number of ETF−STCK pairs	Number of ETF−ETF pairs
Financials	4	1	3	4	–
Technology	5	1	3	3	–
Healthcare	3	1	2	3	–
Consumer Goods	7	2	3	6	–
Energy	3	1	2	2	–
Industrial	3	1	1	3	–
Utilities	2	1	1	2	–
Commodities	–	5	–	–	3
Regional	–	7	–	–	4
All	27	20	15	23	7

Note: STCK: equity stock, ETF: exchange traded fund.

$$Return_t = \frac{(-Nhx)_{t-1}R_{tx} + (Ny)_{t-1}R_{ty}}{|(-Nhx)_{t-1}| + |(Ny)_{t-1}|}$$

$$Nreturn = Return - Transacion\ Cost$$

$$Cumulative\ Return = \prod_{t=1}^{T}(1 + Nreturn_t) - 1$$

$$Sharpe\ Ratio = \sqrt{252} * mean(Nreturn) \Big/ std(Nreturn),$$

where N represents the decision criterion, which is -1 for short and 1 for the long position; h is the normalized hedge ratio; and R is the asset return. Given that the strategy is implemented on highly liquid trading venues (NYSE and NASDAQ) with particularly high liquid stocks and ETFs, we follow Clegg and Krauss (2018) and adjust the total return of a closed position with 10 bps as a round-trip transaction cost. This level is in line with other existing studies on pairs trading. For example, Do and Faff (2012) consider institutional commissions of 0.1% or less between 1997 and 2009 by referring to Jones (2002) who provided an annual time series of estimated trading costs (bid-ask spreads and commission costs) for the stocks in the Dow Jones Index to find that one-way trading costs have consistently declined over the years and reached approximately 0.2% in 2000. However, Prager et al. (2012), report that the bid-ask spread declined to approximately one cent for the S&P 500 constituents. In terms of commission rates, the trend has also been declining, such that it is now charged as zero for the online trading of stocks and ETFs by many trading platforms (Fidelity, n.d.; TD Ameritrade, n.d.).[3] Although the assumption of 10 bps transaction cost seems reasonable, the strategy is also tested for higher transaction cost levels.

[3] Markets in Financial Instruments (MiFID II) Directive introduced by the European Union aiming to enhance investor protection is seen as another source of impact on significant declines in commission rates (Reuters, 2019).

4. Results

The pairs trading strategy is tested on a wide range of securities from different sectors, commodity markets, and country indices (see Table S1, available online); 20 ETFs and 27 stocks are used to test the performance of 45 pairs. The portfolio is composed of 15 stock only pairs, 7 ETF only pairs, and 23 stock−ETF pairs.

We aim to explore the effectiveness of cointegration-based pairs trading together with our set of pre-conditions, which are used for selecting the potential pairs. As the cointegration relation and corresponding hedge ratios are calculated for a period of 13 years for each pair, only the statistics of the cointegrated series are reported in Table 2[4]. The results support the rolling cointegration relationship among the pairs (Kutan & Zhou, 2003), that is, the pairs show a changing comovement behavior with time. Out of the 45 pairs studied, we observe 15 and 28 cointegrated pairs in the first and second year of the hedge ratio formation, respectively, and the number changes over time. Following Clegg and Krauss (2018), it is possible to observe profitable results for partially cointegrated data using a cointegrating pairs trading strategy. For trading applications, we not only use cointegration-based pairs, but also consider the pairs without a significant cointegration relation and use their corresponding hedge ratios. Table 2 shows the percentage of cointegrated pairs in the portfolio together with the corresponding annualized mean returns. The results do not provide evidence of a correlation between the density of cointegrated pairs and trading performance.

The results are based on an optimum parameter, which provides the highest cumulative compound return. One of the parameters is the Z-score, which is tested for the range between 0.1 and 3.1; window size is another parameter that is tested based on its weekly (five days), monthly (20 days), semi-annual (120 days), and annual (250 days) values. Therefore, a total of 64 different parameter combinations are tested to evaluate the best performance, which is the highest cumulative compound return. A sensitivity test is further conducted to detect the parameter with the highest impact on our strategy. The mean of the standard deviations are used to conduct a comparison. Table 3 summarizes the trading statistics for each trading period. In all the periods, majority of the pairs are traded; for example, there are only two out of 45 pairs without any open trading positions during the first year of trading. Although it is observed that the average number of positions is relatively higher for the global financial crisis period and the preceding year (trading period of 2008−2009 and 2009−2010), respectively, it does not lead to a poorer cumulative return performance due to the accumulation of transaction costs; further, over these years, the annual mean returns have attained their highest levels. Maximum drawdown is another trade statistic, which observes the highest loss incurred among the open positions for any pair. The average maximum drawdown is the highest but limited to 11% in the trading year of the global financial crisis. Parameter sensitivity tests consider window size as the critical parameter. Across all sub-periods, the minimum proportion of pairs with window size sensitivity is 70%, suggesting its significance in strategy irrespective of the market condition.

Table 4 summarizes annualized risk and return characteristics of the strategy for the observed data from January 2007 to January 2021. Here, S&P 500 index, with the ticker symbol ĜSPC, is considered as the proxy for the general market, and its annualized mean return and Sharpe ratio are taken as the benchmark criteria for performance evaluation. The annualized mean return for the pairs portfolio is 15% with a Sharpe

Table 2
Cointegration test statistics.

Formation period (FP)	Trading period (TP)	Percentage of cointegrated pairs	Annualized mean return of pairs trading strategy
2007−2008	2008−2009	35%	41%
2008−2009	2009−2010	62%	28%
2009−2010	2010−2011	96%	8%
2010−2011	2011−2012	20%	13%
2011−2012	2012−2013	47%	9%
2012−2013	2013−2014	44%	8%
2013−2014	2014−2015	47%	8%
2014−2015	2015−2016	47%	13%
2015−2016	2016−2017	71%	11%
2016−2017	2017−2018	51%	8%
2017−2018	2018−2019	49%	10%
2018−2019	2019−2020	27%	9%
2019−2020	2020−2021	64%	25%

Note: Formation period considers the hedge ratio calculation. Trading period considers running the pairs trading strategy based on the hedge ratio calculated in the formation period.

Table 3
Trading statistics.

Trading period	Proportion of pairs traded	Average number of positions per pair	Average maximum drawdown	Proportion of pairs with Z-score sensitivity	Proportion of pairs with window size sensitivity
2008−2009	1.00	41	−0.113	0.3	0.7
2009−2010	1.00	30	−0.095	0.2	0.8
2010−2011	0.91	16	−0.042	0.13	0.87
2011−2012	1.00	20	−0.053	0.23	0.77
2012−2013	0.98	12	−0.048	0.11	0.89
2013−2014	0.98	12	−0.044	0.11	0.89
2014−2015	1.00	12	−0.036	0.07	0.93
2015−2016	1.00	13	−0.051	0.11	0.89
2016−2017	1.00	16	−0.054	0.2	0.8
2017−2018	0.93	16	−0.035	0.11	0.89
2018−2019	1.00	18	−0.051	0.18	0.82
2019−2020	1.00	12	−0.049	0.25	0.75
2020−2021	1.00	21	−0.087	0.09	0.91

Note: Formation period considers the hedge ratio calculation. Trading period considers running the pairs trading strategy based on the hedge ratio calculated in the formation period.

[4] ADF stationarity test statistics, trace statistics of Johansen cointegration test, and the corresponding hedge ratios calculated for each pair for each period can be provided upon request.

Table 4
Performance evaluation of the pairs trading strategy.

	FP	TP	Annual average return of PT	Annual return of S&P 500	Average Sharpe ratio of PT	Sharpe Ratio of S&P 500
FP: 1 Year	2007–2008	2008–2009	41%	−38%	1.85	−0.94
TP: 1 Year	2009–2009	2009–2010	28%	20%	1.64	0.80
	2009–2010	2010–2011	8%	11%	1.22	0.67
	2010–2011	2011–2012	13%	−1%	1.39	0.07
	2011–2012	2012–2013	9%	12%	1.43	0.94
	2012–2013	2013–2014	8%	26%	1.30	2.23
	2013–2014	2014–2015	8%	12%	1.53	1.09
	2014–2015	2015–2016	13%	−1%	1.47	0.03
	2015–2016	2016–2017	11%	11%	1.34	0.89
	2016–2017	2017–2018	8%	18%	1.33	2.59
	2017–2018	2018–2019	10%	−7%	1.31	−0.34
	2018–2019	2019–2020	9%	29%	1.28	2.09
	2019–2020	2020–2021	25%	15%	1.47	0.59
Average of annual returns and Sharpe ratios from 2007–2021			15%	8%	1.43	0.82
Alternative backtesting framework:						
FP: 1 Year TP: 4 Year	2010–2011	2011–2015	2%	15%	1.56	0.86
FP: 4 Year TP: 4 Year	2010–2014	2014–2018	2%	11%	1.21	0.84
Alternative transaction costs:						
FP: 1 Year	2007–2008	2008–2009	39%	−38%	1.60	−0.94
TP: 1 Year Transaction Cost: 20 bps	2016–2017	2017–2018	7%	18%	1.19	2.59
FP: 1 Year	2007–2008	2008–2009	33%	−38%	1.45	−0.94
TP: 1 Year Transaction Cost: 30 bps	2016–2017	2017–2018	6%	18%	1.08	2.59
Alternative sequence of formation and trading periods:						
FP: 1 Year	01.2008–01.2009	01.2009–01.2010	38%	20%	1.61	0.80
TP: 1 Year	03.2008–03.2009	03.2009–03.2010	29%	59%	1.73	2.12
	06.2008–06.2009	06.2009–06.2010	14%	14%	1.45	0.79
	09.2008–09.2009	09.2009–09.2010	13%	8%	1.50	0.51

Note: Formation period (FP) considers the hedge ratio calculation. Trading period (TP) considers running the pairs trading (PT) strategy based on the hedge ratio calculated in the formation period.

ratio of 1.41. Considering the total amount of transaction costs for a portfolio of 45 pairs as compared to the buy and hold strategy for the benchmark, this result is not disappointing. For the same period, $\hat{G}$SPC provides an 8% mean return and 0.82 Sharpe ratio on an annualized basis. One interesting observation is the relatively stable performance of the pairs portfolio among all the years studied except during the 2007 to 2009 global financial crisis period. In contrast to the highly volatile behavior of the general market, which fluctuates from 38% drawdown to 29% annual peak through the years, pairs trading produces a minimum of 8% to a maximum of 25% annualized returns, except during the 2007 to 2009 financial crisis. Such a relatively steady performance supplies a close analogy to the portfolio hedging with index futures, such that the return is higher than the risk-free rate (Hull, 2009). The results also confirm the findings of previous studies (Clegg & Krauss, 2018; Do & Faff, 2010; Jacobs & Weber, 2015) on pairs trading at times of financial distress; the strategy exhibits a strong performance in the times of market turmoil; it shows its peak during the global financial crisis with an annualized return of 41% and Sharpe ratio of 1.85. Although the strategy mostly underperforms in the bull markets, it still produces a minimum of 8% annualized returns. This outcome with better Sharpe ratios than that of a benchmark is remarkably satisfying, considering the poor performance of many trading strategies that are used in the bull markets (Clegg & Krauss, 2018; Green et al., 2017).

Table 5 presents results of the sub-group performance analysis. All the group settings result in a better performance than the benchmark in terms of annual mean return and Sharpe ratio. Among the selected ones, the pairs from the regional country ETFs seem to perform the best in terms of annualized risk and return figures, which are 1.73 and 20%, respectively. Another interesting result is the better performance of emerging country pairs in contrast to the developed country ETF pairs. This might be attributable to a relatively high volatility and less newswire or

Table 5
Performance evaluation based on group classifications.

Trading Years	2008−2009	2009−2010	2010−2011	2011−2012	2012−2013	2013−2014	2014−2015	2015−2016	2016−2017	2017−2018	2018−2019	2019−2020	2020−2021	2007−2021 Average
Average annualized return														
S&P500	−38%	20%	11%	−1%	12%	26%	12%	−1%	11%	18%	−7%	29%	15%	8%
Financials	73%	88%	5%	20%	14%	8%	8%	14%	4%	6%	7%	5%	10%	20%
Technology	49%	28%	3%	24%	11%	7%	13%	9%	16%	7%	9%	14%	16%	16%
Healthcare	39%	33%	9%	5%	6%	5%	7%	14%	8%	5%	8%	13%	34%	14%
Consumer Goods	51%	14%	10%	10%	7%	6%	7%	7%	10%	8%	8%	8%	16%	12%
Energy	12%	17%	7%	9%	10%	5%	13%	11%	10%	4%	13%	9%	32%	12%
Industrial	15%	22%	3%	18%	10%	7%	5%	17%	12%	5%	4%	6%	22%	11%
Utilities	65%	8%	8%	5%	3%	5%	3%	5%	5%	3%	3%	0%	21%	10%
Commodities	13%	24%	19%	12%	17%	21%	8%	17%	20%	15%	9%	13%	45%	18%
Regional	53%	17%	12%	14%	8%	7%	13%	19%	17%	20%	31%	16%	27%	20%
Portfolio Average	41%	28%	8%	13%	9%	8%	8%	13%	11%	8%	10%	9%	25%	15%
Average Sharpe ratio														
S&P500	−0.94	0.80	0.67	0.07	0.94	2.23	1.09	0.03	0.89	2.59	−0.34	2.09	0.59	0.82
Financials	1.73	1.67	0.96	1.48	1.68	1.27	1.75	1.66	0.98	1.49	1.27	1.04	1.35	1.41
Technology	2.14	1.90	0.54	1.68	1.20	0.58	1.93	0.88	1.37	1.09	1.30	1.53	1.31	1.34
Healthcare	1.52	1.82	1.55	0.85	1.44	0.91	1.23	1.49	0.71	1.09	1.11	1.71	2.35	1.37
Consumer Goods	2.15	1.30	1.48	1.52	1.19	1.42	1.41	0.98	1.32	1.42	1.34	1.51	1.38	1.42
Energy	1.02	1.57	1.04	1.21	1.84	1.62	2.23	1.29	1.31	0.78	1.64	1.49	1.55	1.43
Industrial	1.10	2.19	0.62	2.06	1.86	1.29	1.60	2.02	1.82	1.13	0.86	0.66	1.09	1.41
Utilities	3.18	1.22	1.51	1.29	1.14	1.33	0.85	1.62	1.38	1.22	1.25	0.28	1.47	1.36
Commodities	1.01	1.58	1.32	1.15	1.28	2.01	1.42	1.44	1.32	2.27	0.80	1.39	1.00	1.39
Regional	2.78	1.55	1.97	1.28	1.26	1.30	1.38	1.80	1.88	1.45	2.20	1.90	1.71	1.73
Portfolio Average	1.85	1.64	1.22	1.39	1.43	1.30	1.53	1.47	1.34	1.33	1.31	1.28	1.47	1.43

Note: The last column shows the 13 year average of each related row.

+ MODEL

analysts' coverage on the ETFs of an emerging country. Jacobs and Weber (2015) observe low pair visibility (less newswire or analysts' coverage) and limits to arbitrage (volatility) as the two parameters that make pairs trading more profitable. Their finding also explains the superior performance of the regional ETF pairs among the selected groups. Among the various industry sets, the financial sector together with the healthcare and technology sectors perform well. Financial pairs work well during the global crisis period as majority of the financial stocks collapse together owing to a significant amount of systematic risk; additionally, the pairs from the healthcare sector perform well during the COVID-19 pandemic for similar reasons. Moreover, pairs trading performance analysis suggests that performance does not vary significantly according to the asset type. Finally, our findings provide a ground to question the semi-strong form of market efficiency, considering the level of returns pairs trading produced from 2007 to 2021 in a highly liquid stock and ETF environment.

We re-examine certain parameters of the model as they had consistently achieved higher than risk-free return levels. The strategy is first tested with higher transaction cost levels. We assume a round-trip transaction cost per trade of 20 and 30 bps. The strategy appears to be robust to transaction cost assumptions, that is, it is common for statistical pairs trading strategies as the performance does not change dramatically for higher transaction cost levels. Among the outliers, during the trading year of 2017−2018 in which the trading strategy exhibits a poor performance, the annualized mean return decreases by 1% with a 10 bps increase in transaction costs and Sharpe ratio changes from 1.11 to 1.08 with the transaction cost level at 30 bps (see Table 4). Another outlier is the trading year with the highest average number of positions per pair where we expect to see a diminishing impact of higher transaction costs on strategy performance. In this particular period, which coincides with the 2007 to 2008 global financial crisis considered as the best performing trading year, return performance decreases by 2% with a 10 bps increase in the transaction costs. An addition of 10 bps decreases the annual mean return to 33%; however, the outcome is still fairly satisfying. Additionally, the strategy is tested with a disrupted sequence of formation and trading periods. We replace the first formation period of January 2007 for hedge ratio calculation with different months of the year and employ the backtesting framework accordingly. A sample of the formation and trading periods with the strategy performance results is shown in Table 4. Although the annualized mean return and Sharpe ratios vary according to different period settings, the strategy still performs better than the risk-free market return with satisfying Sharpe ratio levels. However, the variation in the performance bears the potential to construct a dynamically adaptive pairs trading strategy for future studies.

5. Conclusion

Pairs trading has long been one of the most popular hedge fund strategies. This study focuses on developing the strategy with a cointegration approach and a set of pre-conditions for pair eligibility. The portfolio of 45 pairs achieves a 15% annual mean return after transaction costs; its performance does not change much in a negative direction with fluctuating markets, suggesting a casual hedging strategy for market participants. Performance is particularly strong during times of market turmoil. This finding together with a moderate level of return and risk figures for the bull market conditions can help us to draw a conclusion on potential arbitrage opportunities and question the semi-strong form of market efficiency. Another outcome of the study is the high dependency of strategy performance on the model parameters. This result leads us to study parameter optimization, which has the potential to improve the strategy performance for the next trading period. Two parameters that are found to be critical for strategy performance and need to be optimized are the Z-score value and window size of moving average; window size proves to be the most critical parameter based on sensitivity tests. The finding on the lack of a strong relationship between the cumulative returns and degree of cointegration implies that there exists a possibility to gain profits in the absence of a strong cointegration property and a possible change in the cointegration relation among the pairs. This may be attributed to the change in the dynamics of the environment that provides the strength of cointegration between the two securities, such as a change in marketing targets, one of the company's management, financial market structure, or the country's economic prospect. Further, a steady return performance among fluctuating market conditions draws our attention to the set of qualitative preconditions used for pair eligibility. Therefore, fine-tuning the set of pre-conditions and the further development of a dynamically adaptive algorithm to optimize the hedge ratio should be the focus of future studies.

Funding

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK).

Declaration of competing interest

There is no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bir.2021.08.001.

References

Asymmetry Observations. Full market cycle. n.d. https://asymmetryobservations.com/definitions/full-market-cycle/

Chan, E. P. (2008). *Quantitative trading: How to build your own algorithmic trading business*. New Jersey: John Wiley & Sons.

Chan, E. (2013). *Algorithmic trading: Winning strategies and their rationale*. New Jersey: John Wiley & Sons.

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

Do, B., & Faff, R. (2010). Does simple pairs trading still work? *Financial Analysts Journal, 66*, 83–95.

Do, B., & Faff, R. (2012). Are pairs trading profits robust to trading costs? *Journal of Financial Research, 35*, 261–287.

Dunis, C., Laws, J., & Evans, B. (2010). Trading and filtering futures spread portfolios: Further applications of threshold and correlation filters. *Journal of Derivatives and Hedge Funds, 15*(4), 274–287.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica, 55*(2), 251–276.

Fidelity. Brokerage commission and fee schedule. n.d. https://www.fidelity.com/bin-public/060_www_fidelity_com/documents/Brokerage_Commissions_Fee_Schedule.pdf

Fung, H. G., Liu, Q., & Tse, Y. (2010). The information flow and market efficiency between the US and Chinese aluminum and copper futures markets. *Journal of Futures Markets, 30*(12), 1192–1209.

Gatev, E. G., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies, 19*(3), 797–827.

Gatev, E. G., William, N., Goetzmann, K., & Rouwenhorst, G. (1999). Pairs trading: Performance of a relative value arbitrage rule. *NBER Working Paper*. No. 7032.

Girma, P., & Paulson, A. (1999). Risk arbitrage opportunities in petroleum futures spreads. *Journal of Futures Markets, 19*(8), 931–955.

Green, J., Hand, J. R., & Zhang, F. (2017). The characteristics that provide independent information about average US monthly stock returns. *Review of Financial Studies, 30*(12), 4389–4436.

Huck, N. (2015). Pairs trading: Does volatility timing matter? *Applied Economics, 47*, 6239–6256.

Hull, J. C. (2009). *Options, futures, and other derivatives* (9^{th} ed.). Pearson Education India.

Jacobs, H., & Weber, M. (2015). On the determinants of pairs trading profitability. *Journal of Financial Markets, 23*, 75–97.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. New York: Oxford University Press.

Johnson, R. L., Zulauf, C. R., Irwin, S. H., & Gerlow, M. E. (1991). The soybean complex spread: An examination of market efficiency from the viewpoint of a production process. *Journal of Futures Markets, 11*(1), 25–37.

Jones, C. M. (2002). *A century of stock market liquidity and trading costs*. Working Paper. Columbia University. https://doi.org/10.2139/ssrn.313681.

Krauss, C., & Herrmann, K. (2017). On the power and size properties of cointegration tests in the light of high-frequency stylized facts. *Journal of Risk and Financial Management, 10*(1), 1–24.

Kutan, A. M., & Zhou, S. (2003). Has the link between the spot and forward exchange rates broken down? Evidence from rolling cointegration tests. *Open Economies Review, 14*(4), 369–379.

Liew, R. Q., & Wu, Y. (2013). Pairs trading: A copula approach. *Journal of Derivatives and Hedge Funds, 19*(1), 12–30.

Manning, & Napier. (December 5, 2014). *Defining a market cycle*. https://www.manning-napier.com/insights/blogs/research-library/defining-a-market-cycle.

Prager, R., Vedbrat, S., Vogel, C., & Watt, E. C. (2012). *Got liquidity?* BlackRock Investment Institute. https://www.blackrock.com/us/individual/literature/whitepaper/got-liquidity-us-version.pdf.

Puspaningrum, H., Lin, Y. X., & Gulati, C. (2009). Finding the optimal pre-set boundaries for pairs trading strategy based on cointegration technique. *Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper, 21–09*, 25.

Rad, H., Kwong, R., Low, Y., & Faff, R. (2016). The profitability of pairs trading strategies: Distance, cointegration and copula methods. *Quantitative Finance, 16*(10), 1541–1558.

Reuters. (January 29, 2019). *Stock trading costs lower since MiFID II-study*. https://www.reuters.com/article/eu-markets-mifid-idINL5N1ZT56D.

Schizas, P., Thomakos, D., & Wang, T. (2011). Pairs trading on international ETFs. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1958546

Sipilä, M. (2013). *Algorithmic pairs trading: Empirical investigation of exchange traded funds* [Master's thesis, Aalto University]. Semantic Scholar http://epub.lib.aalto.fi/en/ethesis/pdf/13232/hse_ethesis_13232.pdf.

TD Ameritrade. Standard retail pricing: Commissions and fees. n.d. https://www.tdameritrade.com/retail-en_us/resources/pdf/TDA4075.pdf

Vidyamurthy, G. (2004). *Pairs trading: Quantitative methods and analysis* (Vol. 217). New Jersey: John Wiley & Sons.

Xie, W., Liew, R. Q., Wu, Y., & Zou, X. (2016). Pairs trading with copulas. *The Journal of Trading, 11*(3), 41–52.