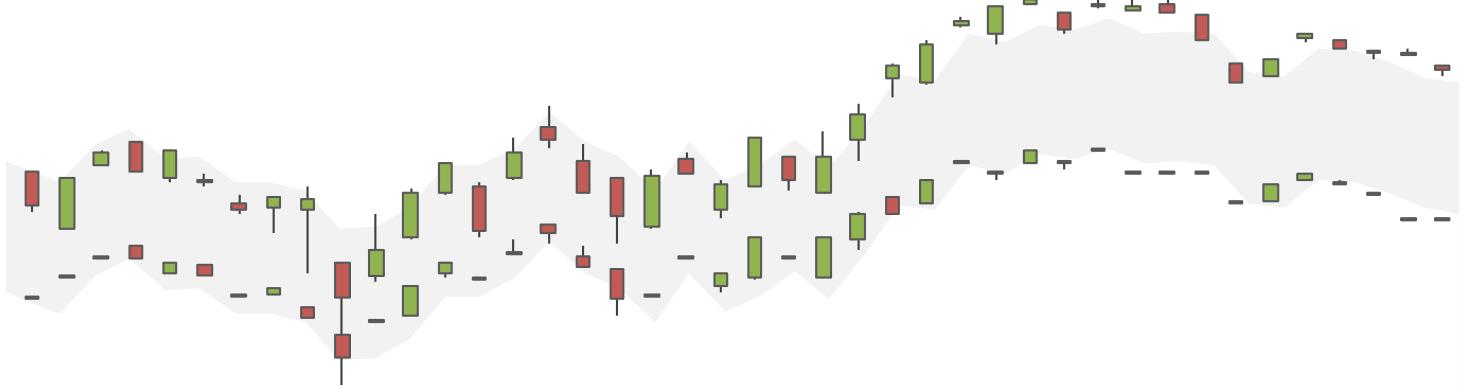

Master thesis

Pairs trading on ETFs

Backtested from 2007-2020



Alexander Ludvig Tommerup, 102495

Cand.merc. in Finance and Accounting

Rasmus Bruun Jørgensen, 102199

Cand.merc. in Applied Economics and Finance

Submission: 15.05.2020

Supervised by: Martin Richter

Characters: 254,746

Pages: 112

Abstract

Pairs trading is a type of statistical arbitrage strategy created to exploit relative mispricings in the price development between two securities. The trading strategy has primarily been conducted on single stocks which have showcased declining profitability in recent years.

This paper examines the application of a pairs trading strategy when substituting the traded securities from single stocks to exchange-traded funds (ETFs). Inferences of this modification is provided through an empirical backtest of the period from 2007 to 2020 by applying the two most prominent methods within the field of pairs trading, namely the distance method and the cointegration method.

The findings of this paper reveal that pairs trading with the use of ETFs provides a more efficient alternative to the utilisation of single stocks when executed with the cointegration method. At the same time, it is demonstrated that a pairs trading strategy based on ETFs differs substantially depending on the parameterisations and method applied. The cointegration method generates robust and reliable trading attributes and statistically significant alpha for five out of six trigger settings tested. On the other hand, the distance method does not provide any statistically significant alpha, a consequence of not providing returns robust to transaction costs, but only generating net profit in highly volatile periods. This has shown to be the case as the pairs composition of the distance method primarily rely on pairs of ETFs tracking the same index, which is associated with realised losses on the traded positions after transaction costs.

Table of Contents

1. Introduction.....	3
2. Philosophy of science	5
3. Methodology	12
4. Literature review on pairs trading.....	19
4.1. Academic literature on pairs trading methodology.....	19
5. Pairs trading and ETFs	25
5.1. Exchange-traded funds	25
5.2. Pairs trading	28
6. Applied pairs trading strategy	35
6.1. Setting up a pairs trading strategy	35
6.2. Formation period.....	38
6.3. Trading period.....	45
6.4. Computation of results	48
6.5. Summary of analytical framework.....	53
7. Empirical results.....	53
7.1. Results of the entire period.....	55
7.2. Determinants of returns	62
7.3. Sub period analysis.....	69
7.4. Factor models.....	92
7.5. Strategy assessment	103
8. Discussion.....	106
9. Further research	109
10. Conclusion	111
11. References.....	113
12. Appendices.....	122

1. Introduction

Pairs trading has long been a popular statistical arbitrage strategy utilised by hedge funds and investment banks ever since Nunzio Tartaglia and his team first implemented the method with Morgan Stanley in the mid-1980s (Gatev, Goetzmann and Rouwenhorst, 1999; Broussard and Vaihkoski, 2012; Vidyamurthy, 2004). The fundamental concept of pairs trading is to identify securities exhibiting similar price development in order to exploit any discrepancies in their relation, with the expectation that this will restore said relation. It was not until 1999 that pairs trading was introduced in the academic literature by Gatev et al. (1999), since then being perceived as the benchmark approach of pairs trading. From this point on, many authors have both tested, verified and further developed the distance method proposed by Gatev et al. (1999). In 2004 a new methodology to execute pairs trading arose when Vidyamurthy (2004) introduced the cointegration method. Whereas the distance method was inspired by Wall Street, Vidyamurthy (2004) introduced a statistically founded model for pairs trading (Gatev et al., 1999). Both methods have since been the two general practices of pairs trading from which other submethods have derived (Krauss, 2015).

Six years after Tartaglia and his team introduced the first pairs trading strategy, the first exchange-traded fund (ETF), namely the SPDR S&P 500 ETF Trust (SPY), was launched by State Street Global Advisors. This new investment vehicle combined the attributes of open- and closed-end funds and thus introduced cost-efficient diversification enabled for intraday trading (Munk, 2018). Since then, ETFs have gained increasing popularity amongst both private and institutional investors and reached new heights ultimo 2019 with a global asset under management of USD 6 Trillion (Kwillia, 2019). The growing popularity of ETFs as an investment vehicle and its unique trading attributes left us with the question whether it would be possible to increase the efficiency of a pairs trading strategy by changing the fundamental idea of which the strategy is founded. This paper thus strives to investigate the performance of a pairs trading strategy by substituting the underlying securities from single stocks with ETFs. With this, it is the expectation that the strategy will yield more stable returns with a reduced amount implied risk; hence the derived effect is a strategy with a better risk-return relation. The presumption is based on the expectation of the synergies obtained by

combining a market-neutral trading strategy with a diversified investment vehicle. To examine whether our expectations are met, we apply the two different methods outlined above, which look to carry out the strategy with different parameters.

Identification of problem area | Our knowledge of interest is based on the findings of the current literature, which has showcased the profitability of pairs trading on single stocks disappearing in recent years (Broussard and Vaihekoski, 2012; Do and Faff, 2010; 2012; Smith and Xu, 2017; Rad et al., 2015). Further, there has only been given limited attention to pairs trading with the use of securities other than single stock. We, therefore, seek to uncover how a pairs trading strategy would have performed from 2007 to 2020 when traded with ETFs. This paper will consequently be structured based on answering the following research question:

**What inferences can be made from backtesting a pairs trading strategy on
US equity ETFs in the period of 2007 to 2020?**

Here *backtesting* is defined as the simulation of a trading strategy during a historical time period (Pedersen, 2015). *Inference* should be understood in the light of the abductive reasoning of pragmatism, which is “*the best explanation*” (Douven, 2017, p. 1); “Best” being founded on empirical, deductive reasoning, and “explanation” as imply something through inductive reasoning (Egholm, 2014).

Intending to uncover the above research question this paper is structured as follows: The first part of the paper covers the perspective and the methodical considerations for which this paper is grounded; this relates to how the paper is established and how it must be interpreted. The second part provides an overview of current literature within the field of pairs trading, with a focus on the most acknowledged papers. The third part will shed light on the fundamental mechanisms of pairs trading and how we apply the strategy, including the reasoning for the choices of ours and how it relates to current literature. The fourth part will synthesise our empirical results by investigating the performance of the strategies and the parameterisation. The last part of this paper

puts forth a discussion for whom this paper might be relevant, which leads to our suggestions of topics for further research.

2. Philosophy of science

This first chapter of the paper has the objective to shed light on the perspective of which this paper is written, and the knowledge it seeks to provide. The essentials of this section thus lies in defining how the reader must interpret the paper and its findings. On the notion of philosophy of science, this paper takes a pragmatic stance as the general practice of pairs trading theory has a pragmatic vision and the aim is to conduct a practical backtest.

Pragmatism as a philosophy of science | First, the history of pragmatism must be emphasized to understand the fundamental ideas of pragmatism. The philosophical movement of pragmatism originates from a discussion club at Cambridge University comprising several philosophers from different schools and movements of philosophy. Amongst these were Charles Peirce and William James, who is considered the founding authors of pragmatism (Egholm, 2014; Kaushik and Walsh, 2019). In their discussion clubs, whenever the debaters were not able to agree on philosophical disputes that seemed dissolvable, William James would settle this with a pragmatic approach (Legg and Hookway, 2019). Rather than looking at the philosophical or metaphysical considerations, James considered the practical consequences of the subject of discussion, which often meant that the opposing philosophical views found no conflict (Legg and Hookway, 2019). The practical consequences are crucial to understanding pragmatism, since the overall knowledge interest of pragmatism is to understand how past experiences are influencing and used in the actions of the present, and what consequences these actions are predicted to lead to (Egholm, 2014). Here, experiences being a broad term of knowledge, individual and social experiences, values, emotions, and praxis, where praxis refers to the iterative and reflective approach of taking action, based on enacting a theory in practice (IGI Global, n.d.; Nekrašas, 2001). Experiences thus represent the considerations of predicting the future and decide on how to act - both consciously and unconsciously. As pairs trading and trading, in general, depends on our

experiences, the concept of a broader definition of experiences makes sense to consider for understanding pairs trading. Actions are thus a dynamic process that depends on the present context and our expectations for the future (Egholm, 2014).

Above also means that learning takes place in the conjunction of our experiences and the present context, as it is necessary to learn from a changing environment to decide on new and more accurate predictions of the future practical consequences. Thus, the starting point of learning and acknowledging something new is our experiences (Egholm, 2014). Here, freedom to choose is not about following the rightful immediate impulse in a particular situation, but to master and consider the consequences that actions predictably will cause and after that act in the appropriate way (Egholm, 2014, p. 180). Mead (1934) describes the process of choosing between rational and irrational actions as a transaction between the “Me” and the “I”. The “Me” is the reflectatory self that considers social logics, rules, past experiences, others expectations towards himself, the actual situation and maintain habits, whereas the “I” is the immediate response to a situation or context (Mead, 1934). The “I”’s actions towards a situation is more or less unknown despite the “Me” defining how to act in the situation (Mead, 1934). Why are the practical consequences and with this the “Me” and the “I” important when discussing a pairs trading strategy? The nature of a statistical rule-based trading strategy infers that an investor should not trust the impulsive “I” to make trading opportunities. The founding father of pairs trading, Tartaglia, argues that pairs trading is a matter of psychology and that “human beings do not like to trade against human nature, which wants to buy stocks after they go up not down” (Gatev et al., 2006, p. 4). In this light, pairs traders are thus disciplined traders taking advantage of un-disciplined investors that create anomalies due to over- or underreaction to security prices (Gatev et al., 2006). A more disciplined trader knows better on how to predict the practical consequences of their actions by knowing when a situation or trade opportunity is similar to past experiences, or a situation requires a thorough investigation. If the latter is true, then one has to cognitively consider the situation and consider possible new explanations of the new and unknown situation (Egholm, 2014). The reasoning of such is carried out in three ways; authoritative reasoning, a priori reasoning, and abductive reasoning (Egholm, 2014). Authoritative reasoning is when an authority states what to believe is true or how to act - this is typically the case with religion. A priori

reasoning is using existing general practices to try to understand a new situation and reason for any similarities there might be. The abductive reasoning - also referred to as the experimental reasoning - is for situations where we cannot fully predict the practical consequences and thus need to conduct experimental tests to fully understand the situation and act after that (Egholm, 2014). It is essential to consider that abductive reasoning, and the results these experimental tests lead to, does not state a new objective truth nor does it seek to be able to comment on the objective truth about this situation or phenomenon (Egholm, 2014). Pragmatic abductive reasoning strives to reflect on new or unknown phenomena or situations and thus contribute new experiences to the existing belief system. It is in this light we conduct the backtest of applying ETFs on pairs trading. The way of reasoning is thus in contrast to the realistic notion of one universal truth. The different reasonings thereby showcase the pragmatic standpoint on truth (Egholm, 2014; Kaushik and Walsh, 2019).

Individuals base their understanding of what is true and the right way to act on their belief system, the conception of truth, and past experiences. For example, two traders will act in the way that they consider the most educated guess on what the practical consequences of a trade will be; one might believe the stock price will go up while the other might believe the stock price will go down. The reasoning behind each belief might be based on an authority (e.g. broker recommendations), earlier trading experiences, or an extensive analysis.

Even though this paper is not focusing on behavioural finance, the general pragmatism helps to understand the existence of pairs trading and the underlying premise of conducting a backtest. Besides, it provides us with an understanding that various belief systems and thus, models can exist simultaneously. The different methods of pairs trading are good examples of such general practices that exist simultaneously. As such, the individual must try to navigate these experiences and decide on the appropriate applicability in a particular situation.

Pragmatism in econometrics | Building upon the ideas of the classic pragmatism as defined above, Granger (2009) defines pragmatism in the light of econometric studies in a more casual way as the “*practical considerations as opposed to theoretical or idealistic ones*” (Granger, 2009 p. 2) meaning that a practical method is the opposite to

a dogmatic method. In one of the first issues of *Econometrica*, Schumpeter (1933) argued that in order for economics to provide positive advice for practitioners, it must be in the form of quantitative work and subsequent exact proofs. Based on this argument, the quantitative approach of this paper should help to consider the practical consequences of applying economic theories (Schumpeter, 1933).

Two key takeaways from Schumpeter's arguments can be made; i) there is a close link between econometricians and practical decision-makers and ii) the matter of exact proof (Granger, 2009). In the case of pairs trading, Gatev et al. (1999) conducted the first pairs trading paper in collaboration with practitioners building on their best practices. Regarding the notion about exact proof from Schumpeter (1933), exact proof is in the light of either a dogmatic or a pragmatic approach (Granger, 2009). As an example of the two approaches, consider the notion of forecasting. A traditional dogmatic synthesis action is going from theory to model to forecast, meaning applying theory to come up with an empirical model and use this for forecasting (Granger, 2009). However, such a model would often be deemed inadequate or suboptimal in subsequent evaluation of such a model for practical implementation (Granger, 2009). For making a model adequate to practical considerations, the dogmatic model is often then re-specified, and the author goes from a dogmatic to a pragmatic approach. An alternative and more pragmatic sequence is to consider a variety of models based on the adequate data, then form alternative forecast models and conclude various forecasts (Granger, 2009). As such, the pragmatic inference is not a single best method but a set of alternatives. In that way, Granger believes econometricians can provide more valuable information to the practical decision-makers (Granger, 2009).

Another consideration of dogmatic and pragmatic approaches to truth is the nature of the distance method of Gatev et al. (2006). Krauss (2015) argues that the use of a minimization of the sum of squared deviations is an analytical suboptimal. In a dogmatic view, the theoretical implications of the models might not necessarily be exact proof. However, Krauss (2015) also acknowledges that the distance method has been widely cited, acclaimed and used by both practitioners and academics, and further states that precisely the easy and nonparametric approach of Gatev et al. (2006) opened the doors for pairs trading to academic literature. This underlines that proof can also be defined otherwise. To this, Granger (2009) concludes that if a model is based on "*careful and*

sound statistical analysis" (Granger, 2009, p. 3) that also includes and considers practical implications, such a model could also be considered as exact proof.

This definition does not mean that all analyses, by definition, are true, but if a finding meets the requirements of validity, reliability and practicality, it can be considered true to the extent it helps predict the practical consequences and thus help on how to act in a situation. It does not mean that there are no single answers to a particular solution or that an equation is not always the same. If a situation might be solvable with a priori reasoning and general practices stated as a model or equation, there are no reasons for further considerations. However, often, when deciding on actions that shall yield a desired practical consequence, it is essential to consider different experiences, both dogmatic and pragmatic. Thus, even though practical and dogmatic approaches are different, a truly pragmatic approach includes both (Granger, 2009). For a pragmatic researcher to discard a general practice, theory or model, they must then consider whether there is a better alternative for solving the particular situation (Granger, 2009). It is in light of this that this paper considers two methods to carry out the pairs trading strategy.

Our position as a pragmatic researcher | As a pragmatic researcher, it is therefore important to act as a fallibilist, which means to accept and seek for anomalies in our existing knowledge in order to continuously falsify and test our belief system (Hetherington, 2020). In this paper, we take a pragmatic epistemological approach based on abductive reasoning. The abductive approach to research is a combination of inductive and deductive research that seeks to present possible explanations and understandings of new phenomena, in our case the application of pairs trading with ETFs (Egholm, 2014). Where deductive reasoning means to falsify whether an inference is true by moving from theory to results, inductive reasoning is moving from results to theory; considering the reality, using one's cognitive reasoning and postulating a probable theory. The abductive approach combines both ideas by generating one or more "best explanations" about the situation and new phenomena (Egholm, 2014). "Best" being founded on empirical, deductive reasoning and "explanation" as it implies something (Egholm, 2014). In other words, abductive reasoning is the "*inference of the best explanation*" (Douven, 2017, p. 1).

The abductive reasoning is reflected in the fundamental idea of backtesting and further reflected in the continuous iterations of inductive and deductive reasoning. We keep a fallibilist approach by continuously being critical to our belief system. We initially consider the fundamental ideas of efficient markets and the current literature within the space of pairs trading. As it is evident in the literature review in chapter 4, pairs trading has only been applied on ETFs to a limited extent, and the existing literature can therefore not fully explain the profitability nor the applicability of pairs trading on ETFs. We therefore believe that we can provide useful results on applying pairs trading to ETFs. However, in order for us to infer the best explanation of the practical consequences of using ETFs and not only presume, we conduct both inductive, deductive and empirical tests based on our presumptions. In the light of pragmatism as defined by Granger (2009), we included different models and theories and considered the findings in the light of the realistic quality criteria of science. As pragmatism and the realistic movements of both positivism and critical rationalism are close, we argue that these quality criteria are valuable considerations for verification of our research - more on these quality criteria in chapter 3.

The aim of this paper is thus to provide useful learnings from the backtest to help investors predict the future practical consequences of using pairs trading on ETFs and in that way on how to act (Egholm, 2014).

Review and criticism of our philosophical approach | Our knowledge of interest has been to uncover the practical implications of applying pairs trading to ETFs, thus seemingly making the pragmatic philosophy of science the appropriate approach for our specific research question. However, the more severe consequence of taking the pragmatic stance as the philosophy of science is that our results can never be concluded as definitive or absolute. Using pragmatism might also beg the question for whom and to what degree our results are useful, despite us arguing that they might (Egholm, 2014). For us to take a more general perspective, other realistic philosophies of science such as positivism or critical rationalism could be considered useful (Egholm, 2014). Positivism and pragmatism have many similarities, and the history of both movements sprung out of similar discussion clubs and similar anti-dogmatic and anti-metaphysical views (Nekrašas, 2001). The theories that positivism and pragmatism propose are

also very similar. Pragmatism finds great importance in the practical considerations of concepts and practices, while positivism seeks what is useful and practical (Nekrašas, 2001). However, two key differences separated the two; the width of the meaning experience and the epistemological approach (Nekrašas, 2001). In the sense of experiences, positivism refers to observations and not experiences, thus stating a more narrow approach than pragmatism. However, more modern movements such as logical positivism resemble more pragmatism than the classic positivism. Logical positivism was more interested in concepts and propositions and not objects and observations, and preferred to discuss “observable consequences” (Nekrašas, 2001). Arguably this concept is close to the concept of practical consequences. As such, the ideas of positivism and pragmatism resemble each other significantly in their fundamental way of thinking. However, positivism uses an inductive epistemological approach, meaning what is observable in reality is considered true. For example, if you see 100 white swans, the positivistic researcher would conclude that all swans are white - using what is observed to postulate theories (Egholm, 2014). As such, positivism is criticized for being a rather naive approach to reasoning. Pragmatism also uses this ampliative way of thinking, however, in the abductive reasoning there are also both implicit and explicit thoughts to the explanatory considerations such as; can it be true that all swans are white? In induction, there is only an appeal to the observed frequencies or statistics, and herefrom generalize the results (Douven, 2017). Observed frequencies and statistics are, of course, also included in abduction but not as the only notion (Douven, 2017). Opposite of the inductive approach is the critical rationalism that builds upon the same principles of positivism, however reason with deduction instead. Deductive research seeks to come up with a logical conclusion based on a statement or a hypothesis. However, it is possible to come up with a logical conclusion that cannot necessarily be generalized if the statement is not true (Bradford, 2017). Again, it must be emphasized that the goal of the deductive reasoning is also the postulate of causal connections and generalizations (Douven, 2017).

Arguably, there are pros and cons of both positivism and critical rationalism, and despite the two movements being similar, the reasoning of inference is different. We argue that the abductive approach as a combination of both is a useful compromise

between the two movements when we are not interested in generalizations or causal connections of our findings.

All in all, we believe the pragmatic approach is the most useful philosophy of science for our research question.

3. Methodology

In order to answer the research question of this paper, a quantitative study has been conducted based on the application of statistical models on economic data, primarily in the form of time series of ETF closing prices. The application of statistical models to analyse economic data is also referred to as econometrics (Stock & Watson, 2015). The argument for doing so is highlighted in the section of the philosophy of science; a quantitative, econometric study is considered a useful research design for economics to give positive advice to practical decision-makers (Schumpeter, 1933). On a more overall note, quantitative research is the practice of quantifying problems through processing data that can be perceived as useful statistics.

Literature and limitations | The statistical models applied in our research derive from different sources of literature of both normative and theoretical character. The relevant literature is collected from various journals, books and other qualitative sources through the available databases of Copenhagen Business School. We have in the process of collecting relevant literature emphasised that academic papers must either be published in a relevant journal and thus have been peer-reviewed or cited or referred to by a published article or book.

However, we have made two exceptions of Schizas, Thomakos and Wang (2011) and Rudy and Dunis (2010) as these are some of the few papers written on pairs trading using ETFs. We take a critical position towards the findings of Schizas et al. (2011) and Rudy and Dunis (2010) and do not use their statistical models, but consider some of the elements of their respective research. For a thorough examination of the literature considered in this paper, see the literature review in chapter 4.

Data and limitations | Our primary data collection comprises time-series of daily US-based ETFs closing prices. The following section will touch upon our applied sample and our considerations about the limitations and sorting of our sample.

The first step of collecting the applied primary data was to identify potential ETFs. We limited our search to ETFs listed in the USA and US dollars, but did not limit the geographical location of the benchmark index; meaning that the ETFs could track all potential indices in the world and thus not limited to the American market. This preliminary sorting gave us a sample of 2,303 ETFs indicated by table 1.

Table 1: Applied filters for the sample selection

Type	Filter	# ETFs
Primary listing country	USA	2,303
Type of ETF	Equity ETF	1,348
Inception date	< 01-01-2019	1,088
Expense ratio	< 1%	1,069
Last year avg. volume	< 100,000	325

Source: www.etfdb.com/etfs/

The reason for limiting our sample to US-listed ETFs was to ensure comparable trading days and eliminate the exposure to currency trends and with this exchange rate risk. We are aware of the implied delimitations this has on the applied data as it reduces the possible combinations of potential tradable pairs. However, the risks of including currency effects in our results and having non-eligible trading days and hours for one of two ETFs in a pair are deemed more problematic than the positive implications it would have yielded otherwise.

Next, we chose to limit our sample to only considering equity ETFs, as the scope of this paper is to focus on equity ETFs. The current literature primarily considers stocks, and as we consider portfolios of stocks by utilising equity ETFs, we do not diverge significantly from the concepts of the existing literature. We, therefore, exclude both bond ETFs and alternative ETFs; such as inverse and leveraged ETFs. We are aware that this also delimits possible combinations. However, it also gives us the possibility to consider and postulate the best explanations about this one asset class. After filtering for equity ETFs only, we get a sample of 1,348 ETFs. After that, we filtered out ETFs with an inception date later than 01-01-2019, as we need 12 months of data for the

formation period to determine the optimal pairs combinations. This gives us a sample of 1,088 ETFs.

Lastly, we filter for both low expense ratio (<1%) and an average daily volume over the last twelve month of more than 100,000 USD. The expense ratio filter seeks to delimit our sample from costly ETFs. Filtering for average daily volume is done to ensure that the ETFs applied comprise a sound level of liquidity. As a pairs trading strategy requires more trades than a long-hold position, liquidity of the included securities in pairs are important as the timing of market entry and exit is essential. Therefore, low levels of liquidity could potentially compromise the execution of the strategy and be a costly trade. Thus, the paper has arbitrarily chosen the level of volume mentioned above. After the above filters, the total number of ETFs is 325.

Data extraction | For the above mentioned ETFs, we have extracted daily dividend-adjusted close prices from the 1st of January 2006 to the 1st of April 2020 using the Capital IQ database. The time period from 2006 to 2020 is chosen to reflect a full economic cycle, including both bear and bull markets and times with high and low macroeconomic growth rates. New ETFs are added after at least six months of trading - starting either the 1st of July or the 1st of January. The waiting time of six months before being included in the model is to mitigate any spurious volatility or returns after the ETF IPO. Here, Rompotis (2019) finds that there are abnormal returns for the first six months after the IPOs of ETFs. Therefore, including the ETF right after the inception date might include misleading price developments in the formation period. Thus, we wait six months, meaning that in the last formation period, we have 322 ETFs. Close prices have been validated by a comparison to prices of Yahoo Finance and Eikon Datastream. The 322 ETFs conclude a number of different sectors and areas of focus; i) international ETFs ($n = 103$), ii) US Large Cap ETFs ($n = 81$), US Industry-specific ETFs ($n = 68$), Small Cap ETFs ($n = 31$) and other sectors ($n = 42$). The data sample used in this paper thus comprise a wide range of different ETFs (See appendix 1 for a full list of the included ETFs and appendix 2 for extracted prices).

Other supporting extracted data elements include bid and ask prices, expense ratios, interest rates, last twelve-month daily trading volume and data for factor models. The data for all factors in the factor models are derived from the WRDS database (WRDS,

n.d.). The data for the liquidity factor was only available as monthly data points; hence we decided to run all of the factors on a monthly basis. Bid and ask prices and volumes were also extracted from Capital IQ. Expense ratios were extracted from etfdb.com, and the interest rates of the money market rate and the LIBOR overnight rate were collected from OECD (OECD, 2020) and the Federal reverse bank of St Louis (FRED, 2020: etfdb.com, 2020). Extracting bid and ask prices were problematic for several reasons. Firstly, the first quarter of 2007 had missing data points for the majority of the ETFs in Capital IQ. Secondly, we experienced several data points throughout the time series that had either a value of zero or the calculated spread would generate a negative value. As this should not be possible in a bid-ask spread, we decided to apply a five trading days rolling average bid-ask spread throughout the period. Arguably, it would have been preferred to have had the adequate data available and not to an adjustment of such, but we would argue that this is the best practical solution. We raised the problem with Capital IQ with the reply being that they were aware of the problem and had initiated a project to solve these issues, with no estimated time of completion (see appendix 3). Therefore, we tried to extract the data from Eikon Datastream, however the data was also faulty. For the first quarter of 2007, we have therefore decided to use the average bid-ask-spread for the second quarter of 2007 to fill out the missing data points.

Data processing | In the process of interpreting and processing the data, we have used Microsoft Excel as our primary tool. Further, we use Python and subsequent statistical modules for the computation of the cointegration tests. For the cointegration tests, we used the statsmodels Python module that provides functions of the estimation of various statistical models (Statsmodels, n.d.). Here, we used their OLS function (Statsmodels b, n.d.; Prettenhofer, 2014; Statsmodels c, n.d.), Augmented Dickey-Fuller test (Statsmodels d, n.d.), VAR (Statsmodels e, n.d.) and Johansen test (Statsmodels f, n.d.; Statsmodels g, n.d.). The results of the statistical models have been verified to other existing statistical models (Statsmodels, n.d.). See appendix 4 for our applied code and ranking of cointegration pairs.

The excel files are located in the appendices and description of the different files can be found in chapter 12. Here, appendix 5 shows the ranking of pairs using the distance method, and appendix 6 and 7 are the applied trading sheets for the two methods.

Assessment criteria | Having presented our methodological considerations and our philosophy of science, we find it essential also to consider the assessment criteria of a research paper. In general, all researchers, more or less, agree that research would be meaningless if there is no way to assess the quality of research and the knowledge it provides. For our paper, we consider the criteria stated by realistic philosophical approaches of positivism, critical rationalism and realism, as pragmatism arguably has noticeable similarities in its reasoning as discussed in the philosophy of science. We argue that the following considerations are important to consider when the reader assesses whether this paper is considered generalisable, valid, reliable, transparent, coherent and consistent (Justesen & Mik-Meyer, 2010).

Generalizability | As touched upon in the philosophy of science, the aim of this paper or the aim of pragmatism, in general, is not to generalise or postulate certain or absolute truths. Therefore, the criteria of generalizability in the light of realism should not be considered. However, this does not mean that our results have no value or cannot provide us with new knowledge, but rather that generalizability should be considered in another light. Instead of the generalizability of realism, we take the definition of analytical generalizability (Kvale and Brinkmann, 2015). The criteria of analytical generalizability refer to the process of determining whether a well-considered assessment and analysis can be indicative on how to act or understand a certain phenomenon or situation (Justesen and Mik-Meyer, 2010). This is consistent with the notion of truth in pragmatism, the idea of backtesting and consistent with the ideas of Granger (2009) that a model based on “careful and sound statistical analysis” would be considered as exact proof (Granger, 2009, p. 3). We believe that this paper can provide useful knowledge and learnings; however, such an assessment is dependent on the reader and not for us to conclude.

Coherence and consistency | Coherence refers to how well the components - philosophy of science, methodology, theory, analysis, discussion and conclusion - cohere in the study. Coherence is closely related to the notion of consistency, meaning that concepts and theories are well-described and used in a consistent manner (Justesen and Mik-Meyer, 2010). We argue that our study comprises both a high degree of coherence and consistency. The pragmatic approach is consistently evident throughout our study, as well as the applied methodologies. The applied methods and concepts are well-defined and consistent with the current literature, and the applied methodology of this paper has been screened in the literature review.

Reliability | Reliability refers to the notion of whether it would be possible to conclude the same findings using the same methodology and data, if the study was conducted by another researcher (Justesen and Mik-Meyer, 2010). As such, reliability is also strongly related to transparency.

Reliability is thus a matter of the degree the study is free from the researcher's personal beliefs, values and opinions, and to what degree biases and subjectivity have been minimised. Here, bias means "systematic errors in data collection or analysis, caused by inadequate technical procedures" (Payne and Payne, 2011). For quantitative studies, reliability is thus an important consideration as these matters can potentially influence the findings and knowledge that the study claims to provide. We have in our paper attempted to be as transparent in our choices, considerations and arguments as possible, in order for other researchers to understand and consider our study in a useful way. In such a way, we have tried to be as objective as possible towards both data collection and analysis. However, we cannot fully segregate our view of the reality from the view on the reality of this study; thus the idealistic notion about the researcher being completely uninfluential on the study might not be practically feasible. However, this does not imply that we disregard the notion of reliability, as we still strive to provide as reliable and objective results as possible.

Concerning the notion about biases, some key areas when conducting a backtest must be considered. Sample bias refers to the notion of whether we randomly have chosen our ETFs in the sample or deliberately selected specific ETFs with certain characteristics (Payne and Payne, 2011). We argue that the sample selection is not deliberately

selected, and the entire process is transparently described in the section “Data and limitations” for readers to see. It might be that other researchers do not agree on our sorting or applied filters, but that does not mean that our sample is not reliable.

Another bias that has been considered by numerous authors in pairs trading is the data-snooping bias (Gatev et al., 2006; Andrade et al., 2005; Smith and Xu, 2019). Data snooping refers to the incongruous inference of data mining to conclude misleading results or relationships in a dataset. The bias arises when data processing and analysis is exposed to an excess amount of parameterisations in order to force certain results. Thus, data snooping also includes when the researcher infer to the analysis that he will perform after looking at the dataset. For every additional parameter applied to a statistical model, the possibility that the statistical inference is based on random results increases (Lo, 1994). Gatev et al. (2006) argues that the reason for their rather limited test of different parameters is due to the risk of data snooping. However, Smith and Xu (2017) concludes that more considerations should be given of the parameterisation of pairs trading methods. The authors understand the reasons for not doing so in the light of the risk of data-snooping, however, argues that hedge funds and other institutions that initially applied pairs trading had most likely conducted thorough analyses and investigations of the optimal parameters. However, these considerations beg the question of the adequate amount of parameterisation tests without being exposed to data-snooping. For this study, we have pre-planned the parameterisations and analyses before analysing and inferring on the empirical results. Here, chapter 6 highlight all the various components of the pairs trading methodologies that have been considered before applying the methodologies to the data. It is on this basis that we deliberately have chosen to investigate some parameters further as these were deemed necessary to understand the characteristics and performance of a pairs trading strategy based on ETFs. These parameters were also chosen before looking at the data and based on the processes of the existing literature. We are thus aware of the risk of data-snooping, however as our process is transparent and every step has a related argumentation, it is argued that this paper remains free from data-snooping. The transparency and idealistic approach of maximum objectivity of this paper enables this study to be considered reliable.

Validity | Validity refers to what degree the study answers the research question; meaning do the researchers conclude what they say they conclude (Justesen and Mik-Meyer, 2004). The validity of a study depends on the philosophy of science, as validity might vary from philosophy to philosophy. The validity of quantitative studies are closely interlinked to reliability, as a study can state a valid answer to a research question and considered valid. However, it would have no knowledgeable use if the answer is not reliable, coherent or consistent. In the light of our pragmatic approach, we would argue that our study is valid in providing useful knowledge about the profitability and applicability of using ETFs on a pairs trading strategy, as it is considered reliable, coherent, consistent and based on “careful and sound statistical analysis” that does not seek to make definitive generalisations (Granger, 2009, p. 3). We have further verified our results by comparing our results to the existing literature, thus attempting to verify whether our results and findings have been noticeably different. However, on a final remark and as described throughout this section, the overall assessment is dependent on the reader and not for us to conclude upon - we can only state our choices and considerations.

4. Literature review on pairs trading

In the following chapter, the paper will examine the existing literature on pairs trading. The section will touch upon both relevant themes, common positions, potential shortcomings and different literate standpoints. The review aims to synthesize relevant methodologies for our paper and potential gaps in the literature. Therefore, the section will take a broad perspective on the literature on pairs trading. The in-depth description of the applied methodologies and theories will be considered later in this paper.

4.1. Academic literature on pairs trading methodology

Despite pair trading being a common trading strategy by investors and hedge funds, the academic literature is rather limited. The first academic paper to introduce pairs trading was authored by Gatev et al. (1999). The paper was revised in 2006 with an updated sample and further considerations, and is today still one of the most cited

papers on pairs trading (Gatev et al., 2006). Gatev et al. (2006) set forth the baseline framework of pairs trading based on a formation period and a trading period. In the formation period, the trader identified pairs that showcased similar price movements, and the best pairs were then traded in the subsequent trading period based on a fixed trading threshold (Gatev et al., 2006). This method has also been defined as the distance method, and the structure of a pairs trading strategy outlined in this paper has been the benchmark approach since then (Krauss, 2015). The authors tested the strategy from 1962-2012 using US stocks (Gatev et al., 2006) and concluded an 11% annual excess return with a Sharpe ratio six times that of the overall market. Despite the strong findings, the authors argued that the profitability of the model was declining in their sample period.

The distance method has since then been discussed several times, both based on the same algorithm as Gatev et al. (2006) and alternative versions of the algorithm.

One of the first articles that was published using the distance method of Gatev et al. (2006) was by Andrade, Di Pietro and Seascholes (2005). The authors tested the strategy on the Taiwanese stock market from 1994 to 2002 and concluded similar results to Gatev et al. (2006) of an annual excess return of 10.2%. Following the above, Papadakis and Wysocki (2007) tested the same sample period as of Gatev et al. (2006) and concluded a 7.7% annual return result. However, the authors only tested the method on a subset of the US equity market. In addition to similar results as of Gatev et al. (2006), the authors found that earnings announcements and analysts forecasts tend to cause divergences, however opening a pair trading based on earnings announcements showcased to be unprofitable. These findings were consistent with the findings of Andrade et al. (2005), who noticed that uninformed demand shocks drove the profits of the strategy. Do and Faff (2010) tested the distance method on the US equity market in the same period but also added the period until 2009, as well as tested Papadakis and Wysocki (2007) parameters on earnings announcements. The authors could not find evidence that earnings announcements had a significant effect on the profitability of the strategy. This questions somewhat the findings of Papadakis and Wysocki (2007) due to the small sample size as well (Kraus, 2015). Do & Faff (2010) also found that the profitability of the strategy had a declining trend, but that the strategy performed strongly in turbulent times. Do & Faff (2012) then wrote an additional article on pairs

trading and included transaction costs as parameters. Gatev et al. (2006) had somewhat considered transaction costs in a very overall fashion, but Do and Faff (2012) changed the view of transaction costs. The authors showcased that transaction costs had a significant impact on the profitability of pairs trading and essential to consider when assessing the practical profitability of a pairs trading strategy. A noticeable finding of Do and Faff (2012) is that the pairs trading was unprofitable from 2002 to 2009. Engelberg et al. (2009) presented a thorough study of the distance method, but as an alternative version of the algorithm of Gatev et al. (2006). The authors further discussed the implications of information flows from both idiosyncratic and common news on the performance of pair trading on the US equity market. In addition to similar profit results as Gatev et al. (2006), their results showed that idiosyncratic news increased divergence risk. In line with both Engelberg et al. (2009) and Papadakis and Wysocki (2007), Jacobs and Weber (2015) also investigated the determinants of profitability for pair trading. The authors concluded that pair trading profitability was correlated with investor inattention, and when there are no limits of arbitrage. Besides, they also supported the findings of Engelberg et al. (2009) that idiosyncratic news had a negative effect on the profitability.

Several other authors have tested the distance method of Gatev et al. (2006) on international stock markets. Broussard and Veihekoski (2012) tested the method on the Finnish stock market in the period of 1987 to 2008 and concluded an average annual return of 12.5%. Perlin (2009) tested the distance method on the Brazilian market and verified the results of Gatev et al. (2006). Borgun, Kurun and Guven (2010) tested the distance method on the Turkish market and concluded an average annual return of 3.35% and better Sharpe ratio than the overall market as well.

All in all, Gatev et al. (2006) framework of pair trading has been one of the most used frameworks of pairs trading. The distance method has been tested, verified and applied by numerous authors in different variations and on different markets. However, several authors conclude that the profitability of the distance method has been declining. As most of the literature on pairs trading was written some time ago, it would be interesting to investigate whether this is still true. Based on that notion and the popularity of the distance, we find it necessary and useful to include the distance method in our paper.

In 2004, a second branch of pairs trading literature began to gain traction when Vidya-murthy (2004) proposed a cointegration process as an alternative to the pairs trading strategy of the distance method. The fundamental structure of the formation and trading period is the same; however, the process of detecting and trading eligible pairs is somewhat different. Instead of ranking pairs based on the least sum of squared deviations, the cointegration approach is based on the statistical cointegration of the pair. Where the distance method is based on practical implications from practitioners, the cointegration method has a stronger statistical foundation (Gatev et al., 1999; Krauss, 2015). Gatev et al. (2006) even added a section about cointegration in their 2006 revision of the distance method. Vidya-murthy (2004) described the methodology of pair trading with cointegration, but did not conduct any empirical tests. This is also the case for Gregory, Ewald and Knox (2011) and Herlemont (2004) that set out similar frameworks but did not conduct any empirical studies. The first paper that applied empirical evidence to the use of cointegration in pairs trading was by Hong and Susmel in 2003; however, their paper was not published until 2013 (From Krauss, 2015; Hong and Susmel, 2013). The paper was based on the spreads between American Depository Receipts and their respective shares in Asia (Hong and Susmel, 2013). In their 2003 version, the authors only assumed cointegration. In their 2013 published version, they argued that the pairs were cointegrated based on Augmented Dickey-Fuller tests and Perron-Phillips tests. The authors concluded an average daily return after transaction costs of 0.9% (Hong and Susmel, 2013). However, the high returns may well be driven by an appreciation of local currencies, casting doubt on the results of the paper (Krauss, 2015).

One of the more cited papers on cointegration that conducted an empirical study is Caldeira and Moura (2013). The authors tested a pairs trading strategy on the Brazilian market using an Engle-Granger two-step approach and a Johansen test to test for cointegration (Caldeira and Moura, 2013). The authors ranked the best pairs based on the highest Sharpe ratio in the formation period. They concluded an average annual return of 16.34% on the Brazilian market with a Sharpe ratio of the portfolio of 1.34 before transaction costs. Li, Chui and Li (2014) tested various cointegration tests on the Chinese stock market and concluded an average annual excess return of 17.6%.

Huck and Afawubo (2015) tested both the distance method and the cointegration method and concluded that the latter of the two provided both higher returns and a more robust strategy. Just as Caldeira and Moura (2013), Huck and Afawubo (2015) used the Engle-Granger two-step and Johansen test but ranked the pairs on the trace statistics derived from the Johansen test and not the Sharpe ratio.

Smith and Xu (2017) also compared the distance method and cointegration, but concluded the opposite as Huck and Afawubo (2015) that the distance method performed best. However, the authors were not able to provide any significant positive results (Smith and Xu, 2017). Also, the authors tested different parameterizations and underlined the importance of these parameters for the profitability of a pairs trading strategy. Rad, Low and Faff (2015) did the most comprehensive study of the different methods with more than 23,000 stocks in their sample. The authors tested the distance method, the cointegration method and a copula method. The authors found that the three methods yielded a monthly average excess return of respectively 91, 85 and 43 bps - meaning that the distance method yielded the best results. They also found that the cointegration test yielded the best results in turbulent periods (Rad et al., 2015). It is thus evident from above that it cannot be fully stated whether the cointegration or the distance method yields the best results.

All in all, the cointegration method has a stronger theoretical background than the distance method, but has fewer empirical tests conducted. Based on the above section, it seems like the cointegration method yields positive results which have also been examined by several authors both theoretically and empirically. Thus, we also consider the cointegration methodology for our paper.

Some other methods must also be mentioned. The copula method, as mentioned in Rad et al. (2015), is one of several other types of methodologies that have been tested or applied on pairs trading. Different types of methodologies are discussed by Krauss (2015) and he amongst others mention the correlation approach, time-series approach and stochastic control approach. The other approaches have not received broader noticeability for the use of pairs trading explicitly, why we do not consider these approaches further. Some other approaches also include multivariate trading models. However, as the scope of this paper is to understand the applicability of pairs trading

using ETFs, these considerations are out of scope. We do not neglect that these methods could have been arguably useful for empirical tests; however, we believe that the distance method and the cointegration remain the most-known and verified methodologies.

Considerations about sample | It is evident from the above academic papers and the discussion paper of Kraus (2015) that the samples of the normative papers of pairs trading strategies almost always comprise single stocks. There are a few exceptions of papers using futures by Dunis, Laws and Evans (2006; 2008); however, these papers focus on spread trading and not directly focused on pairs trading. As such, when reviewing the academic literature on pairs trading using ETFs, only three relevant articles can be mentioned; Schizas et al. (2011) and Rudy and Dunis (2010; 2011). Schizas et al. (2011) propose an alternative version of the distance method on a sample of 26 international ETFs, while Rudy and Dunis (2010; 2011) make a comparison of the 100 most liquid ETFs and 100 most liquid US stocks. In their 2010 paper, Rudy and Dunis (2010) used a cointegration approach, whereas, in the 2011 paper (Rudy and Dunis, 2011), the authors used a correlation approach. However, it is only the Rudy and Dunis paper of 2011 that have been published.

Review of the existing literature on pairs trading | From our review of the existing literature, several noticeable findings were evident. First of all, the distance method and the cointegration method seem to provide the most verified methodologies for our paper. Thus, as mentioned, we have decided to use these two methods. From the cross-method tests that have been conducted, it is evident that it is not possible to determine which of the two that yields the best results. Lastly, the samples of the academic papers are very much focused on stocks. As ETFs have the same tradability as stocks, it is peculiar that very few pairs trading papers have been conducted on ETFs. The only research on ETFs comprises non-published papers on small samples. From our literature review, we see a gap in the existing literature regarding the applicability of pairs trading on ETFs.

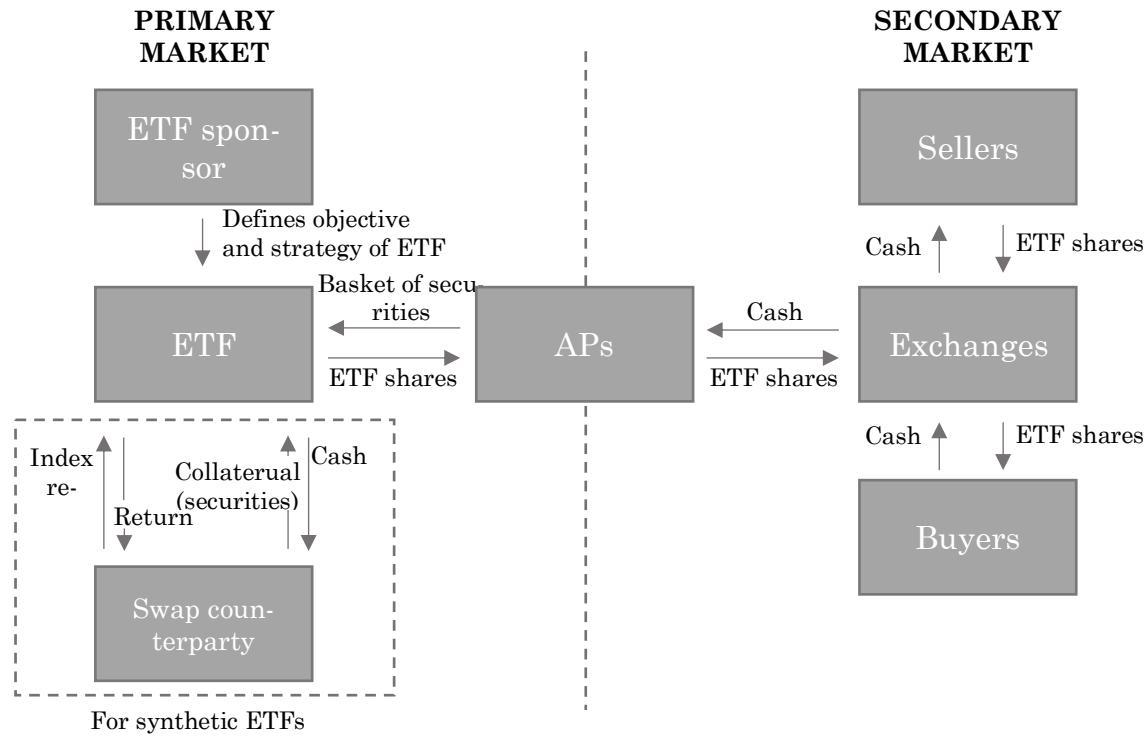
5. Pairs trading and ETFs

The following chapter will introduce pairs trading and ETFs before the methodologies and theories of pairs trading are further examined. The first section will introduce ETFs and their mechanisms. The next section will touch upon the history of pairs trading and discuss why pairs trading has its existence. Lastly, the section will touch upon some of the general characteristics of anomalies and risks of pairs trading.

5.1. Exchange-traded funds

An ETF is a financial instrument with the objective to replicate the return of its corresponding benchmark index (Deville, 2008). The ETF shares trade on the secondary markets, i.e. exchanges, at a transparent price. Here, the number of shares of the ETF is adjustable to respond to the supply and demand as with open-end funds (Pagano, Serrano and Zechner, 2019). As with stocks, ETFs trade on an intraday basis which distinguish them from mutual funds (Ben-Dhavid, Franzoni and Moussawi, 2017; Fidelity, n.d.). The replication of the underlying benchmark is either done with the physical or synthetic method. Physical ETFs try to replicate the returns of the benchmark by holding all or a sample of the index stocks in their portfolio, while synthetic ETFs replicate the returns of the benchmark through derivative trading. The creating and redemption of shares is done using an “in-kind” mechanism, as illustrated below (Deville, 2008).

Figure 1: In-kind creation and redemption



Source: Deville, 2007

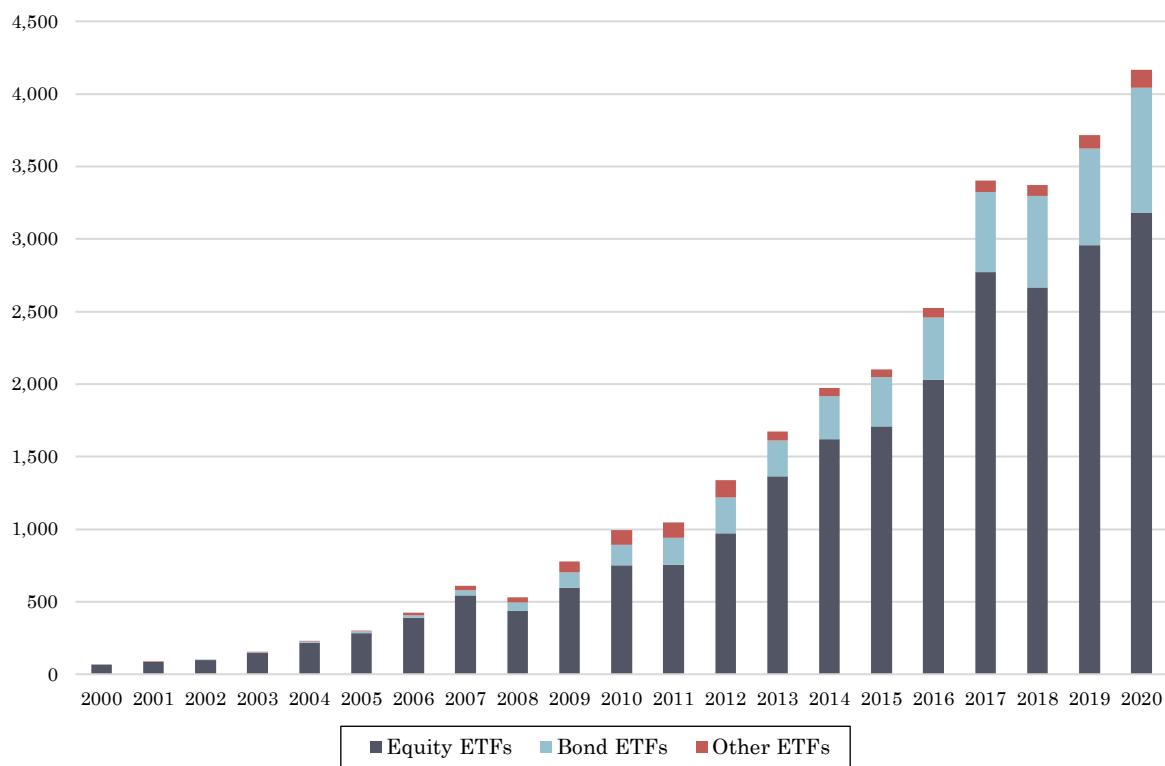
The creation and redemption of new ETF shares are conducted by authorised participants (APs), which are large brokers/dealers assigned by the ETF sponsor (Hill, Nadig, and Hougan, 2015). The ETF price can diverge from the net asset value (NAV) of the ETF due to characteristics of being a product traded on an exchange, where the price is determined by supply and demand. However, these deviations tend to remain relatively small and temporary as these mispricings are closed by the creation and redemption mechanism performed by the APs, as illustrated in figure 1 (Ben-Dhavid et al., 2017).

5.1.1. History of Exchange-traded funds

State Street Global Advisors launched the first ETF (SPY) in 1993 with the goal of tracking the S&P 500 index. The US market for ETFs has since grown exponentially with more than 2,300 active ETFs today, holding a total net asset under management of USD 4,166b (etfdb.com, 2020; Blackrock, 2020). On a global scale, asset under management equals USD 6 Trillion in December 2019 with the expectation of growing to

USD 12 Trillion by 2023 (Kwilla, 2019; Blackrock, 2020). Below is the development of the AUM of ETFs on the US market illustrated:

Figure 2: US ETFs, net asset value under management, 2000-2020



Source: ICI, 2020; Deutsche Bank, 2017

It is evident from the above graph and considerations that ETFs as investment vehicles have gained noticeable traction in both the US and globally. Today, ETFs have become an integrated part of most investors' and traders' daily portfolios, and 78% institutional investors argued in a recent survey that ETFs were their preferred index vehicle (Blackrock, 2020). In the same survey, investors stated that they most often used ETFs for i) tactical adjustments, ii) long-term positions and iii) management of portfolio risks. This illustrates very well both the flexibility and width of the applicability of the ETFs in various areas of the financial markets. Further, 30% of the overall daily trading volume on US stock exchanges are represented by ETFs, as well as 20% of the aggregate short interest (Ben-Dhavid et al., 2017).

5.1.2. Arbitrage trading with ETFs

Relative-value arbitrage trading strategies with the application of ETFs can be performed in three different ways; i) ETF price vs ETF NAV, ii) two ETFs tracking the same index and iii) pairs trading (Ben-Dhavid et al., 2017). Arbitrage trading strategies utilising mispricings between ETF price and NAV is the fundamental mechanism of the in-kind redemption and creation performed by APs. The second trading strategy utilises potential tracking errors and mispricing dynamics between two similar ETFs that should yield the same return. In general, existing literature has shown that there is very little mispricing between the net asset value of equity ETFs and its market price (Ben-Dhavid, 2017). However, some studies have shown that ETFs can produce larger mispricings in more volatile periods (Pagano, Serrano, & Zechner, 2019). However, despite some mispricings occurring, few have been able to generate a robust arbitrage strategy exploiting these mispricings.

5.2. Pairs trading

5.2.1. History of pairs trading

The first pairs trading strategy was introduced by Nunzio Tartaglia in the mid-1980s (Vidyamurthy, 2004; Gatev et al., 2006; Thorne, 2003). Working in Morgan Stanley, Tartaglia put together a team of academics and mathematicians to set forth a statistical, algorithmic trading strategy to expose arbitrage opportunities in the equity markets (Gatev et al., 2006). The team figured out that a useful method to exploit these arbitrage opportunities was trading equities pairwise (Vidyamurthy, 2004). The process relied on finding two equities that “moved together” and then traded these pairs when divergences occurred with the idea that the divergence would correct itself (Vidyamurthy, 2004). According to Gatev et al. (2006), the Morgan Stanley team traded these pairs in 1987 and managed to profit \$50M. However, 1987 is also one of the more remarkable trading periods in US history triggered by the “Black Monday” market crash; one of the largest price drops in the history of the American stock market (Carlsson, 2007). Not only did prices decrease significantly, but market efficiencies were drastically impaired in the following months, and large indices experienced significant mispricings (Carlsson, 2007). According to Edward Thorp (2003) - founder of the first quant hedge fund, simulations of the 1987 market and the following months showcased that

this exact period was one of the best periods for statistical arbitrage. After a very profitable year for the Morgan Stanley team, the following years were not as profitable, and the team was dissolved in 1989. Despite the dissolution of the Morgan Stanley team, pairs trading has since become a common and popular investment strategy for investors and hedge funds (Vidyamurthy, 2004; Gatev et al., 2006; Engelberg et al., 2009).

5.2.2. Why does pairs trading work?

Fundamentally is pairs trading a relative-value trading strategy that exploits inefficiencies in the market by utilising statistical arbitrage (Gatev et al., 2006; Huck and Afawubo, 2015; Gregory et al., 2011; Papadakis and Wysocki, 2007; Jacobs and Weber, 2015). As such, the nature of pairs trading takes a critical stance to the fundamental hypothesis that the market is fully efficient. An efficient market is where all market prices reflect all relevant information, and the market price fully reflects the fundamental value of the security (Pedersen, 2015). This implies that two securities that are substitutable or yield the same return must have the same price (Gatev et al., 2006). This relationship is also referred to as the law of one price meaning the “*proposition (...) that two investments with the same payoff in every state of nature must have the same current value*” (from: Gatev et al., 2006, p. 5). An efficient market also implies that market return reflects the best risk-return relation; hence there would be no need for active investors trying to beat the market (Pedersen, 2015). Nevertheless, there exist active managers, and in connection to this claim, Pedersen (2015) raises the question whether it is then the market or the investors that are inefficient - or maybe both. A third consideration about the causes of inefficiencies in the market is described by Shiller (1992), who argues that inefficiencies are not caused by the market or lack of information, but rather that people make mistakes and are subject to common biases. This notion is somewhat in line with the concepts of the “Me” and the “I” stated in the section of the philosophy of science, as well as the fundamental idea of pairs trading stated by Tartaglia (Gatev et al., 2006). However, if market inefficiencies are only a matter of human errors, then it would entail that beating the market would be easy, which is somewhat far from the truth (Pedersen, 2015). We agree on the notion stated by Pedersen (2015) that all of the reasons stated above have some truth to their names.

Pedersen (2015) summarises these considerations in what he describes as an efficiently inefficient market (Pedersen, 2015). Inefficient enough to make it possible for market makers and arbitrageurs to profit from their provision of liquidity and price discovery, but efficient enough to ensure any arbitrage opportunities due to potential mispricings are quickly closed and thus only exploited by the best (Pedersen, 2015). These characteristics ensure that the market continues to be as close to efficient as possible.

On the matter of an efficient market and the law of one price, Chen and Knez (1995) define two degrees of integrated markets, a weak form where arbitrage opportunities exist and a strong form where pricing in all markets are fully integrated and efficient (Chen and Knez, 1995). For example, the authors find that NYSE and NASDAQ showcase a weak form of closely integrated markets, meaning that arbitrage opportunities would occur. The authors argue that in the weak form of integrated markets, securities with similar payoffs in closely integrated markets must also have close to the same prices (Chen and Knez, 2015). This is an important notion to consider as it allows us to conduct an examination of near-efficient securities and markets through the application of pairs trading (Gatev et al., 2006).

Engelberg et al. (2009), Caldeira and Moura (2013) and Vidyamurthy (2004) further describe pairs trading as statistical arbitrage, meaning exploiting the mean reversion characteristics of two similar securities based on a predefined algorithm. Here, Do and Faff (2012) add to the description and define pairs trading as risk arbitrage, because the arbitrage opportunity is not risk-free. As such, pairs trading does not profit from market trends but on two contrary positions in two securities forming a pair. The trading position is both long and short in the market, and thus considered a market-neutral strategy with no exposure towards the market (Vidyamurthy, 2004; Gatev et al., 2006; Huck and Afawubo, 2015; Schizas et al., 2011). We will touch upon this later in this chapter. Do & Faff (2012) agree on the definition from both Engelberg et al. (2009) and Gatev et al. (2006) and further describe pair trading as a short-term contrarian strategy that exploits violations in the law of one price. As such, pairs trading is not purely a relative-value strategy or purely a statistical arbitrage, but a combination of both. One could state that pairs trading utilises a statistical risk arbitrage model to initiate relative-value trades (Erhman, 2006).

Thus the overall mechanism of pairs trading can be summarised as a strategy that exploits the price divergences of securities with similar payoffs in an efficiently inefficient market. The fundamental idea of betting on price differences between two similar securities and not on the general market trends is that it is considered more challenging to predict the general market than the relative performance of similar securities (Ehrman, 2006).

5.2.3. Why do anomalies occur?

As information does not flow and stick to market prices, information must be collected in order for news or similar to flow into the price of the security (Pedersen, 2015). In order for information to be collected and subsequently infused in the market prices, there must implicitly exist incentives for producing information that initiate traders to do so. These incentives typically come in the type of an inefficiency that can trade into an arbitrage opportunity in the market. The reason that some market players can profit from the collection and subsequent production of information and others cannot profit from the same information suggests that there is a timing aspect of exposure to news and production of information. This means that in securities where many parties are trying to exploit any arbitrage opportunities, the time for information to be included in the market prices is short. For ETFs, this time of opportunity is further shortened with the mechanisms of the authorised participants, ensuring that any mispricings between the net asset value and market price are quickly adjusted.

For ETFs tracking the same index, the spread anomaly should be very close to zero as the APs are closing any mispricings to NAV - which are the same for both ETFs. Here, the flow of information is high. For other pairs of ETFs, the relative value component of the two securities relative to each other is not as prominent, hence the APs of the two ETFs do not necessarily close the cross-mispricing between these two ETFs. In between these two extremes are many pairs of ETFs that track parts of similar indices, with different degrees of efficiencies dependent on the supply and demand of the ETFs. As such, we can conclude that the different flows of information and different degrees of monitoring by both investors and analysts have an effect on how often anomalies occur and for long how they last.

The flow of information is not necessarily the only reason for these anomalies to occur, as this only represents part of the reasons for an efficiently inefficient market as touched upon earlier. Common mistakes or human errors also cause anomalies in the market prices. To this, it is found that there exists a tendency of delayed overreactions and initial underreactions of information before represented in the market prices (Pedersen, 2015). The time it takes the “true” market price to reflect new changes, thus reflecting a period of potential anomalies. In these cases, traders must estimate the effects of the news on the stock price. However, how do traders predict the future cash flow of a security and thus the future price of such? The most fundamental idea is looking at the historical evidence to consider what to expect of the future. In other words, traders must use their experiences to predict the practical consequences to determine whether to buy, sell or hold a security. Common mistakes then refer to the times where a trader either overreacts or under-reacts on news of the market prices based on the respective investor’s interpretation of either idiosyncratic or common news. Shleifer and Vishny (1997) find that anomalies become understood and accepted very slowly, and often investors decide to act too late. When the pattern is not very noisy, the anomalies are quickly corrected. However, in cases of more noisy anomalies, investors tend to delay their reactions because of the risk of an unprofitable arbitrage position (Shleifer and Vishny, 1997). An arbitrage position might potentially become unprofitable due to noise-traders that sell their position irrationally, liquidity shortages or the fundamental risk that one of the two selected securities will continue to diverge (Do and Faff, 2010).

The dynamics of anomalies highlight an interesting aspect of pairs trading; too noisy anomalies have the risk of becoming unprofitable while anomalies that have no noise might also prove to be unprofitable due to stiff competition or transaction costs. The mechanisms of the limits of arbitrage and anomalies are essential matters when trying to understand what determines pairs to diverge and converge.

5.2.4. Risks of pairs trading

Even though pairs trading is conceptually a market-neutral strategy, it does not necessarily mean that the strategy is risk-free (Do and Faff, 2012). One of the critical risk factors associated with pairs trading is the risk that a diverged pair will not be able to

converge before the end of the trading period. This kind of risk is also referred to as horizon risk (Engelberg et al., 2009). Horizon risk is a function of time as the risk increases as time passes and we reach the end of a trading period which gives diverged pairs less time to converge. The magnitude of the horizon risk becomes even more profound when setting a maximum holding period for each trade as it requires a pair to converge within a prespecified time-bound.

In continuation of horizon risk, another key risk to consider is the divergence risk; the risk that after entering a long-short position in the market the pair continues to diverge (Krauss, 2015; Engelberg et al., 2009). The divergence risk is thus similar to fundamental risk. It means that the pair will continue to diverge and thus produce negative returns until the trading period is terminated or the position is liquidated (Do and Faff, 2010). Engelberg et al. (2009) find that the risk of divergence depends on the type of news that causes the divergence. If news contributes to a shift in the value of only one of the two securities, it would often mean that the two prices will drift apart and the comoving relationship between the two securities will be disrupted. Here, Engelberg et al., (2009) find that pair divergences based on idiosyncratic news yield lower profitability compared to those that diverge due to common news. This makes sense given the divergence and horizon risk.

In the light of divergence risk and horizon risk, arguably using ETFs could provide some risk-minimising elements that are not possible to obtain when using single stocks. Because an ETF by its nature tries to replicate the return on an index consisting of multiple securities, an ETF achieves the diversification benefits as it is exposed to the portfolio of stocks which make up the index. The result of this diversification benefit is that the ETF is not exposed to idiosyncratic risk to the same extent as single stocks. Price developments of ETFs reflect a market-weighted development of the single stocks within the index, as long as the ETFs are not mispriced; thus, other things being equal, pairs trading on ETFs reduce both horizon and divergence risk. Changes to the fundamental value of one of the stocks does not have a significant effect when it is included in a portfolio composition via the ETF as if it had been a single stock against another single stock in a pair.

Last but not least, the use of ETFs eliminates bankruptcy risk (Rudy and Dunis, 2011; Schizas et al., 2011). Because the assets owned by the ETF are securities that are not amortised, their value will always represent the market value. Therefore, it will always be possible for the ETF to gain cash by selling its underlying holdings in the market in order to meet its obligations.

5.2.5. Market exposure

A pairs trading strategy is based on the premise that every trade initiated has an equally long and short exposure in the market. This is achieved by investing an equal dollar amount in both the long and short position. By investing an equal dollar amount in both directions, a dollar-neutral strategy is achieved (Ehrman, 2006). For pairs trading, a dollar neutral investment is thus critical for correctly executing the trading. Related to dollar neutrality is market neutrality, where the exposure to the market is hedged away such that the portfolio is unaffected by market movements. A portfolio becomes market neutral by ensuring the fluctuations with the market is uniformly distributed such that any market trend is cancelled out (Ehrman, 2006; Pedersen, 2015)

As such, dollar neutrality does not necessarily entail market neutrality as dollar neutrality implies an equal dollar investment in both directions, whereas market neutrality is achieved by an equal exposure to systemic risk in both directions. However, in the case of pairs trading, both dollar neutrality and to a great extent market neutrality is obtained. The reason for this is that the strategy trades on securities which has exhibited similar price development during the formation period. The fact that they have moved together must imply that the two securities have close to the same market exposure, β . When this is the fact and the investment at the same time is dollar neutral, each opening becomes an equal dollar investment in a long and short position of the two securities with a close to the same market exposure.

By this, the net dollar exposure to market movements is to a large extent eliminated, as the relative price development of the two securities tend to be the same, implying that pairs trading is close to or is market neutral. If dollar neutrality is not maintained, but the opening of a position instead is based on an equal share investment, the exposure towards the market would not be neutral anymore. Instead, the neutral position

would either be a hedged bullish long position or a hedged bearish short position because market movements would not have the same effect on the two securities (Ehrman, 2006). Taking a share-neutral strategy, the investor would be more exposed towards the market trend either going up or down, while hedged towards the opposite trend.

A derived effect of a dollar-neutral investment is that the net returns of such long-short position is argued to be interpreted as an excess return (Gatev et al., 2006; Smith & Xu, 2017; Broussard and Vaihekoski, 2012; Andrade, 2005; Caldeira and Moura, 2013; Do & Faff, 2010; 2012; Huck and Afawubo, 2015). Due to the dollar-neutrality of the long-short position, the strategy becomes self-financing as the long position is financed by cash generated from the sale of the short position. This is feasible because we invest equally in both positions. As we do not assume that the net cash position is “sitting” at a risk-free rate, the return has the interpretation of excess return.

6. Applied pairs trading strategy

The following chapter covers the analytical pairs trading methods applied in this paper. The paper relies on the sum of squares distance method by Gatev et al. (2006) and the cointegration method described by various authors. The following chapter will consider the several steps and parameters of a pairs trading strategy with references to existing literature, with the aim of discovering the optimal parameters for backtesting the pairs trading strategy when applying ETFs.

6.1. Setting up a pairs trading strategy

Before a pairs trading strategy can be initiated, some preliminary setting must be considered. These basic settings should be carefully considered as the choices have consequences for the performance of the strategy executed.

6.1.1. Potential pairs

The first step of initiating a pairs trading strategy is to consider the scope of potential pairs. Gatev et al. (2006) consider an unrestricted approach where all US-based stocks are defined as potential pairs. The authors therefore test all possible combinations of

2,317 US-based stock for their fit of pairing (Gatev et al., 2006). Rad et al. (2015) takes an even wider approach and includes 23,616 stocks. Authors such as Caldeira and Moura (2013), Andrade et al. (2005) or Broussard and Vaihekoski (2012) use a similar unrestricted approach, however focusing on other markets, namely Brazil, Taiwan and Finland respectively. Other authors propose a more restricted approach to the scope of potential pairs. Engelberg et al. (2009) pair US stocks based on their respective industry group. Do and Faff (2012) test different approaches to both unrestricted and restricted pair combinations and conclude that the excess returns of the restricted approach of similar industries yielded better results than the unrestricted approach. A more index-focused approach is taken by Baronyan, Boduroglu and Sener. (2013) who focuses on the Dow Jones 30 Index and Huck and Afawubo (2015) who test the stocks in the S&P 500 index. As described in chapter 3, this paper takes a more unrestricted approach; however, still only covering US-issued ETFs. Potential pairs can thus be created across both industry-specific, country-specific, size-specific or other types of ETFs. As the sample includes indices often replicating part of or an entire industry, we do not believe that it would be suitable to take a restricted approach of potential combinations of pairs (Appendix 1). Before the pairing of securities is carried out, Papadakis and Wysocki (2007) and Vidyamurthy (2004) propose a preliminary sorting to ease the process of pairing. With the risk of sample bias, this paper does not conduct such preliminary selection.

6.1.2. Length of formation and trading period

A pairs trading strategy is a two-period approach consisting of an initial formation period followed by a trading period. The two periods are continuous periods where specific rules are applied respectively to form and rank the most optimal pairs, and thereafter trade the top pairs (Schizas et al., 2011). Gatev et al. (2006) initially presented a 12-month formation period followed by a 6-month trading period, which hereafter has been considered the benchmark approach within pairs trading (Gatev et al., 2006; Yu & Webb, 2014; Huck & Afawubo; 2015; Papadakis and Wysocki, 2007; Smith and Xu, 2017; Do and Faff, 2012; Engelberg et al., 2009). Huck and Afawubo (2015) and Smith and Xu (2017) have conducted empirical tests on the lengths of the two periods. Huck & Afawubo (2015) consider a formation period of 12 and 24 months respectively, and conclude a 12-month formation period yields the highest average monthly return.

Smith and Xu (2017) consider a 9-month formation period versus a 12-months and conclude that a 12-month period provides better results.

Based on the consensus within the pairs trading literature and the empirical tests conducted on different lengths, this paper complies with a 12-month formation period and a 6-month trading period.

By applying a 12-month formation period, a full-year cycle will be taken into account when detecting the optimal pair to be traded in the subsequent trading period, thus eliminating any potential seasonal biases. Furthermore, with a 6-month trading period, it is assumed that it is enough time for the pairs to diverge and converge again while still holding the relationship found in the formation period.

Gatev et al. (2006) further suggest recalibration of the strategy every month, meaning that a new formation period is initiated every month. Thus, six overlapping portfolios will always be active at any given time, besides the first 18 month where the initial formation periods are carried out respectively for the six portfolios (Gatev et al., 2006; Do & Faff, 2010). This monthly recalibration is not used by Papadakis and Wysocki (2007) in order to prevent overstatements of excess returns due to the overlapping trading periods.

In this paper, a new formation period is initiated every six months such that every trading period is followed by a new trading period as illustrated in the below figure 3. The traded portfolio is thus recalibrated every six months without overlapping portfolios as otherwise proposed by Gatev et al. (2006). The reason for not considering the overlapping portfolios is due to the complexity of juggling six different portfolios at any given time and the potential overstatement of excess returns due to potential trading on the same pair in several portfolios (Papadakis and Wysocki, 2007).

Figure 3: Formation and trading periods

	2006		2007		2008		2009		...
Formation periods	H1 2006		H1 2007		H1 2008		H1 2009		...
		H2 2006		H2 2007		H2 2008		H2 2009	...
Trading periods		H1 2006	H2 2006	H1 2007	H2 2007	H1 2008	H2 2008	H1 2009	...

6.2. Formation period

6.2.1. 6.2.1 The distance method

The first step of the distance method in detecting optimal pairs is to normalise the ETF prices to equal unity with the base date as the first day in the formation period (Gatev et al., 2006; Engelberg et al., 2009; Do & Faff, 2010; 2012). The normalised end of day prices of each ETF is calculated as:

$$P_t^i = \prod_{\tau=1}^t (1 + r_{\tau}^i), \quad (1)$$

where P_t^i is the normalised price of ETF i on day t until T_{fp} with T_{fp} being the number of days in the formation period, hence $t = \{1, 2, \dots, T_{fp}\}$. τ is the index value of ETF i between $\tau = 1$ and t , and r_{τ}^i is ETF i 's total return generated between $\tau = 1$ and t (Smith and Xu, 2017).

The reason to employ normalised prices is because the nominal price level itself is not of interest. The distance method aims to identify ETFs which exhibit the same pattern in their daily returns, which essentially makes the price level of less interest. In other words, the nominal prices of two ETFs can be substantially different from each other, but if their prices exhibit identical growth patterns (in relative terms) they are deemed to have a small difference.

With the normalised prices, the next step of the distance method is to compute the sum of squared differences $SSD_{i,j}$ for each formation period as the following:

$$SSD_{i,j} = \sum_{t=1}^T (P_t^i - P_t^j)^2. \quad (2)$$

Let T be the number of trading days in the formation period, P_t^i and P_t^j be the normalised prices of ETF i and ETF j on trading day t . The measure of closeness, defined by $SSD_{i,j}$, is calculated between every possible ETF combination in every trading period (Appendix 5; Huck and Afawubo, 2015; Papadakis and Wysocki 2007). Here the possible number of combinations can be defined as $N_{fp} * (N_{fp}-1) * \frac{1}{2}$, where N_{fp} is the number of ETFs in the particular formation period. This implies that the last formation period consists of 322 ETFs, thus the number of possible combinations totals 51,681.

Based on the associated sum of squared differences, $SSD_{i,j}$, for each possible combination of ETFs, the pairs are ranked from lowest to largest in order to detect those pairs

with the lowest values which become eligible for trading in the subsequent trading period.

6.2.2. Cointegration

The cointegration method detects eligible pairs for trading based on the statistical relationship between the two ETFs concerned. The reason for using cointegration to conduct pairs trading is that the method is useful for identifying a long-run equilibrium between two ETFs. If a long-run equilibrium is established, any deviations from here are expected to be restored by an adjustment in one or both of the ETF prices (Vidyamurthy, 2004).

For two, integrated of order one ($I(1)$), time-series to be cointegrated, there must exist a linear relationship between the two integrated of order zero ($I(0)$) (Caldeira and Moura, 2013; Stock and Watson, 2015). In other words, two ETFs are cointegrated when their prices follow the same stochastic trend and thereby hold a constant relationship on the long run.

To identify whether the relationship between the prices of two ETFs is constant over time, $P_t^i - \theta P_t^j$ must be stationary. Here P_t^i and P_t^j are the normalised prices from equation 1 and θ is the cointegration coefficient, i.e. a constant term estimated to establish a linear relation between the two ETF prices (Vidyamurthy, 2004; Stock and Watson, 2015). If this relationship is stationary, i.e. has a constant mean and standard deviation, P_t^i and P_t^j are said to be cointegrated. In order to determine whether the relationship between two ETFs is stationary, we use the two-step Engle-Granger Augmented Dickey-Fuller (EG-ADF) test (Appendix 4; Stock and Watson, 2015). To perform the EG-ADF test, the two ETF price series must be of the same order of integration (Smith and Xu, 2017). We ensure all ETF price series are of the same order of integration by testing for a unit root with the Augmented Dickey-Fuller (ADF) test (Appendix 4). Because the price of an ETF follows a random walk, it makes it $I(1)$ by definition which imply the ADF test must be performed on the first difference of the ETF prices (Stock and Watson, 2015). The first difference of P_t^i is given by $\Delta P_t^i = P_t^i - P_{t-1}^i$, implying the ADF test is performed on ΔP_t^i . When running the ADF test, we estimate the appropriate number of lags (p) to include for each ETF based on the Akaike information criterion (AIC). AIC is an estimator of the optimal amount of information

to be included based on the minimum information criterion. The AIC is therefore a useful analysis to determine the optimal number of lags (Stock and Watson, 2015). Taking the approach of Huck and Afawubo (2015), the maximum number of lags to be tested is defined as 10. Similar test to the AIC includes the Bayes information criteria (BIC) (Stock and Watson, 2015). The argument for choosing AIC over BIC is that BIC typically returns a fewer number of lags than AIC, and when it comes to the ADF test, it is better the lag length is too long than too short, making AIC the preferred estimator (Stock and Watson, 2015). The ADF test itself tests the null hypothesis that ΔP_t^i has a stochastic trend, i.e. is nonstationary, against the one-sided alternative hypothesis, that ΔP_t^i does not have a stochastic trend, i.e. is stationary (Stock and Watson, 2015). With a statistical significance level of 5%, we reject the null hypothesis and continue to the first step in the EG-ADF test.

The first step of the EG-ADF test is to estimate the cointegration coefficient, θ by the ordinary least squares (OLS) estimation of the following regression:

$$P_t^i = \alpha + \theta P_t^j + z_t \quad (3)$$

here α is the intercept, P_t^i and P_t^j are the normalised prices of the two ETFs on time t and z_t is the residual value on time t (Stock and Watson, 2015).

The outcome of this first step provide us with the slope of the estimated regression line which is the cointegration coefficient, θ and the intercept, which can be interpreted as the premium for holding P_t^i over P_t^j (Vidyamurthy, 2004).

With the estimated values of the cointegration coefficient and the intercept, we are able to construct a linear relationship between the two ETFs in question, which enable us to continue to the second step to test whether this relationship is stationary, i.e. the two ETFs are cointegrated.

As described above, in order for the two ETFs to be cointegrated, a linear relationship between their prices must be $I(0)$. To test if this is true, we perform an ADF test for a unit root on the residuals in equation 3 with the alternative hypothesis that P_t^i and P_t^j is cointegrated with a statistical significance level of 5% (Appendix 4; Stock and Watson, 2015). As with the previous ADF test, the appropriate lag length is estimated by AIC.

It is worth noting that the order in which the pairs are constructed (P_t^i against P_t^j or P_t^j against P_t^i) affects the estimated cointegration coefficient and thus influence the ADF test of the residuals (Caldeira and Moura, 2018; Asteriou and Hall, 2011). Since the order of how a pair is constituted may affect whether the two ETFs are determined to be cointegrated or not, we conduct the EG-ADF test in both directions to prevent missing out on identifying potential cointegrated pairs (Appendix 4).

In an attempt to obtain a more robust argument to determine whether a pair is cointegrated, and in order to meet some of the shortcomings of the EG-ADF test, the Johansen test is furthermore applied. Some of the pitfalls related to the EG-ADF test are that inferences based on the t-statistics can be misleading (Stock and Watson, 2015). Additionally, the order in which the two ETFs are regressed on each other can potentially exhibit contradicting evidence of whether a pair is tested to be cointegrated. Furthermore, the fact that the EG-ADF test is a two-step approach enables potential mistakes or misinterpretations made in the first step, to affect the second step and hereby provide erroneous results (Asteriou and Hall, 2011). These drawbacks are resolved in the Johansen test where the estimator of the cointegrating coefficient is considered more efficient (Stock and Watson, 2015). Part of the reasons is that the model is a unified framework and because all variables are treated as endogenous, the order of the variables in the regression is irrelevant (Naser, 2017; Stock and Watson, 2015). For the Johansen test, the variables can be of different order, but in order to mitigate the problem of spurious regressions, variables of $I(1)$ are preferred (Asteriou and Hall, 2011). The Johansen approach tests for cointegration with the application of a vector autoregressive (VAR) model, which allows the relationship between two or more variables to be tested. In general, a VAR model with k variables, contains k regressions where all k variables and their p lags are the regressors. These k regressions are combined into vector form to construct a VAR model (Stock and Watson, 2015). The general form of a VAR model can be stated as the following:

$$Y_t = \alpha + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + u_t \quad (4)$$

Where Y_t is a $k \times 1$ vector, α is a $k \times 1$ vector containing the intercepts for each regression, A_i is a $k \times k$ coefficient matrices for each corresponding lag and u_t is a $k \times 1$ vector

containing the error terms (Stock and Watson, 2015; Brooks, 2008). As with the ADF test, the appropriate lag length of Y is determined by AIC. This paper will only be centered around a bivariate VAR(p) model going forward, as the scope lies within the relationship between two ETFs. In order to determine whether a pair is cointegrated with the Johansen approach, the VAR model must be transformed to a vector error correction (VECM) model by taking first difference of equation 4 and include $P_{t-1}^i - \theta P_{t-1}^j$ as an additional regressor, which then can be written as:

$$\Delta Y_t = \alpha + \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + \dots + \Gamma_{p-1} \Delta Y_{t-(p-1)} + \Pi (P_{t-1}^i - \theta P_{t-1}^j) + u_t, \quad (5)$$

here ΔY_t is a 2×1 vector containing the two variables; ΔP_t^i and ΔP_t^j , α is a 2×1 vector holding the intercepts for each variable, Γ is a 2×2 matrix with the coefficients to the lagged values of the two regressors where $i = \{1, 2, \dots, p-1\}$, Π is a 2×2 matrix and represents the adjustment effect of any divergence from the long run relationship and θ representing the cointegration coefficient (Stock and Watson, 2015; Asteriou and Hall, 2011). Equation 5 can be broken down further to $\Gamma_i = (I - A_1 - A_2 - \dots - A_p)$ and $\Pi_i = -(I - A_1 - A_2 - \dots - A_p)$ with A_i coming from equation 5 and I being an identity matrix. This first part of the equation represented by the variables in first difference is the short run dynamics, whereas the last term starting with represents the long run dynamics (Asteriou and Hall, 2011). For both the short and long run effects of the VECM model, an intercept and no deterministic component is included. When including an intercept for both the short and long run effects of the model, it is only the intercept for the short run, α , which remains in equation 5 (Asteriou and Hall, 2011). The inclusion of an intercept is, as for the EG-ADF test, an indicator for the premium for holding one ETF over the other. The reason for not including a deterministic trend is the nature of the development in the ETF prices which does not contain a deterministic trend.

From equation 5 it is the Π which the Johansen test is centered around, as whether two ETFs are cointegrated is determined by the rank of Π , denoted r where $r \leq (k - 1)$ (Asteriou and Hall, 2011). With this, it implies that if $r = 1$ for this paper, the pair is cointegrated, as the test is performed on two variables, i.e. $k = 2$. If the rank of Π is zero then the variables are not cointegrated. The degree to which two variables are found cointegrated can be tested by two tests, maximal eigenvalue statistic, λ_{\max} or the

trace statistic, λ_{trace} where both tests are based on the eigenvalues of Π (Asteriou and Hall, 2011). This paper takes the same approach towards the ranking as Huck and Afawubo (2015) and rank the pairs based on their respective trace statistics λ_{trace} . To do this, we must first determine the rank of Π which is done by solving following for λ

$$\det(\Pi - \lambda I_n) = 0, \quad (6)$$

where \det is the determinant of the $\Pi - \lambda I_n$, where Π is estimated from equation 5, containing the unknown values of λ which is about to be estimated and I_n is an identity matrix. Here all matrices are 2x2 matrices (Bergen, 2019).

Likelihood test | The trace tests are likelihood-ratio tests, which on an overall basis assesses the goodness of fit of the model. The trace statistic tests whether the $\text{rank}(\Pi) = 0$, with the null hypothesis stated as $\text{rank}(\Pi) = 0$. The alternative test is defined as $0 \leq \text{rank}(\Pi) \leq r$, where r is the maximum of possible cointegrating vectors which in our case is 1 as stated above. As such, the trace test is a joint hypothesis test and the likelihood ratio test statistic is given as (Asteriou and Hall, 2011);

$$\lambda_{\text{trace}}(r) = - T \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_{r+1}). \quad (7)$$

From λ_{trace} , the Johansen test provides a measurement of the degree of cointegration between the variables which enables a ranking of the most cointegrated pairs. The choice of using the trace statistics λ_{trace} besides consistency with Huck & Afawubo (2015) is consistent with the conclusions of Lütkepohl, Saikkonen and Trenkler (2001), where they find the trace statistic more preferable than the maximal eigenvalue statistic based on their simulations and practical use. However, the authors find that the two tests do not give rise to any major differences. We believe that by including both the EG-ADF test and the Johansen test for cointegration, we are able to provide a more robust model and to a larger extent mitigate false positive results.

6.2.3. Comparison of distance method and cointegration

The two methodologies outlined above represent two fundamentally different approaches to determine pairs that “move together”. Where the distance method has its foundation in practical implications, the cointegration method is anchored in statistical properties. As such, the cointegration method is considerably more comprehensive in

both its specifications and computational efforts to apply than the distance method. Here, the cointegration method requires a more profound assessment and understanding of the characteristics of data input in order to set up and run the method as it is sensitive to the inputs and parameters applied. Oppositely, the distance method is a much simpler approach. The nature of the method is considered by Do et al. (2006) to be model-free, meaning not impacted by any misspecifications or misinterpretations (from Krauss, 2015). The simplicity of the distance method has given rise to several tests of the parameterizations of the model in the search for improvement to the initial model (Papadakis and Wysocki, 2007; Smith and Xu, 2017; Huck and Afawubo, 2015; Do & Faff, 2012). The two methods thus showcase a typical dilemma for economics, choosing between a simpler but somewhat less nuanced model, and a model that is significantly more nuanced with more specifications but more extensive to do (Granger, 2009).

6.2.4. Choosing top pairs that are made eligible for the trading period

After the process of detecting the most eligible pairs for trading, the pairs trader must decide upon the number of pairs to include in the trading period. For the majority of the current literature, the top 20 pairs from the formation period are selected for trading in the subsequent trading period (Andrade et al., 2005; Caldeira and Moura, 2013; Do and Faff, 2010; 2012; Huck, 2013; Huck and Afawubo 2015; Papadakis and Wysocki, 2007; Smith and Xu, 2017; Yu and Webb, 2014). Rad et al. (2015) and Shizas et al. (2011) test, respectively, the top 5 and top 20 pairs and Gatev et al. (2006) also consider the top 101-120. Gatev et al. (2006) and Rad et al. (2015) conclude from their investigations that the top 20 pairs yield marginally better results. Only two other papers consider a traded portfolio of other than 20 pairs, namely Engelberg et al. (2009) with top 200 pairs and Jacobs and Weber al. (2015) with top 100 pairs.

It is thus evident from the existing literature that most papers conclude on the top 20 pairs. As such, we use the top 20 pairs from the formation periods in our trading periods (see appendix 10 and 11 for eligible pairs).

6.3. Trading period

6.3.1. Triggers

After the top 20 pairs have been selected, the 6-month trading period is initiated. In order to rely on the relationships identified for the selected pairs in the formation period, we continue with the price development in normalised prices with base dates as the first day of the associated formation period.

For the distance method, a position is opened when the spread, $P_t^i - P_t^j$, exceeds a threshold “trigger” value. The opening of a pair trade can thus be stated as:

$$P_t^i - P_t^j \geq \text{open trigger}_{i,j} * \pm \text{standard deviation}_{i,j,fp} , \quad (8)$$

and the closing of a pair trade is stated as:

$$P_t^i - P_t^j \leq \text{open trigger}_{i,j} * \pm \text{standard deviation}_{i,j,fp} , \quad (9)$$

where $\text{trigger}_{i,j}$ is a rule-based constant value of either the open trigger or the close trigger, and $\text{standard deviation}_{i,j,fp}$ denote the standard deviation of the price spread between P^i and P^j during the formation period. Gatev et al. (2006) use an opening trigger value of 2 and a closing trigger of 0. These trigger values have no empirical foundation and have been chosen arbitrarily to prevent data snooping (Gatev et al., 2006). Due to the fact that these trigger values are arbitrarily chosen, we seek to find the optimal trigger combination to enter and exit a trade. We therefore investigate the profits generated with the various combinations of opening and closing trigger values, where the opening trigger can take the values of 2, 2.5 and 3 and the closing trigger can be either 0 or 0.5. Smith & Xu (2017) test a number of different trigger values and conclude that for the 2000s, the best opening trigger setting for the top 20 pairs is 3. Huck and Afawubo (2015) test trigger values of 2 and 3, and also conclude that the best trigger value is 3.

For the cointegration method, the z-score of the spread is applied to identify when to open and close a pair, consistent with the method outlined in Caldeira and Moura (2013). The z-score is defined as;

$$Z\text{-score}_t = \frac{x_t - \mu_{fp}}{\sigma_{fp}} \quad (10)$$

Where x_t is the spread defined as $P_t^i - \theta P_t^j$, μ_{fp} is the mean spread in the formation period and σ_{fp} is the standard deviation of the spread during the formation period.

To get comparative results with the distance method, we test the same combinations of opening and closing triggers for the cointegration method. Basically, there is no difference in whether the z-score or a trigger value multiplied by the standard deviation is applied, as the z-score is an indicator of how many standard deviations a given observation is away from the mean of the period. Therefore, a z-score of 2 is equivalent to $2 \times \text{standard deviation}_{i,j,fp}$.

For both methods, if a position is open on the last day of the trading period, the position will be closed regardless of how long the position has been open and the magnitude of the spread between the two ETFs.

6.3.2. One-day delay

In continuation of above paragraph on the trading period and when to open a pair, several papers have imposed a one-day-later rule meaning that the trade is delayed with one day after the trigger is activated (Gatev et al., 2006; Engelberg et al., 2009; Papadakis and Wysocki, 2007; Smith and Xu, 2017; Yu and Webb, 2014). The one day delay is an attempt to mitigate the bid-ask bounce, which may be a source for an upward bias of the return (Gatev et al., 2006; Smith and Xu, 2017). Such bias arises because the price of the two securities in the pair fluctuates within the boundary of the bid and ask prices. The argument for the one-day-delay is that the security to short (buy) could potentially be quoted as an ask (bid) quote at the time the trigger is breached (Gatev et al., 2006; Papadakis and Wysocki, 2007). The argument for delaying the trade is the assumption that closing prices one day later are equally probable to be at the bid or ask price (Smith and Xu, 2017). Gatev et al. (2006) interpret the loss of profitability between waiting and not waiting to be equal to the transaction costs of trade. However, this is arguably a noisy proxy for transaction costs, as the loss of profitability might also be due to other externalities or missed potential profitable openings that have rapidly converged. Instead of waiting one day for initiating a trade, we have decided to pay the entirety of the bid-ask spread as a transaction cost, which will be covered later. This is done because waiting one day might also distort the profitability of the divergence on the first day.

6.3.3. Holding period and stop loss

The divergence risk covered in chapter 5 can to some extent be mitigated by imposing a maximum holding period in the trading strategy, either as a limited number of days a position is open or as a stop-loss closing trigger. The risk of a pair continuously diverging is thus either time or loss bound in order to keep a potential loss to a minimum. However, by imposing a limit for how long a position can be active might create a greater loss than otherwise would have been the case. A threshold for when to exit a position might also remove the potential of convergence at a later time and thus removing the potential positive return. Whether or not to implement a maximum holding period has not reached a consensus in the academic literature. With examples of a 7% stop-loss proposed by Caldeira Moura (2013) and a maximum holding period of 20 days and one month by Schizas et al. (2011) and Jacobs and Weber (2014) respectively, it is challenging to determine what seems to be the best solution. These different attempts to limit the divergence risk might be a result of the risk advergence of the strategy proposed and thus up to the individual investor to decide.

This paper has taken the stance of Andrade et al. (2005), Baronyan et al. (2010), Brous-sard and Vaihekoski (2012), Do & Faff (2010;2012), Gatev et al. (2006), Huck (2013), Huck and Afawubo (2015), Perlin (2009), Rad et al. (2015) and Smith and Xu (2017) and not imposed a maximum holding period for our strategy. The reason for not limiting the potential loss due to continuous divergence lies in the expectation that ETFs do not have the same divergence risk as single stocks, as described in chapter 5. We find prematurely closure - i.e. the realisation of the loss at the threshold level - a greater risk than to wait for the pair to converge. Furthermore, we find it counterintuitive to only rely on the fundamentals of the pairs trading strategy to a certain extent, also when the idiosyncratic risk is considered reduced when utilising ETFs rather than single stocks. We furthermore find support in our deselection of a maximum holding period in the study of Engelberg et al. (2009), who compare a max holding period of 10 days against no max holding period. Here the findings reveal that without a maximum holding period, the return generated per pair is larger than when the subject for the ten days bound (Engelberg et al., 2009). Although the authors find that a great part of the return is generated in the first days after a divergence, profit is still generated in the days following (Engelberg et al., 2009; Jacobs and Weber, 2014). Therefore, we find

the benefit for the potential increased profitability of holding a pair until it eventually converges to outweigh the potential risk associated with a continuous divergence.

6.4. Computation of results

For the distance method the return of a pair, h for day t is calculated as the following:

$$R_t(P^h) = r(l^h) - r(s^h), \quad (11)$$

here $r(l^h)$ is the return for the long position and $r(s^h)$ is the return for the short position (Smith & Xu, 2017). For the cointegration method we employ a return calculation that includes the cointegration coefficient (Caldeira and Moura, 2013). The cointegration coefficient presents the proportional relationship between the two ETFs, thus in order to ensure an equal exposure to the market, the coefficient is multiplied on the short position of the pair. With this the computation of the return can be stated as the following:

$$R_t(P^h) = r(l^h) - \theta r(s^h), \quad (12)$$

here, $r(l^h)$ is the return on the long position, θ is the cointegration coefficient inferred from the cointegration relationship and $r(s^h)$ is the return on the short position.

The daily return on the portfolio comprising the 20 traded pairs is then calculated in the same way for both methods as the value-weighted returns of the pairs in the portfolio (Smith & Xu, 2017; Gatev et al., 2006). The returns are subtracted transaction costs before weighted.

$$R_{port}^t = \sum_{t=1}^{N_t^*} W_t^h R_t(P^h), \quad (13)$$

where $W_t^h = \frac{w_t^h}{\sum_{j=1}^{N_t^*} w_t^j}$ and $w_t^h = [1 + R_{t-1}(P^h)] * [1 + R_{t-2}(P^h)] * \dots * [1 + R_{i,t-1}(P^h)]$.

Here W_t^h is the weight applied to each pair's daily return, calculated as the cumulative return of the trading period at day t relative to the sum of cumulative returns for N_t^* open pairs held in the portfolio on day t $\sum_{j=1}^{N_t^*} w_t^j$ (Smith and Xu, 2017; Papadakis and Wysocki, 2007). The return of the portfolio, R_{port}^t , on day t is then calculated as the sum of the weighted returns of the pairs. Broussard and Vaihekorsi (2012) also test an

equal-weighted computation, which would be problematic as returns are not recalibrating throughout the trading period, implying that if negative returns occur, the losses will not be covered by supplying additional capital. Therefore, the equal-weighted approach will be inappropriate as an equal investment amount in every trade is not the approach for this paper. Refinancing after each open trade could potentially lead to higher returns, but also require that the investor is able to refinance its commitments continuously. Finally, it is worth noticing that when referred to returns throughout the remainder of this paper it is implicitly understood as excess return, i.e. an additional return relative to the risk-free rate, as described in chapter 5.

6.4.1. Transaction costs

One of the more overlooked parameters in pairs trading literature is the matter of transaction costs, despite it being one of the more critical factors for a profitable strategy (Do & Faff, 2012). Pairs trading is associated with a large number of transactions through the trading periods, thus making it a much costlier strategy compared to a traditional long-hold position. Neglecting transaction costs could lead to overestimated results. Do and Faff (2012) is one of the more cited papers on transaction costs and outlines three transaction costs to consider; commission fees, short-sell costs and market impact or bid-ask spread (Huck and Afawubo, 2015; Smith and Xu, 2017). Besides, when applying pairs trading to ETFs, there also exists an expense ratio covering management fees and other expenses related to the ETF sponsor (etfdb.com, 2020).

Bid-ask spread | The bid-ask spread is the spread that exists between the ask and bid price of a security at any given time t . The spread thereby represents the difference between the highest price a buyer is willing to pay for an asset and the lowest price that a seller is willing to sell his asset (Curtis, 2019). Therefore, a share price will move between the bid and ask price, also referred to as the bid-ask bounce (Gatev et al., 2006). The bid-ask bounce as a trading cost represents a possibility that the transaction will occur at a costlier price than the close price, as it might not have been possible to acquire or sell a security at the given close price. Gatev et al. (2006) calculate the transaction costs as the lost profit of waiting one day. Do and Faff (2012) refers to the cost of the bid-ask spread as the cost of the market impact. The authors further conclude that the bid-ask widens on average when divergences arise, and only decline the

subsequent days marginally (Do and Faff, 2012). Despite these findings, the authors argue that a trader would practically not execute the order at the most expensive spread but spread out the transactions over a couple of days. The authors then assume an average market impact cost of 26 bps over the full sample (Do and Faff, 2012). For the backtest, this paper pays the entire bid-ask spread when both opening and closing trade positions. Considering the approach of Gatev et al. (2006), we argue that the one-day-delay does not fully represent the costs of transaction costs. The lower returns when waiting one day could potentially also be due to missed profits from the actual trade position. Considering Do and Faff (2012), we do not consider it to be possible to speculate against the width of the spread nor is such spread betting the scope of this paper. Further, we do not believe a fixed rate of bid-ask trading costs is representable of the practical implications.

Commission costs | Commission costs refer to the service charges a brokerage firm or another advisor charges for executing a proposed trade in the market (Frankenfield, 2019). The commission cost can include both brokerage fees, exchange fees and other costs regarding the execution of trades. Do and Faff (2012) showcase that commission costs have been declining noticeably since the 1960s. Here, an average commission cost for an institutional investor has declined from 70 bps in 1963 to 9 bps in 2009. In their paper, Do and Faff (2012) uses the just mentioned commission rates with a 20% discount. In the period from 2009 to 2019, the commission costs have further declined to 3 bps as per Q4 2019 (ITG, 2019).

For our analytical framework, we apply the commission costs presented by Virtu in their quarterly global cost review (ITG, 2018; Virtu, 2019). However, it was only possible to acquire review reports going back to data from 2009. From 2006 to 2009, we apply the level of commission costs as referred to in Do and Faff (2012).

Short-sell costs | Short selling a security means that a trader sells a security without owning the security. The short-seller borrows the security from a lender and sells the security in the market. The first step of short selling is to find a willing lender of a security. The lenders are typically custodian banks or brokers that act as lending agents to owners of stocks that wish to borrow out their stocks (Charles Schwab, 2020).

It states by legislation that the borrower must post an initial margin requirement to the brokerage firm equivalent to 50% of the market value of the short position (Fabozzi and Asness, 2004). According to US regulation, the short-seller has to post 102% of the borrowed amount as collateral for the loan of the short sell position. The proceeds from the sales are posted as collateral for the security loan. This also means that if the security price goes up, the short-seller has to post additional collateral and vice versa (Fabozzi and Asness, 2004; Pedersen, 2015).

When the borrower then returns the security to the lender, the lender returns the cash collateral plus an interest (Pedersen, 2015). The rate that the borrower receives is called the rebate rate and is often equal to rates such as the broad general collateral rate or the LIBOR overnight rate. If the rebate rate is lower than the money market interest rate, then the lender earns a premium as it was possible to invest a higher amount than must be returned to the borrower (Pedersen, 2015). The spread between the two rates is an implicit cost for the borrower and represents the securities-lending fee or loan fee (Pedersen, 2015; Fabozzi, 2004). In some cases, the loan can also be an actual fee. In practice, the rebate rate is determined by supply and demand, and how easy or hard the securities are to borrow. For harder to borrow securities because of illiquidity or high demand-pressure, the rebate rate goes down; thus, the lender keeps a large amount of the proceeds from the invested collateral. Cases of low rebate rates often occur in financial distress as the demand for short positions increases (Fabozzi and Asness, 2004).

Further, the owner of the security can at any time recall the security from the borrower. In such cases, the borrower has to buy back the security in the market and return the security. However, in cases of easy to borrow securities, the lender is often able to retrieve another security that is then sold to the owner and the short sell position to the borrower can continue (Charles Schwab, 2020).

Alternative ways of short-selling also includes using futures or inverse ETFs. A futures contract is an agreement between two parties where it is stated that the buyer agrees to receive a good or security at price x at time t and the seller agrees to deliver the good or security to price x at time t (Fabozzi and Asness, 2004). This means that when the future contract is agreed upon, there occurs no actual transaction other than an agreement. Here, the buyer of a future takes a long position and seller a short position

(Fabozzi and Asness, 2004). If the security price declines compared to the agreed price at time t , then the short position becomes more valuable as it will be possible to sell the security to the buyer at a higher price compared to the market. If one part of the future agreement wishes to liquidate his position, this can be done by offsetting the original position by either buying or selling new futures contracts (Fabozzi and Asness, 2004). As such, futures contracts can be a useful alternative for borrowing securities. Another possibility of short selling is using inverse ETFs. Inverse ETFs seek to replicate the returns of a short position in an underlying index by using various derivatives as the underlying holdings. As the inverse ETFs comprise these various derivatives with expiration dates throughout, the long-term tracking ability of the ETFs however is not as accurate as the ordinary long ETFs. However, the inverse ETFs provide an easier alternative to short positions in ETFs, as these do not require a margin account with a brokerage firm nor require to pay a securities-lending fee. However, the annual expense ratio of inverse ETFs is on average 1%. As such, the ETFs offer more flexibility to short-selling. However, not all of the ETFs included in our sample has an opposite inverse ETF.

For our backtesting, we propose a securities-lending fee equal to the spread between the short-term interest rate - referred to as the money market rate - and the USD Libor overnight rate (OECD, 2020; iborate, 2020). We use the USD Libor overnight rate as it was not possible to retrieve historical data on the broad general collateral rate from before 2014 (FED, 2020). The USD Libor overnight reflects the overnight interbank lending fee and is an uncollateralized rate, however, the USD Libor overnight rate is closely related to the broad general collateral rate (FED, 2020; iborate, 2020).

We assume that we always can provide the full collateral and thus have no need for further loans to provide for any further funding of margin accounts or potential extra collateral. We acknowledge that other ways of short selling are available for both institutional and private investors; however, as the ETFs included are all liquid and trade on large American exchanges we assume that all ETFs can be borrowed.

Lastly, we do not include the returns of the rebate rate in our returns of the pairs trading strategy, as we consider the dollar-neutral position to be in excess of cash

considerations. As mentioned earlier, the returns of the strategy in this paper is interpreted as excess returns.

Expense ratio | The expense ratio is a proxy for the holding costs of an ETF and represents the costs of the ETF supplier of offering the ETF. The expense ratio primarily comprise management fees and other costs related to the maintenance of the ETF fund (Vanguard, n.d.). The annual expense ratio varies in our sample from 0.03% to 1% (see appendix 1). For our analytical framework, we pay the expense ratio for the ETFs measured as a daily holding cost.

6.5. Summary of analytical framework

For our backtest, we conclude a framework that is based on the distance method and the cointegration method. The vast majority of the current literature on the distance method adopts the method proposed by Gatev et al. (2006). The reason for this paper to follow the distance approach presented by Gatev et al. (2006) is to investigate whether this benchmark approach is also applicable to ETFs. For the cointegration method, we have primarily applied a mix of the methodologies proposed by Caldeira and Moura (2013) and Huck and Afawubo (2015). For the general parameters of the trading strategy, we apply a 12-month formation period and a 6-month trading period with recalibration every 6-month. Also, we apply an unrestricted approach towards potential pairings, and the detection and ranking of pairs are carried out based on above mentioned two methods. Different trigger values of the methods are considered throughout the paper, and we do not apply any max holding periods, stop loss functions or a one-day-later rule. Lastly, we apply a dynamic model for transaction costs to capture the practical consequences of pairs trading further. The trading setup of the two methods applied can be found in appendix 6 and 7.

7. Empirical results

The following chapter outlines the empirical results of the methods outlined in chapter 6. Section 7.1 will present the overall findings of the sample period for the two methods and different trigger values. After the results both before and after transaction costs,

further in-depth analysis of the determinants of profitability and the pair composition of the two methods will be conducted in section 7.2. Section 7.3 breaks down the results in four subperiods; 2007-2010 (crisis), 2010-2012 (post-crisis), 2012-2016 (recovery), 2016-2020 (bull market) to further nuance the results of the total sample period. In section 7.4, the strategies will be compared to different factor models to uncover what might explain the nature of these returns. Lastly, we will summarise and consider the practical implications of the findings.

The empirical results presented below will not include the results of Q1 2020, as an alignment of the first-quarter results to the rest of the periods would be misleading to the overall picture of the strategy. The reason is that the 6-month trading period is cut in half, and the trading is therefore not comparable with the rest of the trading periods. An annualisation relying on the first four months would be unreliable, especially taking the market development into account. An assessment of the results generated in the first quarter of 2020 will, therefore, be examined separately in section 7.3.

Annualised results | The results presented in the following chapter are annualised to provide more comparable and useful results. Because the daily returns are expressed as simple returns, the annualisation of the average daily excess return and the standard deviation is calculated as

$$\text{Excess return}_{\text{annual}} = \text{Excess return}_{\text{daily}} * n, \quad (14)$$

$$\text{Standard deviation}_{\text{annual}} = \text{Standard deviation}_{\text{daily}} * \sqrt{n}, \quad (15)$$

where n is the average number of trading days per year over the period of interest (Pedersen, 2015). Similarly, both the skewness and kurtosis have also been annualised and calculated as the following

$$\text{Skewness}_{\text{annual}} = \text{Skewness}_{\text{daily}} * \frac{1}{\sqrt{n}}, \quad (16)$$

$$\text{Kurtosis}_{\text{annual}} = \text{Kurtosis}_{\text{daily}} * \frac{1}{n}, \quad (17)$$

here, n is again the average number of trading days per year over the period of interest (Duc and Schorderet, 2008).

7.1. Results of the entire period

In the following section, we consider the results of the different trigger values for the period of 2007 to 2020. The descriptive results are summarised both before and after transaction costs in table 2 below.

Table 2: Full sample performance (2007-2020)

Open / close trigger (o)	Distance method						Cointegration method					
	2/0	2/0.5	2.5/0	2.5/0.5	3/0	3/0.5	2/0	2/0.5	2.5/0	2.5/0.5	3/0	3/0.5
Average annual openings per pair	6.94	8.69	4.98	5.95	3.79	4.24	8.71	9.75	5.64	6.10	3.76	4.00
Average length of convergence	21.91	15.77	26.06	19.69	27.96	22.73	10.95	8.58	13.54	11.27	16.25	14.00
Before transaction costs												
Annualized mean excess return (%)	6.56	7.92	6.35	7.55	6.73	7.58	19.04	21.62	17.15	18.15	14.75	15.71
Annualized median excess return (%)	1.75	2.19	1.36	1.72	0.95	0.95	7.62	7.70	0.74	0.00	0.00	0.00
Annualized standard deviation (%)	3.85	4.34	4.08	4.57	4.36	4.86	7.11	7.37	7.32	7.50	7.56	7.76
Daily maximum (%)	9.81	11.20	10.28	11.81	10.96	12.74	5.87	5.96	6.16	6.20	6.17	6.21
Daily minimum (%)	-1.40	-1.25	-1.13	-0.88	-1.15	-0.88	-4.34	-4.34	-4.34	-4.34	-4.34	-4.34
Annualized skewness	1.88	1.89	1.87	1.95	1.75	1.90	0.11	0.11	0.11	0.11	0.10	0.11
Annualized kurtosis	4.36	4.42	4.31	4.57	3.95	4.48	0.10	0.09	0.09	0.08	0.08	0.08
Batting ration	0.90	0.92	0.88	0.90	0.87	0.88	0.93	0.94	0.92	0.92	0.89	0.90
Slugging ratio	3.03	2.51	3.72	3.26	4.63	4.24	0.30	0.27	0.35	0.32	0.43	0.40
Sharpe ratio	1.70	1.82	1.56	1.65	1.54	1.56	2.68	2.93	2.34	2.42	1.95	2.02
After transaction costs												
Annualized mean excess return (%)	0.32	-0.44	0.84	0.49	1.46	1.27	2.54	0.78	3.66	2.89	3.91	3.46
Annualized median excess return (%)	-1.19	-1.96	-0.90	-1.18	-0.58	-0.64	-0.68	-1.86	0.00	0.00	0.00	0.00
Annualized standard deviation (%)	2.78	2.92	2.94	3.11	3.17	3.28	6.37	6.45	6.57	6.62	6.92	7.00
Daily maximum (%)	6.68	6.96	6.94	7.25	6.91	7.23	4.81	4.75	4.63	4.63	4.63	4.63
Daily minimum (%)	-1.24	-1.23	-1.21	-1.16	-1.28	-1.22	-4.35	-4.35	-4.35	-4.35	-4.35	-4.35
Annualized skewness	1.62	1.58	1.57	1.62	1.29	1.40	0.07	0.06	0.08	0.08	0.07	0.07
Annualized kurtosis	3.58	3.48	3.41	3.51	2.52	2.82	0.09	0.08	0.08	0.07	0.07	0.06
Batting ration	0.27	0.20	0.31	0.26	0.35	0.32	0.54	0.47	0.61	0.56	0.65	0.62
Slugging ratio	2.89	3.17	3.13	3.31	3.28	3.31	1.18	1.35	0.99	1.12	0.82	0.89
Sharpe ratio	0.12	-0.15	0.29	0.16	0.46	0.39	0.40	0.12	0.56	0.44	0.56	0.49

7.1.1. Results before transaction costs

For the distance method, table 2 shows that the annualised mean excess return before transaction cost is around 7% with the lowest and the highest return is 6.35% and 7.92% respectively. When comparing the annualised mean excess return of the different triggers before transaction costs, trigger 2/0.5 yields the highest return and is characterised by having both a low opening trigger and a higher closing trigger. A result of such a combination of opening and closing triggers, 2/0.5 also generates the highest number of average openings per pair (8.69 openings) and the lowest number of holding days (15.77 days). Oppositely, the worst performance is found in the opening triggers

of 3, which both generate fewer trades and have longer holding periods. These results before transaction costs infer that trading properties favour triggers trading as much as possible and where the holding period is the shortest amount of time. Intuitively this makes sense as the triggers yielding more trades with shorter holding periods perform better before transaction costs, as no lower boundary exists for when a return positively contributes to the aggregated return. For the cointegration method, the best performing trigger yields an annual mean excess return before transaction costs of 21.62%, while this is 14.75% for the worst performing trigger.

Consistent with the findings in the distance method, the highest annualised mean excess return is also generated by trigger 2/0.5 before transaction costs. Here, the annual average number of trades per pair is 9.75 with an average time to convergence of 8.58 days. These trading statistics are noticeably different from the other triggers. The worst performing trigger measured by annualised mean excess return is 3/0, which has on average 3.76 annual trades per pair and 16.25 days of convergence. The impact of the different triggers on the trading statistics is more visible than the distance method, i.e. as the annualised mean excess returns span from 14.75% to 21.62%. Consistent with the assumption from the distance method, it is also evident in the cointegration method that the shorter the holding time, the more trades and thus the better results before transaction costs.

When comparing the returns of the two methods before transaction costs, the cointegration methods yield higher annualised mean excess returns than the distance method, for all triggers. Comparing the annual number of trades per pair and the time of convergence, the cointegration method generates more trading activity. The shorter holding period for the cointegration method indicates that the traded pairs have stronger mean-reverting properties relative to the distance method. The empirical evidence from the comparison between the two methods hereby shows that shorter holding days entails more trades which implies higher returns, when not accounting for the costs of trading.

The returns before transaction costs in the distance method with the 2/0.5 trigger are distributed such that the mean is noticeably higher than the median thus indicating that some large daily returns have been generated throughout the sample period. These properties are confirmed by the maximum daily return of 11.20% against the

minimum return of -1.25%. For the cointegration method, the difference between the mean and median is especially noticeable for opening triggers of 2.5 and 3 where the medians are 0 or close to 0. This is a result of the triggers' lesser days with an open position which naturally results in more days with a return of zero, thus making the median zero. To further understand the distribution of the returns and not just the magnitude, the skewness and kurtosis must be evaluated.

The skewness for the distance method before transaction costs is both positive and above 1, indicating that the distribution is highly positively skewed. This implies that the distribution is characterised by many small values and fewer very large observations. The high kurtosis shows the same picture for the distance method. Given the highly positive skewness, the kurtosis indicates that extreme observations are most probable having large positive values. This implication is consistent with the fact that the distance method has a high maximum and not that low minimum. These findings imply that there exist few days with large positive returns, but the method to a larger extent produces many small returns.

For the cointegration method, the skewness and kurtosis are both below 0.5 and positive. This indicates a higher degree of symmetry with a small positive skewness, indicating similar tendencies as the distance method but more moderately distributed.

When the comparison across triggers and the two methods are carried out, the risk-reward relation measured by the Sharpe ratios is useful to determine whether a higher excess return is more attractive in relation to the associated risks of the returns (Pedersen, 2015). By using the Sharpe ratio, two strategies that yield different returns can more easily be compared in connection with their respective risks, as the attractiveness can better be ranked. Before transaction costs, the best trigger combination in the distance method yields a Sharpe ratio of 1.82 (trigger 2/0.5) while the worst yields a Sharpe ratio of 1.54 (trigger 3/0). For cointegration, these numbers are respectively 2.93 (trigger 2/0.5) for the best and 1.95 (trigger 3/0) for the worst. Despite having higher standard deviations, the cointegration method outperforms the distance method due to the proportional higher excess return as well. In order to make the magnitude of the Sharpe ratios more relatable and verify these results, we compare our results to the existing literature on their application with single stocks. Huck and Afawubo (2015) achieve an annualised Sharpe ratio of 1.53 for their cointegration method with

a trigger of 3/0. Caldeira and Moura (2013) obtain a similar Sharpe ratio of 1.34 for their cointegration method. For the distance method various Sharpe ratios before transaction costs are obtained as the following; Andrade et al. (2005) get 1.11, Do and Faff (2012) get 1.05, Gatev et. al. (2006) get 0.85, Rad et al. (2015) get 0.75, Huck and Afawubo (2015) get 0.22 and Smith and Xu (2017) get 0.5. Based on these results, the Sharpe ratio generated in this paper for both the distance and cointegration method appear to provide a more efficient risk-return relation than the existing literature on single stocks. Even though the Sharpe ratio is applicable for comparison, the direct comparison must be considered with some reservations as the Sharpe ratio is a relative term that is not independent of the time period in which it is calculated (Sharpe, 1994). Nonetheless, it still gives an indication of the relative performance of the different studies, and the results of this paper are still noticeably higher measured by the Sharpe ratio. In order to determine whether it is the risk or the return which is the driving component of the higher Sharpe ratio of this paper relative to current literature the two elements must be examined separately. The annualised mean excess of the current literature applying the distance method is as the following; Do and Faff (2010) generate 10.2%, Do and Faff (2012) obtain 12.5%, Gatev et al. (2006) obtain the highest with 17.2% and Rad et al. (2015) generate 10.9%. Comparing these results to what is achieved in this paper, all trigger combinations for the distance method produce lower excess return before transaction costs than the results of the existing literature. When further comparing the annualised standard deviations of this paper, which are in the range from just below 4% to just below 5%, to the existing literature, this paper lies below what else has been achieved. From these results, we can conclude that for the distance method, the higher Sharpe ratios of this paper, relative to current literature, is driven by the relatively lower risk aspect which offsets the lower excess return. For the cointegration method, Rad et al. (2015) obtain a 10.2% annual average excess return, Caldeira and Moura (2013) get 16.4%, while Huck and Afawubo (2015) produce higher results of 25%. The excess returns of this paper vary depending on the triggers, but where the lowest lies above that achieved by Rad et al. (2015) and the best performing below that of Huck and Afawubo (2015). For the cointegration method, it can thus not be stated whether it is the risk or return generating the higher Sharpe ratio. As the results of the Sharpe ratios are outperforming both Huck and Afawubo (2015)

and Rad et al. (2015), but the returns of the triggers are within the same range, the reason for the higher Sharpe ratios must also be lower standard deviation. These findings underline why it is essential to consider the generated returns relative to the associated risk when comparing different strategies.

When the number of trades and the average time to convergence is compared to the existing literature, noticeable differences are apparent. The average time of convergence per trade for Gatev et al. (2006) is 57.5, with an annual average of 3.92 openings per pair. For the same 2/0 trigger of this paper, the average time to convergence is 21.9 days with an annual average of 6.94 openings per pair. The 2/0.5 trigger achieves the highest number of trades with the lowest time of convergence in this paper with an average length of convergence of 15.8 days and 8.7 openings. Comparing the results of the cointegration method with Huck and Afawubo (2015), a similar pattern is apparent. For the trigger of 2/0, Huck and Afawubo (2015) hold an open position in 35.5 days on average per trade with on average 3.2 annual trades per pair. These numbers are for this paper 11 days for a pair to converge with 8.7 trades per pair on average. These results infer that when ETFs are applied in a pairs trading strategy, a noticeable quicker time of convergence is achieved and more trades are carried out compared to the existing pairs trading literature on single stocks.

It must be emphasised that analysing the results of the trading strategy before transaction costs has little practical implications, as it implies that there exist no transaction costs. Thereby, if we compare the results to the general market or another benchmark, we would overestimate the profitability of the two methods in comparison to a long-hold strategy. Therefore, we cannot draw any definitive practical conclusions but can only infer the observed results in comparison to the current literature that also considers returns before transaction costs. Here, the comparison to the existing literature might actually be less noisy before transaction costs as the assumptions about the associated costs of trading differ across authors. However, the results considered before transaction costs might also be exposed to the bid-ask bounce resulting in potentially upward biased estimates. This potential bias is, on the other hand, equally present for the compared results of the current literature which maintain the legitimacy of the comparison.

To summarise the findings of the results before transaction costs, both the distance and the cointegration methods provide positive results with high Sharpe ratios across all triggers. Comparing the Sharpe ratios of the two methods internally, the cointegration method consistently outperforms the distance method. When comparing the distance method and the cointegration method to existing literature, both methods of this paper obtain larger Sharpe ratios than what else has been reported.

As the assessment of results before transaction costs only to a limited extent provide useful insights into the practical performance and profitability of the methods, the returns must be considered after transaction costs. After transaction costs the risk of an upwards bias due to the bid-ask bounce is eliminated, and the results will provide an understanding of the practical implications of the strategies.

7.1.2. Results after transaction costs

Table 2 displays a considerable reduction in the profitability for both methods after introducing transaction costs. For the distance method, the two best performing triggers before transaction costs, 2/0 and 2/0.5, are now reporting an annualised mean excess return of respectively 0.32% and -0.44% after accounting for the costs of trading. In opposition to the results before transaction costs, the 3/0 and 3/0.5 triggers are now those that perform the best in terms of annualised excess return. The best performing trigger for the distance method yields an annualised mean excess return of 1.46% and is generated by trigger 3/0. Recalling the concluding remarks derived from the findings before transaction costs with the notion that more trades and shorter holding periods yield higher returns is not true after transaction costs. A striking observation to emerge from employing transaction costs is that the best trigger before transaction costs now is the worst-performing strategy and vice versa. The 3/0 trigger has, on average, 3.79 annual openings which is less than half the amount of the 2/0.5 trigger with 8.69. This means that the findings before transaction costs have turned upside-down when considering the results after transaction costs.

For the cointegration method, a similar development is apparent. The 2/0 and 2/0.5 triggers experience a decline in the annualised mean excess return from respectively 19% to 2.5% and 21.6% to 0.8% as a consequence of the transaction costs. The best performing triggers of the cointegration method are after transaction costs 2.5/0 and 3/0 with annualised mean excess returns of respectively 3.66% and 3.91%. When

comparing the distribution of the returns after transaction costs to those before transaction costs, it is broadly the same tendencies regarding the median, skewness and kurtosis. The skewness and kurtosis for the distance method is still rather high, as especially the daily maximum value is quite extreme. The distribution of the cointegration method is slightly more symmetric measured by the skewness. Most noticeably, all triggers have negative or zero medians which underline the positive skewness of the distribution.

Consistent with the reduction in the returns, the Sharpe ratios have experienced a comprehensive decline relative to before costs. For the distance method, trigger 3/0 and 3/0.5 yields the highest Sharpe ratios of respectively 0.46 and 0.39. For the cointegration method, trigger 2.5/0 and trigger 3/0 generate an almost identical Sharpe ratio of 0.56 with 3/0 being the highest. The results after transaction costs are more comparable with a long-short position in the market. Here, we consider the excess return of the market portfolio, which generates a Sharpe ratio of 0.50 in the corresponding period (Appendix 12; French, n.d.). By comparing the Sharpe ratios, both trigger 2.5/0 and 3/0 in the cointegration method outperforms the general market. However, for the distance methods, no trigger combinations provide a Sharpe ratio above the excess market returns.

Comparing our results after transaction costs to the existing literature is somewhat more difficult as the effects of transaction costs and assumptions hereof are different from author to author. Do and Faff (2012) which is seen as one of the most prominent papers on transaction costs report a decline from 1.05 to 0.28 in the Sharpe ratio for their best strategy when accounting for transaction costs. Rad et al. (2015) and Gatev et al. (2006) obtain Sharpe ratios of 0.3 and 0.59, respectively, when accounting for transaction costs. For the cointegration method, only Rad et al. (2015) provide results after transaction costs. They report a Sharpe ratio of 0.35 after transaction costs, with the best Sharpe ratio of this paper being 0.56 for the 3/0 trigger. Strikingly, from 2000 to 2015, Rad et al. (2015) only obtained positive returns in the period from 2007-2009. The diminishing profitability of pairs trading is consistent with the findings of Gatev et al. (2016), Do and Faff (2010; 2012), Smith and Xu (2017), and Broussard and Vaihekoski (2012). Smith and Xu (2017) cannot conclude that any of their results are significantly different from zero after including transaction costs in the period from

2000-2014. Broussard and Vaihekoski (2012) obtained a Sharpe ratio of 0.31 from 1998 to 2008 and Do and Faff (2010) a Sharpe ratio of 0.32 from 2003-2009. The results of these papers and the thorough analysis of more than 23,000 stocks by Rad et al. (2015) implies that the profitability of pairs trading from 2000 is scarce. We cannot fully state whether our model yields better results than the existing literature, nor do we seek to diminish the results of the existing literature. Nevertheless, based on above results, the results of the cointegration method appear superior to the declining trend in the profitability of stocks-based pairs trading, the excess return of the market and the results of the current literature.

Summary | We have in the above outlined the overall results of the applied pairs trading strategy with ETFs. Generally, the returns before transaction costs appear prominent with promising results compared to the existing literature. The results after transaction costs are remarkably lower, and the optimal triggers are those producing the fewest trades with longer holding periods which is in complete opposition to the results before transaction costs. As a result of employing transaction costs, no more than two triggers, both from the cointegration method, outperforms the excess market portfolio for the same period when measured in Sharpe ratios. Whether our excess returns are statistically significantly higher than the market portfolio will be further investigated in section 7.4. Furthermore, the overall results obtained in this paper surpass those reported in the existing literature after transaction costs. The findings further underline the importance of considering different trigger values when applying a pairs trading strategy, as the results of the two methods vary significantly from trigger to trigger.

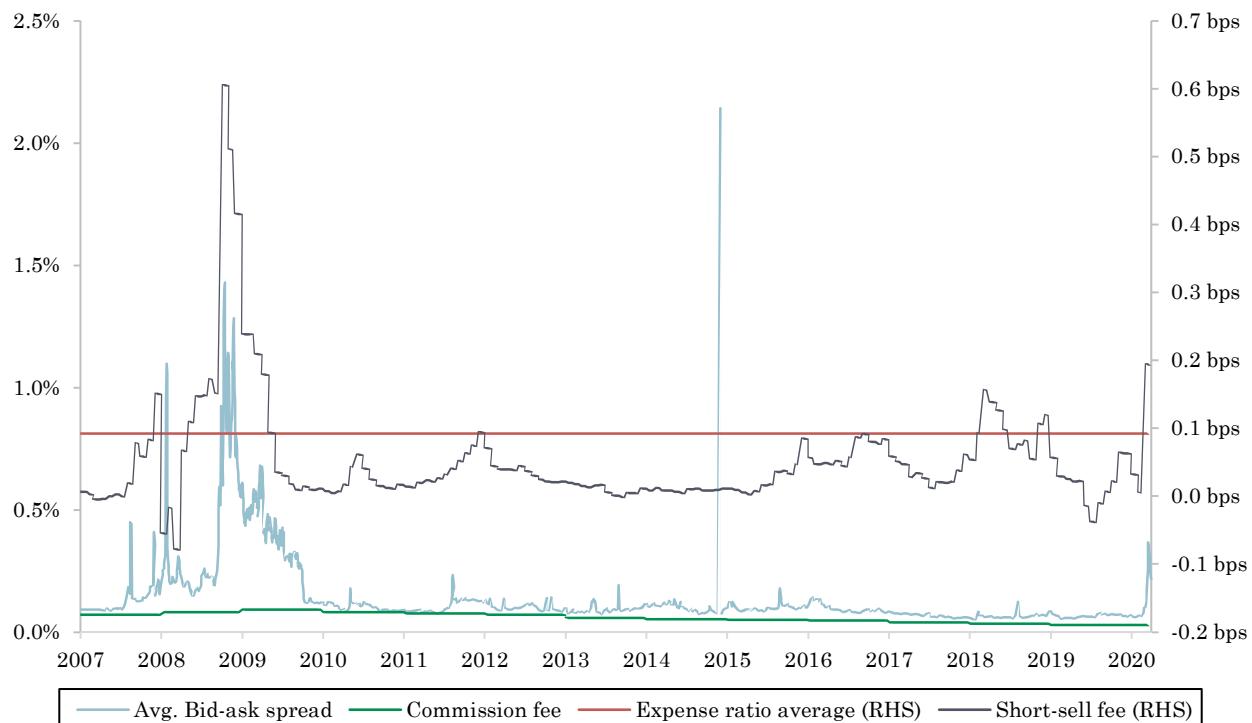
7.2. Determinants of returns

In section 7.1 we outlined the results for the entire sample period. In the following section we will examine the key determinants of pairs trading that have either a direct or indirect effect on the returns of the different triggers to understand the robustness and determinants of these returns.

7.2.1. Effect of transaction costs

From the assessment in section 7.1, it is found that the optimal trigger combination changed when accounting for transaction costs. The primary reason for the shift in profitability is the one-off transaction costs comprising commission costs and bid-ask spread. These costs are much more impactful than the holding costs, which are expense ratios and short-sell costs. This implies that triggers associated with more trades to a larger extent will be affected by the one-offs. On average across all triggers, one-off transaction costs represent 95% of the associated costs of trading for each trade (for the cointegration method: 62% bid-ask spread, 34% commission, 1% short-sell costs, 3% expense ratio) (Appendix 9). The heavily skewed composition of one-off transaction costs explains the shift to higher trigger values after the deduction of these costs. When looking at the transaction costs over the historical period from 2007-2020, the impact of the costs vary across the different trading periods, as illustrated in figure 4.

Figure 4: Average transaction costs (2007-2020)



The figure demonstrates that commission fees have experienced a continuous decline from 7 bps to 3 bps for institutions (ITG, 2019). At the same time, the bid-ask spread exhibits fluctuations in shorter periods, most noticeable in times of high volatility such as 2007-2009. Also, the 2nd of December, 2014 recorded the highest average bid-ask

spread (2.14%) as a result of falling oil prices and economic instabilities in Greece (Fletcher, 2014).

For the short-sell costs, there is a similar trend to that of the bid-ask spread. However, as the short-sell costs are reflected by respectively the money market rate and the LIBOR overnight rate, the trend of the short-sell costs are to a higher degree caused by the general trends in the interest rates. As it is also shown in figure 4, there are periods where the short-sell costs are negative meaning that a borrower receives a return for short-selling. Arguably, the lender would most often ask for a negative rebate rate, however, in other cases, the credit risks of short-sell-lending could potentially make the short-lending a cost for the lender (Pedersen, 2015). Since the drops are immediate and we do not wish to impose any arbitrary assumptions on the characteristics of short selling, the latter is assumed to be the case.

When comparing our transaction costs to the existing literature, the majority applies fixed transaction costs (Caldeira and Moura, 2013; Do and Faff, 2010; 2012; Gatev et al., 2006; Gregory et al., 2011; Huck, 2013; Huck and Afawubo, 2015; Perlin, 2009; Smith and Xu, 2017). The average one-way transaction costs per trade is 0.28% for the existing literature (appendix 8). In comparison, the average one-way transaction costs for this study is 0.22% when assuming a holding period of 15 trading days (see appendix 9). Do and Faff (2012) find in their investigation that transaction costs have been declining from 1963 to the end of their sample period in 2012. In the period 1963-1988, the authors obtained an average one-way transaction cost of 81 bps while this is 28 bps for the period of 2009-2012 (Do and Faff, 2012). When comparing to Do and Faff (2012), the lower costs in our sample are due to the declining commission costs from 7 bps to 3 bps, and an average bid-ask spread for all ETFs over the sample period of 14 bps compared to 20 bps for the stocks investigated by Do and Faff (2012). As such, the lower costs can be explained by the declining trend in transaction costs and by applying ETFs rather than single stocks. Further, figure 4 illustrates that applying fixed one-way transaction costs, like most of the existing literature, will either under-estimate or overestimate the costs of trading.

Our findings are thus in line with the existing literature, arguably slightly lower due to the declining prices and utilisation of ETFs (Huck and Afawubo, 2015; Do and Faff,

2012; Rad et al., 2015; Smith and Xu, 2017). We believe that imposing a dynamic approach improves the practical robustness of our strategy.

7.2.2. Batting and slugging

In addition to considering the profitability after transaction costs, how often the methods are right or wrong when a trade is initiated, together with how right the methods are when right should also be considered. These processes are important as they fundamentally tell us whether it would have been better not to trade at all.

The process of doing so is referred to as batting and slugging and originally derives from baseball (Hakes and Sauer, 2006). In econometrics terms, this refers to the amount of trades that yield a positive return, and the average positive return versus the average negative return when opening a position (Heritage, 2010). In this paper, the slugging ratio is calculated as the average return on a position generating a positive return over the average return on a position generating a negative return presented in absolute terms. With this, the profitability of the strategy can be improved by either increasing the number of winning trades (batting) or increasing the return on position generating positive return (slugging) or both.

Batting ratio | When considering the results of the batting ratio before transaction costs, the cointegration method yields slightly stronger results for all triggers relative to the distance method with trigger 2/0.5 as the best trigger generating 94% positive trades. This is consistent with the findings of Smith and Xu (2017) reporting similar results before transaction costs.

After transaction costs, the batting averages drop considerably. In the cointegration method, trigger 3/0 generates a ratio of 65%, which is the highest batting ratio after transaction costs. Despite the considerable reduction in the batting ratio after introducing transaction costs, only one of the six triggers (trigger 2/0.5) with the cointegration method generates a batting average below 50%, which is a sound basis. Most strikingly in the examination of the batting ratios, all triggers for the distance method drop to below 35% when accounting for transaction costs. In other words, only around every third trade yields a positive return after transaction costs across the entire sample period. Comparing these results to Rad et al. (2015), the authors experienced a batting

ratio of 71% for the distance method and 69% for the cointegration method thus almost two times higher than the results of this paper for the distance method. Even though the results of Rad et al. (2015) are based on a wider time horizon, the comparison still indicates noticeable different trading attributes for the distance method from this paper. Further, the results of the batting averages show some noticeable differences in the trading attributes of the two methods.

Slugging ratio | When comparing the results of the average return on positive over negative openings, the cointegration method is noticeably different from the distance method before transaction costs. Despite having a high batting ratio of 95%, the slugging ratio is at a low level around 0.3 and 0.4 before transaction costs. This implies that the few times a trade yields a negative return, it has a large impact on the profitability. The distance method yields better results, with a slugging ratio of 2.5 as the worst of the six triggers before transaction costs.

After transaction costs, the slugging ratio of the cointegration is more moderate, as it has increased to around one across the various triggers. Noticeably, triggers 3/0 and 3/0.5 have a higher negative average than positive average, but at the same time, a higher offsetting batting ratio. For the distance method, the returns of the positive trades are about 3x that of the negative trades across all triggers after transaction costs. The negative return on trigger 2/0.5 after transaction costs is thus a result of the slugging ratio is not able to compensate for the low batting ratio.

As touched upon in the introduction of this section, the two concepts of batting and slugging are interlinked. A high batting ratio or a high slugging ratio independently does not necessarily imply high returns, but the right combination might do. In light of this, the higher batting ratio of the cointegration method suggests a more reliable strategy due to the higher winning ratio compared to the distance method. However, the slugging ratio must also be taken into consideration as this reveals a much higher average positive return relative to the negative return for the distance method compared to the cointegration method.

7.2.3. Pair composition

The above sections have highlighted a number of determinants for the difference between the generated returns both before and after accounting for transaction costs for the two methods. To further determine what is causing the different returns generated for the two methods, the following section will shed light on the composition of the pairs included in the two methods.

For both methods, the fundamental idea is to find pairs exhibiting co-movement, with this process differing between the two methods. When introducing ETFs to pairs trading, two securities can exhibit identical price development which is very different from stocks. Whereas it is often a challenging task to identify stocks displaying similarities and co-movement, it is considerably easier for ETFs. As we have taken an unrestricted approach to our data sample, ETFs in our sample also include ETFs tracking both the same or similar indices. One example of a traded pair in our sample that tracks the same index is SPDR S&P 500 ETF (ETF1) and iShares Core S&P 500 ETF (ETF28). Other pairs showcasing similarities but not necessarily tracking the same index could be the pair of iShares Russell 1000 ETF (ETF29) and Vanguard Total stock market ETF (ETF60). For this pair, even though the two ETFs track different indices, the two indices are closely interlinked and will, to a large extent, exhibit the same pattern. With the above scenarios being potential outcomes from both the distance method and the cointegration method, the pairs trading strategy could essentially become a strategy trying to exploit mispricing and tracking errors rather than a strategy based on divergences from identified long-run relationships. On the other hand, it can also be argued that these pairs have an identical or close to identical long-run mean, implying that the pairs trading and mispricing arbitrage strategy are two sides of the same coin. For the distance method, a total of 104 different pairs out of 540 potential spots (20 pairs * 27 trading periods) are traded throughout the sample period (Appendix 10). This means that every pair is traded 5.2 times throughout the trading period. From this it becomes clear that several pairs have a consistent low level of sum of squared deviations. Of the 104 ETFs, the top 3 pairs that are traded the most throughout the sample period are the following; SPDR S&P 500 (ETF1) vs iShares Core S&P (ETF28) is traded in all trading periods, iShares Russell 3000 ETF (ETF38) vs Vanguard Total stock market ETF (ETF60) is traded in 23 trading periods. SPDR S&P MidCap 400

ETF (ETF2) vs Ishares Core S&P Mid-Cap ETF (ETF35) is utilised in 21 trading periods (Appendix 10). As the pairs are detected and ranked based on the minimisation of squared deviations, the distance method tends to favour pairs of ETFs that track the same index. Here, 52% of the 540 potential spots in the full sample period comprise pairs of ETFs that track the same index.

Furthermore, out of the 540 traded pairs, 75% of these pairs comprise two ETFs both set to track US Large Cap indices, 7% of the pairs both track Small Cap indices, which is 4% for Foreign large Cap Blend and 14% of other and mixed categories (Appendix 10). The fact that 52% of the pairs comprises ETFs both tracking US large-cap indices is also manifested in the standard deviation of the traded pairs' spread in the formation periods. Across all periods, the corresponding value to a 2x standard deviation opening trigger for the distance method is on average 0.003 (Appendix 10). For comparison, Gatev et al. (2006) report an average opening trigger of 0.053 and Papadakis and Wysocki (2007) present a similar trigger value of 0.057. Above shows that using ETFs provides a high degree of "closeness".

The cointegration method trades 240 different pairs over the entirety of our sample period (Appendix 11). Compared to the distance method, this is more than a doubling in the composition of traded pairs. Out of the 540 potential spots, the number of pairs that comprise of ETFs that track the same underlying index is 35%. Further, out of the 540 potential spots, 46% of the pairs comprising two ETFs both tracking US Large Cap indices. These numbers are 7% for US Small Cap indices and 3% for Foreign Large Cap indices. This leaves 44% of the pairs to comprise ETFs tracking different index categories (Appendix 11). Based on this, it can be argued that the cointegration method yields more diverse and nuanced pairings that are not as limited to large-cap ETFs that track the same underlying index as the case of the distance method. The simpler approach of the distance method only considers the relation between the two securities, but does not consider any explanatory pattern between the two. Recall the three pairs that were traded the most in the distance method; these exact same pairs are respectively traded 16, 2 and 7 times in the cointegration method (Appendix 11). These results suggest that even though two ETFs exhibit a closeness determined by the distance method, it does not necessarily indicate that they are cointegrated. The distance method's assumption that the best information about the relationship between the two

securities is the direct relation at time t might be misleading as there could be additional explanatory information in the historical relationship. This matter is considered in the cointegration method, which is why the detection of pairs can be considered somewhat more nuanced.

The results of this section provide insights into the composition of the selected pairs in the two methods and reveal several key attributes of the two methods. However, we cannot fully draw any conclusions about the impact of the pair compositions on the profitability of the strategies from above. Hence, the next section will break down the sample period in subperiods to understand the impact of the pairs composition on the robustness and profitability of the two methods throughout the sample period.

Summary | The above sections highlighted several important attributes that must be considered and understood when conducting pairs trading. We showed that the reason for the higher trigger values deem the best results after transaction costs is due to the skewness of the costs towards one-off costs, i.e. bid-ask spreads and commission costs. After that, the batting and slugging ratios showed noticeable differences in the underlying nature of the two methods. Where the distance heavily relies on a few large returns, the cointegration method showcases more moderate returns in general. The pair composition reveals some important properties of the two methods on both profitability and trading. The distance method tends to provide a much less nuanced combination of pairs with a high composition of pairs with ETFs both tracking the same index or at least indices both exposed to US large-cap stocks. Despite the cointegration showcasing some similarities, the method has a more diverse composition of different pairings with less reliance on pairs with ETFs tracking the same underlying index.

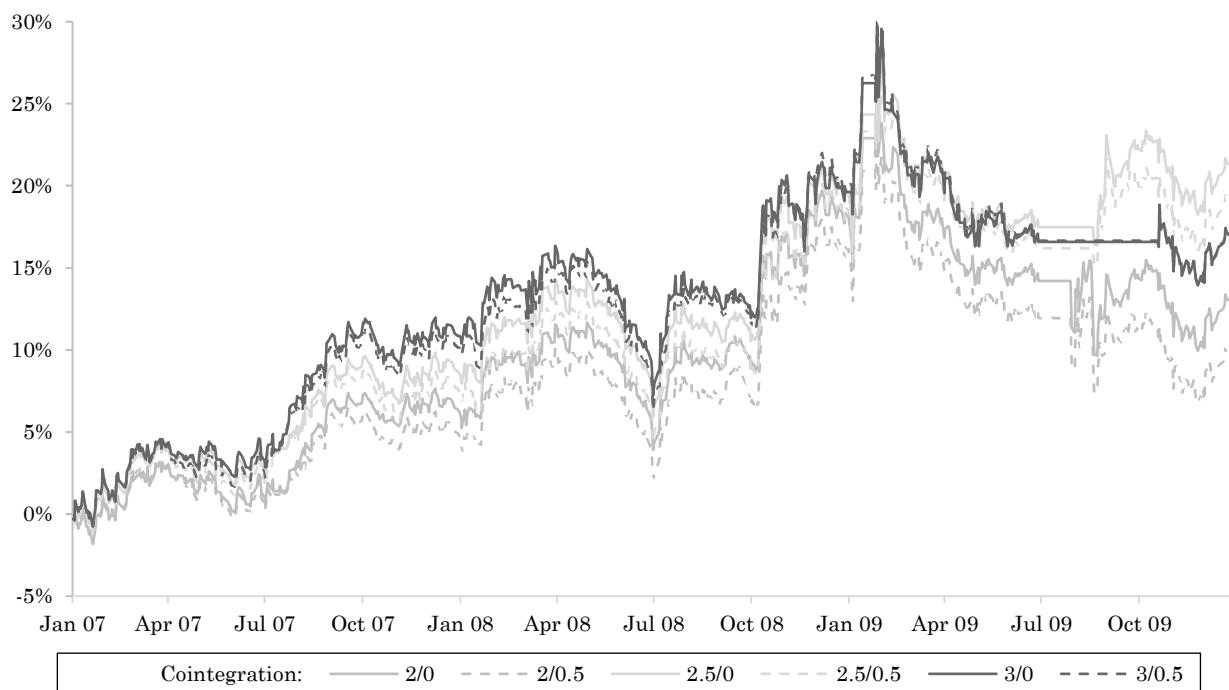
7.3. Sub period analysis

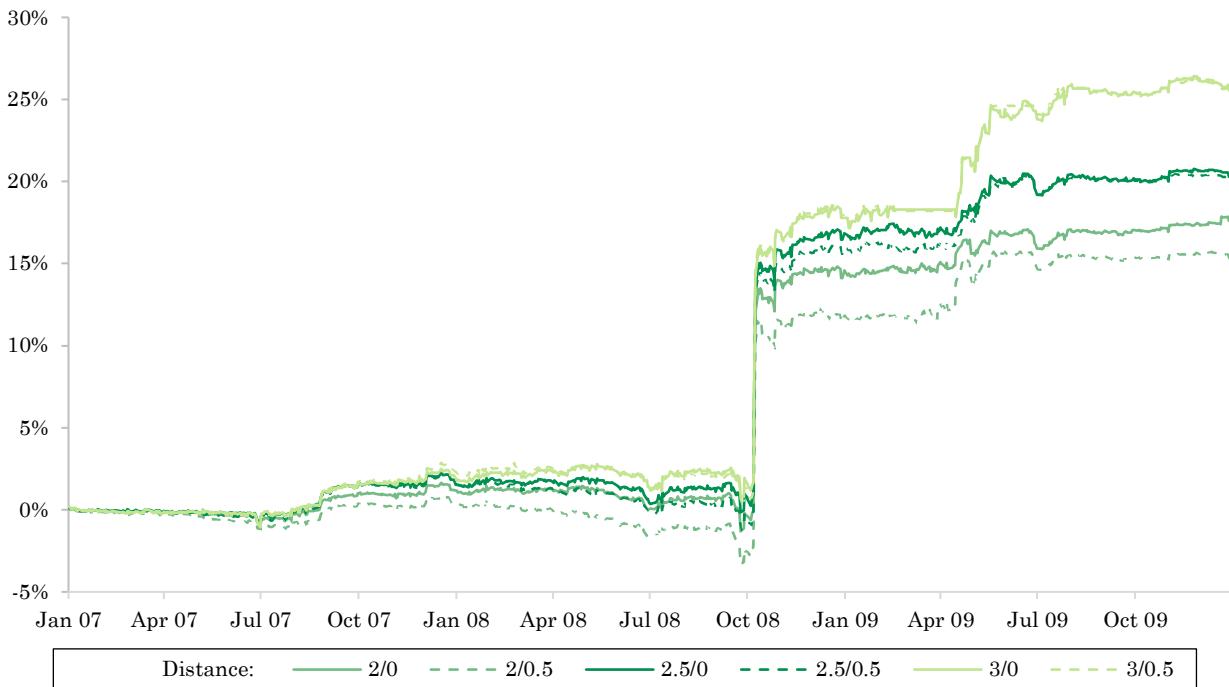
Prior sections have provided an overall assessment and understanding of the dynamics of the two different pairs trading methodologies for the entire period. With this understanding established, this section will investigate the performance and characteristics in different subperiods under different levels of market volatility. The aim of this section is to provide a more nuanced understanding of the two methods and an identification of when and how the returns are generated.

7.3.1. Pairs trading from 2007 to 2010

The first trading period from 2007 to 2010 is characterised by a period of high volatility caused first and foremost by the financial crisis that started to send ripples through the financial sector mid-2007 (CBOE, n.d.; Pedersen, 2015). To illustrate the performance of the two methods and their various triggers, figure 5 showcases the cumulative returns across the triggers after transaction costs. Further, see appendix 13 for descriptive statistics for the subperiods.

Figure 5: 2007-2010 cumulative returns for triggers after transaction costs





This subperiod yields, in general, the best return of the sample period. Before transaction costs, the distance method generates an annualised mean excess return in the range of 19-24%. In contrast, the annualised mean excess return for the cointegration method is in the range of 20-25%. For the distance method, the annualised mean excess return before transaction cost is around three times the value of the entire period across all triggers. This provides a clear indication of this subperiod being the main driver of the overall performance of the distance method. When taking the transaction costs into account, this subperiod still performs relatively better than the performance for the entire period with an annualised mean excess return of 5-8% for the distance method and 3-7% for the cointegration method.

As a consequence of transaction costs, the batting ratio of the distance method drops from around 94% to a range of 45 to 60% (Appendix 13). Comparing these figures to the results of the entire period, the batting ratio is materially higher and with an approximately similar slugging ratio to that of the entire period, it is evident that the profitability of this period is superior. This is equally the case for the cointegration method. Further, the distance method obtains a Sharpe ratio of 1.34 after transaction costs for the best trigger, 3/0, which is noticeably higher than the overall period (Appendix 13). These results again show that this period positively contributes to the total risk-return reward for the entire sample period.

The performance of the cointegration method after transaction costs generate similar annualised mean excess return but with a noticeable higher standard deviation resulting in a Sharpe ratio of 0.68 for the best trigger, 2.5/0 (Appendix 13). This is amongst others explained by the batting and slugging ratios. Despite high batting ratios after transaction costs, the low slugging ratio indicates that there exist some highly negative returns.

Periodical characteristics | When breaking the period even further down after transaction costs, several interesting findings arise when considering the results. For the cointegration method, returns began to increase at the beginning of August, at the same time as the so-called quant crisis of 2007 (Pedersen, 2015; Pastor and Stambaugh, 2019). The quant crisis began when the first indicators of the subprime crisis started to affect quant algorithms of hedge funds (Pedersen, 2015). The effects eventually led quantitative strategies to initiate sales of high-expected-return stocks and purchase of low-expected-returns to close down short positions (Pedersen, 2015). Here, especially value, momentum and quality style strategies were impacted the most (Rao, 2017) Despite the hedge funds having different strategies, it was in broad terms the same securities that were being sold (Pedersen, 2015). The crisis lasted throughout August before regaining traction again at the end of the month. Noticeably, the general market was up by 1.5% in the same period (Pedersen, 2015). When looking at the main drivers of the returns in both methods, two takeaways can be concluded. Firstly, one ETF is repeated in the most profitable trades of both the distance method and the cointegration method, namely Vanguard S&P 500 ETF (ETF 107). For the cointegration method, ETF107 was paired with two total market Cap ETFs; S&P total US stock market (ETF79) and S&P 1500 Composite (ETF56), and two large-cap ETFs; Vanguard Large-Cap ETF (ETF80) and iShares Core S&P 500 ETF (ETF28) (Appendix 11 and Appendix 13). These four pairs produced a large proportion of the positive returns in 2007. For the distance method, ETF107 vs ETF56 and ETF107 vs ETF80 were, as in the cointegration method, the main drivers of profit generated in 2007 (Appendix 13). Secondly, two of the most profitable pairs for the cointegration method in addition to above comprise respectively MSCI Singapore (ETF8) vs S&P 500 Quality ETF (ETF112) as well as Russell 2000 ETF (ETF31) vs Value Line Dividend ETF (ETF76)

(Appendix 13). Having established that it was primarily value and quality stocks that were impacted by the quant crisis, it is no surprise that pairs comprising one of these ETFs produce positive returns. The volatility and movements in the market as a consequence of the quant crisis have thereby contributed to deviations in the established price relations, which form the basis of the profitability in 2007.

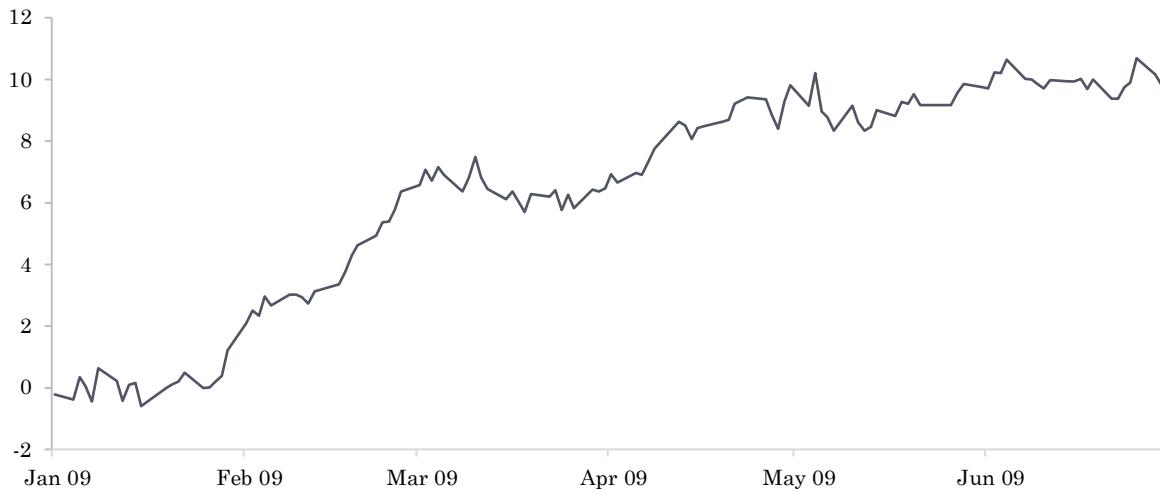
The second half of 2008 is by far the best trading period throughout the entire sample period (Appendix 12 and Appendix 13). This is also illustrated in figure 5 above where the returns of the distance method increased extraordinarily for all triggers in mid-october 2008. The same increase is seen for the cointegration method, however with a smaller magnitude relative to the distance method. The high returns are driven by the large inefficiencies and price drops in the market in the aftermath of the Lehman Brothers crash on the 15th of September, 2008 (Kingsley, 2012). When identifying the traded pairs during this period in both methods, the S&P 1500 Composite ETF (ETF56) is included in all of the most profitable trades (Appendix 13). The corresponding ETFs to ETF56 in the profitable pairs include Vanguard total market ETF (ETF60), iShares Russell 1000 and 3000 ETF (ETF29 and ETF38) and different S&P 500 ETFs (ETF28, ETF107) (Appendix 13). As such, the profit is generated from more fundamental market anomalies between large-cap ETFs and broad US equity markets ETFs or two ETFs tracking different degrees of the broad US equity market.

One of the reasons that the distance method outperforms cointegration in this trading period from mid to ultimo 2008 is amongst others due to the method's favouring of pairs comprising ETFs both tracking large-cap indices which due to the noisy anomalies yield high returns with low risk. Noticeably when comparing the profit generated by the two methods, the distance method generates almost all of the positive return within a week of trading in October 2008, illustrated by the steep increase in figure 5, while the cointegration method generates the profits throughout the remainder of the trading period ending ultimo 2008.

In 2009, the returns were more volatile with both high and very low returns with no immediate pattern in pairs generating either negative or positive returns. However, we find examples, such as ETF134 and ETF86 during the first half of 2009, where a z-score reaches 10 standard deviations away from the mean which is illustrated in figure 6. In this example, Vanguard Mid-Cap ETF (ETF86) and First Trust Dow Jones

Internet Index Fund (ETF134) continue to diverge until the position is closed at the end of the trading period resulting in great losses. This could, to some extent, have been mitigated by imposing either a stop loss or a maximum holding period, which is not introduced in our strategy.

Figure 6: Z-score of ETF134 and ETF86 (H1 2009)

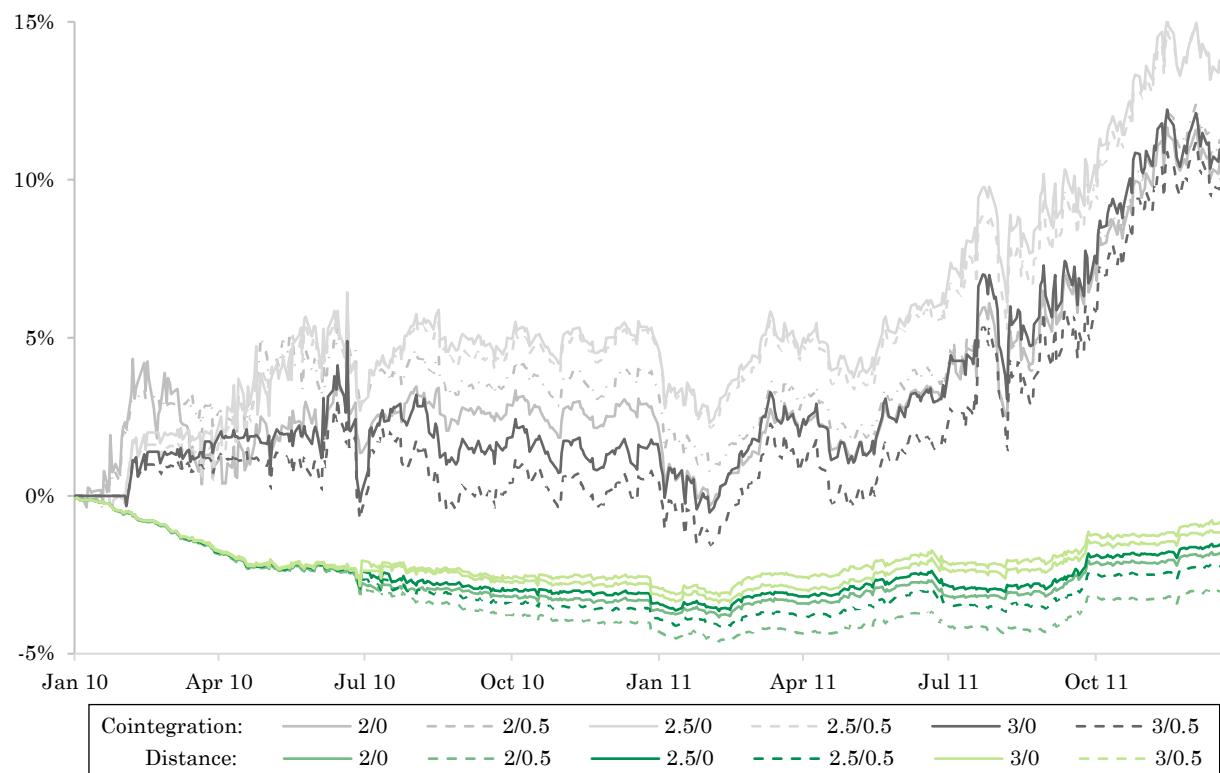


Consistent with the findings of the existing literature and in the history of pairs trading, the highly volatile period from 2007-2010 has been found to display some of the most lucrative conditions for the pairs trading strategy, as a consequence of the market neutrality of the strategy and the exploitation of the noisy anomalies caused by panic in the market (Thorne, 2003; Do and Faff, 2012). Because pairs composition of both methods include large or total market ETFs, both strategies can profit from these fundamental and extensive anomalies in the market, while enjoying the high mean-reverting properties of these ETFs described in chapter 5 and earlier sections. This is especially the case for the distance method with many of the pairs comprising ETFs tracking the same index. From the subperiod breakdown, it is noticeable that ETF107 and ETF 56 were both very active in 2007 and 2008 with high market volatility and high returns generated for the two methods. Both of these ETFs were included in a lot of traded pairs, giving rise to some criticism of both methods and pair composition, but especially the nature of the distance method generating many similar pairs. Whereas the many pairs with ETF56 turned out profitable, the opposite case could potentially arise, causing negative results in the portfolio due to an adverse development in just one ETF.

7.3.2. Pairs trading from 2010 to 2012

This subperiod is characterised by a more calmed financial market covering the aftermath of the financial crisis. However, this subperiod lasting from 2010 to 2012 still experienced two shorter periods of market turmoil with a VIX reaching a level around the 40s (CBOE, n.d.). Compared to 2008, this is a lower level, but still higher than the average VIX for the full period (CBOE, n.d.). The subperiod is set to investigate how well the pairs trading works under post crisis circumstances where anomalies and inefficiencies might still influence the pricing of the ETFs. Below figure 7 is illustrating the cumulative return of the two methods for the different triggers in the subperiod. See appendix 14 for descriptive statistics of the subperiod.

Figure 7: 2010-2012 cumulative returns for triggers after transaction costs



When comparing this period to the previous, the annualised mean excess return for the distance method has declined to a range of 5-6% before transaction costs. The cointegration method has not been subject to such a profound decline, thus yielding a return before transaction costs in the range of 13-20%, slightly lower compared to 2007-2010 (Appendix 14).

Despite the lower returns for the distance method, the risk-reward relation has been improved notably before transaction costs for this period relative to the previous. With the annualised mean excess return before transaction cost being lower for this period than for both the entire and previous period, the determinant of the high Sharpe ratios is the standard deviation. The annualised standard deviation for this period is at a level of around 1%, which is remarkably lower than the previous subperiod's level around 8% - 9% (Appendix 13 and Appendix 14). When the costs are deducted from the gross return, the annualised mean excess return for every trigger value turns negative for the distance method. This is a result of a batting ratio of less than 40% regardless of the trigger. Whereas the low batting average was offset by a high slugging ratio in 2007-2010, the slugging ratios for this subperiod is around or below 1. These results suggest that the lower annualised mean excess return before transaction costs are not high enough to offset the transaction costs associated with the execution of the pairs trading strategy for the distance method.

For the cointegration method, despite the lower annualised mean excess return before transaction costs compared to the previous period, this period has a more stable generation of returns which is indicated by the annualised standard deviation before transaction cost (Appendix 14). The annualised standard deviation has been reduced relative to the previous period and is for this period compared to the entire period, with a level of around 7% (Appendix and Appendix 14). The reduction in the volatility of the returns has a greater impact on the Sharpe ratio than the effect of the lower annualised mean excess return before cost which results in an improved Sharpe ratio. These characteristics before transaction costs are, to a large extent, transferred to the results after transaction costs. After accounting for transaction costs, the batting ratios remain high around 70% and are hereby between 10 and 20 percentage points above this ratio of the entire period. This higher winning rate offsets the slugging ratio between 0.73 and 0.83. The combined effect of these two factors results in an annualised mean excess return after transaction cost which outperforms the results generated over the entire period for all triggers. The higher average excess return after transaction cost counter-balances the generally higher annualised standard deviation relative to the entire period. The outcome of this period's excess return and standard deviation is a Sharpe ratio of 1.00 as the highest, obtained by trigger 2.5/0.5 (Appendix 14). Generally, the

Sharpe ratios obtained during this period is higher than the previous period as well as for the full period across all triggers. This means that this period is generally more lucrative than the other periods of the cointegration method after transaction costs.

Periodical characteristics | The findings of the subperiod presented above reveal less volatility in the daily returns of the two methods and that the cointegration to a great extent outperforms the distance method.

Considering first 2010 and the distance method, no particular pairs are set to be the driver of the negative development (Appendix 14). Instead, it is caused by the fact that many pairs are open almost the entire time of both trading periods resulting in the stable negative returns, as illustrated in figure 7 (Appendix 14).

For the cointegration method, 2010 yielded positive results for the majority of the pairs in both trading periods of the year, but did also generate noticeably negative returns in some specific pairs. Here, all trades that either generated a significant positive or negative return were Europe-only ETFs. Here, Euro Stoxx Dividend Index Fund (ETF163) was included in two pairs that yielded noticeable positive returns and two pairs that yielded noticeable negative returns (Appendix 11; Appendix 14) Considering that the European debt crisis began to gain momentum in the end of 2009 and 2010, this might be an explanation for the higher amount of Europe-only pairs (Kenny, 2019). As with the findings of the first subperiods, there is a tendency that the profitable pairs comprise one ETF tracking a fraction of a broader market against the other ETF tracking this broader market. This is, for example, illustrated by Euro Stoxx or MSCI France (ETF14) against MSCI Eurozone (ETF49) (Appendix 14).

As the distance method does not include any European ETFs and only two pairs comprising the broad world index, the method does not benefit from the profitable development in the pairs of these types, which to a large extent explains the differences in cumulative return development between the two methods.

In 2011, the distance method yielded slightly positive results, however not enough to offset the negative returns of 2010. For the distance method, no individual pairs can be identified to be the main driver of the performance in 2011.

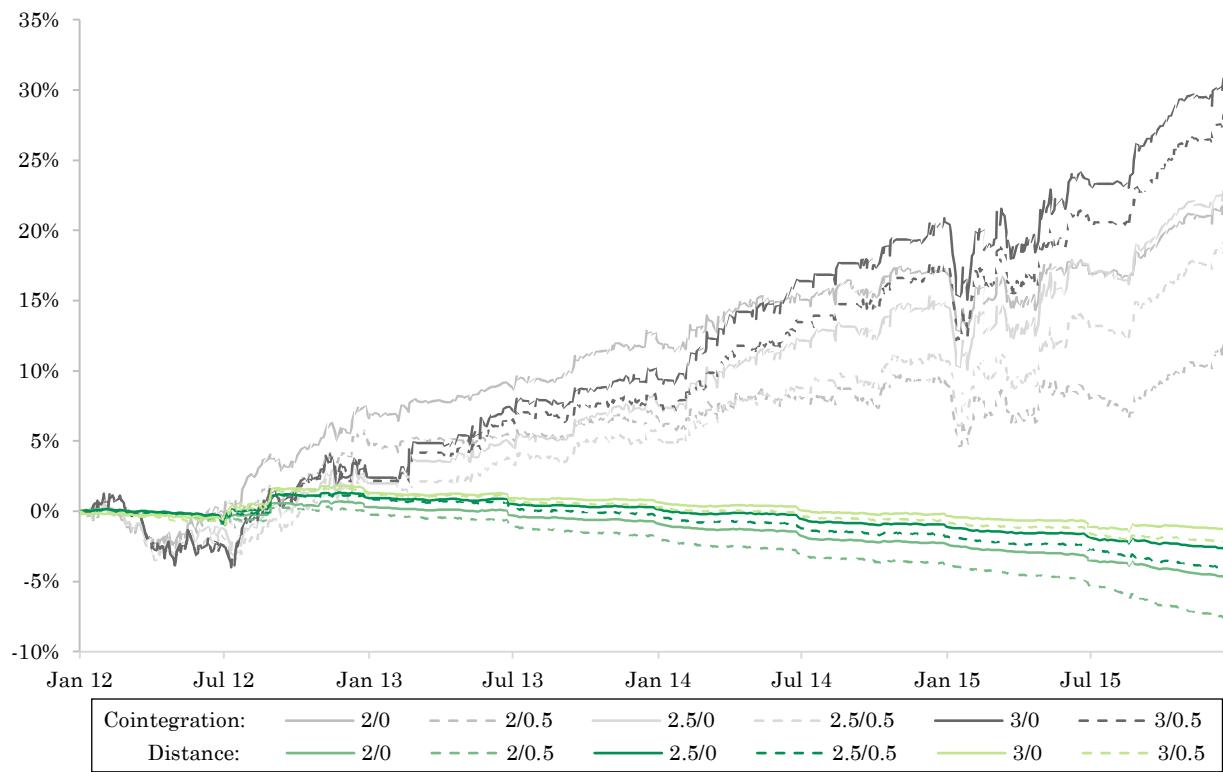
For the cointegration method, the second half of 2011 showcased the best results of the subperiod. The positive results were driven by the turmoil of the subsequent period

caused by the “Black Monday” crash in August (Wearden, 2011). The crash was caused by the downgrading of US treasury bonds which led to large price drops in the largest indices (Wearden, 2011). For the cointegration method, it is again pairs containing ETF107 which yield by far the highest returns. Here, the counterparts to ETF107 are Schwab US large-cap ETF (ETF184) and S&P Total stock market ETF (ETF79) (Appendix 14). These two pairs yield almost the entirety of the profit in this period. The remaining profit was derived from four pairs of various value ETFs with the S&P 400 Mid Cap value ETF (ETF109) as a counterpart in all four. These returns come from several noticeable divergences of ETF109 which therefore influence all four pairs (Appendix 14). With ETF107 and ETF79 being important for the returns in this subperiod, we can again see a similar pattern to the results of the first subperiod, that pairs that yield the greatest profits comprise a more specific ETF, against an ETF tracking the broader market. The relative timing of the black Monday crash and the generated return during the subperiod, implies that the method does not necessarily react directly to large events but the subsequent periods after such events. This is consistent with the results of the second half of 2008. Another noticeable finding is that the distance method did not include either of the pairs of ETF107 or ETF109 in 2011, explaining why the returns of the method is low in this period compared to the cointegration method. Above considerations and the results of this subperiod infer that the distance method lacks diversity as it, to a great extent, relies on pairs comprising ETFs tracking the same index. The result of such pairs composition’s unprofitability is illustrated in figure 7 (Appendix 14).

7.3.3. Pairs trading from 2012 to 2016

In this period, the financial market seems to have moved on from the financial crisis with the volatility being more under control. The VIX index is at its lowest level since the end of 2006 and the market return exhibits steady growth during this period (CBOE, n.d.; French, n.d.). Below figure 8 displays the cumulative returns for the sub-period, with the descriptive statistics of this subperiod being found in appendix 15:

Figure 8: 2012-2016 cumulative returns for triggers after transaction costs



For the distance method, the pattern continues in this period with Sharpe ratios before transaction costs between 2.6 and 4.0 which become negative after transaction costs. The attractiveness of higher opening trigger values after transaction costs in this sub-period is consistent with the findings of previous periods. As the slugging ratio is minimum 1.25 after transaction costs, the negative returns must be related to the batting ratio, which for this period achieves a previously unseen low level between 0.13 and 0.29 thus offsetting the slugging ratio and creating an overall negative return. In general, this period aligns with the trends identified in the previous periods for the distance method, generally having low returns that are not robust to transaction costs (Appendix 15).

Different seems to be the case for the cointegration method which creates an annualised mean excess return before transaction costs between 22% and 32% as a result of a strong mean reversion that causes diverged pairs to return to equilibrium in between 4.5 and 8.5 trading days. This is relatively better than the full period average of 12.5 across all triggers and also an explanation for 98% of the openings fully converging within the trading period. Further, after transaction costs are deducted, the annualised mean excess return for the period lies in the range from 3% up to almost 8%

depending on the triggers. Batting ratios remain high in all triggers after introducing transaction costs, with especially trigger 3/0 standing out with a batting average of 81%. The high batting ratio combined with slugging ratios remaining above 1 and up to 1.9 after transaction cost further underline this period as lucrative for the cointegration method (Appendix 15).

Periodical characteristics | When considering the breakdown of the period there is generally a similar trend for the two methods in 2012, but for the remainder of the sub-period exhibit no common patterns. The low returns for 2012 in the cointegration method is primarily derived from the pair comprising MSCI Spain ETF (ETF10) and Invesco global listed Private Equity ETF (ETF139) in the first half of 2012. The driver of the negative development for this pair is a consequence of a divergence of ETF10 potentially as a reaction to the European debt crisis as mentioned earlier, which not to the same extent impacts ETF139 (Appendix 15; Kenny, 2019).

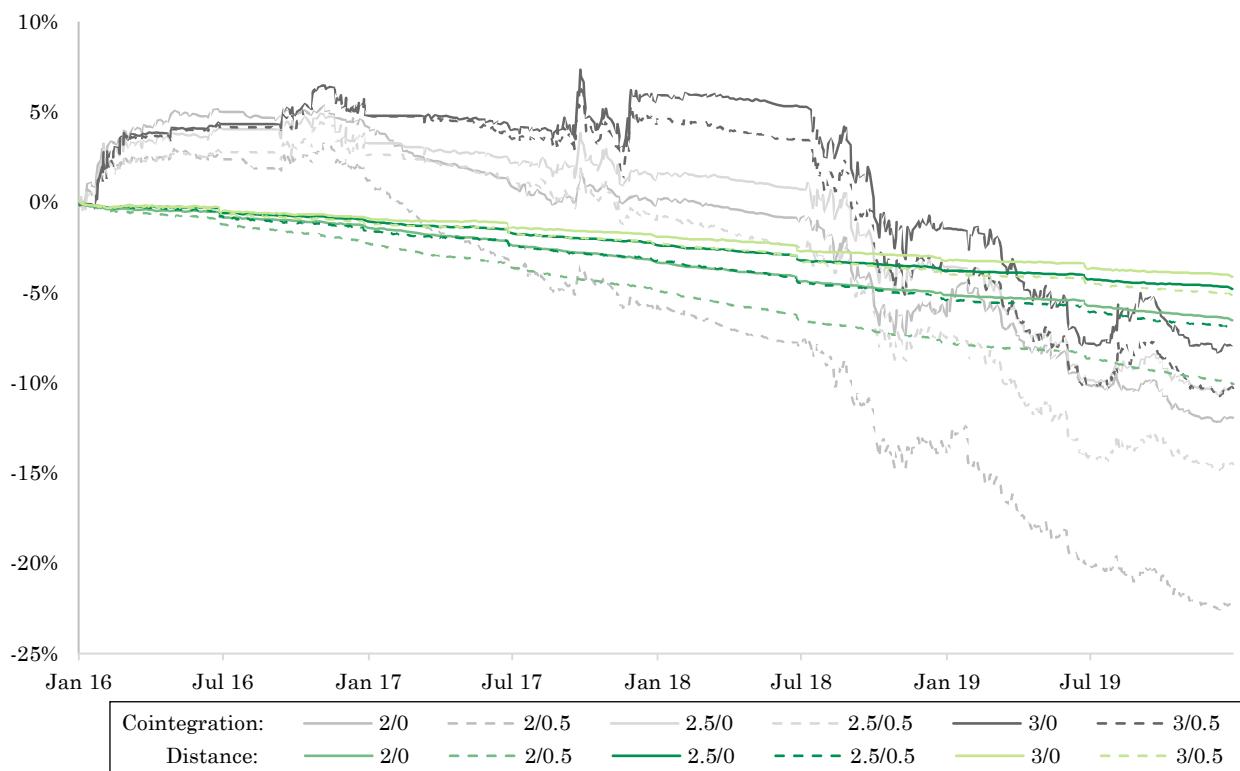
ETF10 did not converge until after the end of the trading period causing a loss during the period. This particular pair is not included in the distance method, but instead, the negative returns are again caused by the reliance of pairs with ETF tracking the same index and thus not generate returns robust to the trading costs (Appendix 15). In the years following 2012 in this subperiod, the distance method yields stable but negative returns with no particular pair(s) being the drivers of either positive or negative returns. In general, all pairs of the distance method all tend to yield a small but negative return after transaction costs.

In 2013, the cointegration method obtained the majority of the profit from pairs containing ETF107 and ETF56 against ETF79, ETF80 and each other. The same picture is evident in the remaining years, with both ETF107 and ETF56 included in the pairs yielding the majority of the most profitable trades (Appendix 15). We again see a general pattern of the profitability between ETFs tracking part of the market against ETFs that track the general market. Compared to the first two subperiods, this period is not affected by major events influencing the returns of the methods, thus making this subperiod more reliable for considering the viability of the two methods.

7.3.4. Pairs trading from 2016 to 2020

The subperiod from 2016-2020 experiences a continuation of the good market trends experienced under the previous subperiod. Furthermore, the VIX index hit an all-time low at the end of 2017. This low range in the volatility is interrupted in 2018 where especially the beginning and the end of 2018 is characterised by increased market turmoil (CBOE, n.d.). Below is the cumulative returns of the subperiod illustrated by figure 9, and the descriptive statistics can be found in appendix 16.

Figure 9: 2016-2020 cumulative returns for triggers after transaction costs



The distance method continues to follow the pattern identified in the previous periods for this period. The 2/0.5 trigger obtains the best Sharpe ratio before transaction costs, whereas the highest Sharpe ratio is obtained by trigger 3/0 when the costs of trading are included. The batting ratio for the distance method is between 79% and 91% before transaction costs, and annualised mean excess return after transaction costs turns negative in this subperiod (Appendix 16). What stands out in this period is the batting ratio after accounting for transaction cost reaches rock bottom of 2% as the lowest and 7% as the highest (Appendix 16). The matter is worsened by a low slugging ratio as well. This extremely low robustness against transaction costs results in a Sharpe ratio

of -2.44 as the best for the period. Thus, this period is not very favourable for pairs trading with ETFs using the distance method. Noticeably, 81% of the traded pairs in this subperiod for the distance method comprise pairs of ETFs tracking the same index, ranging from 75% to 90% in the respective trading periods (see appendix 20).

The cointegration method exhibits for the first time a negative annualised mean excess return after transaction cost, making this subperiod the worst in our sample period. The negative excess return after transaction costs is a result of too low annualised mean excess return before transaction costs to withstand the related trading costs. When comparing the annualised mean excess return before transaction costs for this period to the previous ones, the returns for this period are particularly lower.

Periodical characteristics | The distance method displays the same characteristics as the previous periods, with a consistently negative return after subtracting the trading costs. 2016 is for the cointegration method consistent with the pattern presented in the previous subperiods, with pairs comprising ETF56 and ETF107 generating the majority of the profitability.

For 2017, the cointegration method yields an annualised mean excess return between -7.32% and 0.85% after transaction costs. Here the period also showcases the highest amount of pairs comprising ETFs tracking the same index, which amounts to 60% (Appendix 21).

Second half of 2018 and the first half of 2019 are causing the majority of the losses for these two years. For the second half of 2018, the same picture as 2017 is evident with 55% of pairs comprising ETFs tracking the same index causing stable negative returns (Appendix 21). This further worsened by a number of more noisy pairs. The worst performing pair is MSCI Global Metals and Mining Producers ETF (ETF254), and Kraneshares CSI China Internet ETF (ETF231) with the pair on its own is associated with an almost 20% loss (Appendix 16). In general, the pairs are more diverse than the pairs of the former trading periods. However, it is difficult to fully state whether the pairs composition of this trading period is the only reason for the negative returns.

When looking at the trading period for the first half of 2019, the reason for the negative returns is easier to state. For the trading period, 83% of the negative returns derive from five pairs that all include the S&P Retail ETF (ETF129) (Appendix 16). For all

five pairs, there is the same diverging trend in the z-score to somewhere between 6 and 12 at the end of the trading period caused by ETF129 (Appendix 16). As such, we can conclude that this ETF drives the majority of the negative returns in this trading period.

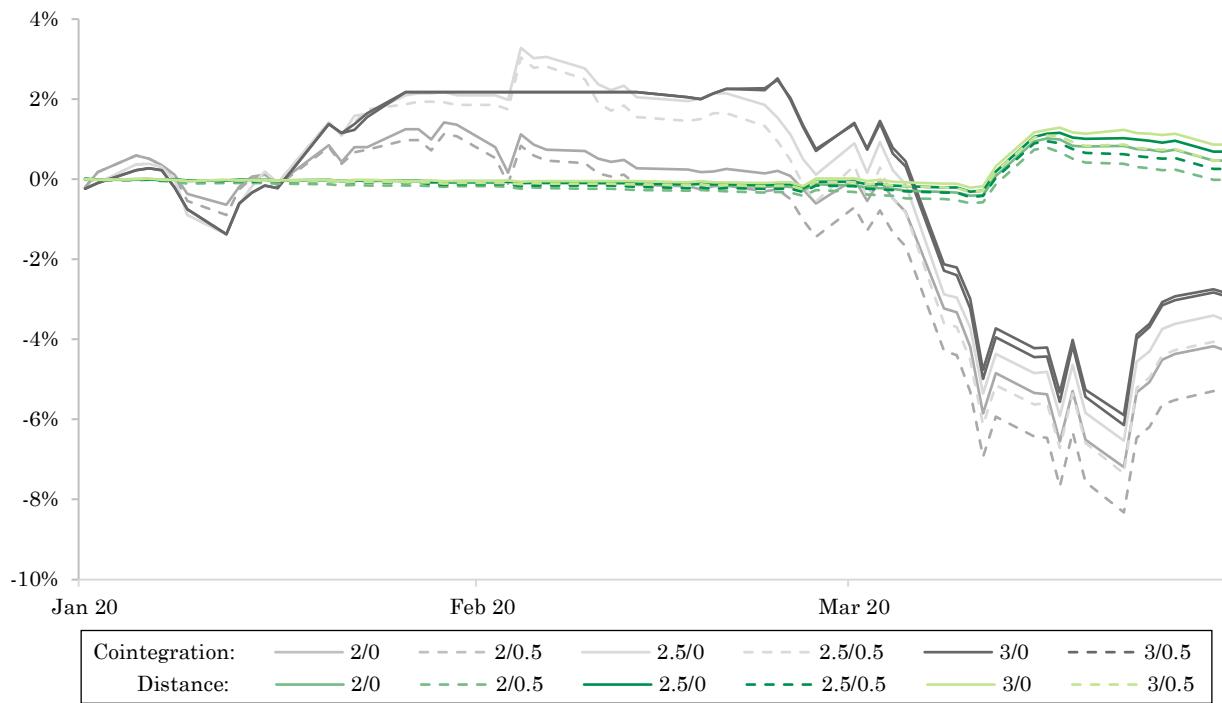
The subperiod from 2016-2020 exemplified the notion about pairs either not diverging enough or diverging too much. For the cointegration method, 2017 and 2018 showcased negative results due to pairs comprising ETFs tracking the same index, while 2018 also and 2019 showcased negative returns because pairs comprising ETF being too noisy causing continuous divergence in some relations. Lastly, whereas the properties of having one or few ETFs present in several traded pairs can yield positive returns as seen in the previous subperiods, an adverse effect can also be the outcome as seen in 2019 with ETF129.

7.3.5. Pairs trading in Q1 2020

As indicated by the section name, this subperiod is only covering the first three months of 2020. The presentation of the results obtained during this period are not annualised because, as previously described, this will distort the assessment of the period. Therefore, the results below display quarterly results, obtained in the first quarter of 2020. This period is thus not directly comparable to the remaining periods as it would be naive to expect the pattern of the first quarter will continue the rest of the year. The argument for including the period and still to some extent comparing the results to the rest of the periods, lies in the desire to examine how the use of pairs trading with ETFs responds to the turbulence occurred as a result of the market's reaction to COVID-19 pandemic. Furthermore, in spite the results only tell us something about the development of the first quarter of 2020, the period still provides insights and increases our understanding of the pairs trading strategy which is in line with the general foundation of our pragmatic stand.

From figure 10 below, the cumulative returns of the subperiod illustrated, and the descriptive statistics can be found in appendix 17.

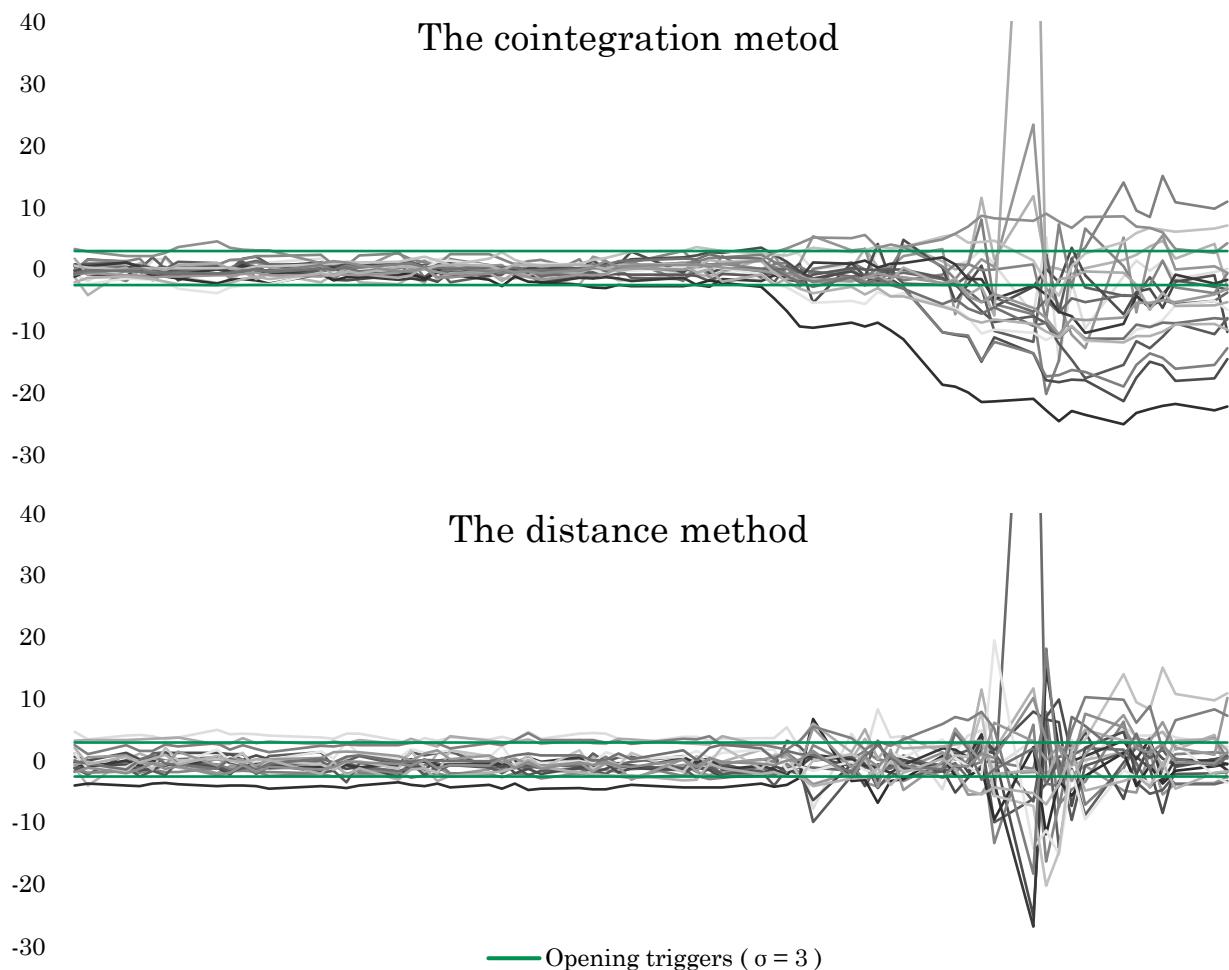
Figure 10: Q1 2020 cumulative returns for triggers after transaction costs



Even though this period has a duration corresponding to a quarter of the time period usually compared with, the number of openings for the distance method is on average across the triggers 6, which is almost at the same level as the annual average openings for the entire period. We are thus dealing with a period characterised by a great amount of market fluctuations, illustrated in Figure 11. The figure shows that the beginning of 2020 only leads to a few minor openings for the distance method, while a more chaotic picture is drawn in March. Here, some major divergences occur but are more or less restored within a few days. This also appears from the length of convergence, which range from 5 to around 6.5 is considerably lower for this period relative to what has previously been experienced. This can also be seen in the high batting ratio before transaction costs. With a quarterly excess return before transaction cost around 5% for the distance method, this period has already generated a return larger than the annualised average of the two preceding subperiods (Appendix 17). When the cost of trading is deducted, much of the negative pattern previously identified still applies, as the profitability is considerably reduced. This period stands out in the sense that the excess return after transactions is positive except for the 2/0.5 trigger. Primo 2020 thus showcases some of the same characteristics for the distance as with the 2007-2010 subperiod (Appendix 12 and Appendix 17).

Figure 11 also shows that the cointegration method exhibits a high degree of comovement between the traded pairs during the first two month of 2020, but as March is reached, this is sat out of play due to large divergences. These divergences do not seem as chaotic and random as for the traded pairs with the distance method. Although the spreads between the different pairs do not exhibit massive fluctuations, they exhibit a continuing divergence causing the quarterly excess return before transaction costs to lie in the range between -0.5% and -1.4% for the cointegration method in this period. The continuous divergence also appears in the batting ratio up of 85% before transaction costs. After transaction costs, the quarterly mean excess return is between -2.8 and -5.4% (Appendix 17). Whether divergences of this period will reach full convergence and hereby generate a profitable period is still yet to be uncovered.

Figure 11: Z-score of from Q1 2020 trading period



Note: The figure has been reduced in size to better illustrate the key takeaways.

Summary | The above breakdown of the subperiods has touched upon many different trends and conditions of the entire sample period. Overall, we can infer that except for the subperiod of 2007-2010, the distance method yields steady negative returns. The profitability of the subperiod from 2007-2010 indicates that large anomalies occurred in the market generally causing the price relations to diverge, but also converge afterwards, due to the noise of the financial crisis preventing other market players from exploiting these arbitrage opportunities quickly. With the highly mean-reverting properties of the selected pairs for the two methods, the subperiod from 2007-2010 provides some good results associated with high returns.

Throughout the sample period, the returns before transaction costs of the distance method produce positive results and strong Sharpe ratios but have, in general, shown to be too low to withstand the transaction costs. Oppositely, the cointegration method has shown to produce much more robust results with only the last subperiod delivering negative returns. Even though it is difficult to draw any definitive conclusions about the profitability of the specific pairs and which perform the best, there is a tendency that pairs including specific large-cap ETFs, against total stock market ETFs, yield the best returns over the full sample period. These types of pairs perform both well in the wake of large financial downturns and also in more stable periods as seen in the period from 2012-2016. It has further been noticed, that especially ETF107 and ETF56 are involved in the majority of the returns of the different subperiods except the last from 2016-2020 (Appendix 16). The last subperiod included a higher number of pairs comprising ETFs tracking the same index for both methods. For the cointegration method, negative returns were also derived from pairs that continued to diverge. These two findings underline the dilemma of locating pairs that diverge enough to cope with transaction costs, but neither too much.

Whereas the repeatedness of ETF56 and ETF107 provided positive results in most trading periods, the opposite was also the case with ETF129 in the first half of 2019. As such, it should also be considered whether the same ETF(s) should be included in more than one pair during each trading period. On the one hand, it might entail great profits as with ETF107 and ETF56, while on the other hand, this one ETF may exhibit divergence from the established relation and thereby negatively influence all the pairs

it enters, as with ETF129. Thus, the classic term of putting all eggs in the same basket is a risk aspect that should be carefully considered when performing pairs trading.

7.3.6. Attributes of the distance and cointegration method

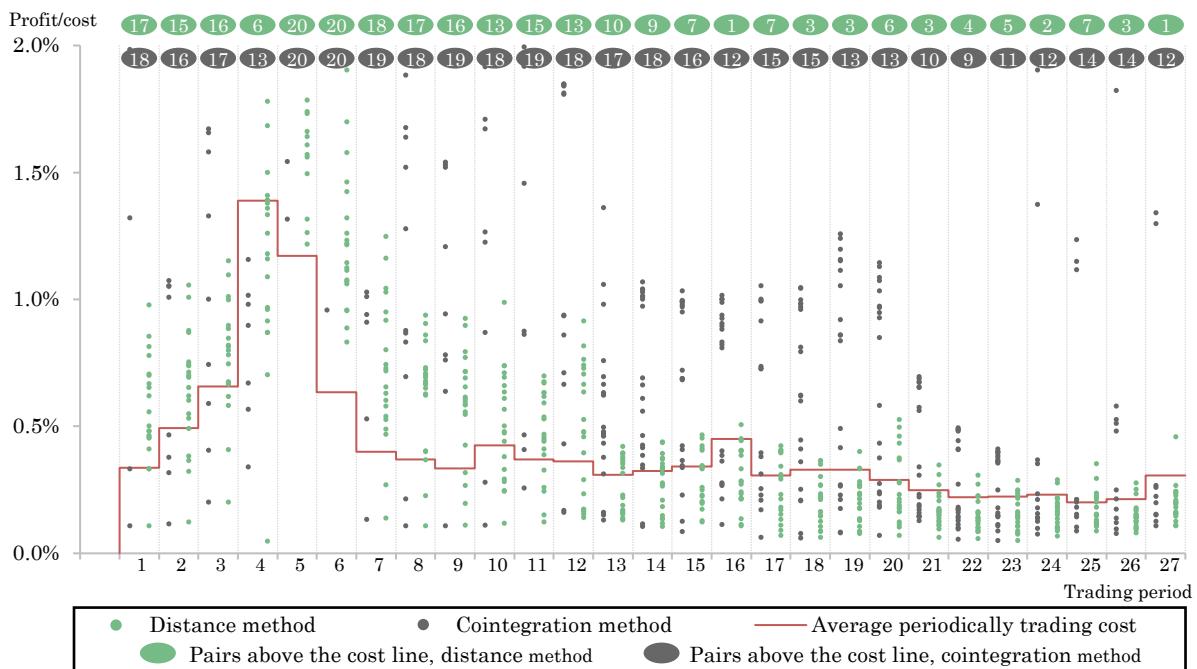
In the attempt to understand the robustness of the two methods against transaction costs and why this robustness fluctuates over the course of the sample, this section will examine the two methods' profit opportunities. This is done by investigating the relationship between the z-score and the associated profit generated by a given convergence. By determining this relationship, we should be able to identify the profit for a convergence equivalent to one standard deviation for a pair. This enables a comparison of the fundamental profitability of the two method's traded pairs. In order to determine the relationship between the z-score and profit we apply the following formula:

$$\Pi_t^{i,p} = \gamma^{i,p} \Delta Z_t^{i,p}. \quad (18)$$

Here i is a traded pair of ETFs in the trading period p . Thus, $i = \{1, 2, \dots, 20\}$ and $p = \{1, 2, \dots, 27\}$. Further, t denotes the trading days in the trading period, p . Π is the profit of a long-short position from $t-1$ to t , γ is the slope coefficient and the parameter we are estimating and ΔZ is the change in the z-score between $t-1$ and t (Appendix 19). The development of the profit and the z-score is calculated on a daily basis. By applying the z-score for both methods, the spread measurement of the distance method has been transformed to be expressed in z-score for comparison (Appendix 19). Based on equation 18, we are able to find the associated profit of a unit change in the z-score. After estimating the slope coefficient for all pairs in the 27 trading period, we multiply the slope coefficient by 3. This is done in order to obtain the profit associated with a convergence of 3x standard deviation and thus simulate the convergence of the trigger values of 3/0. The reason why we are interested in examining the profit of a convergence equivalent to 3x standard deviations is to test whether the trigger value of 3/0 generates a high enough return to offset the associated transaction costs, when the opening and closing triggers are exactly crossed. This is relevant as this is the trigger requiring the greatest divergence before opening a position. Thus, if a z-score of 3 is not enough to offset the trading costs, neither will it be enough for 2 or 2.5. This would also imply that larger trigger values are required.

Plotting the associated returns of a convergence equivalent to 3x standard deviation and the average associated trading cost for each trading period, we obtain 12.

Figure 12: Profit for convergence of 3x standard deviations



Note: Average periodical trading cost comprises two round trips and 15 days holding period.

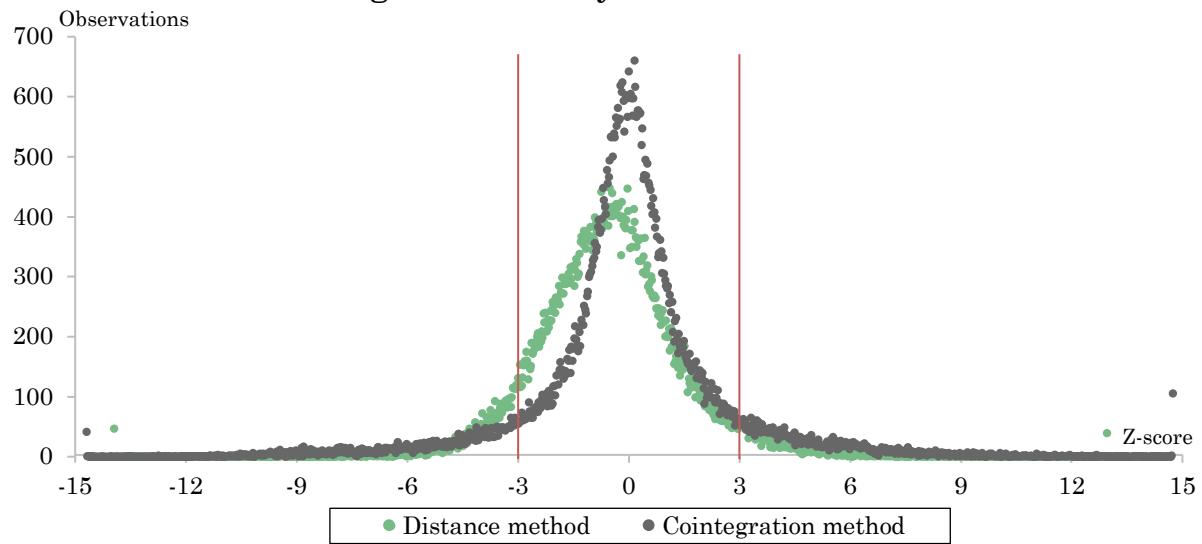
The figure has been reduced in size to better illustrate the key takeaways but indicate the number above the cost line by the circles in the top of the figure.

Figure 12 gives an illustrative indication of how profitable a convergence of 3x standard deviations is in relation to the trading costs for the respective pairs. Here, the pairs from the distance method most often lie below the average cost line for the trading periods. In itself, it does not imply that a position in a pair below the line imposes a loss, as the respective position could exhibit a convergence larger than 3x standard deviations. This could be the case if a pair below the line opens with a z-score of e.g. 9 and fully converge. Then the convergence will be of 9x standard deviations which potentially could provide enough profit to offset the trading costs. In addition to the inference that the pairs of the distance method, in general, require larger divergences in order to produce profitability, it can be seen from figure 12 that both methods have experienced a decreasing profitability for the traded pairs.

For the cointegration method, in the period from 2007 to 2014 (trading period 1 to 16), around 16-20 pairs are located above the cost line except for trading period 4, corresponding to the second half of 2008. The lower number of pairs above the cost line for the cointegration method in the last couple of trading periods are consistent with the negative results presented for the last subperiod.

For the distance method, there is a clear declining trend from a more lucrative period up until around 2010 (trading period 8) down to only one pair above the line in the last trading period. Here, we find a strong connection between the negative performances of the distance method and the number of pairs which do not generate a profit that exceeds the associated costs of trading with a convergence equivalent to 3x standard deviations.

This review of the profitability of a convergence equivalent to 3x standard deviations for the two methods suggests that the pairs in the distance method require a greater convergence to achieve profitability. In order to make this greater convergence, the price relation between the two ETFs in the pair must exhibit a great divergence first. Even though the pairs in the distance method generally need a greater convergence to counterbalance the trading costs, it does not imply that they do not achieve the convergence needed to be profitable. In order to identify the magnitude of the divergences that occur for the traded pairs in the distance and cointegration methods, an additional assessment is required. For this, figure 13 is provided. The figure illustrates the distribution of the daily z-scores for the two methods with the two red lines representing a z-score of ± 3 to align with the characteristics of figure 12. Because it is daily z-scores, all the values are thus not tradeable, as it may be the case that we already had a position in a given pair which continued its divergence. Because we already had an open position in the given pair, the greater profit from the better z-score cannot be utilised.

Figure 13: Daily z-score distribution

	Mean z-score	Observations outside ± 3	Absolut mean z-score outside ± 3
Distance method	-0.512	11.75%	4.526
Cointegration method	0.010	14.88%	5.536

Note: The figure has been reduced in size to better illustrate the key takeaways as the distance method has 3,725 observations and the cointegration method has 3,999 observations in the transition from negative to positive z-score.

Figure 13 shows that almost 15% of all daily z-scores from all 540 traded pairs over the entire sample have a z-score either above or below 3 for the cointegration method, which for the distance method corresponds to almost 12%. However, the number of observations outside the ± 3 standard deviation band should be seen in relation to the magnitude of these observations. Without consideration to the sign of the z-score value, the average daily z-score above 3 is 4.5 for the distance method and 5.5 for the cointegration method. These results suggest that the selected pairs of the cointegration method more often deviate by more than 3x standard deviations from their established mean compared to the pairs of the distance method. Further, when the z-scores of the traded pairs exceed 3x in absolute terms, they also tend on average to be larger in magnitude for the cointegration method relative to the distance method. However, it must be emphasised that larger z-scores does not necessarily imply greater profit as it is the convergence, not the divergence which generates the profit. In other words, in order for a large divergence to generate profit, it must also converge.

As such, the distance method generally produces lower return and fewer pairs above a 3x standard deviation threshold than the pairs for the cointegration method, which suggest that the reason for these lower attributes must originate from the distance method's selection of pairs. It became clear from the pairs composition and the sub-period breakdown that the traded pairs of the distance method to a larger extent comprise pairs of ETFs tracking the same index.

Through an investigation of the associated profit of convergence, it is found that pairs containing ETFs tracking the same index generate a smaller profit (appendix 20). For the distance method, 42 individual pairs are tracking the same index. These 42 pairs take up 51% of all tradeable spots for the entire sample period, which equals to 540 (27*20). Out of these tradeable spots taken by pairs with ETFs tracking the same index, 73% of these do not generate a high enough return to offset the associated costs of trading if the convergence is 3x standard deviations (see appendix 20). This implies that 38% (51% * 73%) of all traded pairs for the distance method throughout the sample period constitute pairs with ETF tracking the same index and do not generate a positive profit if they converge equivalent to 3x standard deviation. For the cointegration method, 35% of the 540 tradable spots contain ETFs tracking the same index, of which 55% of these do not generate a positive return when converging 3x standard deviations. This means that 19% (35% * 55%) of all traded pairs throughout the sample period constitute pairs with ETF tracking the same index and do not generate a positive profit if they converge equivalent to 3x standard deviation for the cointegration method. These results suggest that not only does the cointegration have fewer pairs of ETFs tracking the same index, but the profitability is also higher for those pairs. In addition to above, it has further been found that pairs with ETFs tracking the same index to a large extent hold their linear relationship established in the formation period; hence only to a limited extent diverge more than three standard deviations (Appendix 20). Having established that ETFs tracking the same index both do not achieve a high enough gross profit at a convergence equivalent to 3x standard deviation to offset the associated trading costs and also that these pairs rarely obtain a z-score outside ± 3 , it can be inferred that almost every position in such pairs is associated with a loss. This does not only apply to the distance method but also to the cointegration method. The difference between the two methods is that the distance method to a greater degree

comprises pairs of ETFs tracking the same index. These fundamental dynamics of pairs trading with ETFs tracking the same index are consistent with the generally negative returns for the distance method after transaction cost in every subperiod from 2010 and beyond.

We can conclude that the pairs composition of the two methods has a noticeable impact on the robustness of the two methods throughout the sample period after transaction costs. Fundamentally the cointegration method selects pairs that diverge more and more often, but also yield higher returns when converging, which makes the pairs much more robust to trading costs. These results are thus consistent with the results outlined in the first section of this chapter regarding the cointegration method having better results, more trades and shorter time to convergence. The negative returns of the subperiods for the distance method are a consequence of the method's selection process, which tend to favourise pairs comprising ETFs tracking the same index. The fundamental characteristic of such pairs is that they only to a limited extent experience divergences, and when they do, the associated profit is not sufficient to offset the related trading costs. The majority of traded pairs of the distance method are thus not robust to the costs of trading with the consequence that the openings the method makes, will result in losses when the costs of trading are taken into account. As with the cointegration method, these results are consistent with the results outlined in the first sections of this chapter.

Even though the attributes of the z-score distribution and the profitability associated with convergence equivalent to 3x standard deviation do not necessarily explain the profit, the findings strongly support the determinants of profitability of the two methods outlined in the above sections.

7.4. Factor models

Above sections have highlighted the profitability of the overall sample period, the determinants of the profitability and broken the results down into subperiods. In this section, the nature of the results will be examined with the objective to determine what underlying factors might explain the returns of the two methods. Various factor models

will be conducted to determine the foundation of the profitability and to understand the underlying premises of the excess returns.

In general, factor models assume that the excess return of an asset or portfolio of assets can be explained by a number of common factors (Munk, 2018). The list of different factors as explanatory variables is long and spans from macroeconomic factors to industry-specific factors or more general market factors. In the following we will consider the Capital Asset Pricing Model (CAPM), a liquidity factor by Pastor and Stambaugh (2003; 2019), Fama-French three-factor model, a momentum factor and lastly the volatility index (VIX) (Appendix 22; Ang, 2014).

7.4.1. CAPM

The first factor model to introduce the concept of relationships between the risk of an asset and the risk of an external component was CAPM (Ang, 2014, p. 195). CAPM is a simple single-factor model which decomposes the risk of the pairs trading strategies in two parts; a systemic part that is explained by the market portfolio and an idiosyncratic risk that is not explained by the market. The form of the CAPM can be stated as the following when an alpha is included (Pedersen, 2015)

$$R_t^e = \alpha + \beta R_t^{M,e} + \varepsilon_t . \quad (x)$$

Here, R_t^e is the excess return on the strategy, α is the intercept of the regression and represents the excess compensation for taking on systemic risk, β represents a measure of market exposure, $R_t^{M,e}$ is the excess market return of the market portfolio and ε_t is the residuals also referred to as the idiosyncratic risk. As allocated for in chapter 5, pairs trading is close to or a market neutral strategy, which implies the beta should be close to or equal to zero (Pedersen, 2015, p. 29). Table 3 summarises the results of the CAPM regression model for the two methods and different triggers:

Table 3: CAPM

	Distance method						Cointegration method					
	2 / 0	2 / 0.5	2.5 / 0	2.5 / 0.5	3 / 0	3 / 0.5	2 / 0	2 / 0.5	2.5 / 0	2.5 / 0.5	3 / 0	3 / 0.5
<i>Before transaction costs</i>												
Alpha	0.007	0.008	0.007	0.008	0.007	0.008	0.017	0.020	0.016	0.017	0.014	0.015
	<i>4.141***</i>	<i>4.365***</i>	<i>3.960***</i>	<i>4.164***</i>	<i>3.966***</i>	<i>3.969***</i>	<i>8.508***</i>	<i>8.975***</i>	<i>7.645***</i>	<i>7.868***</i>	<i>6.344***</i>	<i>6.640***</i>
Beta	-0.181	-0.202	-0.184	-0.212	-0.180	-0.208	-0.192	-0.203	-0.190	-0.207	-0.169	-0.194
	<i>-4.834***</i>	<i>-4.780***</i>	<i>-4.814***</i>	<i>-4.919***</i>	<i>-4.507***</i>	<i>-4.601***</i>	<i>-4.138***</i>	<i>-4.106***</i>	<i>-4.062***</i>	<i>-4.290***</i>	<i>-3.477***</i>	<i>-3.888***</i>
<i>After transaction costs</i>												
Alpha	0.001	0.000	0.001	0.001	0.002	0.002	0.002	0.001	0.003	0.003	0.004	0.003
	<i>0.920</i>	<i>0.165</i>	<i>1.365</i>	<i>1.012</i>	<i>1.799*</i>	<i>1.605</i>	<i>2.104**</i>	<i>0.876</i>	<i>2.888***</i>	<i>2.356**</i>	<i>2.615***</i>	<i>2.405**</i>
Beta	-0.079	-0.071	-0.083	-0.082	-0.080	-0.081	-0.050	-0.051	-0.054	-0.056	-0.040	-0.046
	<i>-3.734***</i>	<i>-3.337***</i>	<i>-3.787***</i>	<i>-3.595***</i>	<i>-3.497***</i>	<i>-3.444***</i>	<i>-1.856*</i>	<i>-1.902*</i>	<i>-1.989**</i>	<i>-2.061**</i>	<i>-1.288</i>	<i>-1.519</i>
<i>Annualized alpha values</i>												
Before transaction costs	8.175%	9.711%	7.987%	9.432%	8.330%	9.428%	20.740%	23.426%	18.836%	19.987%	16.253%	17.436%
After transaction costs	1.020%	0.185%	1.579%	1.217%	2.173%	1.996%	2.982%	1.231%	4.142%	3.387%	4.263%	3.872%

Note: Critical values (t-statistics) for significance levels; 10% (*) = 1.6548, 5% (**) = 1.9753, 1% (***) = 2.6077.

The table shows that all triggers provide statistically significant positive alpha values at a 1% level and beta values statistically significant at a 5% level before transaction costs. This means the alpha and beta values are both different from zero. With beta values different from zero, it cannot be stated that the two methods before transaction costs are market-neutral in the light of the CAPM regression. These results of the CAPM regression before transaction costs are consistent with the findings from the subperiod breakdown, that the strategy performs well when the market prices decline. As with the results of the annualised mean excess return, the annualised alpha values of the strategies decline noticeably after considering transaction costs. For the distance method, none of the triggers generate a statistically significant alpha after transaction costs. This means that we cannot reject the null hypothesis that the true values are zero. It can therefore not be determined whether the positive alpha from the regression is just represented by noisy trends or the distance method actually yields a positive excess return (Pedersen, 2015). The statistical significance of the beta values are unchanged with all triggers generating a statistical significant beta value at a 1% level after transaction costs. The estimated coefficients are however changed such that after

transaction costs the correlation with the market has been reduced. The statistical insignificant alpha for the distance method after transaction costs aligns with the findings of this paper with respect to the robustness of the distance method after transaction costs being very low.

For the cointegration method, all triggers except 2/0.5 generate a statistically significant alpha with a minimum level of 5% after transaction costs. Consistent with previous findings of this paper, trigger 2.5/0 and 3/0 yield the best results and thus generate positive alpha values at a statistical significance level of 1%. This provides strong evidence for these triggers obtaining an excess compensation for taking on systemic risk. For the beta values after transaction costs, the 2.5 opening triggers generate statistical significant beta values with t-statistics very close to the threshold. As such, we cannot reject the null hypothesis for the opening triggers of 2 and 3 that the true beta values are zero which indicate that these triggers are either close to or market-neutral after transaction costs.

In general, if applied transaction costs are fixed, the impact of imposing these costs would only affect the intercept i.e. alpha. This is due to the fact that by subtracting a constant the relative relationship is unchanged. However, as we have imposed a dynamic set of transaction costs, the relationship to the market portfolio is differently affected over time. The derived effect is that the loading of the beta coefficient changes. The dynamic set of transaction costs are visible from figure 4 in section 7.2. In this connection, it is noticeable that the statistical significance of beta after transaction cost for the distance method does not change to become insignificant. The negative correlation with the market can be found in the constantly negative mean excess return around 1-2% of the method after the first subperiod and the steady increase experienced in the market over the same period (Appendix 22; French, n.d.). This combined with an almost linear price increase in the market since 2009 and the negative but stable profit for the method forms the relationship between these returns, being statistically different from zero.

Oppositely, because the development of the cointegration method does not exhibit the same degree of constant return development throughout the sample period, the beta coefficients become statistically insignificant implying no relationship between the market development and the return generated.

We can thus conclude that the cointegration method yields an alpha statistically significant at a 1% level for trigger 2.5/0 and 3/0. In addition we find the distance method to be slightly negatively correlated with the market with a statistically significant level of 1% without any evidence of an excess return beyond the market return.

To further investigate whether other factors can explain the excess returns, we test the CAPM including a liquidity factor.

7.4.2. Liquidity risk

A number of authors have touched upon the determinants of profitability and generally reached the conclusion that the driving factors are news and the provision of liquidity (Engelberg et al., 2009; Jacobs and Weber, 2014). Pedersen (2015) argues that the main sources of profitability for hedge fund strategies can overall be divided into compensation for liquidity risk and/or information which aligns with the conclusions of Engelberg et al. (2009) and Jacobs and Weber (2014). As we have already touched upon the effect of news and information in chapter 5, we will in the following consider the liquidity risk of the two methods.

Liquidity can be explained as the ease of trading an asset without notably affecting its price (Munk, 2018). The liquidity is closely linked to the limits of arbitrage. Therefore, in order to identify the liquidity risk of the two methods, we must understand how market liquidity affects the excess return. The further motivation for looking into the liquidity aspect of the two methods is due to the high returns experienced in 2008 which is one of the most remarkable periods in regard to illiquidity in the market, where almost all asset classes experienced enormous drops in liquidity (Pastor and Stambaugh, 2019). A somewhat similar relation can be found in the large liquidity drain in 1987, due to the black monday crash, which was the period in which Tartaglia's team in Morgan Stanley generated high returns (Gatev et al., 2006). With this in mind, we are encouraged to look into the liquidity aspect of pairs trading. Pastor and Stambaugh (2003; 2019) propose a non-traded liquidity factor capturing innovations in the aggregate liquidity measure used for determining the risk towards market liquidity (Pastor and Stambaugh, 2019). This factor is applied with the CAPM model and thus creates a two-factor model. The non-traded liquidity factor measures the relationship between the return of an asset and the general market liquidity based on the covariance of the excess return of the portfolio and the estimated trading costs of

the market (Acharya and Pedersen, 2005; Pastor and Stambaugh, 2019). The use of this liquidity factor is consistent with both Do and Faff (2012) and Rad et al. (2015). The output of the two-factor model is summarised in table 4:

Table 4: Figure 3: CAPM + liquidity factor

<i>CAPM + liquidity factor</i>		Distance method						Cointegration method						
Open / close trigger (o)		2 / 0	2 / 0.5	2.5 / 0	2.5 / 0.5	3 / 0	3 / 0.5		2 / 0	2 / 0.5	2.5 / 0	2.5 / 0.5	3 / 0	3 / 0.5
<i>Before transaction costs</i>														
Alpha		0.007	0.008	0.007	0.008	0.007	0.008		0.018	0.020	0.016	0.017	0.014	0.015
		4.285***	4.518***	4.092***	4.285***	4.073***	4.053***		8.627***	9.107***	7.7756***	8.017***	6.460***	6.765***
Beta		-0.170	-0.188	-0.173	-0.200	-0.170	-0.198		-0.178	-0.188	-0.176	-0.191	-0.156	-0.179
		-4.441***	-4.377***	-4.438***	-4.553***	-4.166***	-4.289***		-3.771***	-3.725***	-3.688***	-3.894***	-3.136***	-3.526***
Liquidity Factor		-0.042	-0.049	-0.040	-0.042	-0.036	-0.035		-0.049	-0.055	-0.051	-0.057	-0.049	-0.052
		-1.414	-1.478	-1.318	-1.247	-1.152	-0.980		-1.343	-1.421	-1.391	-1.495	-1.276	-1.341
<i>After transaction costs</i>														
Alpha		0.001	0.000	0.001	0.001	0.002	0.002		0.003	0.001	0.004	0.003	0.004	0.003
		0.991	0.250	1.456	1.093	1.877*	1.661*		2.293**	1.073	3.022***	2.481**	2.679***	2.469**
Beta		-0.075	-0.067	-0.079	-0.078	-0.076	-0.078		-0.040	-0.041	-0.046	-0.049	-0.035	-0.042
		-3.496***	-3.093***	-3.508***	-3.342***	-3.242***	-3.228***		-1.467	-1.496	-1.673*	-1.761*	-1.099	-1.326
Liquidity Factor		-0.012	-0.013	-0.016	-0.014	-0.015	-0.012		-0.036	-0.037	-0.028	-0.026	-0.018	-0.018
		-0.704	-0.778	-0.906	-0.793	-0.814	-0.622		-1.697*	-1.779*	-1.305	-1.216	-0.746	-0.739
<i>Annualized alpha values</i>														
Before transaction costs		8.485%	10.076%	8.282%	9.746%	8.599%	9.688%		21.104%	23.838%	19.216%	20.409%	16.617%	17.827%
After transaction costs		1.107%	0.282%	1.696%	1.323%	2.283%	2.082%		3.248%	1.508%	4.350%	3.582%	4.398%	4.005%

Note: Critical values (t-statistics) for significance levels; 10% (*) = 1.6548, 5% (**) = 1.9753, 1% (***) = 2.6077.

The results of table 4 suggest that we cannot reject the null hypothesis of the liquidity coefficient being zero for any of the above triggers either before or after transaction costs for either of the methods. We can thus conclude that the results of this paper do not support the findings of the existing literature which find a relation between the return generated and the liquidity in the market (Rad et al., 2015; Do and Faff, 2012). The results of the two-factor model are thus very similar to the results of the simple CAPM regression, with no noticeable changes in the alpha values of either methods. However, the significant beta coefficients of trigger 2.5/0 and 2.5/0.5 in the cointegration method under the CAPM regression are no longer significant, underlying that the coefficients of the CAPM were only marginally significant. From the above results we

can only conclude that the market and liquidity factors do only to a small extent or no extent explain the excess return of the pairs trading strategy outlined in this paper.

7.4.3. Fama-French + momentum + liquidity

To further investigate other potential factors that might explain the variations of the excess return, a Fama-French three-factor model including an additional momentum factor and the liquidity factor applied above, will be carried out below. Fama-French three-factor model is one of the best-known multi-factor models for explaining asset returns (Ang, 2014; Munk, 2018). The multi-factor model comprises the traditional CAPM factor and two additional factors; small-minus-big (SMB) and high-minus-low (HML). SMB represents the return of a portfolio of stocks in small companies subtracted the returns of a portfolio comprising large companies (Munk, 2018). Here, the size of the companies are determined by their market capitalization (Ang, 2014). HML represents the return on a portfolio of stocks with a high book-to-market value subtracted the returns of a portfolio of low book-to-market value (Munk, 2018). The book-to-market value is calculated as the book value of the company divided by its market capitalization (Ang, 2014). The momentum factor represents buying stocks that have gone up over the last six month and selling stocks that have declined over the same period. The fundamental idea is that “winner stocks continue to win and losers continue to lose” (Ang, 2014, p. 235). Lastly, the liquidity factor is the same as described above. The conclusions made by other authors within pairs trading show different results; Gatev et al. (2006) conclude only a significant momentum factor, Huck (2013) conclude significant excess market, HML and momentum factors, Huck and Afawubo (2015) conclude only a significant momentum factor, Smith and Xu (201) have no significant factors and Rad et al. (2015) conclude significant momentum and liquidity factors. As such, the existing literature does not give evidence to any definitive conclusion in the explanatory factors. This encouraged us to test this multi-factor model with the results of the regression is summarised in table 5 below:

Table 5: Multifactor model

	Distance method						Cointegration method					
	2 / 0	2 / 0.5	2.5 / 0	2.5 / 0.5	3 / 0	3 / 0.5	2 / 0	2 / 0.5	2.5 / 0	2.5 / 0.5	3 / 0	3 / 0.5
Before transaction costs												
Alpha	0.007	0.009	0.007	0.008	0.007	0.008	0.018	0.020	0.016	0.017	0.014	0.015
	<i>4.326***</i>	<i>4.594***</i>	<i>4.107***</i>	<i>4.296***</i>	<i>4.114***</i>	<i>4.080***</i>	<i>8.742***</i>	<i>9.306***</i>	<i>7.947***</i>	<i>8.167***</i>	<i>6.455***</i>	<i>6.740***</i>
Excess Return on the Market (beta)	-0.191	-0.219	-0.189	-0.218	-0.193	-0.222	-0.211	-0.224	-0.211	-0.226	-0.177	-0.202
	<i>-4.323***</i>	<i>-4.420***</i>	<i>-4.181***</i>	<i>-4.304***</i>	<i>-4.126***</i>	<i>-4.179***</i>	<i>-3.895***</i>	<i>-3.887***</i>	<i>-3.862***</i>	<i>-4.027***</i>	<i>-3.096***</i>	<i>-3.442***</i>
Small-Minus-Big Return	0.025	0.026	0.009	0.010	0.008	0.003	0.083	0.097	0.112	0.126	0.097	0.105
	<i>0.326</i>	<i>0.297</i>	<i>0.116</i>	<i>0.118</i>	<i>0.102</i>	<i>0.030</i>	<i>0.875</i>	<i>0.965</i>	<i>1.184</i>	<i>1.285</i>	<i>0.972</i>	<i>1.026</i>
High-Minus-Low Return	0.017	0.015	0.016	0.013	-0.003	-0.005	0.103	0.153	0.123	0.116	0.044	0.031
	<i>0.244</i>	<i>0.190</i>	<i>0.230</i>	<i>0.169</i>	<i>-0.034</i>	<i>-0.057</i>	<i>1.207</i>	<i>1.694*</i>	<i>1.436</i>	<i>1.320</i>	<i>0.490</i>	<i>0.333</i>
Momentum Factor	-0.031	-0.054	-0.025	-0.033	-0.053	-0.058	0.002	0.023	0.020	0.023	0.011	0.007
	<i>-0.757</i>	<i>-1.181</i>	<i>-0.603</i>	<i>-0.699</i>	<i>-1.230</i>	<i>-1.176</i>	<i>0.045</i>	<i>0.430</i>	<i>0.398</i>	<i>0.435</i>	<i>0.209</i>	<i>0.122</i>
Liquidity factor	-0.038	-0.044	-0.037	-0.039	-0.032	-0.031	-0.043	-0.048	-0.044	-0.050	-0.045	-0.049
	<i>-1.281</i>	<i>-1.317</i>	<i>-1.215</i>	<i>-1.141</i>	<i>-1.023</i>	<i>-0.864</i>	<i>-1.163</i>	<i>-1.232</i>	<i>-1.205</i>	<i>-1.315</i>	<i>-1.165</i>	<i>-1.232</i>
After transaction costs												
Alpha	0.001	0.000	0.001	0.001	0.002	0.002	0.003	0.001	0.004	0.003	0.004	0.003
	<i>1.007</i>	<i>0.304</i>	<i>1.456</i>	<i>1.102</i>	<i>1.913*</i>	<i>1.687*</i>	<i>2.476**</i>	<i>1.268</i>	<i>3.224***</i>	<i>2.647***</i>	<i>2.704***</i>	<i>2.468**</i>
Excess Return on the Market (beta)	-0.085	-0.082	-0.087	-0.088	-0.090	-0.092	-0.061	-0.061	-0.068	-0.069	-0.045	-0.050
	<i>-3.405***</i>	<i>-3.300***</i>	<i>-3.348***</i>	<i>-3.281***</i>	<i>-3.352***</i>	<i>-3.323***</i>	<i>-1.958*</i>	<i>-1.979**</i>	<i>-2.165**</i>	<i>-2.191**</i>	<i>-1.229</i>	<i>-1.377</i>
Small-Minus-Big Return	0.000	0.000	-0.010	-0.014	-0.014	-0.018	0.081	0.096	0.097	0.110	0.069	0.068
	<i>-0.004</i>	<i>-0.007</i>	<i>-0.226</i>	<i>-0.297</i>	<i>-0.294</i>	<i>-0.384</i>	<i>1.480</i>	<i>1.787*</i>	<i>1.757*</i>	<i>1.996**</i>	<i>1.080</i>	<i>1.089</i>
High-Minus-Low Return	-0.019	-0.028	-0.021	-0.028	-0.030	-0.037	0.055	0.058	0.063	0.051	0.023	0.011
	<i>-0.488</i>	<i>-0.712</i>	<i>-0.507</i>	<i>-0.668</i>	<i>-0.716</i>	<i>-0.847</i>	<i>1.114</i>	<i>1.200</i>	<i>1.276</i>	<i>1.025</i>	<i>0.408</i>	<i>0.200</i>
Momentum Factor	-0.030	-0.047	-0.032	-0.041	-0.051	-0.056	0.009	0.019	0.018	0.023	0.017	0.016
	<i>-1.295</i>	<i>-2.030**</i>	<i>-1.319</i>	<i>-1.666*</i>	<i>-2.068**</i>	<i>-2.182**</i>	<i>0.312</i>	<i>0.659</i>	<i>0.610</i>	<i>0.770</i>	<i>0.499</i>	<i>0.481</i>
Liquidity factor	-0.010	-0.011	-0.015	-0.013	-0.013	-0.010	-0.032	-0.033	-0.024	-0.023	-0.017	-0.017
	<i>-0.623</i>	<i>-0.656</i>	<i>-0.839</i>	<i>-0.718</i>	<i>-0.710</i>	<i>-0.527</i>	<i>-1.511</i>	<i>-1.601</i>	<i>-1.119</i>	<i>-1.052</i>	<i>-0.669</i>	<i>-0.678</i>
Annualized alpha values												
Before transaction costs	8.706%	10.378%	8.461%	9.944%	8.806%	9.896%	21.627%	24.515%	19.789%	20.957%	16.867%	18.045%
After transaction costs	1.141%	0.344%	1.719%	1.349%	2.339%	2.122%	3.535%	1.788%	4.657%	3.831%	4.506%	4.063%

Note: Critical values (t-statistics) for significance levels; 10% (*) = 1.6548, 5% (**) = 1.9753, 1% (***) = 2.6077.

As with the results outlined in both the single-factor and two-factor model, the multi-factor model presented in table 5 does not provide further statistical significant factors i.e. does not provide additional explanation of the return generated by the pairs trading strategy applied. For the distance method, it is solely the general market movements, i.e. beta with a statistical significant level of 5% across, that to some degree explain the excess return after transaction costs. Additionally, the momentum factor is a

statistically significant explanatory variable for the 2/0.5, 3/0 and 3/0.5 triggers in the distance method. This is in line with the findings of the existing literature on the distance method.

For the cointegration method, the alpha term remains the best explanatory factor for the generated return, except for the 2/0.5 trigger. In addition to the alpha term, the beta term of the excess market return is also statistically significant for the 2/0.5, 2.5/0 and 2.5/0.5 after transaction costs, when the multifactor model is applied. Also, trigger 2.5/0.5 is marginally significant with SMB. Our findings on the opening triggers of 3 with no significant factors are thus in line with Smith and Xu (2017) that also find no significant explanatory factors for their generated return in the 2000s.

7.4.4. Volatility

When considering the findings of the overall result and sub-period breakdowns, there are many implications of market volatility having an effect on performance of the applied pairs trading strategy on ETFs. In the attempt to investigate this potential relation, we regress the generated return on VIX. VIX is a measure of the option volatility of the S&P 500 index and designed to measure the 30-day expected volatility of the equity market, thus making it a useful proxy for the general volatility (Ang, 2014, p. 218; CBOE, n.d.). The results of the regression is summarised in table 6:

Table 6: VIX factor model

VIX	Distance method						Cointegration method					
	2 / 0	2 / 0.5	2.5 / 0	2.5 / 0.5	3 / 0	3 / 0.5	2 / 0	2 / 0.5	2.5 / 0	2.5 / 0.5	3 / 0	3 / 0.5
<i>Before transaction costs</i>												
Intercept	-0.020	-0.023	-0.020	-0.022	-0.021	-0.024	-0.004	-0.002	-0.005	-0.007	-0.004	-0.006
	<i>-5.180***</i>	<i>-5.416***</i>	<i>-5.056***</i>	<i>-5.152***</i>	<i>-5.276***</i>	<i>-5.292***</i>	<i>-0.747</i>	<i>-0.391</i>	<i>-1.068</i>	<i>-1.326</i>	<i>-0.728</i>	<i>-1.078</i>
VIX Index	0.001	0.002	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001
	<i>7.228***</i>	<i>7.623***</i>	<i>6.997***</i>	<i>7.197***</i>	<i>7.293***</i>	<i>7.301***</i>	<i>4.269***</i>	<i>4.075***</i>	<i>4.250***</i>	<i>4.622***</i>	<i>3.324***</i>	<i>3.808***</i>
R ²	0.253	0.274	0.241	0.252	0.257	0.257	0.106	0.097	0.105	0.122	0.067	0.086
Correlation	0.503	0.523	0.491	0.502	0.507	0.507	0.325	0.312	0.324	0.349	0.259	0.293
<i>After transaction costs</i>												
Intercept	-0.012	-0.012	-0.012	-0.013	-0.013	-0.013	-0.004	-0.007	-0.004	-0.006	-0.001	-0.003
	<i>-5.471***</i>	<i>-5.827***</i>	<i>-5.396***</i>	<i>-5.491***</i>	<i>-5.539***</i>	<i>-5.577***</i>	<i>-1.536</i>	<i>-2.390**</i>	<i>-1.439</i>	<i>-1.953</i>	<i>-0.349</i>	<i>-0.791</i>
VIX Index	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	<i>6.102***</i>	<i>6.164***</i>	<i>6.223***</i>	<i>6.179***</i>	<i>6.621***</i>	<i>6.572***</i>	<i>2.472**</i>	<i>2.853***</i>	<i>2.706***</i>	<i>3.023***</i>	<i>1.440</i>	<i>1.812*</i>

R^2	0.195	0.198	0.201	0.199	0.222	0.219	0.038	0.050	0.045	0.056	0.013	0.021
Correlation	0.441	0.445	0.448	0.446	0.471	0.468	0.195	0.224	0.213	0.237	0.115	0.145

Note: Critical values (t-statistics) for significance levels; 10% (*) = 1.6548, 5% (**) = 1.9753, 1% (***) = 2.6077.

From the above regression, it is evident that the VIX index has some explanatory power to the excess returns of the two methods. For both methods, the VIX coefficient is statistically significant at a 1% level except for the opening triggers of 3 in the cointegration method after transaction costs. The statistical significance of the VIX coefficients after transaction costs are somewhat similar, however slightly lower t-statistics and R^2 values. Intuitively, these results are somewhat related to the beta coefficients of the CAPM model, as the VIX index is based on the S&P 500 index. Nevertheless, the findings of the statistically significant explanatoriness of the VIX index confirms to some extent, not for opening trigger values of 3, the findings in the subperiod assessment of higher performance in more volatile periods. In continuation of this, Ang (2014, p. 218) displays the correlation between the VIX, stocks and bonds and concludes the correlation to returns of stocks is -0.39 and 0.12 for bonds. For the distance method, the correlation to VIX is around 0.5 before transaction costs and in the range between 0.45-0.47 after transaction costs. For the cointegration method, the correlation to VIX is around 0.30 before transaction costs and around 0.22 after transaction costs for the significant triggers, meaning we see an opposite pattern for this pairs trading strategy relative to the general characteristics of the stock market. Here it is important to distinguish between correlation and causation, and just because we find some correlation between the return of the two methods and the VIX, this does not imply the VIX is an explanatory variable for the return.

The negative correlation between the return and volatility of the single stock is also referred to as the leverage effect. The leverage effect is based on the idea that when the share price of a company drops, the financial leverage will increase proportionally as the equity value declines (Ang, 2014, p. 218). This makes the company more risky and increases its volatility (Ang, 2014, p. 218). The time for the leverage effect to decay is similar to the time of convergence for the pairs in our model. Bouchard, Matascz and Potters (2008) find that the leverage effect on stock indices decay after 10 days, whereas the average time to convergence of the cointegration method ranges between

9 and 15 days. As the time it takes for the leverage effect of single stocks to decay is on average 50 days, the leverage effect thus explains why the time to convergence of pairs trading with ETFs is much shorter than pairs trading with single stock (Bouchard et al., 2008). 50 days is consistent with the 57.5 days of convergence for Gatev et al. (2006). Whereas the price of single stocks is more sensitive to idiosyncratic risks in order for the market prices to drop, the stock indices require larger panic-like incidents for the leverage effect to be fully activated (Bouchard et al., 2008). This is also evident in our findings from the subperiod breakdown that we often generate profits after large events where the volatility of the assets are still high.

The positive correlation to the VIX index is thus consistent with the fact that we trade when the leverage effect is “activated” and close the trade when the effect has decayed. The quicker decay of the leverage effect for indices confirms our assumption about the divergence risk of the ETFs is lower than the divergence risk of stocks in pairs trading. The fact that the amplification of the leverage effect is higher for stock indices and more short-lived than stocks might also explain why the pairs trading with ETFs perform well in comparison with other pairs trading literature. As such, it is not necessarily the VIX index in itself that explains the returns of our methods, but rather it tells something about the market characteristics which the pairs trading strategy profit from.

Summary | We have in the above section examined the two methods in relation to the market and investigated potential explanatory factors for the excess returns of our strategies. CAPM shows that we generate statistically significant positive alpha values for 5 out of 6 triggers in the cointegration method after transaction costs, of which 2 are significant at a significance level of 1%. Further, CAPM shows that the cointegration method produces statistically insignificant beta values for 4 out of 6 triggers in the cointegration method after transaction costs, with the last two marginally significant. The distance method does not produce any statistically significant alpha, and we can thus not conclude whether the strategies yield a better result than the market portfolio and thus being compensated for taking on systemic risk. We had an assumption that liquidity was key to understanding the nature of our returns, but this factor is statistically insignificant to explain the excess returns. The same generally applies

for the SMB, HML and the momentum factors. Only the momentum factor could provide some evidence on the distance method. The lack of explanatory power, in general, of these factors, indicates that there must exist a number of other factors that can explain the alpha values generated. However, it must be taken into consideration that the sample period examined in the paper consists of periods of fundamentally different market conditions, for which it can be difficult to obtain a linear relationship between a factor and our returns.

Even though it is not measurable to the same extent as the factor models, the VIX index and the subsequent leverage effect provide some explanation for our findings about profiting from anomalies throughout the paper. The leverage effect could also explain the statistically significant correlation with the market for some triggers; when the market drops, the leverage effect is set in play, thus increasing the volatility. This generally produces more inefficiencies in the market, which is the foundation for returns of the pairs trading strategy applied. As such, it might not necessarily be the direct relationship to the return of the market or the market volatility that explains the returns of this paper, but rather the interconnection between these market factors which explain the underlying mechanisms of profiting from anomalies in the market.

7.5. Strategy assessment

Several conclusions have emerged from above analyses and will in this section be synthesised in order to conclude on the robustness and inferences of pairs trading relying on the application of ETFs carried out with the distance and cointegration method.

The distance method | The distance method is considered the benchmark approach of pairs trading but is also the simpler method of the two assessed in this paper. The simplicity of the distance method also becomes an obstacle for the method's profitability when using ETFs as underlying securities. The basic idea of identifying and trading pairs exhibiting a low distance between the return development is not as applicable when using ETFs compared to single stocks. The method provides satisfactory results with high Sharpe ratios before transaction costs, but the lower magnitude of the return on the individual positions is not robust to the related transaction costs. Here, we find

the reason to be the pairs composition of the distance method that favour ETFs tracking the same index.

A consequence of the reliance on pairs comprising ETFs tracking the same index is that the pairs require a larger divergence in order to generate a positive net profit. To this, the above analysis shows that 38% of all traded spots do not generate a gross profit large enough to withstand the costs of trading with a convergence equivalent to 3x standard deviations. This implies that the fundamental trading attributes of the distance method, when applied with ETFs, are not robust to the cost of trading. Further, we find that the period 2007-2010 creates all the profit and that the subsequent years unambiguously have imposed losses for the distance method. The problematic trading attributes mentioned above also come to light in the batting and slugging ratios, where the average “winning” percentage is below 30% for all triggers despite the positive results of the overall period. The high slugging ratio for the entire period is solely derived from the results of the first subperiod. The distance method generates a profit over the full period, but the robustness of this profit is not persistent. The fact that the profit of the method is generated in a short period of time, rather than over the course of the entire period makes it hard to rely on the method as a general trading strategy.

The lack of robustness is also present in the application of the factor models. Here, we find the distance method to generate a statistically significant alpha for all triggers before transaction costs but none after accounting for these. The factor models further reveal that the method has a statistically significant relation to the movements in the market, but no significant alpha. All in all, we cannot conclude that applying the distance method would yield any better results than applying the market portfolio.

The cointegration method | The cointegration method is the more complex method of the two, as it seeks to identify pairs exhibiting a statistically founded relationship during the formation period. Here, it is required that at least one of the ETFs explains the development of the other with the inclusion of a cointegration coefficient. By selecting the 20 traded pairs in this way, it is possible to identify more nuanced combinations of ETFs as the selection is based on an established long-run relationship. Naturally, two ETFs tracking the same index will thus also be selected under the

cointegration method, but the reliance of these ETFs is much lower for this method than the distance method. The cointegration method's ability to identify more nuanced combinations of ETFs is expressed by its pair composition where only 46% of the traded pairs comprise ETFs both tracking a US-based large-cap indices, and no more than 35% tracks the same index. The consequences of the pairs composition is a much more robust and persistent method that performs well after transaction costs. The batting ratios of the triggers in the cointegration method throughout the subperiods remain high except for the last subperiod. This generates a high batting ratio across the full period. The persistent characteristics of the cointegration method is also revealed by the gross profit for the traded pairs to a great extent are robust to the costs of trading with a convergence equivalent to 3x standard deviations. It is however noticed that the associated gross profit of a convergence equivalent to 3x standard deviation has experienced some variation in the second half of the overall sample period.

All in all, the cointegration method yields positive returns after transaction costs that both outperform the existing literature and the market portfolio for two out of six trigger settings. Despite 2009 and the last subperiod, the cointegration method has shown a strong performance. With this, the selected pairs of the cointegration approach have generally exhibited a high degree of cointegration and robustness to transaction costs. This robustness of the return is also evident in by statistically significant alphas after transaction costs for all triggers except trigger 2/0.5. Besides the significant alpha, we find a statistically significant explanatory variable in the beta term for 2/0.5, 2.5/0 and 2.5/0.5 after transaction costs. By this, we do not find any explanation of the return generated for the cointegration method in the more classic factor variables. When regressing the returns of the cointegration method on the VIX index, we find a statistically significant relation with a correlation of 20% for the 2 and 2.5 opening trigger and a lower correlation for the opening trigger of 3. This does not necessarily imply a causal relation, but there is a connection to the leverage effect that emerges from declining market prices.

We can thus conclude from the backtest of the two methods and triggers that the cointegration method exhibits superior performance to the distance method when applying pairs trading on ETFs throughout the sample period from 2007-2020. However, the

negative profit generated in the recent subperiod is worth keeping an eye on not becoming a new normal for the cointegration method.

8. Discussion

Above chapters conducted a thorough investigation of the application of pairs trading on ETFs and outlined the inferences related to the two pairs trading methods throughout various market conditions. We found that the cointegration method produces more robust results and thus more useful for any practical implementation, whereas the distance method is not able to produce any robust results in the form it currently takes. It is thus evident that the simplicity and nature of the distance method is practically suboptimal (Krauss, 2015).

The way this paper has conducted the backtest is subject to our belief systems and should therefore not be considered as a definitive true way of applying pairs trading on ETFs. The paper and subsequent empirical evidence provide a benchmark backtest for practitioners, investors and other academic readers to either apply or further accommodate to their respective preferences and risk aversion before implementing. As we have examined and synthesized in chapter 6, there exist many assumptions and parameterizations to consider when setting up a pairs trading strategy, and there might not exist an universal correct way of applying these parameters and methods. We have based the paper on the existing literature and synthesized what appeared to be the most optimal parameters based on our understanding. We have strived to keep the strategy as simple as possible for practical considerations and to mitigate any biases (Pedersen, 2015).

The results of this paper and backtest therefore give rise to the question for whom the findings of this is useful. On an overall note, the trading strategy has the potential to be implemented by both private investors, hedge funds or institutional investors. However, there are some practical obstacles that might complicate the process of applying the trading strategy.

Overall, the usefulness for different investors depends on the willingness to allocate the necessary time for a pairs trading strategy. Most private investors tend to not be willing to allocate the necessary time to gather and produce the required information to conduct the trading strategies, thus most often outsourcing the process to fund

managers for a fee (Pedersen, 2015). Despite the lack of willingness, it should be feasible to set up an automated procedure for conducting pairs trading as it is a rule-based algorithmic trading strategy.

However, the pairs trading strategy would practically require a significant amount of capital in order to execute the strategy. A practitioner must be able to trade an equal amount in 40 ETFs at any given time while ensuring the requirements of dollar-neutrality. With share prices up to USD 300 this would most likely require large investment to be made. For private investors with the access to the required capital there are no further complications in this regard, but for others - most likely the majority of private investors - this might prove to be a restriction for the practical feasibility of performing the strategy. For large institutional investors or hedge funds that most likely already have implemented numerous similar trading strategies, this should not prove to be a restriction of applying the model but rather a new perspective on how a pairs trading can be modified (Pedersen, 2015).

The assumption that regardless of the price development of the traded ETFs, we would always have the necessary capital to uphold collateral in our short-position might also prove to be another important consideration in regard for whom the trading strategy is useful. Evidently in the quant crisis from 2007, hedge funds and other arbitrage traders also had to exit their positions due to funding issues related to their short-positions (Pedersen, 2015). It is thus evident that institutional investors are also exposed to the risk of not being able to provide collateral when prices of securities in the long-short positions continue to diverge. For private investors the margin requirements would potentially be problematic well before the requirements would be problematic for institutional investors. Nevertheless, the assumption of these matters are important to consider for both institutional and private investors when implementing the trading strategy.

The biggest difference in practical implementation is the transaction costs. The expense ratios of the ETFs are the same regardless of the pairs trader. In regard to the commission costs, the institutional investors are exposed to lower commission than the commission paid by private investors. However, Do and Faff (2012) showcased that the gap between commissions paid by institutions and private investors has narrowed. The authors more specifically show that the gap in commission between all trades and

institutional trades was 16 bps in 1963 while this dropped to 2 bps in 2000. Their analysis does not go any farther, but indicates that the spreads between the institutional and general commission costs are narrowing. The commission thus makes the pairs trading strategy more favourable for institutional investors relative to private investors due to different transaction costs paid. With this, the findings of this paper should be assessed with caution for the private investor as it relies on the commission costs associated with an institutional investor. Regarding the short-sell costs, we assume that the investor is able to receive the LIBOR overnight rate as a rebate rate, which is an interbank rate that does not reflect the borrowing rates to private investors. Typically, brokerage firms charge further fees including fees for applying a margin account (Fidelity-b, n.d.). Hence, there might be a risk that the short-selling costs are higher for private investors compared to what is outlined in this paper. On the matter of short-selling, there might also be some practical obstacles for private investors as not all brokerage firms permit short-selling in the sense outlined in this paper. When it comes to the bid-ask spread this cost must be paid by both institutional investors and private investors. Throughout the paper it has been emphasized that the bid-ask spread is a significant portion of the transaction costs. A lesser exposure to the cost of the bid-ask spread would further enhance the profitability of the pairs trading strategy. Here larger institutional investors have an edge relative to the private investor. Larger banks or similar institutions typically have their own credit line of transactions that can provide a buyer or seller for some of the ETFs (Smith, 2010; Hartmann, Huang, and Schoenmaker, 2018). As such, the pairs traders in large banks are not exposed towards the same bid-ask costs as some of the costs are then paid in-house. Further, the timing of when to enter a position with regard to the bid-ask spread might be more suitable for institutional investors. As such, the application of this strategy intraday or with an actual trader allocating the necessary time to monitor the strategy will further improve the results. As we have learned in chapter 5 and 6, exploitation of arbitrage opportunities is a timing game. The market movement throughout the trading day could enable the realisation of profit, unable to be exploited by only relying on close prices as these only reflect the price at one point in time. Having this in mind, it makes sense that trying to compete in cases where anomalies are obvious to all can be difficult on a daily basis only. The full effect of the divergence will most likely have diminished

to a lesser degree at the end of the day, or only being apparent because it is not a profitable anomaly to trade on anymore. Thus, there may be an advantage for institutional investors who to a larger extent are able to monitor the pairs trading strategy intraday and thus harvest the profitability arising within a short window of time which a private investor only to a limited extent is able to.

We can thus conclude that the outlined pairs trading strategy of this paper is practically implementable for both private and institutional investors and hedge funds. However, the above considerations showcase that the implied usefulness of the trading strategy increases with the economies of scale, as transaction costs can be lowered by large institutions compared to private investors. For the private investors there might be several obstacles in setting up such a pairs trading strategy that would otherwise not be problematic for institutional investors. As such, we believe the most useful application of this trading strategy would be for hedge funds or other institutional investors that already have experience in trading such arbitrage strategies. However, more sophisticated private investors that have the required capital will also be relevant employers of our proposed trading strategy.

9. Further research

The above findings and considerations have contemplated a number of further research topics that potentially could improve the outlined pairs trading strategy using ETFs, but however needs a separate examination.

First and foremost, as we have conducted a backtest of a trading strategy on historical data, the most immediate and important further research is to put the backtest into action, either by actually conducting a real-time trading or testing the results of the paper on an out-of-sample dataset. This would result in a similar approach to that carried out by Gatev et al. (2006) that tested their publication from 1999 again in 2006. Further, this will provide the most optimal type of verification of the results of this paper. Apart from this, it would be interesting to see other backtests of the pairs trading strategy on ETFs from other markets to identify whether the same trends and results are evident.

Further, this paper is carried out starting from predefined parameterizations which we subsequently could not diverge from or refine due to the risk of data snooping or data mining. As a consequence of this, there exist a number of further opportunities to further examine alternative parameterization based on the knowledge put forth by this paper which was also touched upon at the end of chapter 7. As stated in the possible improvement of chapter 7, it would be interesting to further investigate a more dynamic opening and closing trigger setup that could provide more profitable trades after transaction costs. Further, to mitigate that ETFs might be too close or too cointegrated, an initial screening of the ETF universe could potentially provide a universe of ETFs that could be closer to the “sweet spot” of anomalies. The latter however has a high risk of sample bias.

As we have shown that ETFs perform well in comparison to the findings of the existing literature that use single stocks, it would be valuable to conduct a comparison analysis of the performance of respectively ETFs and stocks. Rudy and Dunis (2010) have attempted to do this, but their paper has not been published, why it cannot fully be stated whether the findings are representative of the performance of the two asset classes. Nonetheless, such a comparison could very well further test a number of the findings outlined in this paper. A cross-ETF-stocks universe could potentially provide interesting views on the relationship between stocks and ETFs.

As pairs trading is a practical methodology we emphasize that there exist no correct way of sampling, detecting and trading pairs and we advocate that practitioners and pairs traders should not just implement a normative method, but consider and research the parameters and methods that fit their risk aversion, time and capital at their disposal.

10. Conclusion

This paper has investigated the application of US equity ETFs on a pairs trading strategy by backtesting the strategy in the period of 2007 to 2020. Overall, our results demonstrate that while on the one hand, the use of ETFs fosters a good and efficient execution of the pairs trading strategy, simultaneously enables pairs to comprise ETFs tracking the same index which has shown to be associated with realised losses of the positions taken in the market due to the costs of trading. As a consequence of the fundamentally different procedures of the two applied pairs trading methods, the cointegration method is more subject to the good attributes related to the use of ETFs and thus provides better results compared to the distance method, as well as current literatures relying on single stocks.

In general, the cointegration method generates returns robust to the high transaction costs associated with the execution of a pairs trading strategy. This is reflected in higher batting ratios and positive statistically significant alpha for all of the tested trigger combinations, besides the one associated with most trading activity. Complete market neutrality is not achieved for the cointegration method across all triggers, but rather a minor negative correlation with the market is identified for the opening trigger of 2 and 2.5, whereas we cannot deny that the method is market neutral with an opening trigger of 3. The distance method generates both a statistically significant alpha and beta before transaction costs, and whereas the beta remains statistically significant, the alpha becomes insignificant when the practical implications for the method's application are taken into account. The backtest thus reveals the distance method is practically inappropriate for pairs trading with the use of ETFs and only in highly volatile periods generates a positive net return. In a similar fashion, periods with more market volatility have also provided better returns for the cointegration method. The profitability associated with these periods lies in the pairs trading's fundamental dynamics of exploiting anomalies in the price development which are particularly present in a market characterized by more volatility. The application of ETFs is particularly attractive, as pairs of ETFs tend to converge much quicker and more often compared to the current literature's examination of pairs trading on single stocks. However, the use of ETFs also introduces a suboptimality as a consequence of the ability to generate pairs of ETFs exhibiting identical price development. Especially the

distance method tends to favour these pairs. To this, the backtesting finds that pairs consisting of ETFs tracking the same index negatively influence the performance, as these pairs generally do not allow for a large enough divergence to offset the costs of trading. Unlike the distance method, the cointegration method generally showcases a greater reliance on pairs comprising different types of ETFs, or pairs comprising one ETFs tracking a fraction of the general market, while the other ETF tracks the broader market. Here, the latter provides the majority of the returns for the cointegration method. The nature of the selection processes thus differentiate the performance of the two methods, with the distance method tending to favour pairs with adverse attributes, and the cointegration method generally relying on pairs characterised by higher lucrativity.

Our initial expectations about this new way of applying a pairs trading strategy is met for both methods when measured without considerations to the associated costs of executing the strategy. This is expressed by a better Sharpe ratio than both the existing literature and general market portfolio, primarily driven by a lower risk aspect. After transaction costs, two trigger settings for the cointegration method still provide satisfactory results for the sample period by outperforming the current literature measured by Sharpe ratio and the market measured both by alpha and Sharpe ratio.

The results of this paper provide evidence to the applicability and efficiency of pairs trading relying on ETFs when executed through the cointegration method. As such, the findings of this paper infer that pairs trading with ETFs provides a better alternative to the pairs trading strategy with single stocks and the general market when executed with the cointegration method. Under the current setting, the results of the distance method are consistent with the low profitability reported by the current literature on single stocks. The backtest however also reveals that the trading strategy provided in this paper should be further refined to mitigate the drawbacks of the pairs composition and through improvements of the proposed parameters.

11. References

- Amadeo, K. (2019, November 20). 2008 Financial Crisis Timeline. Retrieved from <https://www.thebalance.com/2008-financial-crisis-timeline-3305540>
- Andrade, S., Di Pietro, V., & Seasholes, M. (2005). Understanding the Profitability of Pairs Trading. Working Paper - UC Berkley. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.466.1045&rep=rep1&type=pdf>
- Ang, A. (2014). Chapter 6, 10, 13. In Asset Management (p. 193). Oxford, United Kingdom: Oxford University Press.
- Asteriou, D., & Hall, S. G. (2011). Chapter 15-17. In Applied Econometrics (2nd ed., pp. 319–390). London, United Kingdom: Palgrave Macmillan.
- Baronyan, S., Boduroglu, İ., & Sener, E. (2010). Investigation of stochastic pairs trading strategies under different volatility regimes. *The Manchester School*, 78, 114–134. <https://doi.org/10.1111/j.1467-9957.2010.02204.x>
- Ben-David, I., Franzoni, F. A., & Moussawi, R. (2017). Exchange Traded Funds (ETFs). *Annual Review of Financial Economics*, 9, 16–64. <https://doi.org/10.2139/ssrn.2865734>
- Bergen, S. (2019, November 6). Eigenvalues of a 2x2 Matrix. Retrieved from <https://www.vcalc.com/wiki/SavannahBergen/Eigenvalues+of+a+2x2+Matrix>
- Blackrock. (2020, January 21). The Evolution of ETFs. Retrieved from <https://www.visualcapitalist.com/the-26-year-history-of-etfs-in-one-infographic/>
- Borgun, E., Kurun, E., & Guven, S. (2009). Dynamic Pairs Trading Strategy For The Companies Listed In The Istanbul Stock Exchange. Working Paper - University of Munich. Retrieved from <https://mpra.ub.uni-muenchen.de/id/eprint/19887>
- Bouchaud, J., Matacz, A., & Potters, M. (2008). The leverage effect in financial markets: retarded volatility and market panic. *The Research Division of Capital Fund Management*, 1–12. Retrieved from <https://arxiv.org/pdf/cond-mat/0101120.pdf>
- Bradford, A. (2017, July 25). Deductive Reasoning vs. Inductive Reasoning. Retrieved 17 April 2020, from <https://www.livescience.com/21569-deduction-vs-induction.html>
- Brooks, C. (2008). Chapter 7. In *Introductory Econometrics For Finance* (2nd ed., pp. 318–378). Cambridge, United Kingdom: Cambridge University Press.
- Broumandi, S., & Reuber, T. (2012). Statistical arbitrage and FX exposure with South American ADRs listed on the NYSE. *Financial Assets and Investing*, 5–18. <https://doi.org/10.5817/fai2012-2-1>

Broussard, J. P., & Vaihekoski, M. (2012). Profitability of pairs trading strategy in an illiquid market with multiple share classes. *Journal of International Financial Markets, Institutions and Money*, 22(5), 1188–1201.
<https://doi.org/10.2139/ssrn.1729182>

Caldeira, J., & Moura, G. V. (2013). Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy. *SSRN Electronic Journal*, 49–80.
<https://doi.org/10.2139/ssrn.2196391>

Carlson, M. A. (2007). A Brief History of the 1987 Stock Market Crash With a Discussion of the Federal Reserve Response. *FEDS Working Papers*, 1–25.
<https://doi.org/10.2139/ssrn.982615>

CBOE. (n.d.). Vix-Index. Retrieved 2 May 2020, from <http://www.cboe.com/vix>

Charles Schwab. (2020). 9 Frequently Asked Questions About Short Selling. Retrieved from <https://www.schwab.com/resource-center/insights/content/9-frequently-asked-questions-about-short-selling>

Chen, Z., & Knez, P. (1995). Measurement of Market Integration and Arbitrage. *The Review of Financial Studies*, 8(2), 287–325. Retrieved from
<https://www.jstor.org/stable/2962274?seq=1>

CRSP. (n.d.). CRSP U.S. Total Market Index | CRSP - The Center for Research in Security Prices. Retrieved 30 April 2020, from <http://www.crsp.org/products/investment-products/crsp-us-total-market-index>

Curtis, G. (2019, June 9). The Basics of the Bid-Ask Spread. Retrieved from <https://www.investopedia.com/trading/basics-of-the-bid-ask-spread/>

DeFranzo, S. E. (2019, May 3). Difference between qualitative and quantitative research. Retrieved 17 April 2020, from <https://www.snapsurveys.com/blog/qualitative-vs-quantitative-research/>

Deutsche Bank. (2017). ETF monthly . Retrieved from <https://etf.dws.com/CHE/DEU/Download/Research-Europe/efcf86d9-cd06-4e91-a072-2126095cb982/ETF-Monthly-Europe.pdf>

Deville, L. (2008). Exchange Traded Funds: History, Trading, and Research. *Handbook of Financial Engineering*, 67–98. https://doi.org/10.1007/978-0-387-76682-9_4

Do, B., Faff, R., & Hamza, K. (2006). A New Approach to Modeling and Estimation for Pairs Trading. Working Paper.

Do, B., & Faff, R. (2010). Does Simple Pairs Trading Still Work? *Financial Analysts Journal*, 66(4), 83–95. <https://doi.org/10.2469/faj.v66.n4.1>

Do, B., & Faff, R. (2012). Are pairs trading profits robust to trading costs? *Journal of Financial Research*, 35(2), 261–287. <https://doi.org/10.1111/j.1475-6803.2012.01317.x>

- Douven, I. (2017, April 28). Abduction (Stanford Encyclopedia of Philosophy/Summer 2017 Edition). Retrieved from <https://plato.stanford.edu/archives/sum2017/entries/abduction/>
- Duc, F., & Schorderet, Y. (2008). Market Risk Management for Hedge Funds. In Annualization problem (pp. 155–179). Hoboken, NJ, United States: Wiley.
- Dunis, C. L., Laws, J., & Evans, B. (2006). Trading futures spreads: An application of correlation and threshold filters. *Applied Financial Economics*, 16(12), 903–914. Retrieved from <https://doi.org/10.1080/09603100500426432>
- Dunis, C. L., Laws, J., & Evans, B. (2008). Trading futures spread portfolios: applications of higher order and recurrent networks. *The European Journal of Finance*, 14(6), 503–521. Retrieved from https://econpapers.repec.org/article/tafeur-jfi/v_3a14_3ay_3a2008_3ai_3a6_3ap_3a503-521.htm
- Egholm, L. (2014). Videnskabsteori (1st ed., Vol. 1). Copenhagen, Denmark: Hans Reitzels.
- Ehrman, D. S. (2006). The Handbook of Pairs Trading. Hoboken, NJ, United States: Wiley.
- Engelberg, J., Gao, P., & Jagannathan, R. (2009). An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1330689>
- etfdb.com. (2020). ETFs. Retrieved 24 April 2020, from <https://etfdb.com/etfs/>
- Fabozzi, F. J., & Asness, C. (2004). Chapter 7. In Short Selling (pp. 9–16). Hoboken, NJ, United States: John Wiley & Sons.
- FED. (2020). Broad General Collateral Rate - Federal Reserve Bank of New York. Retrieved from <https://apps.newyorkfed.org/markets/autorates/bgr>
- Fidelity. (n.d.-a). CEFs versus ETFs and Mutual Funds - Fidelity. Retrieved 22 April 2020, from <https://www.fidelity.com/learning-center/investment-products/closed-end-funds/cefs-mutual-funds-etfs>
- Fidelity. (n.d.-b). Trading Commissions and Margin Rates - Fidelity. Retrieved 6 May 2020, from https://www.fidelity.com/trading/commissions-margin-rates-c?version=v2&s_tnt=123583:2:0
- Fletcher, N. (2014, December 12). World stock markets tumble as FTSE suffers worst weekly fall since 2011. Retrieved from <https://www.theguardian.com/business/2014/dec/12/world-stock-markets-tumble-and-ftse-suffers-worst-fall-since-2011>
- Frankenfield, J. (2019, August 20). Commission. Retrieved from <https://www.investopedia.com/terms/c/commission.asp>

FRED. (2020, April 16). FRED Economic Data. Retrieved from <https://fred.stlouisfed.org/series/USDONTD156N>

French, K. (n.d.). Kenneth R. French - Data Library. Retrieved 14 April 2020, from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

G. Rompotis, G. (2019). IPO ETFs: An Alternative Way to Enter the Initial Public Offering Business. *Linear and Non-Linear Financial Econometrics -Theory and Practice*. <https://doi.org/10.5772/intechopen.90269>

Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs Trading: Performance of a Relative-Value Arbitrage Rule. *Review of Financial Studies*, 19(3), 797–827. <https://doi.org/10.1093/rfs/hhj020>

Gatev, E., Goetzmann, W., & Rouwenhorst, K. G. (1999). Pairs Trading: Performance of a Relative Value Arbitrage Rule. *NBER Working Paper No. 7032*, 797–827. <https://doi.org/10.3386/w7032>

Granger, C. W. J. (2009). In Praise of Pragmatics in Econometrics. *The Methodology and Practice of Econometrics*, 255–267. <https://doi.org/10.1093/acprof:oso/9780199237197.003.0010>

Gregory, I., Ewald, C.-O., & Knox, P. (2011). Analytical Pairs Trading Under Different Assumptions on the Spread and Ratio Dynamics. *Economics*, 1–40. <https://doi.org/10.2139/ssrn.1663703>

Hakes, J. K., & Sauer, R. D. (2006). An Economic Evaluation of the Moneyball Hypothesis. *Journal of Economic Perspectives*, 20(3), 173–185. <https://doi.org/10.1257/jep.20.3.173>

Hartmann, P., Huang, H., & Schoenmaker, D. (2018). Will human forex dealers become extinct? in *The Changing Fortunes of Central Banking*. In (pp. 241–258). Cambridge, United Kingdom: Cambridge University Press.

Heritage, C. (2010, August 5). Four Factors That Determine Your Results. Retrieved from <https://www.nasdaq.com/articles/four-factors-determine-your-results-2010-08-05>

Hetherington. (n.d.). Fallibilism | Internet Encyclopedia of Philosophy. Retrieved 15 April 2020, from <https://www.iep.utm.edu/fallibil/>

Hill, J., Nadig, D., & Hougan, M. (2015). A Comprehensive Guide to Exchange Traded Funds. CFA Institute Research Foundation, 1. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2616223

Huck, N. (2013). The high sensitivity of pairs trading returns. *Applied Economics Letters*, 20(14), 1301–1304. <https://doi.org/10.1080/13504851.2013.802121>

Huck, N., & Afawubo, K. (2015). Pairs trading and selection methods: is cointegration superior? *Applied Economics*, 47(6), 599–613.
<https://doi.org/10.1080/00036846.2014.975417>

iborate. (2020). USD Libor rate. Retrieved from <http://iborate.com/usd-libor/>

ICI. (n.d.). ETF. Retrieved 10 April 2020, from <https://www.ici.org/research/stats/etf/>

IGI Global. (n.d.). What is Praxis | IGI Global. Retrieved 17 April 2020, from <https://www.igi-global.com/dictionary/the-power-of-praxis/23190>

iShares. (n.d.). iShares Russell 1000 ETF. Retrieved 30 April 2020, from <https://www.ishares.com/us/products/{portfolioId}/{seoSlug}>

ITG. (2018). Global Cost review (2018). Retrieved from <https://www.virtu.com/uploads/2019/02/ITG-Global-Cost-Review-4Q18.pdf>

Jacobs, H., & Weber, M. (2015). On the determinants of pairs trading profitability. *Journal of Financial Markets*, 23, 75–97. <https://doi.org/10.1016/j.finmar.2014.12.001>

Joyner, A. (2020, March 24). Investors eye cracks in \$4.4 trillion U.S. ETF market as sell-off rages. Retrieved from <https://www.reuters.com/article/us-health-coronavirus-etfs/investors-eye-cracks-in-4-4-trillion-u-s-etf-market-as-sell-off-rages-idUSKBN21B1IQ>

Justesen, L., & Mik-Meyer, N. (2010). Kvalitative metoder i organisations- og ledelsesstudier (1st ed., Vol. 1). Copenhagen, Denmark: Hans Reitzels Forlag.

Kaushik, V., & Walsh, C. A. (2019). Pragmatism as a Research Paradigm and Its Implications for Social Work Research. *Social Sciences*, 8(9), 255.
<https://doi.org/10.3390/socsci8090255>

Kenny, T. (2019, November 19). What is the European Debt Crisis? Retrieved from <https://www.thebalance.com/what-is-the-european-debt-crisis-416918>

Kim, T. (2018, October 5). Experts — including Warren Buffett — say short selling can be beneficial for markets. Retrieved from <https://www.cnbc.com/2018/10/05/experts-including-warren-buffett--say-short-selling-can-be-beneficial-for-markets.html>

Kingsley, P. (2012, August 7). Financial crisis: timeline. Retrieved from <https://www.theguardian.com/business/2012/aug/07/credit-crunch-boom-bust-time-line>

Krauss, C. (2015). Statistical arbitrage pairs trading strategies: Review and outlook. *IWQW Discussion Papers* 09/2015, 09, 1–55. Retrieved from <https://www.econstor.eu/handle/10419/116783>

Kvale, S., & Brinkmann, S. (2015). Interview. Copenhagen, Denmark: Hans Huber.

Kwilla, S. (2019, December 2). Global ETF AUM Touches \$6 Trillion: What Are the Hot Picks? Retrieved from <https://www.nasdaq.com/articles/global-etf-aum-touches-%246-trillion%3A-what-are-the-hot-picks-2019-12-02>

Legg, C., & Hookway, C. (2019, March 14). Pragmatism (Stanford Encyclopedia of Philosophy/Spring 2019 Edition). Retrieved from <https://plato.stanford.edu/archives/spr2019/entries/pragmatism/>

Li, M. L., Chui, C. M., & Li, C. Q. (2014). Is pairs trading profitable on China AH-share markets? *Applied Economics Letters*, 21(16), 1116–1121.
<https://doi.org/10.1080/13504851.2014.912030>

Lo, A. (1994). Data-snooping biases in financial analysis. *Harris & Harris Group Professor of Finance*, 59–66. Retrieved from https://www.hillsdaleinv.com/uploads/Data-Snooping_Biases_in_Financial_Analysis,_Andrew_W._Lo.pdf

Lütkepohl, H., Saikkonen, P., & Trenkler, C. (2001). Maximum eigenvalue versus trace tests for the cointegrating rank of a VAR process. *SFB 373 Discussion Papers*, 83. Retrieved from <https://ideas.repec.org/p/zbw/sfb373/200083.html>

Mead, G. H. (1934). The 'I' and the 'Me' in Mind, Self, and Society. (pp. 173–178). Chicago, USA: The university of Chicago Press.

Munk, C. (2018). Financial Markets and Investments. Copenhagen, Denmark: Lecture notes.

Naser, H. (2017). On the cointegration and causality between oil market, nuclear energy consumption, and economic growth: evidence from developed countries. *Energy, Ecology and Environment*, 2(3), 182–197. <https://doi.org/10.1007/s40974-017-0052-0>

Nekrašas, E. (2001). Pragmatism and positivism. *Problemos*, ISSN 1392-1126, 41–52. Retrieved from https://www.researchgate.net/publication/330627797_Pragmatizmas_ir_pozityvismas/fulltext/5c4b3014458515a4c73ffc41/Pragmatizmas-ir-pozityvismas.pdf

OECD. (2020, April 16). Interest rates - Short-term interest rates - OECD Data. Retrieved from <https://data.oecd.org/interest/short-term-interest-rates.htm>

Pagano, M., Serrano, A., & Zechner, J. (2019). Can ETFs contribute to systemic risk? (Reports of the Advisory Scientific Committee). Retrieved from https://www.esrb.europa.eu/pub/pdf/asc/esrb.asc190617_9_canetfscontributesystemicrisk~983ea11870.en.pdf

Herlemont, C. (2003). Pairs Trading, Convergence Trading, Cointegration. (2004). Working Paper, YATS Finances & Technologies, 1. Retrieved from https://www.researchgate.net/publication/249912444_Pairs_Trading_Convergence_Trading_Cointegration

- Papadakis, G., & Wysocki, P. (2007). Pairs Trading and Accounting Information. Working Paper - MIT Sloan and Boston University. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.151.5583&rep=rep1&type=pdf>
- Pástor, L., & Stambaugh, R. F. (2003). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy*, 111(3), 642–685. <https://doi.org/10.1086/374184>
- Pastor, L., & Stambaugh, R. F. (2019). Liquidity Risk After 20 Years. Chicago Booth Research Paper No. 19-13, 1–24. <https://doi.org/10.2139/ssrn.3371948>
- Payne, G., & Payne, J. (2011). Key Concepts in Social Research. Bias, 1–5. <https://doi.org/10.4135/9781849209397>
- Pedersen, L. H. (2015). Efficiently inefficient: how smart money invests and market prices are determined. Princeton, United Kingdom: Princeton University Press.
- Perlin, M. S. (2009). Evaluation of pairs-trading strategy at the Brazilian financial market. *Journal of Derivatives & Hedge Funds*, 15(2), 122–136. <https://doi.org/10.1057/jdhf.2009.4>
- Pisani, B. (2015, September 25). What happened during the Aug 24 'flash crash'. Retrieved from <https://www.cnbc.com/2015/09/25/what-happened-during-the-aug-24-flash-crash.html>
- Prettenhofer, P. (2014, February 7). Ordinary Least Squares in Python. Retrieved from https://blog.datarobot.com/ordinary-least-squares-in-python?fbclid=IwAR1iaGcmIsAMsep0eQIdfXhNZ-rEt6gcQcR3kxCC9DrqRxrtcZLi5Wi6-_a0
- Rad, H., Low, R. K. Y., & Faff, R. (2015). The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quantitative Finance*, 16(10), 1541–1558. <https://doi.org/10.1080/14697688.2016.1164337>
- Rao, A. (2017, August 2). Fast-moving Markets: Revisiting the August 2007 Quant Crunch in Real Time. Retrieved from <https://www.msci.com/www/blog-posts/fast-moving-markets-revisiting/0696968896>
- Reklaitis, V. (2015, August 25). what may have caused the flash crash in some big ETFs. Retrieved from <https://www.marketwatch.com/story/heres-what-may-have-caused-the-flash-crash-in-some-big-etfs-2015-08-25>
- Rudy, J., Dunis, C., & Laws, J. (2010). Profitable Pair Trading: A Comparison Using the S&P 100 Constituent Stocks and the 100 Most Liquid ETFs. SSRN, 1–19. <https://doi.org/10.2139/ssrn.2272791>
- Rudy, J., Dunis, C., & Laws, J. (2011). Mean Reversion Based on Autocorrelation: A Comparison Using the S&P 100 Constituent Stocks and the 100 Most Liquid ETFs. *ETF Risk*, 36–41. <https://doi.org/10.2139/ssrn.2272794>

Schizas, P., Thomakos, D. D., & Wang, T. (2011). Pairs Trading on International ETFs. SSRN, n.d. <https://doi.org/10.2139/ssrn.1958546>

Schumpeter, J. (1933). The Common Sense of Econometrics. *Econometrica*, 1(1), 5. <https://doi.org/10.2307/1912225>

Sharpe, W. (1994, April 1). The Sharpe Ratio. Retrieved 28 April 2020, from <http://web.stanford.edu/%7Ewfsharpe/art/sr/SR.htm>

Shiller, R. J. (1992). Market Volatility. Boston, USA: MIT Press.

Shleifer, A., & Vishny, R. W. (1997). The Limits of Arbitrage. *The Journal of Finance*, 52(1), 35–55. <https://doi.org/10.1111/j.1540-6261.1997.tb03807.x>

Smith, C. (2010). Transaction costs In How to Make a Living Trading Foreign Exchange (pp. 5–10). Retrieved from <https://books.google.dk/books?id=qAJrD-wAAQBAJ&pg=PA7&lpg=PA7&dq=large+banks+no+bid-ask+spread+cost&source=bl&ots=YVLVHSSqCq&sig=ACfU3U3inwBl7v9Qz6cUZ-tRhvwMc7AL-LoQ&hl=en&sa=X&ved=2ahUKEwjsu4PX0rDpAhWC2qQKHZ6LBH0Q6AEwCX-oECAcQAQ#v=onepage&q=large%20banks%20no%20bid-ask%20spread%20cost&f=false>

Smith, R. T., & Xu, X. (2017). A good pair: alternative pairs-trading strategies. *Financial Markets and Portfolio Management*, 31(1), 1–26. <https://doi.org/10.1007/s11408-016-0280-x>

Statsmodels. (n.d.-a). Introduction to Stattools. Retrieved 20 March 2020, from <https://www.statsmodels.org/devel/index.html>

Statsmodels. (n.d.-b). statsmodels.regression.linear_model.OLS. Retrieved 20 March 2020, from https://www.statsmodels.org/devel/generated/statsmodels.regression.linear_model.OLS.html

Statsmodels. (n.d.-c). statsmodels.tools.tools.add_constant. Retrieved 20 March 2020, from https://www.statsmodels.org/stable/generated/statsmodels.tools.tools.add_constant.html

Statsmodels. (n.d.-d). statsmodels.tsa.stattools.adfuller. Retrieved 20 March 2020, from <https://www.statsmodels.org/stable/generated/statsmodels.tsa.stattools.adfuller.html>

Statsmodels. (n.d.-e). statsmodels.tsa.vector_ar.var_model.VAR. Retrieved 21 April 2020, from https://www.statsmodels.org/stable/generated/statsmodels.tsa.vector_ar.var_model.VAR.html

Statsmodels. (n.d.-f). statsmodels.tsa.vector_ar.vecm.coint_johansen. Retrieved 20 March 2020, from https://www.statsmodels.org/stable/generated/statsmodels.tsa.vector_ar.vecm.coint_johansen.html

Statsmodels. (n.d.-g). statsmodels.tsa.vector_ar.vecm.JohansenTestResult. Retrieved 20 March 2020, from https://www.statsmodels.org/stable/generated/statsmodels.tsa.vector_ar.vecm.JohansenTestResult.html#statsmodels.tsa.vector_ar.vecm.JohansenTestResult

Stock, J. H., & Watson, M. W. (2015). Chapter 1, 14, 16. In *Introduction to Econometrics, Update, Global Edition* (3rd ed., pp. 47, 568–634, 684–716). Harlow, England: Pearson Education Limited.

Thorne, E. (2003). A Perspective on Quantitative Finance: Models for Beating the Market. *Quantitative Finance Review* 2003, 33–38. Retrieved from <http://www.edwardothorp.com/wp-content/uploads/2016/11/thorpwilmottqfinrev2003.pdf>

Vanguard. (n.d.). Vanguard ETF Profile | Vanguard. Retrieved 30 April 2020, from <https://investor.vanguard.com/etf/profile/portfolio/vti>

Vidyamurthy, G. (2004). *Pairs Trading*. Hoboken, NJ, United States: Wiley.

Virtu. (2019). Global cost review. Retrieved from https://www.virtu.com/uploads/documents/Virtu_GlobalCostReview3Q19.pdf

Wearden, G. (2011, August 10). Debt crisis sends financial markets into turmoil – Monday 8 August 2011. Retrieved from <https://www.theguardian.com/business/2011/aug/08/stock-market-turmoil-us-downgrade-eurozone-crisis>

WRDS. (n.d.). Fama-French Portfolios and Factors. Retrieved 28 April 2020, from <https://wrds-web.wharton.upenn.edu/wrds/ds/famafrench/index.cfm?CFTREEITEMKEY=Fama%2520&%2520French,%2520Liquidity%2520Factors,%2520and%2520Risk-Free%2520Rate>

Yahoo Finance. (n.d.). CBOE Volatility Index (^VIX) Charts, Data & News - Yahoo Finance. Retrieved 2 May 2020, from <https://finance.yahoo.com/quote/%5EVIX/>

Yu, & Webb. (2014). The Profitability of Pairs Trading Strategies Based on ETFs. Working Paper, 1. Retrieved from http://swfa2015.uno.edu/B_Asset_Pricing_III/paper_196.pdf

12. Appendices

Appendix 1 - List of all ETFs

See excel file “Appendix 1,2,3,8,9,10,11”

Appendix 2 - ETF prices

See excel file “Appendix 1,2,3,8,9,10,11”.

Appendix 3 - Mail from CapitalIQ, missing data points

See excel file “Appendix 1,2,3,8,9,10,11”

Appendix 4 - Python code

See zip file “Appendix 4 - code and ranking of cointegration”. Included in the zip file are two input excel files, “Formation_period_ultimo” and “Formation_period_medio”. Here the Python code “ETF_COINTEGRATION_FINAL” is run for every formation period starting either in July or January. The excel file “Formation_period_ultimo_2007” shows an example of the output of the formation period in 2007 when running the python code. Excel file “Formation_period_ultimo_2007_ranking” showcase the ranking procedure of the output - this has been done in the same excel file, but for better comparison this is seperated.

Appendix 5 - Ranking of pairs - distance method

See excel file “Appendix 5 - Distance_ranking_Formation_period_2006”. The excel file is an example of a ranking file which has been conducted on all formation periods. The excel file comprises two sheets; “Spread data formation” is the spread between the ETFs in the respective pair in the formation period, and “detecting pairs” calculates the SSD and the standard deviations. The pairs are then ranked based on the SSD.

Appendix 6 - Cointegration trading sheet

See excel file “Appendix 6 - Cointegration trading sheet”. This excel file provides the basis upon which the trading strategy has been conducted. With the input sheet it is possible to choose the desired trading period and trigger setting, and the sheet “Returns incl. costs” depicts the weighted portfolio return including transaction costs. This sheet is run for all trading periods and triggers used in this paper.

Appendix 7 - distance trading sheet

See excel file “Appendix 7 - Cointegration trading sheet”. This excel file is similar to appendix 6 for the distance method. This sheet is run for all trading periods and triggers used in this paper.

Appendix 8 - Summary of transaction costs in existing literature

See excel file “Appendix 1,2,3,8,9,10,11”

Appendix 9 - Average transaction costs

See excel file “Appendix 1,2,3,8,9,10,11”

Appendix 10 - ETF pairings distance method

See excel file “Appendix 1,2,3,8,9,10,11”

Appendix 11 - ETF pairings cointegration method

See excel file “Appendix 1,2,3,8,9,10,11”

Appendix 12 - Full period descriptive statistics

See excel file “Appendix 12-18” and sheet “Appendix 12 - Full table”. The data for the full period table can be found in sheet “data for appendix 12” in the same excel file. Data regarding trading statistics can be found in the sheets “additional data to tables” in the same excel file.

Appendix 13 - 2007-2010 descriptive statistics

See excel file “Appendix 12-18” and sheet “Appendix 13 - 2007-2010”. The appendix shows the table for the subperiod, annual tables and the returns of the 20 pairs in the trading periods. The data for the periodical table can be found in sheet “data for appendix 13 - 2007-2010” in the same excel file and data for the annual tables can be found in the sheets “Data for annual tables”. Data regarding trading statistics can be found in the sheets “Additional data to tables” in the same excel file.

Appendix 14 - 2010-2012 descriptive statistics

See excel file “Appendix 12-18” and sheet “Appendix 14 - 2010-2012”. The appendix shows the table for the subperiod, annual tables and the returns of the 20 pairs in the trading periods. The data for the periodical table can be found in sheet “data for appendix 14 - 2010-2012” in the same excel file and data for the annual tables can be found in the sheets “Data for annual tables”. Data regarding trading statistics can be found in the sheets “Additional data to tables” in the same excel file.

Appendix 15 - 2012-2016 descriptive statistics

See excel file “Appendix 12-118” and sheet “Appendix 15 - 2012-2016”. The appendix shows the table for the subperiod, annual tables and the returns of the 20 pairs in the trading periods. The data for the periodical table can be found in sheet “data for appendix 15 - 2012-2016” in the same excel file and data for the annual tables can be found in the sheets “Data for annual tables”. Data regarding trading statistics can be found in the sheets “Additional data to tables” in the same excel file.

Appendix 16 - 2016-2020 descriptive statistics

See excel file “Appendix 12-18” and sheet “Appendix 16 - 2016-2020”. The appendix shows the table for the subperiod, annual tables and the returns of the 20 pairs in the trading periods. The data for the periodical table can be found in sheet “data for appendix 16 - 2016-2020” in the same excel file and data for the annual tables can be found in the sheets “Data for annual tables”. Data regarding trading statistics can be found in the sheets “Additional data to tables” in the same excel file.

Appendix 17 - Q12020 descriptive statistics

See excel file “Appendix 12-18” and sheet “Appendix 13 - Q12020”. The appendix shows the table for the subperiod and the returns of the 20 pairs in the trading periods. The data for the periodical table can be found in sheet “data for appendix 17 - Q12020” in the same excel file. Data regarding trading statistics can be found in the sheets “Additional data to tables” in the same excel file.

Appendix 18 - All returns

See excel file “Appendix 12-18” and sheet “Appendix 18 - all returns” for a full overview of all returns in the sample period.

Appendix 19 - Z-score distribution

See excel file “Appendix 19 - z-score distribution”. The excel file contains the calculations for the z-score distribution.

Appendix 20 - Attributes of profitability for distance method

See excel file “Appendix 20-21”

Appendix 21 - Attributes of profitability for cointegration method

See excel file “Appendix 20-21”

Appendix 22 - Factor models

See excel file “Appendix 22 - Factor models”. The excel file showcases the various factor models conducted in this paper.