Pairs Trading in Crypto Currencies - A cointegration based application

Sebastian Setzer

Dissertation written under the supervision of Professor Eva Schliephake

Dissertation submitted in partial fulfilment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, 2022-06-01.

Abstract

This thesis shows the possibilities of statistical arbitrage in the crypto currency market. Within the field of statistical arbitrage a pairs trading approach is chosen and an overview of different pairs trading strategies is given. Finally the cointegration approach in the Engle-Granger two step framework is applied. The approach is clustered into a formation period of 210 days and a trading period of 75 days. Although, it can be shown that the pairs trading outperforms a simple Buy-and-Hold strategy for three independently chosen periods, the returns can not be considered as market neutral in most cases. Furthermore, two strategies in the cointegration setting are compared. One strategy based on daily data and one strategy based on hourly data. It can be shown that the strategy based on daily data is superior to the strategy based on hourly data. This suggests that the crpyto market is not as inefficient as assumed for higher frequency data after all and contradicts results from previous research. Following these results, it can be assumed that the crypto market is still partially not fully efficient.

Keywords: Pairs Trading, Crypto Currencies, Cointegration

Contents

List of Figures

List of Tables

1 Introduction

In Financial Markets profits can be made through buying an undervalued asset or shorting an overvalued asset. There are several approaches to find over- or undervalued assets for investors. Although, investors believe that they are doing a profit with their investments, it is not certain if they do so. Therefore it seems appealing to investors to obtain profits with a high certainty. Obtaining riskless profits in the financial markets is called arbitrage. This for example can be achieved through buying and selling the same assets at different locations at different prices. Suppose that the market for Euro is in the USA at €1 = \$1.10 and in Switzerland at €1 = \$1.10. If all of a sudden there was a high demand for Euro in the USA, the price there would rise to €1 = \$1.15. But the market in Switzerland would react with a time lag. Then one would buy Euro in Switzerland and sell them immediately in the USA to generate arbitrage. The classical form of arbitrage exploits market inefficiencies that violate the law of one price. In addition to this classical definition of arbitrage, there is also a newer version of arbitrage, in a more statistical context. This so-called statistical arbitrage is based on trying to find miss-valuations based on stochastic trends. One of the fairly unknown approaches of statistical arbitrage is pairs trading. Pairs trading investigates, whether at least two assets have shown an equilibrium relationship in the past. The basic idea behind pairs trading is to trade deviations from equilibrium and to bet on a return to equilibrium. This strategy had a lot of success in the stock market in the 90s and in the early 2000s, but has been characterized by declining profitability since then. Although it was shown that significant abnormal returns could still be achieved in more illiquid bear market periods (Do and Faff, 2012). According to (Fama, 1970) this kind of generating abnormal returns should not be possible, since all information should be fully priced in. A market can not be called fully efficient if such opportunities show up. This strategy is therefore predestined for markets that are considered inefficient and often show declining market periods. One market that has long been considered as not particularly efficient is the crypto market, which (Urquhart, 2016) showed using the crypto currency Bitcoin as an example.

The present work investigates with a pairs trading strategy if statistical arbitrage opportunities show up for the relatively new crypto currency market. This is especially interesting under the assumption that crypto currency markets tend to increase their efficiency in recent years, but still show inefficiencies when using higher frequent data than daily data (Aslan and Sensoy, 2020). As a method for investigating the efficiency of the crypto currency market, a pairs trading strategy based on cointegration according to (Vidyamurthy, 2004) is chosen. In order to determine whether pairs trading strategies are profitable in the crypto market, which would mean that the market is not fully efficient, this paper will primarily focus on the following two research questions:

1. Do cointegrated crypto currency pairs outperform a simple Buy-and-Hold strategy?

2. Do pairs which are formed on a higher data frequency outperform pairs which are formed on a lower, daily frequency?

Therefore, the present work, mainly contributes to the topic if crypto markets can be considered as efficient and whether efficiency has further increased compared to previous studies (Urquhart, 2016), (Aslan and Sensoy, 2020), (Fil and Kristoufek, 2020). Furthermore, a distinction between daily and hourly data is made for investigating if the efficiency is higher when using higher frequent data.

If the present work finds that the pairs trading strategies are superior to the benchmark Buy-and-Hold strategy, it can be assumed that market efficiency is still not completely present in the crypto currency markets and they are prone to abnormal returns. If the pairs trading strategy with the higher data frequency is superior to the strategy which is based on daily data, it can be assumed that market inefficiencies tend to increase on the short term. This thesis shows that the applied pairs trading strategy always outperforms the underlying benchmark strategy and that a strategy which is based on daily data is superior to the strategy which is based on a higher data frequency. Although, it must be mentioned that both strategies not always perform market-neutral as one would expect from a statistical-arbitrage strategy. The strategy based on daily data achieved returns between [-0.249, 0.129] percent as an average daily return and the strategy based on hourly data had results between [-0.284, -0.045] on average. Both strategies are tested for 3 trading periods with respectively 75 trading days. The findings of this research are especially appealing for investors, which are not trading in a high frequency framework and are considering to perform a pairs trading strategy in a cointegration-based framework.

Figure 1: Market Capitalization of the Crypto Currency Market: June 2016 - May 2022

Source: Own presentation with data from Statista[a] which was retrieved on the 2022-05-28

[a]https://www.statista.com/statistics/730876/cryptocurrency-maket-value/

Exploiting market inefficiency using pairs trading was first done, as far as known, by Nunzio Tartaglia in the mid-1980s. He was hired by the investment bank Morgan Stanley in order to build a team packed with physicists and mathematicians, to develop quantitative arbitrage strategies. The team around Tartaglia managed to build a groundbreaking strategy, which today is known as pairs trading, it was already automated to the point that it could execute trades on its own (Vidyamurthy, 2004). Years after the first practical use of pairs trading, the first wider recognised academic publication on the subject of pairs trading was written by (Gatev, Goetzmann, and Rouwenorst, 1998). This showed an easy-to-follow two-step concept, with the division of a formation and a trading period. This was followed by several other pairs trading approaches from (Vidyamurthy, 2004) and (Elliott, Van Der Hoek*, and Malcolm, 2005). So far, academic research has focussed mainly on pairs trading in traditional asset classes such as equities – however, there is not much research on how pairs trading strategies perform in the crypto currency market. The crypto currency market really picked up speed for the first time in mid-2017, before cooling off again at the beginning of 2018 which can be seen in Figure 1. After a period without much movement in the market, the crypto market began to rise sharply again at the end of 2020. Considering the novelty of the market, it is not surprising that there is not much research on the topic of pairs trading in crypto currency markets so far. One of the few papers on this field is from (Fil and Kristoufek, 2020), they tested two pairs trading strategies for 26 crypto currencies from January 2018 to September 2019 which mostly underperformed the benchmark Buy-and-Hold strategy. Furthermore, they found evidence that strategies with higher frequent data outperform strategies which are based on daily data. The present research investigates if this still holds for the period from 2020-10-14 to 2022-05-13.

The following parts of this paper are structured as follows: In section Literature the existing literature about pairs trading will be discussed. This literature is complemented in section Applicable Methodology for Pairs Trading by the description of different pairs trading methods. Section Empirical Application: Cointegration Approach illustrates the chosen cointegration approach and describes the practical implementation of this work in Python. Subsequently, the results are discussed in section Results and the work is concluded with a final evaluation in section Conclusion.

2 Literature

Pairs trading first gained academic attention with the the publication of (Gatev et al., 1998). In his approach he used the so-called distance method to find tradeable pairs. For finding the minimum distance between stocks, first a cumulative return index is constructed – with dividends reinvested. Afterwards the prices of each security are normalized and in the next step the different indices are tested against each other in order to find the pairs with the minimum sum of the squared distances. This was done for daily data and in a so called 12-months long formation period. This was followed by a 6-months trading period, in which trading signals were generated when the prices diverged more than 2 standard deviations from the mean that was observed in the formation period. They did their research based on daily CRSP data from 1962 till 1997 for all liquid US-Stocks. The minimum distance method showed annualized excess returns up to 12 percent while being robust to transaction costs and uncorrelated to S&P 500 returns. The results of a newer version of the original study from (Gatev, Goetzmann, and Rouwenhorst, 2006) were confirmed by (Do and Faff, 2012). Furthermore they extended the observation period to 2009, but also showed decreasing profitability over time, particularly since 2002. They attribute the decline in profitability to two main reasons: First of all, the increasing hedge fund trading activity, which exploites the statistical arbitrage opportunities. And secondly the increased number of pairs that deviate from the equilibrium relationship and never return.

In addition to the distance approach, (Vidyamurthy, 2004) introduced a cointegration-based approach. This approach tries to find a statistical relationship between two assets, as a condition both assets have to be integrated of the same order. Furthermore, he proposed a multivariate strategy, which does include more than two assets. He also showed, that testing the cointegration of assets is way more effective than testing their correlation, for finding pricing anomalies which are influenced by shared underlying factors. In addition to that (Rad, Low, and Faff, 2016) introduced a copulas approach to further expand the cointegration approach. However, this approach was inferior to the other approaches presented, according to the rate of return.

Complementary to the distance and cointegration approach, (Elliott et al., 2005) introduced a more sophisticated approach which has the modelling of the spread , which generates the trading signals, in its focus. It can be categorized to the stochastic approaches. They model the spread with an Ohrnstein-Uhlenbeck process, which is an autoregressive model of order 1. Meaning that the future values of the estimated spread solely depend on the historical values. Furthermore, an advantage of the model is that it shows the property of mean reversion. This allows to calculate the time until the spread of the pair returns to their historical mean. Another work in this area was done by (Do, Faff, and Hamza, 2006). He states that the approach from (Elliott et al., 2005) has three advantages: Firstly, as described above, the spread is modelled reverting to the mean. Second, this approach is based on a continuous time model, which is particularly suitable for forecasting. Secondly, the time until the spread returns to the mean can be estimated. The last advantage he mentions is the calibration traceability, where the true state of the spread is estimated with a Kalman-filter. The true state gives an estimate of the actual spread.

All the above-mentioned works were investigating pairs trading strategies in traditional markets like stock markets. With the upcoming of crypto currencies, many investors, including private investors, have become involved with the topic of trading in crypto currency markets, since the entry barriers to install a setup for systematic strategies are very low and far better than on traditional stock-exchanges. However, on the topic of pairs trading on crypto markets, there appears to be very little research so far.

(Fil and Kristoufek, 2020) found that in-line with equity pairs trading strategies the profits are decreasing over time. They show that at least on a daily basis crypto currencies are acting inefficient to a high degree, furthermore they show that the cointegration approach is superior to the distance approach especially when using intraday data. Their results back an inefficient behaviour of crypto currencies on a daily basis, but they are stating that these inefficiencies are time variant. These results are in-line with the results of (Wei, 2018). He attributes the inefficiencies in the crypto market mainly to the liquidity situation of the respective currencies. He states that as the liquidity of the currency increases, increasing efficiency can be seen. Since the publication of these studies, the crypto market has gained quite some momentum, see Figure 1, so it is interesting to know if the described inefficiencies still can be found in the market, or if the crypto market has gained further efficiency.

3 Applicable Methodology for Pairs Trading

3.1 Methodology

In this section I explain different approaches for pairs trading. First of all the Distance Approach which can be considered as the most simplistic one of the presented concepts. It is characterized by a division of the time period into two phases: A formation period in which a relationship between the assets is found, and the following trading period. In the trading period deviations from the mean are traded and bets are placed on a return to the mean. Although, I do not use the method of the minimum squared distance for finding the pairs in this work, this approach illustrates quite well what pairs trading is all about: Finding somehow a relationship between two assets. Furthermore, I used the two-step approach of (Gatev et al., 2006) as a blueprint to combine it with the cointegration approach of (Vidyamurthy, 2004). The formation period in this work is based on the Cointegration Approach, and the trading period is based on the idea of calculating a spread between the pairs like (Gatev et al., 2006) did it, although the spread in this work is calculated in a different way, as it shown in section Trading Rules and Strategy. The concepts are rounded out in section Other Approaches by the presentation of other possible pairs trading approaches, like time-series and machine learning concepts, which are marked by higher complexity. In Section Choice of Approach the chapter is completed by the selection of the approach which is used in the present work.

3.1.1 Distance Approach

In distinction from the other approaches, the minimum distance approach is characterized by its simplicity and therefore it is widely used in pairs trading. This approach is divided into two phases. Firstly, there is a so-called formation period in which equally moving pairs are selected and secondly a trading period in which trading signals are generated and used to take a position in the underlying stocks and subsequently to unwind them.

Formation Period: (Gatev et al., 2006) use daily data from 1962 to 2002 on all liquid US stocks. For every stock i a cumulative total return index $P_{i,t}$ is formed and normalized for the 12 months formation period. On the first day of the formation period each index has a value of 1 US-Dollar (El-Oraby, 2019).

$$P_{i,t}^{norm} = \prod_{j=1}^{t}(1 + r_{i,j}) = \frac{P_{i,t}}{P_{i,1}} \tag{1}$$

Where $P_{i,t}$ describes the price of stock i at time t and $r_{i,t}$ the discrete return with $r_{i,1} = 0$. Without this, prices could not be measured relatively to each other and hence stocks with a lower price could not be compared to stocks with a higher price. The next step then, is to find a relationship for the constructed price indices. For this purpose the Euclidean distance between the stocks is calculated and the closest pairs are selected (El-Oraby, 2019).

$$V(P_{i,t} - P_{j,t}) = \frac{1}{T}\sum_{t=1}^{T}(P_{i,t} - P_{j,t})^2 - (\frac{1}{T}\sum_{t=1}^{T}(P_{i,t} - P_{j,t}))^2 \tag{2}$$

Where $V(P_{i,t} - P_{j,t})$ describes the spread variance of the pair $P_{j,t}$ and $P_{i,t}$. This can now be solved for the sum of the squared distances (Krauss, 2017):

$$\overline{SSD_{i,j,t}} = \frac{1}{T}\sum_{t=1}^{T}(P_{i,t} - P_{j,t})^2 = V(P_{i,t} - P_{j,t}) + (\frac{1}{T}\sum_{t=1}^{T}(P_{i,t} - P_{j,t}))^2 \tag{3}$$

Trading Period: According to (Gatev et al., 2006) the Top 20 pairs which have the smallest historic distance for the 12-months formation period are now chosen for the following 6-months trading period. In addition to that they also applied sector filter criteria to filter out pairs which belong to the same industry. This also adds a qualitative component to the quantitative approach. For the trading period the price series of the securities are normalized again to the first day of the period (Krauss, 2017). And positions are taken when the historic spread of the pairs diverges more than two standard deviations from their mean and closed again when they revert to the historic mean (Krauss, 2017).

Although the strategy has not performed badly historically, it also comes with some drawbacks. (Krauss, 2017) mentioned three major reasons. The best possible pair according to the distance approach would be a pair with a zero squared spread, meaning that no profit could be generated as no trading opportunities would open up. Although this specific case should not occur in real life, he also found that the top pairs with a minimal distance do also have a low spread volatility – subsequently less profit generating opportunities are generated. Secondly, the above-mentioned approach tests for long run relationships, therefore the pairing of the stocks cannot be justified statistically as 12-months are a too short observation period to find mean-reverting properties of the paired underlyings. Thirdly, the Top 20 pairs are always formed, this could lead to the result that there are enforced pairs that are not suitable for trading and this could subsequently lead to losses.

3.1.2 Cointegration Approach

One of the most recognised pairs trading methods was developed by (Vidyamurthy, 2004) and is based on cointegration which was first shown by (Engle and Granger, 1987). Cointegration tests for long term statistical equilibrium relationships between two or more securities. An important method for investigating if two time series are cointegrated is stationarity. Stationarity checks if a time series has a constant mean and variance over time, meaning that it is probability distribution is constant over time. Stationarity can be tested with the Augmented-Dickey-Fuller test (ADF-test), developed by (Dickey and Fuller, 1979): When a stationary time series is found, mean-reversions can be traded, as over time the time series always diverges back to the mean. What becomes difficult here is that as stated by (Neusser, 2016) most economic time series are non-stationary and become only stationary after some transformation. Log-returns for example are often stationary but cannot be traded, as only prices can be traded and not returns. Therefore, the goal is two find two securities which are stationary together, although they are not stationary on their own. To test for cointegration it is essential that the two tested time series are non-stationary. If they are non-stationary then they are tested on cointegration. Firstly, an equilibrium relationship is estimated with an ordinary least squares regression:

$$P_{i,t} = a + ßP_{j,t} + u_t \tag{4}$$

Here, a represents the estimated intercept of the regression equation and ß the estimated coefficient. The spread in pairs trading is denoted by the residuals in the equation u_t. The results of the regression are only meaningful if the pairs are cointegrated, otherwise this could result in a spurious regression, meaning that the found relationship is only random (El-Oraby, 2019). Secondly, the residuals are tested on stationarity. This is done with above mentioned ADF unit root test:

$$\Delta\hat{u}_t = \psi * \hat{u}_{t-1} + \sum_{i=1}^{p-1} \psi_i \Delta_{t-i} + \varepsilon_t, \tag{5}$$

where the error term ε_t is expected to be white noise. The null hypothesis, $\psi* = 1$, is that the residuals are non-stationary, and therefore can be discarded. Thus, $-1 < \psi* < 1$, which means that the residuals are stationary then this leads to the result that the time series is cointegrated. Additionally, it is important to test each pair twice, since $Coint(I, J) \neq Coint(J, I)$. Therefore two regressions have to be made:

$$P_{i,t} = a_1 + ß_1 P_{j,t} + u_{1,t}$$
$$P_{j,t} = a_2 + ß_2 P_{j,t} + u_{2,t}$$

And the residuals of both regressions have to be tested for unit roots, like mentioned above. For investigating cointegration on more than two time series the Johansen test would be appropriate (El-Oraby, 2019), since this is a multivariate approach for finding cointegration of multiple assets.

3.1.3 Other Approaches

There are other approaches which are however are marked by a higher complexity of the respective approach. The time series approach for example is characterized by finding mean reverting properties in the security pairs without using either cointegration and a formation period. Furthermore, the time series approach is trying to optimize the trading period by finding the most accurate and most profit generating trading signals. The most acknowledged academic work in this field was done by (Elliott et al., 2005). They model the spread of the pairs with a mean-reverting Gaussian Markov chain model. Another approach is the stochastic control approach. Like the time series approach this approach is also not focused on the formation period rather than the trading period. It focusses on building the optimal tradeable portfolio. One of the better known academic works on this field was done by (Liu and Timmermann, 2013). They build on a cointegration framework and trying to find the optimal holdings for the currency which is shorted and the currency which is bought, with a power utility over terminal wealth function. In addition to that they also include the benefits of diversification in an arbitrary environment (Krauss, 2017). The latest approaches in the field of pairs trading mainly deal with machine learning approaches, which even integrate data from social media (Li, Chamrajnagar, Fong, Rizik, and Fu, 2019).

3.2 Choice of Approach

(Krauss, 2017) showed in his meta study the performances of different approaches. For the distance approach the annualized return is between 7 and 11 percent for equities. Although the most cited paper for the cointegration approach (Vidyamurthy, 2004) does not provide empirical results, the paper of (Caldeira and Moura, 2013) shows an annualized return of 16 percent for the cointegration method for equities. But a comparison is actually inadmissible here, since (Caldeira and Moura, 2013) only investigate the time range from 2005 to 2010 for Brazilian stocks and the papers on the distance approach examine a time range of at least 40 years, on US stocks. However, there are several indicators that the cointegration approach is superior to the distance approach. This is due to the simple calculation methodology of the distance approach, which simply measures the distance of price indices. Econometrically, the approach which uses a regression and a stationarity test to select the tradeable pairs, appears to be much more meaningful. Therefore, in the framework of the present work the cointegration approach is chosen in order to find the tradeable pairs.

3.3 Buy-and-Hold as Benchmark Approach

In this chapter the approach which serves as a benchmark for the pairs trading strategy is briefly discussed. For each identified currency pair a corresponding Buy-and-Hold pair is formed. The Buy-and-Hold pairs are equally weighted from the beginning of the formation period. For each currency an index is formed with 0.5 as a starting value in t_0. Afterwards for each observation period the $1 + logreturns$ are multiplied with the index of each currency. Then both indices are added together and the returns for the combined index is calculated for each point in time. Based on this index the average daily return and all the other summary statistics are calculated.

4 Empirical Application: Cointegration Approach

4.1 Data Source

To determine if a trading strategy was profitable in the past, it is necessary to backtest the strategy. In order to do so, historical data is needed to simulate the strategy. The data used in this paper comes from Binance[1] and is gathered with their API for Python. Binance is one of the biggest crypto exchanges in the world and one of the few providers that offer short frequency data for a variety of crypto currencies for free. They provide different frequencies from daily data down to 1-minute data. For every data frequency they provide open, close and volume data. For the purposes of this research the closing spot price of the hourly and daily data is used. The closing price refers to the spot price at 2:00 UTC for daily data. The first observation of the hourly data also takes place at 2:00 UTC and is then supplemented at one-hour intervals. Furthermore, every currency is denoted against the stable coin Tether which is pegged to the US-Dollar and should always provide a value close to 1 US-Dollar. This is necessary since Binance does not allow direct trading with fiat currencies. Fiat currencies are currencies which are issued by governments and not pegged to a commodity like gold and silver. For trading crypto currencies first fiat money has to be changed into stable coins and afterwards the stable coin can be traded against a crypto currency. Tether is always used as the currency in the denominator to ensure comparability between the individual currencies.

[1] https://www.binance.com/en

Table 1: Crypto Currencies Market Data - 15.05.2022 15:15 CET

Name	Ticker	Price	Market Cap	Volume (30d)	Market Share
Bitcoin	BTC	30,302.320	575.717	1,152.769	44.629
Ethereum	ETH	2,079.470	250.749	1,504.265	19.438
BNB	BNB	301.950	49.111	52.918	3.807
XRP	XRP	0.433	20.824	48.402	1.614
Cardano	ADA	0.556	18.736	31.255	1.452
Solana	SOL	53.160	17.902	59.220	1.388
Dogecoin	DOGE	0.090	11.865	24.828	0.920
Polkadot	DOT	11.050	10.885	37.737	0.844
Avalanche	AVAX	34.520	9.256	22.898	0.717
TRON	TRX	0.070	6.840	27.242	0.530
Polygon	MATIC	0.689	5.475	19.856	0.424
Litecoin	LTC	69.070	4.855	21.776	0.376
NEAR Protocol	NEAR	6.750	4.649	20.416	0.360
Bitcoin Cash	BCH	208.580	3.977	105.872	0.308
Stellar	XLM	0.140	3.465	8.338	0.269
Cosmos	ATOM	11.270	3.228	9.816	0.250
Algorand	ALGO	0.465	3.167	4.661	0.245
Monero	XMR	162.680	2.942	10.766	0.228
Ethereum Classic	ETC	21.150	2.851	30.862	0.221
Hedera	HBAR	0.103	2.130	2.555	0.165
VeChain	VET	0.032	2.050	14.630	0.159

Price: Price of the currency in US-Dollar, Market Cap: Market Capitalization in billion US-Dollar, Volume (30d): Traded Volume over 30 days in trillion US-Dollar, Market Share: Market Cap of the currency divided by the total crpyto market capitalization denoted in percent. They data was retrieved from coinmarketcap.com on the 15.05.2022 at 15:15 CET.

4.2 Data Filtering

The choice of crypto currencies is restricted on currencies, which are tradeable on Binance, since the data source for this work is Binance. To get a list of the most important crypto currencies, a liquidity- and market cap filter is applied on the currencies. The data for the market cap and liquidity were retrieved on 15th May 2022 15:15 CET from the website Coinmarketcap[2]. The threshold for liquidity is set to 2 trillion USD for the last 30 trading days, the threshold for the market cap is set to 2 billion USD. Especially, the market cap filter was set to a relatively high value in order to guarantee that a long enough data history is available, since newly issued coins tend to have a lower market capitalization than the ones that are already established on the market. As mentioned above, only coins which are tradeable at Binance were considered. Furthermore, so-called stable coins,

[2] https://coinmarketcap.com

which should always provide the same value to a fiat currency, are not considered. If a currency is pegged to another fiat currency, the time series can be considered as stationary as they do show a constant mean and standard deviation over time. Subsequently, these currencies cannot be integrated of order one in the Engle-Granger framework, meaning that they would not pass the stationarity test (Van den Broek and Sharif, 2018). After the filtering process, 21 currencies were retrieved, they stand for 78 percent of the total crypto market capitalization as of 15th May 2022, this can be seen in Table 1. The total volume of the crypto market was valued at 1.29 trillion USD on that day. The historic spot price data on Binance for all in Table 1 mentioned coins are available from 14th October 2020 onwards.

4.3 Trading Rules and Strategy

The pairs for the trading period are derived from the Engle-Granger 2 step cointegration test (Engle and Granger, 1987), and they are considered for the trading period, when their p-value from the augmented Dickey-Fuller test, which tests the residuals of the regression for cointegration, is below the 2.5 percent significance-level, as it was described in the section Cointegration Approach. For the calculation of the spread, the Z-score is used in order to normalize the spread between the pairs. If a normal spread would be considered the absolute price-difference between Coin A and B would be calculated. Since the prices of the coins can differ significantly, this would not be meaningful. For this purpose it is important to normalize the spreads. First the spreads are calculated with an Ordinary-Least-Squares regression with the currency which is in the denominator as the dependent variable. So in the case of BTC/ ETC, ETC would be the dependent variable. The regression is made on a rolling basis with a rolling window of 30 observation points. So in case of daily data, the regression would always calculate the slope and the mean of the regression based on a 30-day time window. It must be mentioned, that there is no empirical justification for the length of the window. The spread of the pair is calculated with the following formula:

$$S_t = p_{2,t} - \beta p_{1,t} \tag{6}$$

where S_t denotes the spread of the assets p_1 and p_2 at time t and β is defined as the slope from the Ordinary-Least-Squares regression. When calculating the spread for the whole observation period, a stationary time series with a constant mean and standard deviation is obtained. This is the key part of the approach, since the strategy can only be traded on a time series with a mean reverting behaviour (El-Oraby, 2019). In order to retrieve trading signals from the spread, it has to be normalized. The normalization is done with the help of a Z-Score:

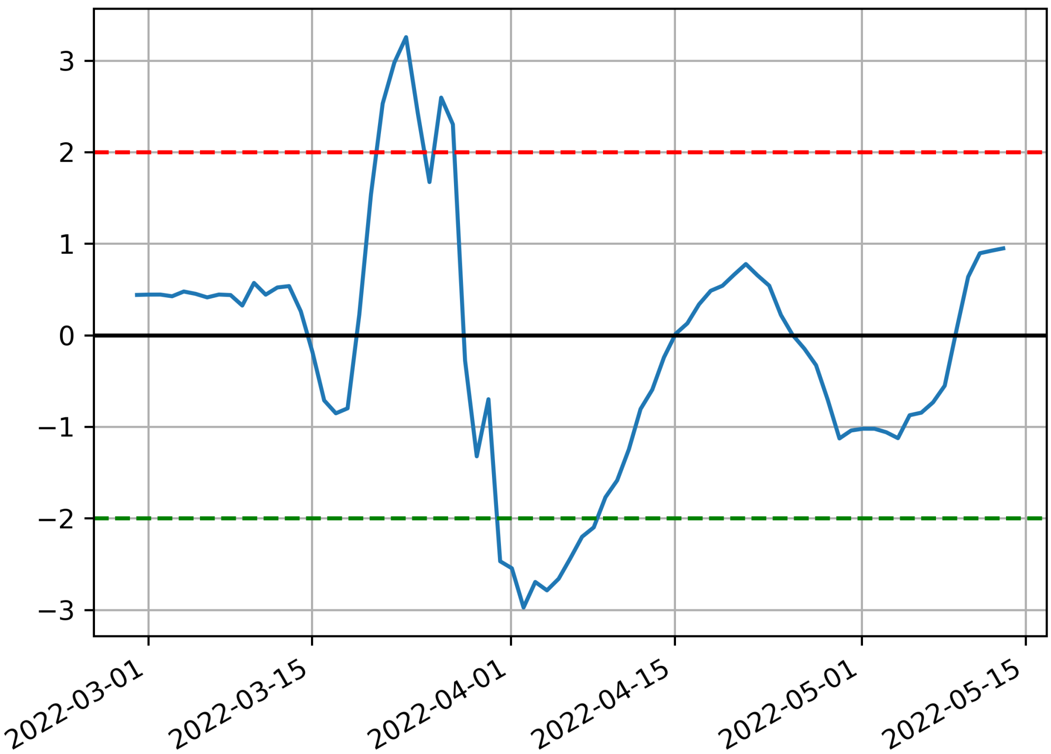
$$z_t = \frac{S_t - \mu_f}{\sigma_f} \tag{7}$$

The mean and the standard deviation of the spread are calculated for a rolling window of 30 observation points.

After this step, the Z-Score only needs to be translated into trading signals: If the spread diverges ± 2 standard deviations from the mean, trading signals are triggered. A long position for Coin A is taken and a short position in Coin B/β if the spread diverges positive 2 standard deviations from the mean. If the spread reaches the threshold of negative 2 standard deviations, the opposite is done. Subsequently a short position in Coin A and a long position in Coin B/β is taken. Both positions are closed when the spread moves back to zero which is equal to the mean of the spread. Although these limits can obviously be optimized, it is in-line with current research in this area (Fil and Kristoufek, 2020), but offers room for improvement. For example if the spread never returns to the mean, the position is still open and will never be closed, therefore it could be useful to set limits at for instance ± 5 standard deviations, to close the position.

In Figure 2 a normalized spread of a crypto currency pair can be seen, for the calculation daily data was used. The position in the present work is opened and closed directly when the spread reaches the thresholds of ± 2 standard deviations. Although, tick data is not incorporated in this work, it seems legit to trade directly on the price used for the calculation of the spread. Since in the real world crypto markets are open 24/7 and it could be traded directly to the next price tick, which should not deviate much from the last tick on which the Z-Score was calculated. Furthermore, the results of the strategy are also more accurate, compared to using a delay of one hour or one day for the price. For calculating the returns log-returns are used, and in contrast to discrete returns they are additive, which makes the calculations of the summary statistics easier. The profit or loss for each strategy is calculated by summing up the log-returns of the periods when the strategy was invested.

Figure 2: Exemplary Z-Score for the crypto currency pair LTC/ VET for the trading period from 2022-02-28 to 2022-05-13



In the figure above the Z-Score for the crypto currency pair LTC/ VET for the trading period from 2022-02-28 to 2022-05-13 can be seen. The pair was formed in the formation period from 2021-08-02 to 2022-02-27, based on a daily data frequency. Source: Own representation with spot price data from Binance

4.4 Formation Period

For each formation period a time window of 210 days is chosen. Since the academic literature does not offer an optimized window for the formation period, it seems legit to use a length of 210 days. Subsequently for the daily data 210 oberservations are observed and for the hourly data 5,040. Furthermore, three formation periods are defined, to reflect different market environments for the pairs trading strategy. The first formation period ranges from the 14^{th} October 2020 to the 11^{th} May 2021, the second from 9^{th} March 2021 to the 5^{th} October 2021 and the third from the 2^{nd} August 2021 to the 27^{th} February 2022.

4.5 Trading Period

Each trading period consists of 75 days, resulting in 75 observations for the daily data and 1800 for the hourly data. The first trading period starts on the 12$^{\text{th}}$ May 2021 and ends on the 25$^{\text{th}}$ July 2021, the second period begins on the 6$^{\text{th}}$ October 2021 and ends on the 19$^{\text{th}}$ December 2021, the last period ranges from the 28$^{\text{th}}$ February 2022 to the 13$^{\text{th}}$ May 2022.

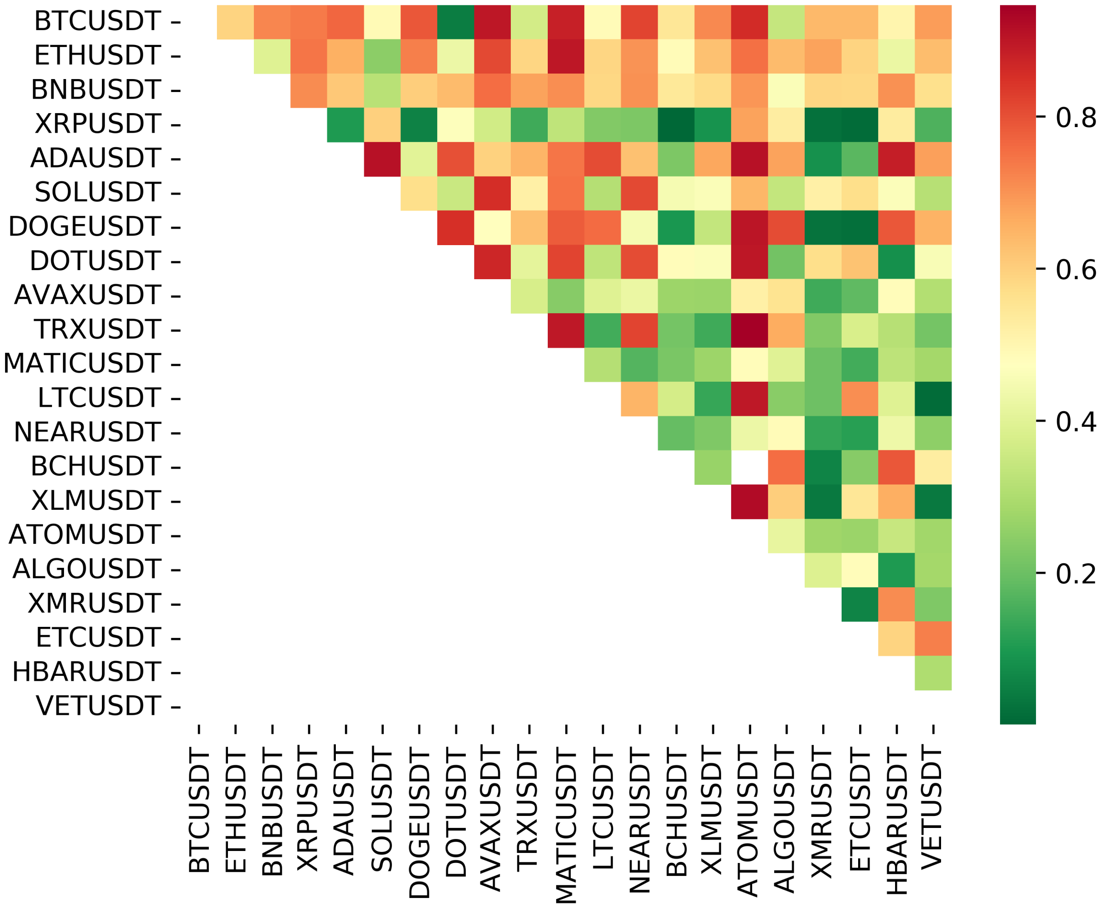
4.6 Practical Implementation of the Cointegration Approach

In this chapter the practical methods for identifying the pairs and calculating the returns are described. For all the calculations the programming language Python (Version 3.7.11) is used. First of all several libraries are installed, including libraries like Numpy, Pandas and SciPy. In the next step the crypto currency data, which is gathered with the Binance API[3] for python, is imported. This is done for the interval type of 1 day and 1 hour, for each observation point the closing price and the according time stamp is retrieved. Furthermore, the data is downloaded with a *for* loop, which allows to split the request of datapoints, since Binance currently does not allow to retrieve more than 1,000 datapoints per currency within a single request. After this step each crypto currency time series is tested for stationarity with the Augmented Dickey-Fuller unit root test. For this purpose the 'adfuller'[4] function from the package statsmodels was used and a significance level of 1 percent applied. If a time series is stationary according to the test, it is going to be filtered out. In the next step the time series are tested for cointegration with the 'coint'[5] function from statsmodels. For this step a significance level of 2.5 percent was chosen. Exemplary for this step the results of this method can be seen in Figure 3. If a pair is marked in dark red it means that the pair is not cointegrated at all, whereas a dark green colour signals that the pair is cointegrated.

Since now the pairs are retrieved they can be regressed for the formation and trading period in order to calculate the spread for each point in time in the trading period. For this purpose an Ordinay Least Squares linear regression was used from the 'scikit-learn'[6] library. Also here, a *for* loop was used in order to calculate the slope and the intercept in

[3]https://api.binance.com/api/v3/klines
[4]https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.adfuller.html
[5]https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.coint.html
[6]https://scikit-learn.org/stable/index.html

Figure 3: Heatmap of p-values for the cointegration test



This heatmap shows the coloured p-values of the cointegration test for the formation period from 2020-10-14 to 2021-05-11 for the for hourly data set. Source: Own representation with data from Binance

each case for a 30 observation points time window. After getting the slope and intercept for each point in time, the spread is calculated as described in section Trading Rules and Strategy in Equation 6. In the next step the Z-Score was calculated in order to retrieve the trading signals, this was done with a *for* loop as well and for every cointegrated crypto currency pair, based on Equation 7. The mean and the standard deviation were calculated for a 30 observation points time window as well. In order to calculate the return for every pair, I aligned the prices and the Z-Score for every time series in a DataFrame. Furthermore, I wrote two functions to get the trading signals. I defined one function for the upper limit when the spread is $\geq$ 2 standard deviations, and one for the lower limit when the spread is $\leq$ -2 standard deviations, as described in section Trading Rules and Strategy. These functions created a new column in the DataFrame where a 1 was written in the row when a position was opened and kept doing so until the spread converged to

0 again. Since this is still not sufficient to calculate the returns for the trading period, I wrote another function, which takes care that the log-returns are inserted in another column as long as a 1 is given from the trading signals functions. And since there can be no return at the first 1 because the position was only opened, it was started at $t + 1$. Afterwards all summary statistics were calculated based on the log-returns. If there was no open position subsequently no return can be calculated than a 'NaN' value was written into the DataFrame column. Afterwards I calculated all the summary statistics. I did all calculations under the assumption that 'NaN' values are ignored and not treated as a 0.

5 Results

The following section discusses the results for every observation period and for each period I provide a graph with the normalized prices for Bitcoin and Ethereum. This classifies the market situation for the formation respectively the trading period and can be seen as proxy for the crypto market. Furthermore, I provide a comprehensive results table with the average mean and standard deviation for each strategy. The detailed statistics for each strategy can be found in the result tables in the appendix.

5.1 Discussion of the First Period (2021-05-12 to 2021-07-25)

Daily Data: In the formation period 13 crypto currency pairs are cointegrated for the approach with daily data, thereof 11 pairs show trading signals in the trading period. Not showing a trading signal means that in the trading period the spread of the pair is not sufficiently diverged from the mean of ± 2 standard deviations. The observed trading signals of the pairs are in the range of [0, 3]. Although for the remaining pairs only one pair shows a positive return for the trading period, however 10 out of 11 pairs beat the simple Buy-and-Hold strategy. On average they have a daily return of -0.25 percent versus -0.79 percent for Buy-and-Hold. For all traded pairs the standard deviation was well below the standard deviation of the Buy-and-Hold strategy. This also holds for the min and max values.

Hourly Data: Ten pairs have been identified to be cointegrated in the formation period from which all show better returns than the simple Buy-and-Hold strategy, resulting in an average return of -0.29 percent versus -1.03 percent. The trading signals of the pairs were in a range of [38, 50]. And 5 of the 10 pairs are identical to the pairs of the daily data set. One of these pairs had no trading signals for the daily frequency. For the remaining four pairs, three pairs showed a better return when traded on hourly data. As one would expect, the trading activity is also much higher for the hourly data frequency. Furthermore, it also can be seen that the average standard deviation for the pairs formed on hourly data was well below the standard deviation of the benchmark Buy-and-Hold strategy.

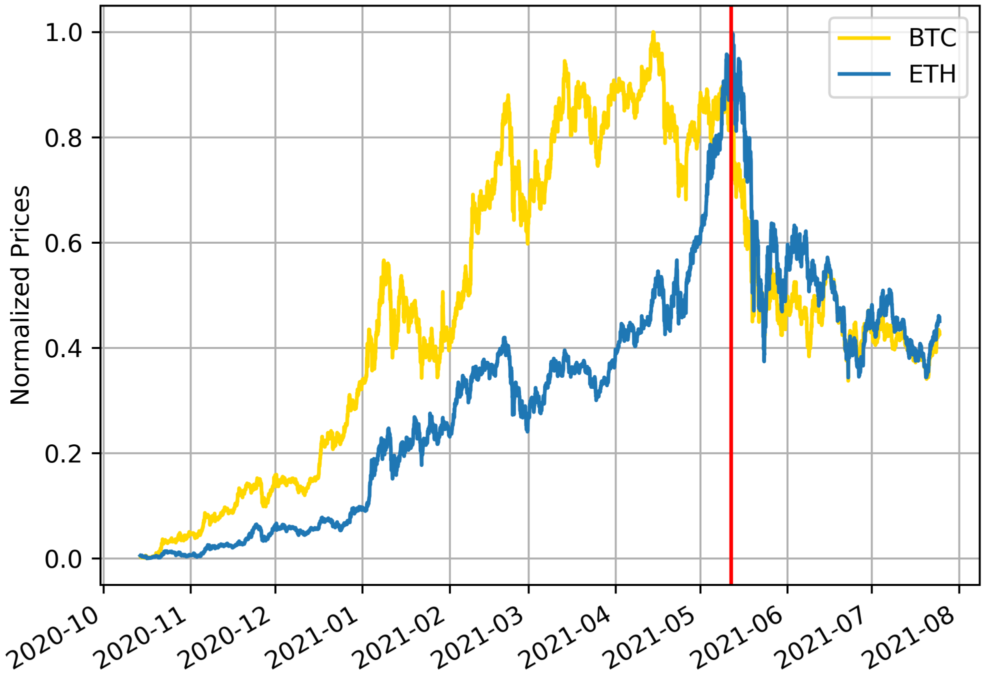
The formation period as it can be seen in Figure 4, is marked by rising valuations in the crypto market. For the trading period a sharp decline in crypto valuations can be seen. Although, there was a decline in crypto prices one would expect that since pairs trading is a market neutral strategy on average positive returns can be seen. On the other hand one could argue that the formation period of the pairs took place in a completely different market environment than the trading period. This is also shows a weakness of the approach, since the pairs were no longer tested for cointegration in the trading period. If they would have been tested on a rolling window basis, probably some pairs would have been no longer cointegrated for the trading period, which would not have resulted in so many negative returns. In terms of return the pairs trading strategy with daily prices had a slightly better return compared to the strategy with hourly prices.

Table 2: Results for the First Period: 2021-05-12 to 2021-07-25

Strategies	Pairs	Tradeable Pairs	Mean	Std Dev
Daily	13	11	-0.246	2.605
Daily: Buy-and-Hold	-	-	-0.793	9.474
Hourly	10	10	-0.284	3.115
Hourly: Buy-and-Hold	-	-	-1.030	9.331

Pairs: Number of cointegrated pairs for the observation period between 2020-10-14 to 2021-05-11, Tradeable Pairs: Number of pairs which showed trading signals in the trading period, Mean: Average daily mean return in percent, Std Dev: Average daily standard deviation, multiplied by 100 for better visualization.

Figure 4: Normalized Prices of BTC and ETH for the first observation period



Source: Own representation with data from Binance

5.2 Discussion of the Second Period (2021-10-06 to 2021-12-19)

Daily Data: In total 11 crypto currency pairs are cointegrated and all of them had trading signals in the range between [1, 3]. The average daily return in this period was slightly positive with 0.009 percent average return over all pairs. And once again the pairs trading strategy could beat the Buy-and-Hold strategy, which delivered a return of -0.22 percent. Furthermore, the performance of 8 out of 11 pairs was superior in terms of return compared to the benchmark strategy.

Hourly Data: For the pairs trading strategy based on hourly data, on average a slightly negative return of -0.04 percent could be observed. But the strategy could also beat the Buy-and-Hold strategy which delivered a return of -0.21 percent. 9 pairs were traded and 6 of them showed a better return than the simple strategy. The trading signals were in a range of [39, 49], which is in-line with the trading signals of the first period.

In the second observation period the crypto market declined till the middle of July

2021, in the following period there was a sharp rise until the end of 2021 as it can be seen in Figure 5. For this period from the 9 cointegrated pairs of the hourly data, 7 pairs are identical to the pairs which were formed based on the daily data set. When comparing the identical pairs, 5 pairs performed better when using the daily data frequency.

Table 3: Results for the Second Period: 2021-10-06 to 2021-12-19

Strategies	Pairs	Tradeable Pairs	Mean	Std Dev
Daily	11	11	0.009	2.605
Daily: Buy-and-Hold	-	-	-0.221	4.561
Hourly	9	9	-0.045	1.921
Hourly: Buy-and-Hold	-	-	-0.212	4.999

Pairs: Number of cointegrated pairs for the observation period between 2021-03-09 to 2021-10-05, Tradeable Pairs: Number of pairs which showed trading signals in the trading period, Mean: Average daily mean return in percent, Std Dev: Average daily standard deviation, multiplied by 100 for better visualization.

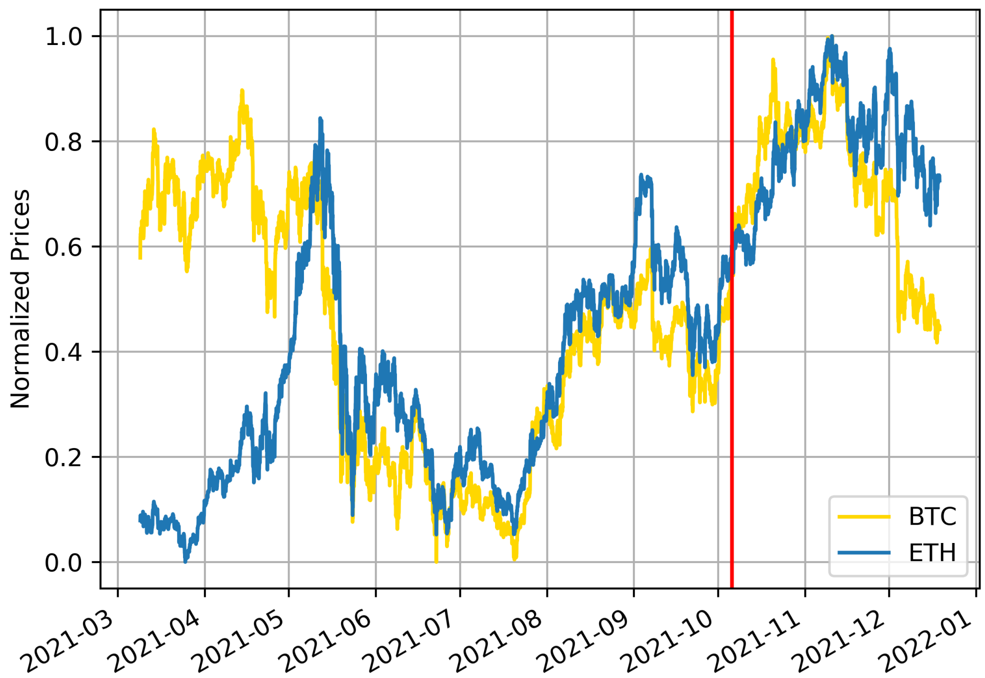
5.3 Discussion of the Third Period (2022-02-28 to 2022-05-13)

Daily Data: The pairs trading strategy based on daily data has 5 cointegrated pairs, which have a range of trading signals in between [2, 5]. In this period the daily strategy generates a return which is well above 0, for the first time, summing up for an average daily return of 0.13 percent. The Buy-and-Hold strategy however performed very weakly with -0.64 percent. Furthermore, the returns are on average skewed slightly positive, whereas the returns of the benchmark strategy are skewed negatively, which is inline with the number of negative trades. The number of negative trades never exceeds 50 percent for the pairs trading strategy compared to the Buy-and-Hold strategy where 4 out of 5 pairs had a number of negative returns well above 50 percent.

Hourly Data: For the pairs trading strategy 12 pairs have been found to be cointegrated. The pairs which are cointegrated show trading signals in the range of [38, 49], which is in-line with the previous two observation periods. From the 12 pairs which passed the Engle-Granger test, 11 beat the Buy-and-Hold strategy. Although the return of the strategy is below the return of the strategy with daily data, it could once again beat the Buy-and-Hold strategy with -0.15 percent average daily return versus -0.54 percent.

From which 5 pairs overlap with the strategy based on daily data. In this period it was the first time that more pairs were cointegrated based on a hourly data, than on

Figure 5: Normalized Prices of BTC and ETH for the second observation period



Source: Own representation with data from Binance

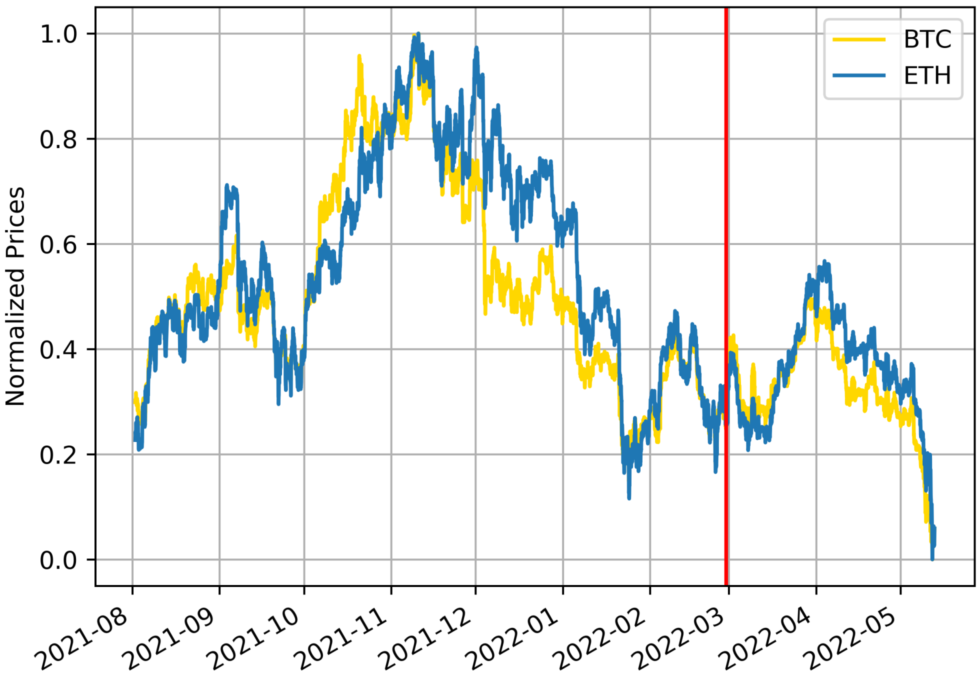
daily data. All of the 5 daily pairs can also be found in the hourly frequency and all of them performed better on a daily basis. Also the kurtosis in general is much higher on an hourly basis than on a daily, meaning that more extreme returns can be found for hourly data. Furthermore, the number of returns which are negative is smaller for the daily data. The third period is marked first by rising prices until the middle of November 2021 and followed by a sharp downtrend until May 2022 which can be seen in Figure 6. As one would expect this time, the pairs trading strategy achieved market-neutral positive returns, although it was only for the strategy which is based on daily data. Interestingly it can be seen that for none of the periods Bitcoin was involved in a pairs trading strategy and Ethereum only once – in the first period. Since Bitcoin and Ethereum are the largest crypto currencies in terms of market capitalization and liquidity it can be assumed that they also the most efficient ones. Therefore it can be assumed that Bitcoin and Ethereum are already too efficient to function as a pairs trading strategy.

Table 4: Results for the Third Period: 2022-02-18 to 2022-05-13

Strategies	Pairs	Tradeable Pairs	Mean	Std Dev
Daily	5	5	0.129	1.426
Daily: Buy-and-Hold	-	-	-0.646	5.144
Hourly	12	12	-0.151	1.749
Hourly: Buy-and-Hold	-	-	-0.542	5.195

Pairs: Number of cointegrated pairs for the observation period between 2021-08-02 to 2022-02-27, Tradeable Pairs: Number of pairs which showed trading signals in the trading period, Mean: Average daily mean return in percent, Std Dev: Average daily standard deviation, multiplied by 100 for better visualization.

Figure 6: Normalized Prices of BTC and ETH for the third observation period



Source: Own representation with data from Binance

6 Conclusion

In this research I have investigated the application of the cointegration approach to the relatively new crypto currency market. For providing the context of the applied methodology I have given an overview of relevant pairs trading strategies. The research was guided by the research questions as outlined in the Introduction.

Do cointegrated pairs in the crypto currency market outperform a simple Buy-and-Hold strategy?

For answering this question, three periods with 285 days each were examined and for filtering out the most important currencies, filters were adapted. In the next step the cointegration method according to (Vidyamurthy, 2004) was applied. I showed that, for all investigated periods the pairs trading strategies on average, regardless if daily or hourly data was used to form the pairs, could beat their benchmark strategy. Although it must be mentioned that most of the time the returns can not be considered as market neutral. Furthermore, the two biggest crypto currencies – Bitcoin and Ethereum – were only cointegrated one time. This suggests that the market for both currencies is already very efficient. After finding out that both strategies were superior to the Buy-and-Hold strategy, the next question I have investigated:

Do pairs which are formed on a higher data frequency outperform pairs which are formed on a lower, daily frequency?

The observed results are pretty clear regarding this. The pairs trading strategy which is based on daily closing prices, outperforms the pairs trading strategy based on hourly data, in every period. This contradicts the results, which were observed by (Fil and Kristoufek, 2020). They observed that hourly data performed better than their daily counterpart. The reason for this could be that (Fil and Kristoufek, 2020) worked with a slightly larger sample size, which included 26 currencies, whereas the present work examines 21 currencies. But it also could be the case that short term efficiency of the crypto currencies has increased. Also this could be because I did not use a one period execution lag like in (Fil and Kristoufek, 2020). Furthermore I did not incorporate trading costs which also lowers the profit, although this should lower the profit of the hourly strategy even more, since more trades are done on the hourly data based strategy.

The cointegration approach used in the present work depends on the chosen settings. For example the following parameters could be adjusted to improve the outcome. Firstly, the rolling window of the regression in the Engle-Granger 2 step method, could be adjusted, to another level. Secondly the rolling window of calculating the mean and the standard deviation for the Z-Score could also be adjusted. Both windows, the one from the regression and the one from the Z-Score, are set to 30 observation points. Thirdly, the thresholds for the trading signals for opening a trade are set to ± 2 standard deviations. Here could be the highest potential of optimization for this method. Further research should focus on optimizing these settings. This could be done with the approach chosen by (Elliott et al., 2005), which models the spread with the help of an Ohrnstein-Uhlenbeck process, which allows to calculate the time in which the spread returns to the mean value. Furthermore, it would be good to also test for cointegration in the trading period. The pairs which are not cointegrated anymore can be removed from the sample as soon as this is no longer the case. Apart from the settings of the chosen approach, it should be highly beneficial to add more coins to the sample, but this would have gone beyond the scope of this work. Coins with a lower market capitalization and a lower liquidity tend to have a lower market efficiency according to (Wei, 2018). Furthermore, transaction costs should be incorporated in further research, since with a high trading activity this can significantly reduce the profit. It would also be conceivable to add a qualitative component and to subdivide pairs that operate in the same subject area according to sectors. For example, pairs that act in the smart contracts sector could be put into a bucket.

Appendix

A Results: Trading Period

Table 5: Results for the first Period for the Strategy based on Daily Data: 2021-05-12 to 2021-07-25

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs <0	Trades
BNB/ SOL	-0.407	4.471	-1.364	8.545	-22.166	13.228	58.730	8
BNB/ MATIC	-0.456	4.692	-0.110	8.831	-20.673	19.256	65.000	12
SOL/ TRX	-0.239	2.774	-2.012	9.118	-12.926	6.549	47.500	8
SOL/ MATIC	-0.128	2.528	0.072	1.060	-6.649	7.468	51.064	12
SOL/ XMR	-0.494	3.321	-1.879	6.087	-13.666	5.349	50.000	8
DOGE/ ETC	0.305	1.563	1.689	1.712	-1.184	4.497	61.538	4
DOT/ AVAX	-0.039	2.275	-0.309	-0.153	-5.840	4.935	38.298	12
DOT/ ALGO	-0.313	1.253	-0.670	-1.331	-2.347	0.823	33.333	4
AVAX/ ALGO	-0.383	2.579	-1.811	5.792	-10.783	4.811	51.429	8
TRX/ VET	-0.006	1.497	0.580	2.883	-3.354	5.010	51.613	8
MATIC/ XMR	-0.547	1.702	0.633	-0.276	-2.850	3.287	62.500	4

Two crypto currency did not show trading signals for the trading period: SOL/ VET and ETC/ VET, Pair: Cointegrated crypto currency pair for the formation period from 2020-10-14 to 2021-05-11, the results are based on the trading period from 2021-05-12 to 2021-07-25, Mean: Average daily mean return for the trading period, Std Dev: Average daily standard deviation multiplied by 100, Skewness: Skewness of the pair for the trading period, Kurtosis: Kurtosis of the pairs for the trading period, Min: Minimum return of the strategy for the formation period, Max: Maximum return for the trading period, % Obs <: Percentage of returns which are below 0 in the trading period, Trades: Number of trades in the trading period. When a position is opened and closed 4 trades occur. 2 trades for the currency which is shorted and 2 trades for the currency which is bought.

Table 6: Results for the Buy-and-Hold Strategy for the First Period for the Strategy based on Daily Data: 2021-05-12 to 2021-07-25

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs < 0
BNB/ SOL	-0.752	9.453	-1.162	5.661	-44.332	26.456	45.946
BNB/ MATIC	-0.533	9.934	-0.149	5.454	-41.345	38.511	48.649
SOL/ TRX	-0.783	9.005	-1.398	5.975	-42.941	22.486	50.000
SOL/ MATIC	-0.371	10.964	-0.302	3.965	-44.006	36.772	44.595
SOL/ XMR	-0.714	9.763	-1.592	8.977	-50.847	29.568	44.595
DOGE/ ETC	-0.841	9.270	-0.864	5.230	-38.321	26.761	60.811
DOT/ AVAX	-1.358	9.486	-1.499	6.496	-47.640	23.933	47.297
DOT/ ALGO	-0.876	9.088	-1.282	5.461	-43.288	22.168	50.000
AVAX/ ALGO	-0.965	8.468	-1.589	6.622	-42.627	19.971	51.351
TRX/ VET	-1.033	8.701	-1.232	5.650	-40.916	23.292	47.297
MATIC/ XMR	-0.495	10.082	-0.913	6.478	-47.860	30.799	45.946

Descriptions for the summary stats can be found in Table 5

Table 7: Results for the Second Period based on Daily Data: 2021-10-06 to 2021-12-19

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs < 0	Trades
BNB/ XRP	0.299	0.953	-0.321	0.095	-2.204	2.227	33.333	8
BNB/ TRX	0.080	1.233	-1.988	6.305	-4.951	2.126	38.235	8
BNB/ VET	0.097	1.583	0.054	1.083	-4.642	3.463	56.410	12
AVAX/ ALGO	-0.393	1.441	-0.457	-0.553	-3.767	1.938	56.667	12
LTC/ XLM	-0.043	1.825	-0.048	2.442	-5.687	5.561	51.852	12
NEAR/ ALGO	-0.213	2.122	-0.648	1.070	-6.415	3.885	46.154	8
NEAR/ HBAR	-0.334	2.154	-0.455	-0.077	-5.267	4.246	47.619	8
BCH/ VET	0.406	1.637	0.031	1.078	-3.235	4.705	48.000	8
ATOM/ ALGO	-0.252	1.784	-0.621	2.222	-5.547	4.920	51.111	12
ALGO/ HBAR	0.075	1.246	0.977	2.898	-2.634	4.586	48.718	8
XMR/ VET	0.374	1.555	-0.091	0.890	-3.261	3.928	31.250	4

Descriptions for the summary stats can be found in Table 5

Table 8: Results for the Buy-and-Hold Strategy for the Second Period for the Strategy based on Daily Data: 2021-10-06 to 2021-12-19

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs < 0
BNB/ XRP	-0.053	3.570	-0.531	0.076	-9.495	7.661	44.000
BNB/ TRX	-0.007	3.607	-0.826	0.506	-9.901	7.087	41.333
BNB/ VET	-0.110	4.379	-0.538	-0.064	-10.504	8.297	48.000
AVAX/ ALGO	0.097	4.973	-0.612	0.200	-13.060	9.950	44.000
LTC/ XLM	-0.237	4.717	-0.452	0.316	-13.202	11.122	45.333
NEAR/ ALGO	-0.162	5.274	-0.018	1.994	-15.086	18.537	49.333
NEAR/ HBAR	-0.065	5.550	0.012	0.437	-13.753	15.128	48.000
BCH/ VET	-0.442	4.415	-0.858	0.586	-11.760	9.411	48.000
ATOM/ ALGO	-0.580	4.661	-0.833	0.853	-14.354	9.715	45.333
ALGO/ HBAR	-0.389	4.808	-0.446	0.380	-13.857	11.470	50.667
XMR/ VET	-0.488	4.221	-0.932	0.519	-11.466	7.856	45.333

Descriptions for the summary stats can be found in Table 5

Table 9: Results for the Third Period based on Daily Data: 2022-02-28 to 2022-05-13

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs < 0	Trades
XRP/ BCH	0.150	0.718	1.553	2.989	-0.859	2.526	44.000	8
XRP/ XMR	-0.101	1.333	-0.590	2.192	-4.426	3.050	46.341	8
XRP/ ETC	0.136	1.508	0.641	1.789	-3.610	4.424	47.500	20
DOGE/ ETC	0.112	1.910	0.113	2.076	-5.064	5.107	50.000	8
LTC/ VET	0.346	1.661	1.057	0.516	-1.713	4.691	50.000	8

Descriptions for the summary stats can be found in Table 5

Table 10: Results for the Buy-and-Hold Strategy for the Third Period for the Strategy based on Daily Data: 2022-02-28 to 2022-05-13

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs < 0
XRP/ BCH	-0.763	4.545	-1.283	3.540	-18.940	6.853	59.459
XRP/ XMR	-0.528	4.739	-1.057	3.816	-18.689	12.688	51.351
XRP/ ETC	-0.683	5.173	-0.649	1.800	-19.272	11.051	56.757
DOGE/ ETC	-0.550	5.664	-1.000	2.433	-20.734	11.488	54.054
LTC/ VET	-0.708	5.597	-1.056	2.276	-21.411	10.426	50.000

Descriptions for the summary stats can be found in Table 5

Table 11: Results for the First Period for the Strategy based on Hourly Data: 2021-05-12 to 2021-07-25

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs < 0	Trades
ETH/ BCH	-0.651	2.849	-1.822	18.449	-5.492	2.383	52.768	164
BNB/ SOL	-0.066	2.428	-0.467	6.121	-2.920	2.431	50.685	188
ADA/ ATOM	-0.586	3.089	-1.836	13.013	-4.822	2.800	52.408	152
SOL/ MATIC	0.044	3.982	-0.316	9.302	-5.637	4.972	49.324	184
SOL/ XMR	-0.050	3.335	-1.688	14.557	-6.451	2.597	48.872	176
SOL/ VET	0.229	3.396	-1.176	10.435	-4.995	3.609	48.027	188
DOT/ NEAR	-0.013	2.708	-0.508	4.505	-3.147	2.376	49.095	188
AVAX/ ATOM	-0.934	3.489	-1.954	15.193	-5.739	3.042	50.000	176
AVAX/ ALGO	-0.698	3.130	-1.943	14.139	-5.260	2.815	48.634	184
ATOM/ ALGO	-0.113	2.745	0.053	3.212	-2.483	2.567	50.178	200

Descriptions for the summary stats can be found in Table 5

Table 12: Results for the Buy-and-Hold Strategy for the First Period for the Strategy based on Hourly Data: 2021-05-12 to 2021-07-25

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs < 0
ETH/ BCH	-1.304	7.710	-2.010	24.235	-21.539	7.676	50.394
BNB/ SOL	-0.868	8.989	-0.898	10.463	-18.814	11.497	49.324
ADA/ ATOM	-0.859	9.144	-1.758	33.704	-27.876	16.157	49.437
SOL/ MATIC	-0.408	10.632	-0.284	8.811	-19.516	15.105	50.957
SOL/ XMR	-0.858	9.524	-1.804	32.924	-28.666	17.730	48.649
SOL/ VET	-1.069	10.115	-0.841	13.667	-23.258	14.088	49.718
DOT/ NEAR	-1.270	9.868	-3.053	45.675	-33.514	9.216	49.718
AVAX/ ATOM	-1.430	9.593	-1.966	29.970	-28.698	14.564	48.930
AVAX/ ALGO	-1.236	8.708	-1.581	21.889	-23.625	13.521	48.930
ATOM/ ALGO	-0.996	9.024	-1.622	30.273	-26.421	17.262	48.592

Descriptions for the summary stats can be found in Table 5

Table 13: Results for the Second Period based on Hourly Data: 2021-10-06 to 2021-12-19

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs < 0	Trades
BNB/ XRP	0.111	1.497	-0.453	7.977	-2.327	1.683	49.277	192
BNB/ VET	-0.089	1.680	-0.137	3.351	-1.449	1.638	50.529	156
AVAX/ ATOM	0.439	2.378	0.037	2.972	-2.181	2.486	49.259	192
LTC/ XLM	-0.016	1.868	-0.398	5.762	-2.001	1.847	49.873	184
NEAR/ ALGO	-0.378	2.272	0.547	8.729	-1.837	3.954	50.490	180
NEAR/ HBAR	-0.166	2.230	-0.217	10.442	-3.004	3.750	48.370	168
BCH/ XMR	-0.327	1.326	-0.167	4.304	-1.515	1.361	53.180	188
ALGO/ HBAR	-0.082	2.291	1.083	14.166	-2.085	4.552	48.285	164
XMR/ VET	0.108	1.744	0.070	3.376	-1.624	1.763	49.942	196

Descriptions for the summary stats can be found in Table 5

Table 14: Results for the Buy-and-Hold Strategy for the Second Period for the Strategy based on Hourly Data: 2021-10-06 to 2021-12-19

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs < 0
BNB/ XRP	-0.052	4.027	-0.652	7.134	-5.883	5.370	47.748
BNB/ VET	-0.078	4.721	-0.263	4.139	-5.581	5.888	48.761
AVAX/ ATOM	0.080	5.912	-0.209	2.619	-6.793	5.597	48.423
LTC/ XLM	-0.256	4.878	-1.528	18.069	-12.713	4.053	48.198
NEAR/ ALGO	-0.169	5.718	0.083	7.833	-9.099	9.545	48.705
NEAR/ HBAR	-0.130	5.632	-0.042	4.996	-8.104	7.501	48.705
BCH/ XMR	-0.491	3.902	-1.648	22.289	-11.025	3.376	50.113
ALGO/ HBAR	-0.430	5.463	-0.192	6.419	-8.016	9.104	49.381
XMR/ VET	-0.480	4.739	-0.823	7.915	-9.802	3.973	50.282

Descriptions for the summary stats can be found in Table 5

Table 15: Results for the Third Period for the Strategy based on Hourly Data: 2022-02-28 to 2022-05-13

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs < 0	Trades
XRP/ DOGE	0.082	1.545	-2.643	48.285	-4.400	1.745	50.778	176
XRP/ BCH	-0.489	1.735	-0.746	33.557	-3.877	3.588	54.457	184
XRP/ XMR	-0.242	1.584	0.690	15.906	-2.151	3.050	49.941	188
XRP/ ETC	0.089	1.434	1.063	6.259	-1.395	1.906	52.283	192
DOGE/ XLM	-0.141	1.786	-0.019	14.394	-2.591	3.233	49.356	196
DOGE/ XMR	-0.514	1.518	-0.154	4.412	-1.654	1.863	50.075	164
DOGE/ ETC	-0.114	2.449	-1.483	44.006	-6.593	5.096	50.772	188
LTC/ VET	-0.196	1.561	0.095	5.326	-1.372	1.946	48.264	192
BCH/ XMR	0.369	1.382	1.211	13.644	-1.249	2.806	46.782	188
XLM/ VET	-0.172	1.627	-0.004	5.204	-1.608	1.977	51.007	152
ALGO/ HBAR	-0.426	2.476	-2.777	110.748	-8.359	6.625	48.683	180
XMR/ ETC	-0.060	1.886	0.008	4.558	-1.975	1.854	51.839	176

Descriptions for the summary stats can be found in Table 5

Table 16: Results for the Buy-and-Hold Strategy for the Third Period for the Strategy based on Hourly Data: 2022-02-28 to 2022-05-13

Pair	Mean	Std Dev	Skewness	Kurtosis	Min	Max	% Obs < 0
XRP/ DOGE	-0.655	5.029	-1.193	43.961	-15.412	11.680	49.803
XRP/ BCH	-0.681	4.588	-0.830	33.564	-12.558	10.378	49.859
XRP/ XMR	-0.463	5.025	-0.193	33.054	-12.681	12.859	48.958
XRP/ ETC	-0.644	5.120	-0.783	24.784	-13.332	10.337	50.366
DOGE/ XLM	-0.499	4.891	-0.796	28.401	-13.365	10.225	49.183
DOGE/ XMR	-0.293	5.374	0.095	24.849	-12.536	12.714	49.070
DOGE/ ETC	-0.475	5.457	-0.529	18.591	-13.187	10.191	49.972
LTC/ VET	-0.613	5.303	-0.720	24.728	-14.114	10.736	49.014
BCH/ XMR	-0.320	4.999	0.181	18.955	-9.682	11.411	49.070
XLM/ VET	-0.568	5.276	-0.566	18.340	-12.889	9.969	48.901
ALGO/ HBAR	-1.008	5.767	-1.034	35.776	-16.717	13.251	48.789
XMR/ ETC	-0.283	5.515	0.083	13.853	-10.456	11.370	50.479

Descriptions for the summary stats can be found in Table 5

References

Aslan, A. and A. Sensoy (2020). Intraday efficiency-frequency nexus in the cryptocurrency markets. *Finance Research Letters 35*, 101298.

Caldeira, J. and G. V. Moura (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. *Available at SSRN 2196391*.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association 74*(366a), 427–431.

Do, B. and R. Faff (2012). Are pairs trading profits robust to trading costs? *Journal of Financial Research 35*(2), 261–287.

Do, B., R. Faff, and K. Hamza (2006). A new approach to modeling and estimation for pairs trading. In *Proceedings of 2006 financial management association European conference*, Volume 1, pp. 87–99.

El-Oraby, K. (2019). *Similarity Approaches for High-Dimensional Financial Time Series-with an Application to Pairs Trading*. Ph. D. thesis.

Elliott, R. J., J. Van Der Hoek*, and W. P. Malcolm (2005). Pairs trading. *Quantitative Finance 5*(3), 271–276.

Engle, R. F. and C. W. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251–276.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance 25*(2), 383–417.

Fil, M. and L. Kristoufek (2020). Pairs trading in cryptocurrency markets. *IEEE Access 8*, 172644–172651.

Gatev, E., W. Goetzmann, and K. Rouwenorst (1998). Pairs trading: Performance ofa relative value arbitrage rule. Technical report, Working Papers ysm26, Yale School of Management.

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

Krauss, C. (2017). Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys 31*(2), 513–545.

Li, T. R., A. S. Chamrajnagar, X. R. Fong, N. R. Rizik, and F. Fu (2019). Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. *Frontiers in Physics 7*, 98.

Liu, J. and A. Timmermann (2013). Optimal convergence trade strategies. *The Review of Financial Studies 26*(4), 1048–1086.

Neusser, K. (2016). *Time series econometrics.* Springer.

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

Urquhart, A. (2016). The inefficiency of bitcoin. *Economics Letters 148*, 80–82.

Van den Broek, L. and Z. Sharif (2018). Cointegration-based pairs trading framework with application to the cryptocurrency market. *Bachelor Thesis, Erasmus University Rotterdam*.

Vidyamurthy, G. (2004). *Pairs Trading: quantitative methods and analysis*, Volume 217. John Wiley & Sons.

Wei, W. C. (2018). Liquidity and market efficiency in cryptocurrencies. *Economics Letters 168*, 21–24.