Lappeenranta-Lahti University of Technology LUT

School of Business and Management

Graduate Thesis

Master's Programme in Strategic Finance and Business Analytics

1st Examiner: Mikael Collan

2nd Examiner: Jyrki Savolainen

Pairs trading: An application of pairs selection and outranking in Norwegian stock market

2.10.2020

Anssi Virtanen

Abstract

Author:	Anssi Virtanen
Title:	Pairs Trading: An application of pairs selection and outranking in Norwegian stock market
Faculty:	LUT School of Business and Management
Master's Programme:	Strategic Finance and Business Analytics
Year:	2020
Examiner:	Professor Mikael Collan

Keywords:	Pairs trading, combined forecasts, Elman neural network, TOPSIS

Statistical arbitrage is a known research topic with a wide body of research with various methodologies to apply. Pairs trading is a part of statistical arbitrage research, where the purpose is to identify co-moving assets of which prices have spread and are expected to co-move back to their equilibrium state with statistical models. In Pairs trading the co-moving assets are bought according to the expected co-movement of the assets back to the equilibrium state so that the one expected to decrease in price is shorted and the one expected to increase is bought or longed, and these positions closed after the co-movement to equilibrium.

In this thesis we apply a combined forecasts approach to pairs trading, where we first forecast spreads with (77*76)/2=2926 possible stock pairs of the market with Elman Neural Networks. We continue to employ a TOPSIS ranking to rank the stocks according to the spread forecasts for each possible pair, and trade every day according to the rankings with portfolio sizes ranging from one pair to the maximum amount of individual pairs possible with the amount of stocks. The results of this method are promising, but solutions should be developed for the opening and closing signals of the trade so that the performance of the portfolios would rely on trades that could have actually been done, instead of over-optimistic assumptions such as being able to trade on daily data using the last days price for both forecasting and opening the position, instead of using data for a signal first and trading after the signal.

Tiivistelmä

Tekijä: Anssi Virtanen

Aihe: Pairs Trading: An application of pairs selection and outranking in Norwegian stock market

Yksikkö: LUT School of Business and Management

Koulutusohjelma: Strategic Finance and Business Analytics

Vuosi: 2020

Tarkastaja: Professori Mikael Collan

Avainsanat: Pairs trading, combined forecasts, Elman neural network, TOPSIS

Tilastollinen arbitraasi on tunnettu tutkimusaihe, josta löytyy laajasti kirjallisuutta ja sovellettavia metodologioita. Pareittain kaupankäynti (Pairs Trading) on osa tilastollisen arbitraasin tutkimusta, jossa tarkoitus on tunnistaa yhteisliikkuvia varallisuuskohteita, joiden arvostukset hajaantuvat tasapainotilasta ja joiden odotetaan yhteisliikkuvan takaisin tasapainotilaan tilastollisten mallien perusteella. Arvopaperi, jonka hinnan odotetaan nousevan ostetaan ja arvopaperi, jonka hinnan odotetaan laskevan lyhyeksimyydään. Positiot suljetaan, kun arvopaperiparin hintojen ajatellaan palautuneen tasapainotilaan.

Tässä pro gradussa sovellamme yhdisteltyjen ennusteiden menetelmää pareittain kaupankäymiseen siten, että ensin ennustamme tuottohajontoja kaikille (77*76)/2=2926 mahdolliselle osakepareille markkinalla Elman neuroverkoilla. Yhdistämme pariennusteet TOPSIS-menetelmällä yhdeksi sijoittelujärjestykseksi kunkin osakkeen odotettujen parihajontojen mukaan suurimmista positiivisista suurimpiin negatiivisiin hajontoihin. Portfolioita mudostetaan yhdestä parista (ääripäät) niin moneen pariin kuin arvopapereita on siten, että kaupat tapahtuvat päivittäin kullekin portfoliolle sijoitusjärjestyksen mukaisesti. Tulokset ovat lupaavia, mutta menetelmää pitäisi kehittää siten, että se käyttäisi signaaleja joiden mukaan kaupat tehtäisiin. Tällaisenaan menetelmä sisältää ylioptimistisia odotuksia kaupankäyntien toteutushinnoista, erityisesti että edellisen päivän hintoja voisi käyttää samanaikaisesti seuraavan päivän hintojen ennustamiseen ja positioiden avaamiseen.

Acknowledgements

I would like to thank professor Mikael Collan for suggesting Pairs Trading as the topic for the thesis and guiding me to get started with it and for providing good suggestions and thorough examination of this paper. Special thanks to Andreas Mikkelsen for providing data and answering my questions regarding the data in a thorough and efficient manner. I would also like to thank postdoctoral researcher Christoph Lohrmann for answering my questions about programming regarding this paper, I was not experienced with programming neural networks before doing it for this thesis. I also appreciate the flexibility and freedom I had while writing the thesis.

I am also grateful to my family and friends who have supported me throughout my studies and when making this thesis. Studying has provided me academic growth. The people around me during this time have helped me grow personally. Thanks to my family, friends, staff of LUT and everybody I met during my studies! You made this journey possible for me in the first place, you made it enjoyable even though it was sometimes demanding, and most of all you helped me grow. Thanks to all of you!

Veikkola, 2.10.2020

Anssi Virtanen

TABLE OF CONTENTS

List of figures

List of tables

List of appendices

List of abbreviations

ANN	Artificial neural network
ECM	Error correction model
ENN	Elman neural network
FFN	Feedforward network

GARCH	Generalized autoregressive conditional heteroskedasticity
HF data	High-frequency data
HFPT	High-frequency pairs trading
MCDM	Multi-criteria decision making
ML	Machine learning
OSE	Oslo Stock Exchange
PCA	Principal components analysis
PT	Pairs trading
RNN	Recurrent neural network
RSNNS	Stuttgart Neural Network Simulator package for R
RSS	Residual sum of squares
SSD	Euclidean squared distances
SSE	Sum of squared estimate of errors
TOPSIS	Technique for order of preference by similarity to ideal solution

1. Introduction

This thesis examines pairs trading (PT) and the various approaches in literature to PT. An experimental methodology is applied to Norwegian Stock data which can be typed as a combined forecasts approach, but also as an unsupervised machine learning (ML) model developed for the sole purpose of PT. Combining forecasts means that multiple forecasts are combined into a single forecast, unsupervised machine learning refers to statistical computing algorithms that provide results without testing for accuracy before forecasting. This thesis reviews relevant literature in Pairs Trading which could be of interest to practitioners and researchers of various fields such as finance, econometrics and data science. Methods in data science keep evolving and becoming more accessible as more and more researchers and practitioners are using, testing and developing new and existing methods. Machine learning, artificial neural networks (ANN) and statistical methods in general are constantly applied to new kinds of problems in various fields of study and business. Econometrics studies a wide variety of problems in economics and finance with all kinds of approaches to the data available, and in todays world the data is more available than ever. This study is about a niche in finance and econometrics, pairs trading, which is a relatively new concept as itself. In econometrics co-integration and co-movement of equities and indices have been researched and such relations of various econometric time series have been identified with statistical methods, but research in PT as such is still slim. It seems that traders and practitioners have first adopted PT strategies and peer-reviewed academic research is following behind. The contents of this paper are essentially introduction to pairs trading, continuing with a literary review of PT and finally examining and applying the chosen methodological approach to PT.

1.1. Background: What is pairs trading?

Pairs Trading is a form of statistical arbitrage, in other words trading to exploit anomalies in capital markets. These anomalies could be due to investor over- or underreactions, inefficacies in the markets, or just due to imperfect information. This paper does not try to explain the reasons behind these anomalies or quantify their existence, but rather review literature and methodologies in PT and apply a certain empirical methodology and compare it to existing research. In a sense this paper is continuing Mikkelsen’s (2018) and Mikkelsen & Kjaerland’s (2018) research on PT in

Oslo Stock Exchange (OSE), using the same data and applying ML. In this paper we apply a Combined Forecasts Approach, to be precise the same methodology that Huck (2009) used.

The idea behind pairs trading is based on finding a pair of assets, typically stocks, the prices of which co-move and are assumed to have a mutual equilibrium that the prices return to, in other words mean reversion occurs. When the prices diverge from the equilibrium the strategy is to buy the relatively underpriced security and sell the relatively overpriced security. In most cases the overpriced security is shorted. When the prices revert to the equilibrium the trade positions are closed and a profit is made. Typically companies from within a same industry are expected to co-move. A special case of this is when a company has more than one class of stocks and the underlying reality is by default the same. Portfolios of assets vs. stocks can also be used. Trading is typically done in the very short term, real-world applications would be automated algorithmic trading.

Pairs trading is generally done in two parts, Pairs formation and Trading period. First find two converging assets (Pairs formation). When they diverge open a short and long position on both assets, and when they return to convergence close the positions (Trading period). See figure 1 below. Methods and practices are various in both parts and the strategy can be implemented to various markets and assets. Use cases for PT could be commodities, stocks competing in the same industry, or stocks in the same market as in this paper. PT strategies can be applied to any securities, as long as there is a possibility to short sell.

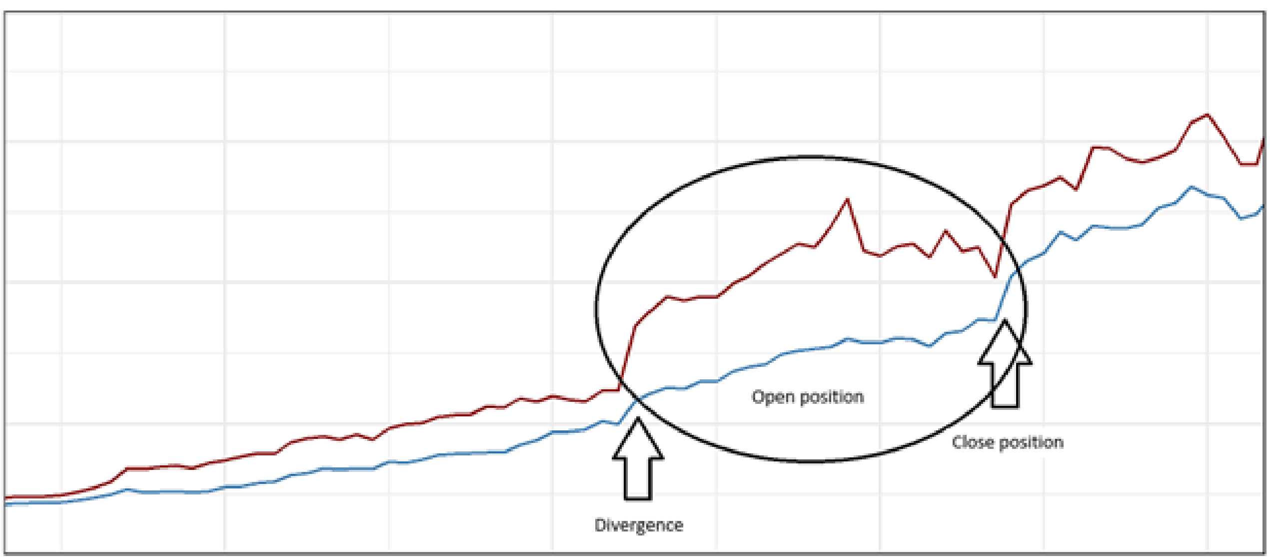

Figure 1. Pairs trading simply defined: When divergence occurs open positions and when prices revert to mean close positions and make a profit.

One of the first notions of PT in the literature is about a group led by Nunzio Tartaglia in Wall Street in mid-1980s which consisted of physicists, mathematicians and computer scientists in Morgan Stanley. They had great success in 1987 when the group reportedly made 50m$ profit for the firm. The group was disbanded at 1989 after a couple of years of bad performance, but Pairs Trading strategies are still used by some traders and hedge funds. The profits of quant-strategies like Pairs Trading are generally declining, with academics hypothesizing different explanations. The strategies becoming better known and technological advancements making them easier, cheaper and thus more viable to use for even more traders could be one explanation for the strategies to be less profitable. A lot has yet to be researched, documented and proven academically. Pairs trading seems to be more popular among traders and institutions than academics, so not too much peer-reviewed information can be found of the topic – after all, why would a trader share their information to others if keeping it to yourself will give you an advantage in earning more money or making a better career at private investing? (Gatev et al., 2006).

1.2. Why research pairs trading? Why this approach?

Traditional pairs trading strategies have been profitable, with declining profits. (Gatev et al., 2006; Stübinger & Bredthauer, 2017; Krauss, 2017). It is unclear how well strategies work and in which conditions. The biggest markets with traditional methodologies seem to have dropped in profitability, but new methodologies are being tested. Also the traditional methodologies could be developed further and different conditions such as smaller markets or volatile market conditions or bear markets could still provide opportunities for these strategies. The research goes on and mixed results for different use cases are reported across the board. Few things can be said for certain, but what we can conclude at this point is that PT does have potential for research and profitable strategies seem to exist.

Huck (2009) describes that the methodology used in this paper could be applied to select some pairs among a large number of securities, and also that the methodology could be modified in various ways. The combined forecasts method used in this paper is not common in academic literature and therefore it is further tested in this paper to see if it works similarly as with Huck (2009) or if it doesn't. You could say that the method has similarities to an unsupervised ML model, maybe it could even be used in other time series ranking applications than just financial time series.

Krauss (2017) made a thorough literature review in the context of Pairs Trading and categorizes 5 approaches. The combined forecast methods, such as the empirical methodology used in this paper, are in the ”other approaches” category. In this paper we test similar methodology to Huck (2009 & 2010) with Norwegian stock data from 06/2006-06/2016. The original goal of this work is to continue Andreas Mikkelsen’s work on Pairs Trading, using the same Norwegian stock data as he was using. Instead of trying to further develop and test the methods used by Mikkelsen (2018), we apply Huck’s (2009 & 2010) methodology on Mikkelsen’s daily dataset.

1.3. Positioning of this research

This is a financial study, which is a part of economics. In academic terms PT is examined with econometric tools, which can be used for both financial and economic research. In this research we examine PT with an econometric viewpoint, which means we use the statistical methods typically used for financial or economic data. This also means that this research has elements of data science, since it is emerging as the used term for the science of employing different types of methodology to analyse phenomena with data. See a brief illustration of the positioning of this paper in Figure 2 below.

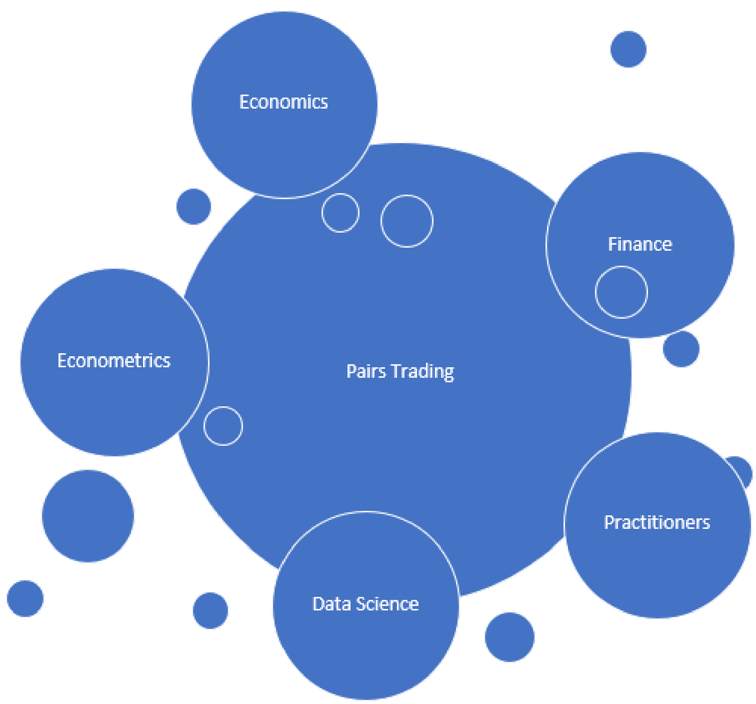

Figure 2. Positioning

Statistical analysis models are never perfect when applied to data – some can be spurious. There is a possibility that the type of forecasting done in this paper is spurious and there is also a possibility that the model could be developed further. This paper reviews PT and empirically tests an uncommon experimental approach. This paper is positioned to scientific literature in various fields and practitioners. Sometimes researchers cut corners in a data science viewpoint, certain models can be proven or disproven to bring information or useful results. If the methodology used in this paper can be found valid in econometrics, then the implications and points of interest of this paper is different to all fields of study.

Econometrics consists of everything used to measure economic phenomena – also financial. Figure 3 below very briefly illustrates some core concepts of econometrics in finance. All of the concepts below could be applied to PT. Volatile markets are seen as the best opportunities for PT. PT usually assumes a long run relationship, co-movement or cointegration. In this paper we apply forecasts to our ranking system, and surely a simulation approach could be taken to certain kinds of problems that could be formulated into PT problems.

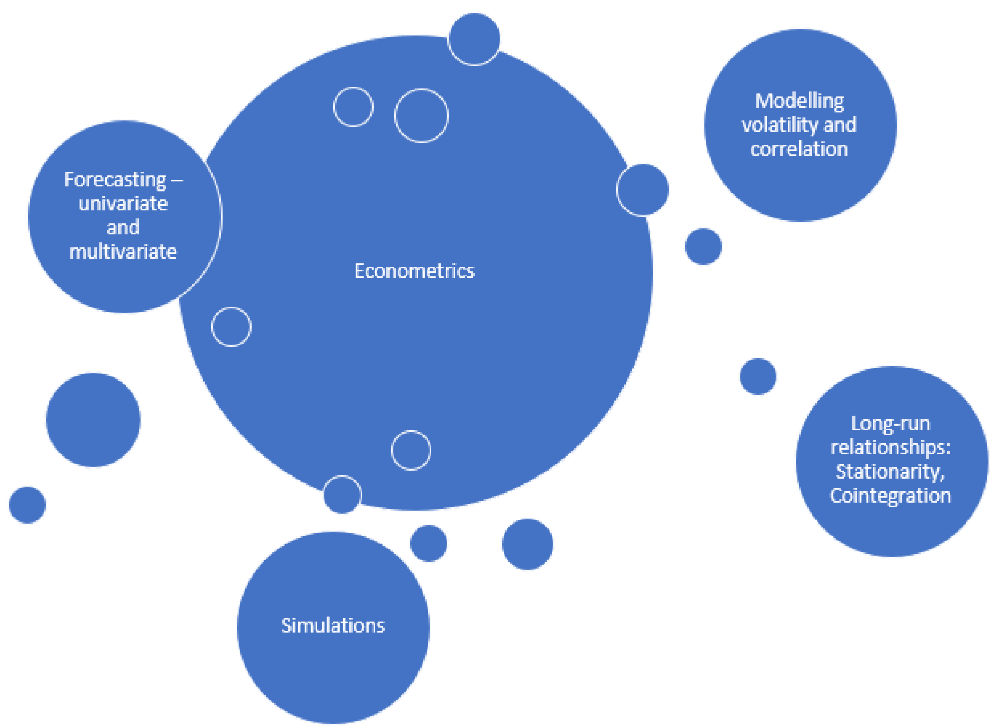

Figure 3. Some core concepts of econometrics (Brooks, 2008)

1.4. Focus of the paper

In this paper we focus on PT and applying Huck's (2009) methodology to Norwegian stock data. The methodology is uncommon and experimental – the applicability and robustness of Huck's (2009) base model hasn't been proven, and its shortcomings or possibilities for applicability might have not been properly discovered. There is a wide body of literature in econometrics, about co-moving assets and their co-integrative relationships, forecasting econometric time series with different kinds of methods and multi-criteria decision making. Practitioners and academics have similar goals and it is hard to find specific information about practices used in real life PT. In short, the focus of this paper is to review the existing PT literature and examine an empirical approach to PT.

Next the research questions of this thesis are examined, the main research question is about the methodology applied in this paper, how does it apply to the data, how the results of this research compare with Huck's (2009 & 2010)?

Sub-question: What is known about pairs trading in academic literature?

We need to make a thorough review with a focus on PT. There is a wide body of econometric research, in this research we will focus on PT and the most relevant findings in this relatively fresh research topic. We will also briefly review research most related to the methodology we are applying in this paper.

Sub-question: How does Huck's (2009 & 2010) combined forecasts method compare with other approaches in pairs trading?

We will review literature for PT approaches and examine if they are comparable and try to make sense of the existing body of literature in order to examine approaches. This thesis focuses on trying to explain how and why results with the approach chosen in this thesis could be compared and if a fair comparison is even possible.

Main Research Question: How does Huck's (2009 & 2010) combined forecasts methodology work with OSE daily data 6/2006-6/2016?

It is interesting to see if the combined forecasts approach Huck (2009 & 2010) used in his research works with this data too since he had such promising results. Huck (2009 & 2010) used weekly data and we use daily which might bring new insights. See figure 4 below for a concept flowchart and focus in a glimpse.

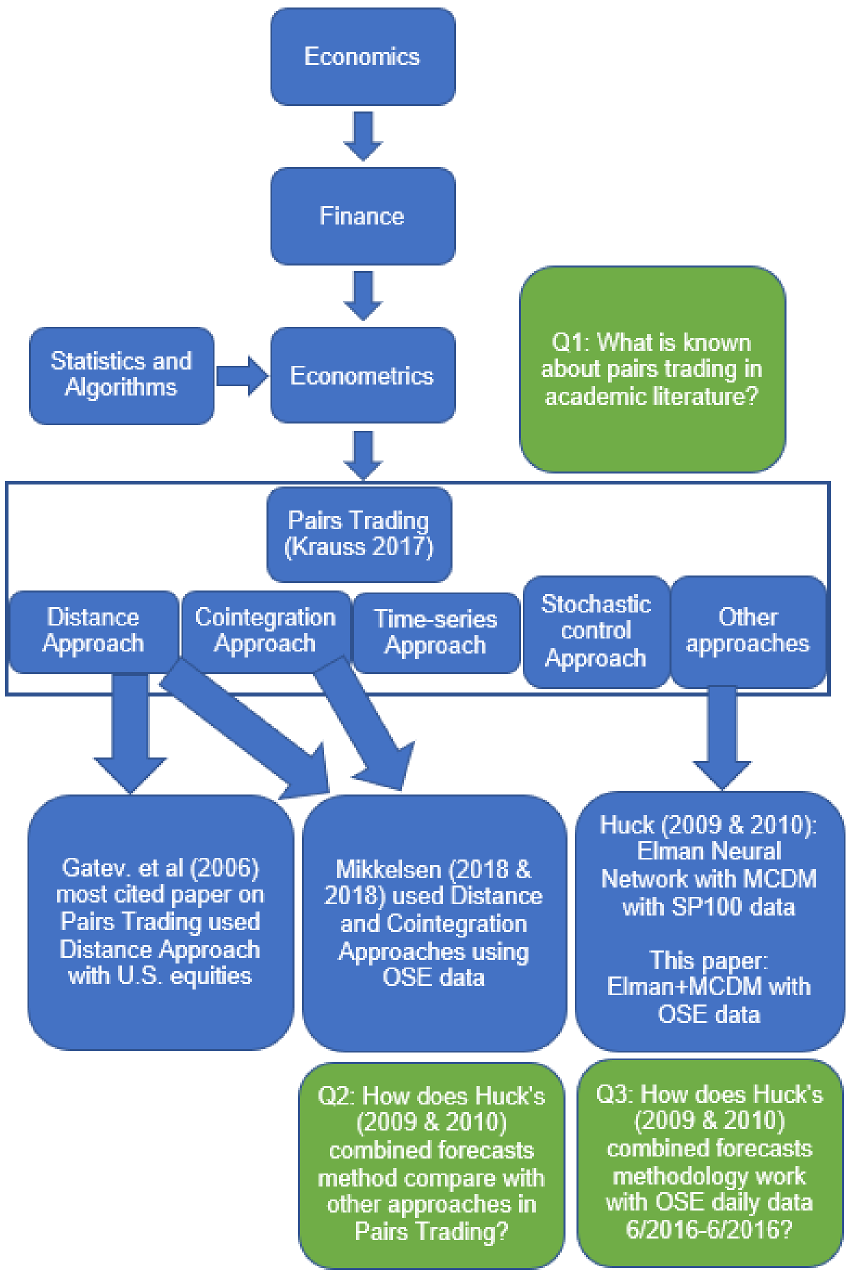

Figure 4. Concept flowchart and focus

1.5. Structure of the thesis

We have briefly defined and explained what Pairs Trading is and the reasons why this study is done. In the following part 2. Literature and methods in the literature, we will review Pairs Trading approaches defined by Krauss (2017) and reviewing Pairs Trading literature. The literature used in this study was found in LUT Finna which includes multiple databases of peer-reviewed papers. Most of the cited papers in this study were found by using PT and methodology related search words and going through Krauss (2017), Mikkelsen (2018) and Mikkelsen & Kjaerland (2018) since they are highly relevant papers for understanding the base concepts and recent findings in PT.

After going through literature and methods in the literature we examine the methodology used in this study, followed by the empirical case of OSE data. PT literature is examined before the methodology, because Neural Networks and combined forecasts are a niche in PT research with only a thin body of research. Examining PT literature first makes it easier for readers that are new in PT to understand the methodology used in this paper. We end the paper with chapter 5. Conclusions, which also includes answering to our research questions, comparison to existing research, examining lessons learned from this study, examining future research directions and ending with critique.

2. Literature and methods in the literature

2.1. Pairs trading methodologies in literature

In Pairs Trading the formation period and trading period can be done in various ways. Krauss (2017) has done a thorough review of Pairs Trading literature, which is the best source of information at the time. Another broad literary review that focuses on high frequency pairs trading (HFPT) is done by Stübinger & Bredthauer (2017). Krauss (2017) categorizes 5 approaches and their typical practices in pairs formation and trading period. The methodology used in this paper is a combined forecasts approach, which Krauss (2017) categorizes in other approaches. See Krauss's categorized PT Approaches in table 1 below.

Table 1. Krauss's (2017) categorized pairs trading approaches

	Pairs formation	**Trading period**
Distance approach	Distance metrics	Nonparametric threshold rules
Cointegration approach	Cointegration tests	Most based on Gatev et al. (2006) threshold rules
Time-series approach	Ignored, established by prior analyses	Optimized trading signals by different methods of time-series analysis
Stochastic control approach	Ignored, established by prior analyses	Stochastic control theory is used to define optimal pairs trading portfolio holdings
Other approaches	ML, combined forecasts approach, copula approach, PCA	ML, combined forecasts approach, copula approach, PCA

Distance and cointegration approaches are the most common approaches in PT literature, you could say that they are the standard approaches to PT. In pairs formation the distance approach measures the distances of time series and the least distant are chosen as pairs. Cointegration uses cointegration tests to find equities that comove and form pairs according to cointegration tests of pairs. In the trading period both of these approaches typically use similar rules basing on deviations, which provide similar results.

Good research follows a certain a priori process and in typically in PT statistical tests are used to find comoving equity pairs. Data snooping could be a problem in any statistical research, in PT some data snooping is expected, as the threshold values for choosing pairs are more like rules of thumb rather than established standard thresholds. Each approach has its own questions to address both in pairs formation and the trading period. The biggest differences between approaches occur in what kind of statistical models are used to form the pairs. There are many variables to tweak with each methodology, for example with cointegration approach what kind of p-values are acceptable with cointegration tests or even what kind of cointegration tests to use and should an error correction model be formulated. In the trading period threshold rules basing on deviations are typically applied, but different rulesets could be formulated. All of the five approaches typed by Krauss (2017) could be further researched and developed, both in the pairs formation and trading periods.

One of the pros of the combined forecasts approach used in this thesis is that it doesn't require statistical testing because it combines vast amounts of neural networks iterating forecasts that are combined similar to unsupervised ML models that can be applied to many kinds of data and their accuracy measured only by results instead of testing for co-movement or tweaking forecasting models to fit the data in the best way. Combining forecasts in the way it is done in this thesis takes a lot of computing power. Artificial neural networks (ANNs) are considered a universal approximator in a wide variety of nonlinear patterns which makes them very appealing in combining forecasts in PT context. The way ANNs are used in this thesis is efficient in providing adequate forecasts with all possible pairs that could be formed with the data without tweaking every single model separately. In this thesis we make spread forecasts for all possible pairs simultaneously. If forecasts were done for each equity separately, this would be typed as a time series approach where each model should be improved to find the best possible statistical accuracy or least errors.

In this paper the best effort is applied in following Huck's (2009) protocol and methodology. Similarities and differences to Huck (2009) are compared and documented. The literary review and research cited in this paper are mostly the ones cited by Krauss (2017), since his paper is a thorough review of the basic methodologies in Pairs Trading. Thanks to Krauss' (2017) thorough research on the brief history of Pairs Trading research, we can see that a multitude of methodologies have been applied, with differing results, advancements and shortcomings, finding indications to different kinds of datas and use cases. In this paper we use the same data as Mikkelsen & Kjaerland (2018), so we have some comparability to the results of Mikkelsen &

Kjaerland (2018) and Huck (2009). In the next chapters we will briefly examine the Pairs Trading approaches defined by Krauss (2017).

2.1.1. Distance approach

Gatev et al. (2006) is the most cited paper using the distance approach. His results are often used as a benchmark and trading rules in PT litearature are typically similar to the ones he used. A more recent trend has been using high frequency data, like Mikkelsen & Kjaerland (2018), Stübinger & Bredthauer (2017). In this thesis we are using the same market and timeframe as Mikkelsen & Kjaerland (2018), but only daily data instead of HF trading applications. Distance approach is explained below, with some examples from the most relevant research to this paper.

Distance approach is typically done in the following way. First returns of equities are normalized. Then sums of euclidean squared distances (SSD) are calculated the and equity pairs with the smallest distances from each other are chosen for the trading period. Mikkelsen & Kjaerland (2018) used HF data and used top 1% pairs, which translates to about 20 pairs in every Trading Period, similar to Gatev et al. (2006) with top 20 pairs. Gatev et al. (2006) and Mikkelsen & Kjaerland (2018) used 2 historical standard deviations as the threshold to open positions for the corresponding Pairs Trades for the top 20 pairs. Trades are closed upon mean-reversion, at the end of the trading period, or upon delisting.

$$SSD = \frac{1}{n}\sum_{t=1}^{n}(P_t^1 - P_t^2)^2$$

Formula 1. Euclidean squared distances SSD

This method has resulted to statistically significant risk-adjusted excess returns, but declining profits with time, at least with S&P 500, which probably holds true with more liquid markets – most likely due to traders and hedge funds adopting similar strategies. (Gatev et al., 2006; Do & Faff, 2010; Stübinger & Bredthauer, 2017). Keep in mind that this is the methodology used in the most cited paper in the field of Pairs Trading, so it can be considered as the default methodology or standard benchmark methodology in Pairs Trading. Huck (2013) and Mikkelsen & Kjaerland (2018) have experimented with different time frames for the formation and trading period.

Stübinger & Bredthauer (2017) present a thorough literature review in the context of HFPT, and test different kinds of thresholds for the trades with various approaches. Krauss (2017) summarizes that distance approach is a simple and transparent strategy that studies have found to be profitable with a wide variety of markets, securities and time frames.

2.1.2. Cointegration approach

Vidyamurthy (2004) is the most cited author for PT with cointegration approach. The basic univariate cointegration methodology has three steps. First pairs are selected either with statistical or fundamental similarity measures. Second, tradability can be assessed with an adapted version of Engle-Granger cointegration test. Third, optimal entry and exit thresholds are chosen with nonparametric methods, such as deviations. (Vidyamurthy, 2004; Krauss, 2017).

$$\varepsilon_{ij,t} = P_{i,t} + \gamma P_{j,t}$$

Formula 2. Spread $\varepsilon_{ij,t}$ between two stocks

Pi and Pj denote I(1)-nonstationary price processes of stocks i and j with intercept neglected. Cointegration coefficient γ is a nonzero real number, which makes the spread $\varepsilon_{ij,t}$ as linear combination of Pi and Pj is I(0)-stationary and therefore mean-reverting. Then the Pi and Pj price processes are cointegrated or an error correction form such as Error Correction Model framework (ECM) should be assumed. (Krauss, 2017)(Rad et al., 2016)

Rad et al. (2016) used similar SSD threshold rules with cointegration approach as Gatev et al. (2006) in their distance approach, which limited the number of pairs chosen, and had very similar profits to Gatev et al. (2006). Puspaningrum et al. (2010) fit an AR(1)-process to cointegration errors and obtain estimations that they use to estimate the number of trades and decide thresholds for their cointegration model. These kinds of advancements in methodology will result in further differing results from the default distance approach and should be further tested empirically.

Distance and cointegration approaches are the two most commonly applied in literature. Both have promising but mixed results, so it is unclear which performs better. In the trading period similar trading rules are typically applied, so results are in some cases very similar between these two methodologies. Mikkelsen & Kjaerlands (2018) results with high frequency data comparing the two approaches favor a shorter time strategy, in other words shorter formation and trading periods. Huck (2013) also finds that different time strategies provide mixed results with daily data. Mixed results are provided over the board with all methodologies in PT literature, including the ones mentioned in following chapters. Most research hypothesizes and confirms bear markets to be good for PT, but Mikkelsen & Kjaerland (2018) and Bowen et al. (2010) have contradicting results with the cointegration and distance approaches. (Gatev et al., 2006; Kim, 2011; Do & Faff, 2010; Rad et al., 2016).

2.1.3. Time-series approach

Krauss (2017) names Elliott et al. (2015) and Bertram (2010) as the most relevant authors in time series approach. Elliott et al. (2015) models spread in state space. They examine PT as spread measured in Gaussian noise, as a mean-reverting Gaussian Markov Chain. Gaussian process is a stochastic process, in which a finite linear combination of random variables is normally distributed. These definitions make it an Ornstein-Uhlenbeck process. Elliott et al. (2015) apply an expectation-maximization algorithm. If the spread of the data follows the Gaussian process, this method can explicitly answer to questions like expected holding times and expected returns. Do et al. (2006) note that Elliott et al. (2015) model is only applicable to securities in return parity. Bertram (2010) also notes that this approach applies Gaussian OU-process to non-Gaussian financial data. Despite the critique the approach does give good insights on spread dynamics. Therefore this application should be used mostly when the data used is as Gaussian and in as much return parity as possible, like Bertram (2010) on a pair of dual-listed securities and Cummins & Bucca (2012) on 861 energy futures. Other time-series applications have also been made, but the OU-approach is the most relevant in literature.

2.1.4. Stochastic control approach

Krauss (2017) also brings up stochastic control approach as one of the main approaches in PT. In this approach asset pricing dynamics are modeled using either Ornstein-Uhlenbeck process or error correction models (ECM).

Liu & Timmermann (2013) apply ECM and use a Hamilton-Jacobi-Bellman equation to find closed-form solutions for the valuation and trader policy. The equation includes a loss function and a nonlinear partial differential equation form value function, which enables it to be used as a maximizer or minimizer and to obtain optimal control of data applied. The strategy identifies boundaries for a stabilization region in which PTs are made and outperforms the simple threshold rule strategies when spreads are highly mean-reverting, but high amount of trades may hinder the superiority of this strategy. (Liu & Timmermann, 2013)

The ECM approach has a cointegration framework in which the market index is seen to follow a geometric random walk, so in the formulas modelling the market Brownian motion is incorporated, which in statistics is the random walk measured from nature translated into numbers. The equation is econometrically reasoned into defining that the used assets' log prices are cointegrated and the market is stationary. Lei & Xu (2015) and Liu & Timmermann (2013) find promising results with the Stochastic Control ECM Approach, but large scale empirical applications are yet to be made.

2.1.5. Other approaches

Krauss (2017) categorizes the three main types of research in other approaches as ML & combined forecasts approach, copula approach and lastly principal components analysis approach.

In ML & combined forecasts approach Krauss (2017) identifies Huck (2009) and Huck (2010) to be the main research. In this paper we use the same approach, with nearly identical methodology as Huck (2009), which is based on forecasting (with Elman neural networks), outranking (with MCDM) and trading. The approach used in this paper will be thoroughly explained in chapter 3 and compared to Huck (2009) in the following chapters. Huck (2010) expands the methodology into multi-step forecasts and testing different kinds of trading rules for the portfolios.

“Copulas are a relatively new field of statistical analysis that offer a new way to study equity spreads for those not intimidated by this exciting frontier” (Ferreira 2008). Copulas are multivariate probability distributions in which the probability distribution of each variable is uniform on the interval [0,1]. They can be used to model marginals and copulae separately and there are many parametric copula families available, some have parameters to control the strength of dependence. In PT copula approach Krauss (2017) identifies two substreams: Return-based versus level-based copula methods.

Return-based copula method means that you input returns into copula. Ferreira (2008), Liew and Wu (2013) and Krauss & Stübinger (2017) have published research with the approach. In this method pairs are built using correlation or cointegration criteria. Then researchers choose which distributions and copula they apply, for example Ferreira (2008) uses one copula, Liew & Wu (2013) use five different copulas and evaluate and choose the best fitting with information criteria. Krauss & Stübinger (2017) apply return-based copula PT methodology to daily S&P 100 data from 1990 to 2014 and have great results with Sharpe ratios above 1,5. Sharpe ratio evaluates portfolio returns and the formula includes comparison to risk free rate and also takes into account the volatility of the portfolio. Values of over 1 are considered winning the market.

Level-based copula method gets its name from a “mispricing index” which is constructed and to which securities are compared to and PTs are made accordingly. Krauss (2017) highlights this approach appealing, because it incorporates mispricings over multiple periods. However he also notes that empirical research made with the approach has had its flaws – Rad et al. (2016) used SSD for pairs selection, instead of copula and their research provides just alternative Pairs Trading signals in the trading phase comparing to more common methods. Xie et al. (2014) developed the framework that authors have been using and derived in under which conditions the cumulative mispricing indices are mean-reverting, but Rad et al. (2016) do not test if theirs are. Also Krauss (2017) notes that authors should also differentiate mean-reversion and momentum pairs, like in Krauss and Stübinger (2017).

Principal components analysis (PCA) approach is the last approach Krauss (2017) presents under other approaches. PCA intakes a collection of points in at least two dimensional space in which a line can be defined that minimizes the average squared distance from a point to the line, which is called a best-fitting line. The next best-fitting line can be chosen from directions perpendicular to the first, and repeating this yields an orthogonal basis where different individual dimensions of the data are uncorrelated. These lines, or vectors, are called principal components. PCA can be applied by singular value decomposition of a design matrix or calculating the covariance matrix of the original data or by eigenvalue decomposition applied to the covariance matrix. Avellanada and Lee (2010) apply the PCA method to large US companies and achieved a Sharpe ratio of 1,44 after transaction costs. Krauss (2017) criticizes Avellanada and Lee’s (2010) analytical solutions and implies that the methodology could be further developed.

3. Methodology: Combined forecasts - neural network forecasts and outranking with multi-criteria decision making model

In this section we examine and review the methodology used in this thesis. While the traditional PT approaches defined by Krauss (2017) could be tested and developed further, Huck's methodology appealed because to my knowledge it hasn't been tested by other researchers than Huck (2009 & 2010), and he had good results. The use of combined forecasts is not mainstream in econometrics and definitely not in PT, as a researcher it is appealing to try new methodologies with only a little research done on them. Unsupervised machine learning methods are continuously developed, new algorithms keep popping up with their various use cases. While new algorithms are not always better, and they can be harder to justify, they still have their use cases and possible new findings to uncover. Various kinds of statistical models can be defined and applied to problems and to answer research questions. In the context of PT this is a new approach with very little existing research, which makes it appealing to apply and research.

This methodology categorizes into other approaches in Krauss's (2017) PT approaches. It is a combined forecasts approach, where in the Pairs Formation Period we use Elman Neural Networks to first obtain forecasts of each possible stock pair, so we obtain (n*n-1)/2 forecasts and we continue to make a combined forecast of these simultaneously obtained forecasts. We make our combined forecasts using a Multi Criteria Decision Making (MCDM) model. Huck (2009) was the first to use this methodology, and he used Electre III as MCDM in his research, he also tested forecasting and trading more than one day in a follow up paper. (Huck, 2010). In this paper we use the technique for order of preference by similarity to ideal solution (TOPSIS) as our MCDM. See Table 2 on next page that illustrates the methodology with comparisons and possibilities for future research in a glimpse.

n stocks		Techniques used		
n financial time series (returns)	n*(n-1)/2 possible pairs	**This paper**	**Huck (2009 & 2010)**	**Other possibilities**
Construction of n*(n-1)/2 bivariate information sets (possible pairs)		Daily data (OSE)	Daily data (S&P100)	Tick-by-tick data applications, HFPT
Selection of pairs that move together, based on statistical test for comovement	Return forecasts on all bivariate sets, calculate spreads	Elman	Elman	Any forecasting method
If prices diverge, take long-short position on pair	Decision matrix that ranks stocks based on of anticipated spreads	TOPSIS	Electre III	Any ranking method
Close position after convergence of prices	Long top ranked, short bottom ranked, close positions after one time step	Trade every timestep	Huck (2009) trade every timestep, Huck (2010) trades up to 4 timesteps	Signals for trades, holding periods, stop-loss etc.
Traditional PT Approaches	**This paper**	**This paper**	**Huck (2009 & 2010)**	**Possibilities for future research**

Table 2. Presentation of the methodology

3.1. Artificial neural networks

Artificial neural networks (ANN) are part of machine learning (ML) and artificial intelligence (AI). ML methods can be scarcely divided into supervised and unsupervised methods. When you first use data to train and then data to test, or as in our case, to forecast, the method is typically considered a supervised method. However, in chapter 3.4. methodology summarized below we establish that the way we first apply ANNs and then MCDM effectively makes this an unsupervised ML method. Machine Learning is effective because it doesn't require the user to define too many rules or thresholds – generally the algorithm itself will figure out the best way to comprehend the data. Also ML methods have less prerequisites for the data to be able to produce reliable results. ML algorithms are easy and fast to run and can produce robust results in a dataset that would be hard or unreliable to fit in other algorithms. (Chong & Zak, 2001)

ANNs were developed on the base idea of brain neuron operation. ANNs have synapses and neurons, which belong to a layer of typically same kind of neurons, in which through the data is processed step by step. In biology neurons communicate with junctions called synapses. Usually ANNs begin with input neurons, and end in output neurons. In simplified terms different types of ANNs have different kinds of neurons in between the input and output neurons. ANNs and the multitude of developed ANN-models have a versatile applicability to many kinds of problems and data. (Basheer & Hajmeer, 2000; Chong & Zak, 2001)

Countless applications of ANNs have been developed with various use cases and configurations. Perceptron is the simplest ANN, other models like feed forward, radial basis network, deep feed forward, recurrent neural network, long/short term memory etc., have also been defined and constructed. The differences between ANN-models can be different kinds of neurons or rules for the ANN to process the data and produce results. The neurons are in a formation of layers, in which the data goes from input to output through whatever neurons are in between the input and output – and sometimes the data is recirculated through the network, sometimes with weights. Sometimes the data doesn't flow "straight through" vertically, but also affects horizontally. Sometimes ANN-models have hidden layers which can work for example as a memory of sorts or for the ANN to learn or process the data in a different way. What is between the input and output, and the rules of data processing are what define the different types of ANN-models. (Hagan et al., 2014)

Brooks (2008) notes that in econometrics, ANNs have little theoretical motivation in finance and they are typically termed as a "black box", Huck (2009) uses the same metaphor. Still, ANNs have been widely used in finance for their applicability to a wide range of analytical problems, in econometrics typically time series and classification. Brooks (2008) criticizes the use of ANNs in econometrics for multiple reasons. The coefficient estimates don't have any real theoretical interpretation, no diagnostics or specification tests are available for estimated models. Brooks (2008) also notes, that in econometrics ANNs typically provide excellent fits in-sample, but typically poor out-of-sample forecast accuracy, meaning that the models tend to overfit. Despite ANNs shortcomings in out-of-sample forecasting, Basheer & Hajmeer (2000) point out that ANN's have remarkable information processing charasteristics, such as nonlinearity, robustness, fault and failure tolerance and an ability to handle imprecise and fuzzy information in the context of a biological system, but in a context of an econometric time-series model we can view these as qualities as good in predicting stock prices since the fitness of different econometric analysis models have their limitations too. Huck (2009) summarizes that "ANNs are considered to be a universal approximator in a wide variety of nonlinear patterns, including regime switches and other nonlinearites and they are good predictors. Nonlinearities are common features in financial data, ANNs are thus especially interesting. ANNs are considered to be a universal approximator."

The most common class of ANNs used in finance are feedforward network models. If a feedforward network model had no hidden layers, it would be just a standard linear regression model. Brooks (2008) points out, that ANN applications work best where financial theory hasn't defined the likely functional form of the relationship of variables. In this paper we use an Elman neural network (ENN) which is belongs to ANN class of recurrent neural networks. In chapter 3.2. we examine what ENN is and the reasoning why it is the chosen ANN in this thesis. You could use other than ANN methods for forecasting and choosing the pairs, such as the ones described before in this paper, but since we are using a combined forecasts approach with 77*(77-1)/2=2926 amount of pair data, the neural network approach with forecasts is justified because it is efficient in computing massive amounts of data at once. It takes a long time to run these forecasts on a computer with 77*(77-1)/2=2926 pairs and 5 years of training data and 5 years of trading data.

Every methodology has its own shortcomings and advantages – The ANN handles vast amounts of data fast and efficiently. Also, the application of ANN frees the methodology from having to perform tests of fitness and continuously tweaking analytical models to best fit each of the (77*76)/2=2926 datasets trained yearly. ANN is might not the best fitting econometrical model for the data we are using, but the vast amount of data and inclusion of a MCDM justifies the use of ANN, because it is expected to fit well enough for our needs and it is efficient to use since it requires very little statistical hypothesis testing. Also despite the critique, Huck (2009 & 2010) had promising results with Elman network and Quek et al. (2007) had good results predicting prices of various commodities with an RNN. Keep in mind the purpose of this paper: To compare the methodologies on same data as Mikkelsen & Kjaerland (2018) and the same methodology on different data, in other words compare to Huck (2009). (Brooks, 2008).

3.2. Introduction to Elman Neural Networks

Similar to a Feedforward Network (FFN), the ENN has a hidden layer. But in addition to that, the ENN also has a context layer, which in simplified terms re-enter lagged inputs through a function so that the model has a memory aspect of sorts. The ENNs operation can be briefly described as follows (see Figure 5 and Figure 6): First input units are entered into the model, the data continues to hidden units, and from there continue from context units back into hidden units. The context layer stores information, so it recirculates information to itself also. Then from hidden units the cycle finally ends into output units. All of the units in between input and output perform a function. The model does not only learn when it is trained, it also learns when it is predicting. (Huck, 2009; Zell, 1998)

Huck (2009) further explains that “The Elman network has no known linear model characterization and can be viewed as a nonlinear dynamic latent variable model in econometric terms. In the specific case of a recurrent Elman type network, it is characterized by a dynamic structure where the hidden layer output feeds back into the hidden layer with a time delay”

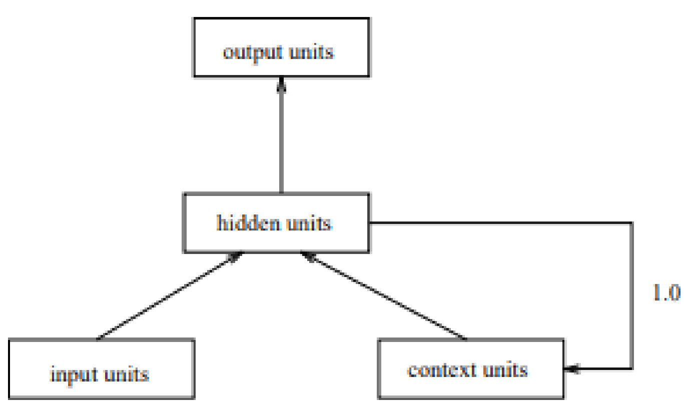

Figure 5. Elman network simplified (Zell, 1998)

Below we can see the formula notations for a single hidden layer Elman network with two outputs as represented by Huck (2009), where $L(n_{k,t})$ represents the tansigmoïd transfer function. i^* the number of input variables. k^* the number of neurons. $w_{k,i}$ the input weights. φ_k the weights between the hidden units evaluated in t and t-1. $w_{k,0}$, $\gamma_{1,0}$ and $\gamma_{2,0}$ are constant terms. The set of k^* neurons are combined in a linear way with the coefficients $\gamma_{1,k}$ and $\gamma_{2,k}$. The outputs $y_{1,t}$ and $y_{2,t}$ are in our application the anticipated returns of two stocks. The parameters of the network are estimated by minimizing the sum of the squared-error loss. Error backpropagation is the most widely used estimation method, which can also be thought as the learning rule of the network. See Figure 6 for a generalized visual representation of Elman network with different notations below.

$$n_{k,t} = w_{k,0} + \sum_{i=1}^{i^*} W_{k,i} Z_{i,t} + \sum_{k=1}^{k^*} \varphi_k n_{k,t-1} \, , \tag{1}$$

$$N_{k,t} = L(n_{k,t}) \, , \tag{2}$$

$$= \frac{2}{1+e^{-2n_{k,t}}} - 1 \, , \tag{3}$$

$$y_{1,t} = \gamma_{1,0} + \sum_{k=1}^{k^*} \gamma_{1,k} N_{k,t} \, , \tag{4}$$

$$y_{2,t} = \gamma_{2,0} + \sum_{k=1}^{k^*} \gamma_{2,k} N_{k,t} \, , \tag{5}$$

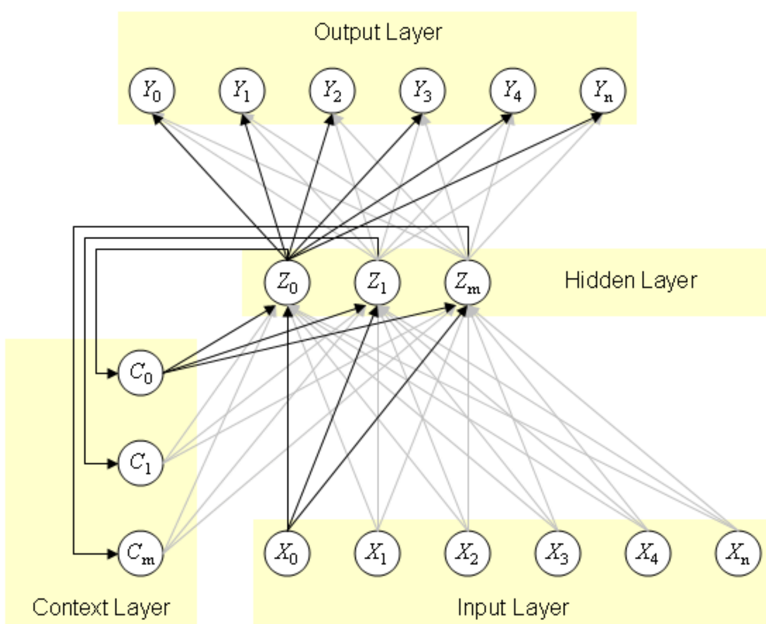

Figure 6. Elman network, general representation (McCullock, Mnemosyne Studio)

3.3. Introduction to Multi-Criteria Decision Making – ranking with TOPSIS and trading accordingly

Multi-criteria decision making (MCDM) models are models that incorporate multiple numerical criteria and compute the alternatives into a ranking order of preference. MCDM models can have different weights for different criteria. In this paper we use Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). In TOPSIS you can define positive and negative criteria that are used for computing a ranking of alternatives. TOPSIS ranks the alternatives based on the shortest geometric distance to the positive ideal solution and longest to negative ideal solution. For example if you were examining options of which car to buy, price and mileage could be negative criteria, since you want to get the cheapest and the one that uses least fuel or energy. A positive criteria could be how many seats the car has or how big trunk, since those are criteria a buyer typically wants to maximize. In our application we use TOPSIS to rank stocks (alternatives) using forecast spreads to another stock as criteria. Then our top candidates are the expected highest positive spreads and lowest are the expected highest negative spreads. We want to incorporate the highest expected positive spread stocks in the stocks we buy and the highest expected negative spreads as the stocks we short in our trading period.

Krauss (2017) criticizes Huck's choice of Electre III as the MCDM because it is a complex MCDM method that has been reduced to only rank alternatives by one dimension: Anticipated spreads of a each possible pair for a stock. Huck (2009) used equal weights for each criteria and didn't drop any alternatives. While Hucks (2009) application of Electre III might provide different results than TOPSIS, the fundamental idea of ranking is the same. Therefore in this paper we use TOPSIS because it is simple to use and easy to replicate for further research and doesn't make the model more complicated than it has to be. After all, we are using our MCDM to rank our stocks by expected spreads of pairs which are calculated from forecasts. Huck (2009) proposes many possible modifications or additions to the application of MCDM in the model, which gives possible ideas for future research for this Pairs Trading methodology.

Our goal is to Pairs Trade according to our TOPSIS rankings of anticipated spreads, so we will choose to buy the stocks with highest anticipated pairs spreads combinations and short the stocks with the most negative anticipated pairs spreads combinations. Huck (2009) also notes that this methodology for Pairs Trading chooses the trades in a single analysis, instead of choosing them one after one. See Table 3 on the next page for a general structure of the decision matrix.

		Criteria (Performances = Spreads between stock return forecasts)			
Alternatives (Stocks)		C_1	C_2	$\cdots$	C_n
	X_1	0	$\hat{x}^{1,t+1}_{\mid X_{1,t},X_{2,t}} - \hat{x}^{2,t+1}_{\mid X_{1,t},X_{2,t}}$	$\cdots$	$\hat{x}^{1,t+1}_{\mid X_{1,t},X_{n,t}} - \hat{x}^{n,t+1}_{\mid X_{1,t},X_{n,t}}$
	X_2	$\hat{x}^{2,t+1}_{\mid X_{1,t},X_{2,t}} - \hat{x}^{1,t+1}_{\mid X_{1,t},X_{2,t}}$	0	$\cdots$	$\hat{x}^{2,t+1}_{\mid X_{2,t},X_{n,t}} - \hat{x}^{n,t+1}_{\mid X_{2,t},X_{n,t}}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	X_n	$\hat{x}^{n,t+1}_{\mid X_{1,t},X_{n,t}} - \hat{x}^{1,t+1}_{\mid X_{1,t},X_{n,t}}$	$\hat{x}^{n,t+1}_{\mid X_{2,t},X_{n,t}} - \hat{x}^{2,t+1}_{\mid X_{2,t},X_{n,t}}$	$\cdots$	0

Table 3. General structure of the decision matrix used in the paper

$X_{j,t}$ represents the past returns of asset X_j until date t.

$\hat{x}^{j,t+1}_{\mid X_{j,t},X_{k,t}}$ is the one period ahead forecast of the return X_j conditionally to $X_{j,t}$ and $X_{k,t}$.

3.4. Methodology summarized

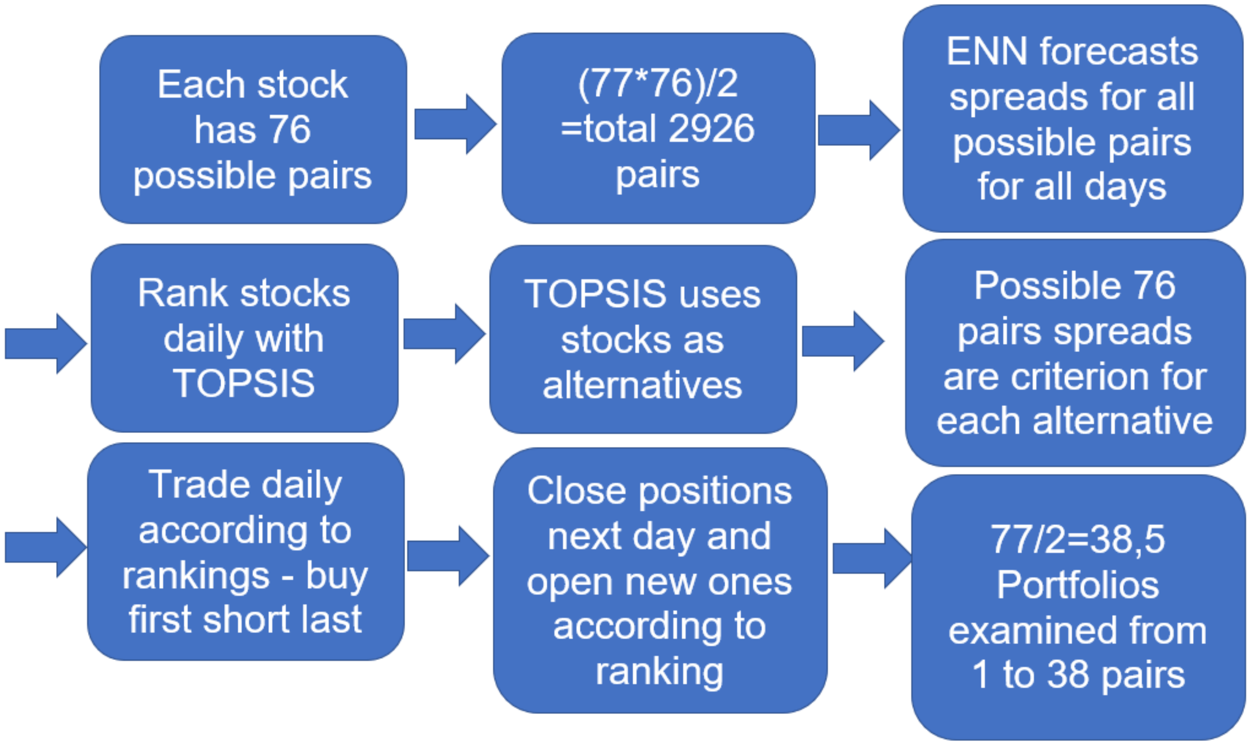

Figure 7. Methodology summarized

The methodology used in this paper is done as follows, also see figure 7 above. First stock return data is obtained and made it into a 77*(77-1)/2=2926 matrix of pairs. ENN is applied to these pairs, trained for 5 years and pairwise spreads of returns for each day for the next 5 years are predicted. The stocks are ranked using expected spreads with TOPSIS. Stocks are alternatives, and all possible pairings are viewed as criteria. PTs are done by opening long position(s) for the best expected alternative(s) and short the worst expected alternative(s). Portfolios are made according to rankings daily. The first portfolio consists of the best alternative of the day longed and worst alternative of the day shorted, the second portfolio is two best alternatives longed and two lowest ranked shorted. Portfolios are made until all possible pairs are used, which means we have 38 portfolios from one pair to 38 pairs named from p1 to p38. Portfolio holdings change daily according to rankings. We also calculate an index portfolio of simply holding all stocks in the dataset.

The standard PT methodologies reviewed in chapter 2 have their limitations and requirements for the data to be applicable. ANNs have less requirements for the data, but also less theoretical basis in econometrics to argue for its use. We could say that ANNs are a “jack of all trades” -solution, which fits the combined forecasts approach applied in this paper. In the context of Pairs Trading and this particular application with MCDM the fitness of ANN has a limited amount of research. But the way we can handle a large dataset without null hypotheses and assumptions of the data to test is fast and computationally efficient. Effectively this method is an Unsupervised Machine Learning method. Huck (2010) notes that this approach to Pairs Trading is still in development. Introducing further complexities in the model will only make it harder to interpret and compare results meaningfully with other methodologies so we hold on to the base methodology but apply it to daily data instead of weekly. But Huck (2010) also notes that testing different trading rules for the model is a meaningful way to go on and further testing and proving the models fitness could decrease the status of a black box when/if relevant trading rules can be parametrized, tested and compared fairly to other methodologies.

4. Case: Testing Huck's (2009) Combined Forecasts Method with OSE daily data

In this chapter we describe how we applied the methodology defined in previous chapter to our data from Oslo Stock Exchange (OSE) with R. In this chapter we go through the empirical part of the research – what we did to the data and how we applied the methodology with references to some of the R scripts programmed in appendices. See figure 7 on page 33 for a summarization of the methodology.

4.1. Data

The data used in this paper is Oslo Stock Exchange (OSE) daily price data downloaded from TITLON database. The prices are adjusted for dividends and corporate events by the TITLON database. The original dataset consists of 214 companies that were active in OSE during the time period of 1980 to July 2016. We examine the time period of 2006/6-2016/6 and make a subset accordingly. We examine a 10 year timeframe to have comparability with Huck (2009) and also to include the same timeframe as Mikkelsen & Kjaerland (2018), they examine January 2012 to march 2016. Because we want to start training on June 2006 and trading on June 2011 we also include 3 days before June because we incorporate 3 lags so our first training forecasts are on June 1st. Stocks with over 10 days of missing values are removed, so we instantly rule out stocks that have either listed after our starting date or delisted before our end date. This also rules out the most illiquid stocks that have many missing trading days. We are then left with 77 stocks. We fill the missing values with previous values since the neural network wont work with missing values. 10 years of trading days means that our dataset is 2535 trading days long. A maximum of 10 missing values have a very limited impact on the end results. We calculate daily returns to use on our model, the first 5 years is used to train the neural networks and last 5 years is used to make spread forecasts and apply the methodology on.

4.2. Applying the methodology to data

The data came in a xts/zoo data structure, for simplicity the data we need is extracted into simple matrix and array data structures. For the Elman neural network the RSNNS-package is used, see appendix 1 for the ENN script used in this paper. In the ENN model we have a total of 6 inputs, so for each pair of stocks we have 3 lagged inputs per stock. Using lagged inputs further incorporates the time aspect in to the model. The model has 2 outputs, 1 timestep ahead for each stock. Huck (2009) avoids overfitting according to empirical rules proposed by other research: He uses twice the amount of hidden nodes compared to the amount of input nodes. We follow the same protocol and use 12 hidden nodes for our 6 inputs. We train the neural networks for the first 5 years including three days before June to have a forecast on 1st of June because of the three lags and to start the trading on the first day of June in 2011. Basically, the loop trains Elman networks to all possible (77*76)/2=2926 pairs, then calculates the spreads of the forecasts for each day and saves them to an array called spreads in the scripts in appendices. These spread forecasts are next used in our TOPSIS MCDM to rank stocks to PT with, which is what and how we combine the use of multiple daily forecasts used in our model.

In Appendix 2 we use topsis-package and run our spreads through a TOPSIS loop that saves our daily rankings to a matrix. In Appendix 3 we make a trading simulation, where we calculate the returns for each 38 pairs portfolios formed according to rankings and column 39 for an index portfolio of all the 77 stocks returns we use for our ENN model for comparison. Appendix 4 also includes 0,039% trading costs, same as Mikkelsen & Kjaerland (2018). While the trading costs could be lower in practice, that was the lowest offer Mikkelsen & Kjaerland (2018) could find at the time from Norwegian internet brokers. Also it might be worth keeping in mind, that short-selling sometimes has other costs than the trading fee itself. Appendix 3 also calculates directions of forecasts, basically calculating if the direction (+-) forecasted by the model was right or wrong for each stock each day used in the simulation.

4.3. results

We examine direction statistics, simple returns and standard deviations for comparison to Huck (2009) and we also include an index of all the 77 stocks in our dataset. All our results statistics are averages of daily results for the whole trading period of 5 years from June 2011 to June 2016.

The direction statistic is sometimes referred to as sign, but essentially it measures if we forecasted the directions (+-) of next days returns correctly. The direction statistic is intuitively interpreted when comparing to the 50% threshold. If the model is over 50% correct it means that most of our trades are profitable and it could mean that the model performs better than random guessing. This model aims to predict profits – not only to predict directions. The direction statistic doesn't account for the magnitude of the profits, so it doesn't distinguish between miniscule or massive correct or incorrect predictions.

Many studies evaluate econometric models' forecasting success using statistical loss functions. (Brooks, 2008). Leitch & Tanner (1991) highlight the importance of direction statistics for evaluating the profitability of statistical models in finance if profits are not observable, because they find that conventional statistical models forecast error magnitudes are only marginally related to profits, and that direction demonstrates a higher degree of statistical association to profits.

If the trading simulation outperforms the index and achieves a direction statistic of over 50% simultaneously it will seem very promising. The results are expected to follow a decreasing trend as portfolios grow bigger, because the first portfolios have the most extreme forecasts and the bigger portfolios include more mixed forecasts. Results are examined below and conclusions are made in the following parts.

Direction statistics show an impressive 56% for the portfolio of one pair, but the correct direction estimates fall down as portfolios go bigger. From portfolio 22 onwards the statistic falls below the 50% threshold, see Figure 8 below.

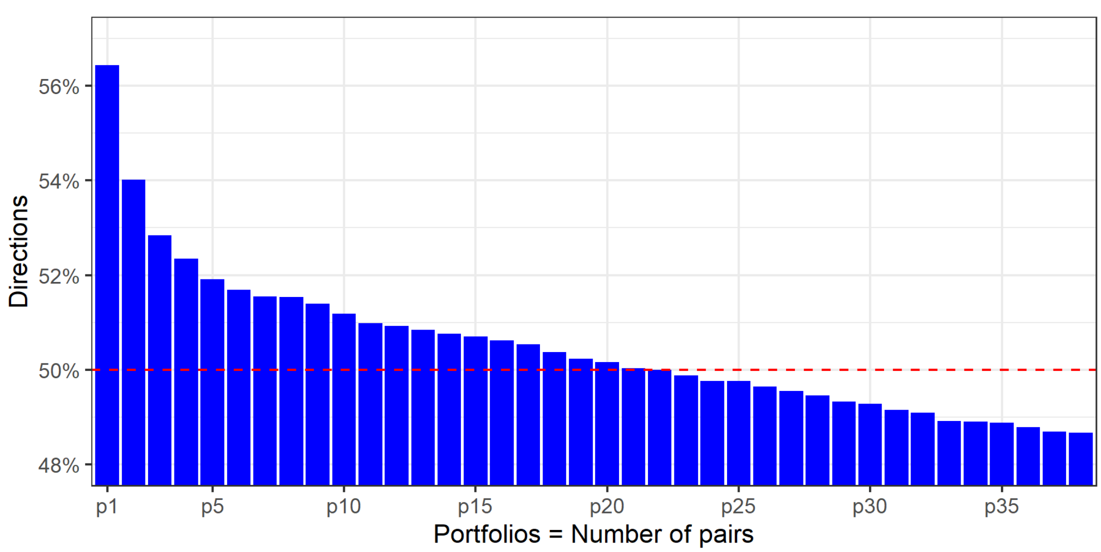

Figure 8. Directions

Daily average returns are average returns for each portfolio of all the trading days in our data. The one pair portfolio has a massive 1,2% average daily returns in the trading period 6/2011-6/2016. All 38 PT portfolios exceed the returns of the index even when trading costs are accounted for. It seems like the model is better at forecasting return spreads when it matters, than evaluating direction for a seemingly good fitting statistical model. See figure 9 below.

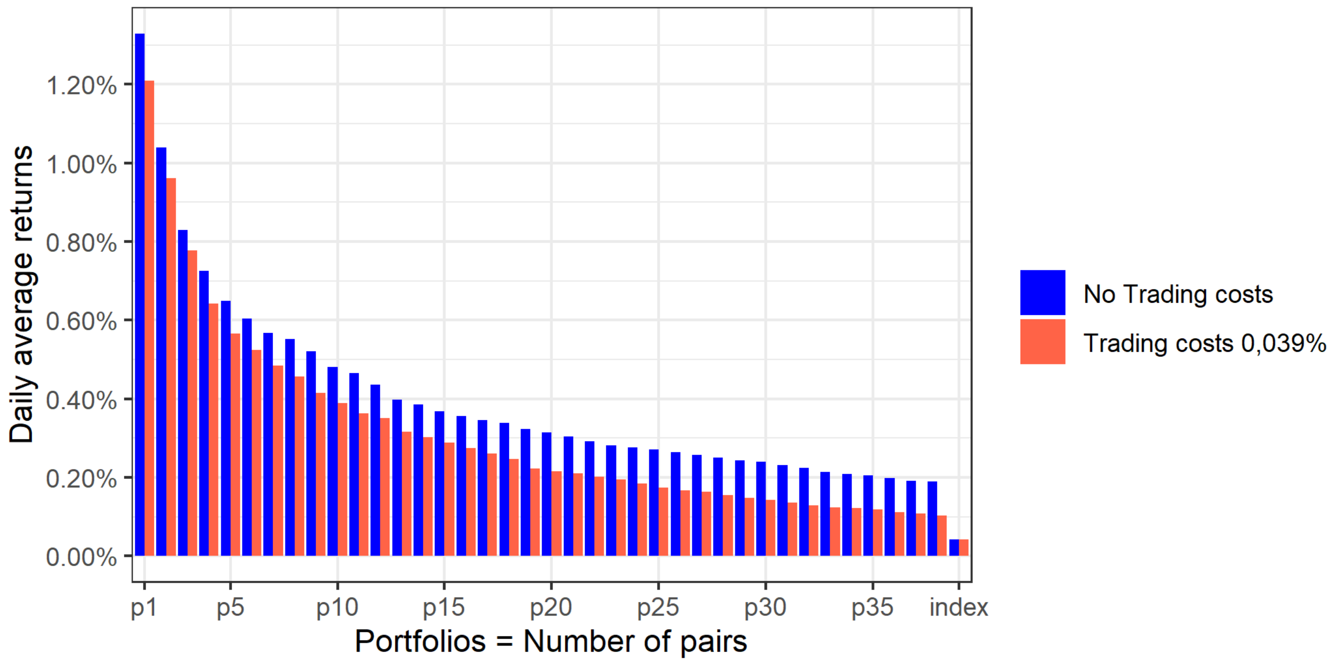

Figure 9. Daily average returns

The standard deviations are bigger for smaller portfolios and smaller for bigger portfolios as expected. Portfolios 13 and 14 make an exception as they have a higher standard deviation than portfolio 12, see figure 10 on the next page.

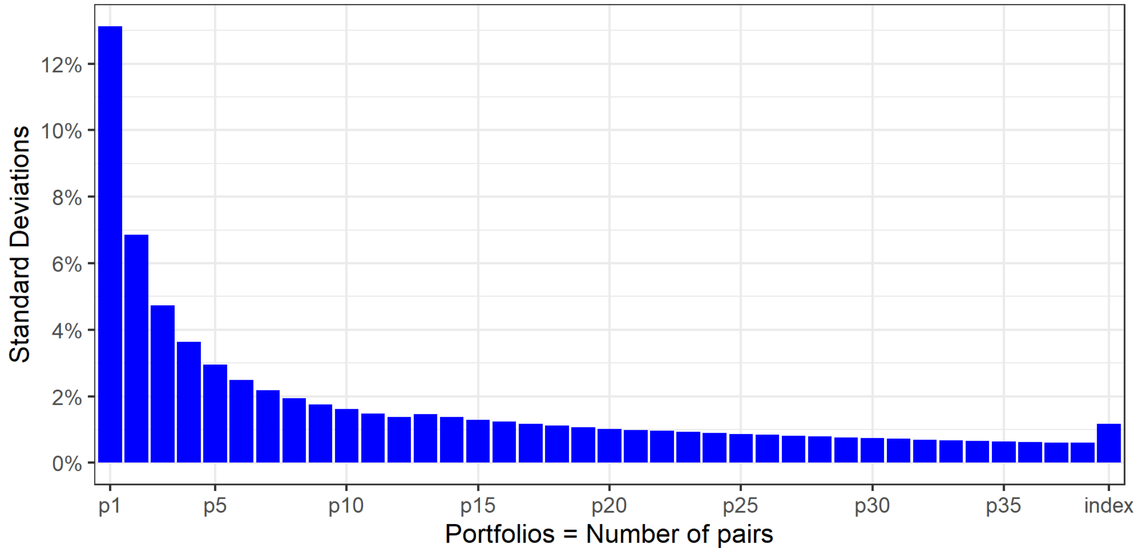

Figure 10. Standard deviations

Tables 4 and 5 below summarize the results for portfolios 1 to 5 and bigger portfolios. The model clearly outperforms the index, and the decreasing performance trend with bigger portfolios is evident in both the graphs and tables.

Table 4. Results, first five portfolios

	p1	p2	p3	p4	p5	index
Directions	56,43 %	54,02 %	52,84 %	52,35 %	51,91 %	
Standard deviation	13,13 %	6,85 %	4,74 %	3,63 %	2,96 %	1,16 %
Avg. Daily Returns	1,33 %	1,04 %	0,83 %	0,73 %	0,65 %	0,04 %
Avg. Daily Returns incl. Trading costs	1,21 %	0,96 %	0,78 %	0,64 %	0,57 %	0,04 %

Table 5. Results, bigger portfolios

	p1	p10	p20	p30	p38	index
Directions	56,43 %	51,19 %	50,16 %	49,28 %	48,67 %	
Standard deviation	13,13 %	1,61 %	1,02 %	0,74 %	0,60 %	1,16 %
Avg. Daily Returns	1,33 %	0,48 %	0,31 %	0,24 %	0,19 %	0,04 %
Avg. Daily Returns incl. Trading costs	1,21 %	0,39 %	0,22 %	0,14 %	0,10 %	0,04 %

4.4. Comparison to existing research

In this chapter we compare the results to existing research. Our results consist of the direction statistics, daily average returns (note that returns are sometimes referred to as excess returns in PT research) and finally standard deviations. These results are computed for all possible pairs portfolios we could have. In this paper we have portfolios from one pair to 38 pairs which is the maximum number of individual pairs with 77 stocks. We also compare our results to the index of all 77 stocks in the dataset.

The direction statistic measures the accuracy (+-) of the forecasts for each portfolio. It is the most relevant measure of accuracy for this model, because of the bloated theoretical returns due to the shortcomings of the model. Therefore we will first address the direction results. Comparing to Huck (2009) our direction statistics decline more consistently as the portfolios grow bigger. Huck (2009) has slight variations, where bigger portfolios directional accuracy exceeds their smaller portfolio counterparts. In our results the portfolios directional accuracy lowers as the portfolios grow bigger. In general, the trends look similar but with our data the directional accuracy falls faster from the top portfolios and falls below the 50% line at portfolio 22 and continues to fall until 48.7% whereas in Huck (2009) all portfolios exceed 51.5% and the decay in accuracy is smaller as portfolios grow bigger.

Christoffersen & Diebold (2006) examine sign dependence with GARCH volatility modelling simulations and find that sign dependence is more likely to be found in intermediate return horizons, in other words weekly and less likely to be found in higher frequency, like the daily data we are using or lower frequency, for example annual data. These findings with GARCH are in line with the direction results of this thesis and Hucks (2009), despite the different statistical methods used.

The errors of the neural networks can be measured in sum of squared estimate of errors (SSE), also know as residual sum of squares (RSS). The error goes down as the networks make more iterations and the networks fit the training data better. We do not optimize the training of these networks or test them, we just simply train and forecast because of the great number of networks. SSE as a training criterion penalizes bigger errors more heavily, than direction would. This could also be the key to another explanation as to why our direction statistics fall lower than Hucks (2009), there is also the slight possibility that we do more iterations than Huck (2009) and overtrain the networks. Huck (2009) doesn’t specify how many iterations he trained his networks, in this paper we train each network for 1000 iterations (Brooks, 2008).

As mentioned before, when comparing results, especially the returns of this model to existing methods is not fair because of the over-optimistic assumptions ingrained in the model as it is. With our first portfolio performing a massive 1,2% daily average return during the whole 5 year trading period we could unjustly claim that it performs better than almost any other method. After just one year of trading with the over-optimistic assumptions, but trading costs accounted for in the simulation, the portfolio with one pair is over 65-fold of the original value and the portfolio with 38 pairs has over doubled in value and the success continues throughout the whole 5 year trading period. However, in the context of returns it is interesting that Huck (2009) had a 1,2% weekly average return on his best performing portfolio. He had a different market and weekly timestep, while in this paper the timestep is daily. Maybe 1,2% is roughly the best returns that this model could provide per timestep, or maybe the same return per timestep is just a coincidence.

Huck (2009) had his best weekly average returns with a 3 pair fortfolio, rising from 1 pair to 2 pairs and then a decreasing trend after 3 pairs. This paper has the best daily average returns on a one pair portfolio. While the average returns are steadily decreasing in this paper, in Huck (2009) they seem to have slightly more fluctuation, but still a decreasing trend is evident in both papers as can be expected. The decreasing trend is expected because the forecasts should be more accurate for smaller portfolios that include the higher and lower ranked alternatives that have more extremely predicted spreads and inaccuracy increase as portfolio sizes grow bigger where mid parts of the rankings are included, in which combined spread predictions can be more mixed and less extreme.

Lastly we have the standard deviations, which are very similar in both papers. It makes sense, that when the portfolios get more diversified, the standard deviations decrease. It is interesting to note, that Huck (2009) had a clearly decreasing trend in standard deviations without fluctuations and in our results we see an upward bump on portfolios 13 and 14.

We can conclude that the results are in line with what can be expected from previous research. Differences between Huck (2009) and our results are minor and could be explained with various reasons, such as different market conditions. The methodology seems promising for further research, especially if the problems mentioned in this paper regarding the signals and making trades could be addressed.

4.5. Conclusions from the tests

While the results are impressive, they are still quite theoretical with very optimistic assumptions especially on how the initial open trades are done, but also for the closing trades. Trading volumes might not translate into 0,039% trading costs in practice. Also the model could be developed to use trading signals, of which after the model opens positions and closing trades should also take into account the real tick trades in the market. These results don't yet apply into practice, but they add to the relevancy of the model both in literature and practice. Keeping in mind Leitch & Tanners (1991) findings of statistical models forecast error magnitudes being only marginally related to profits, the returns being positive even when directions fall below 50% makes sense when referring to existing literature.

We can conclude that the model seems to be able to capture massive returns, at least with the data we are using in this paper. Further examining the results with portfolio performance measures would be misleading, because of the shortcomings of the model. These massive returns are bloated by the over-optimistic assumptions of the model. However, we are not making a model applicable to real world, we are merely testing Huck's (2009) methodology on a different dataset, and the results are in line with existing research with only minor differences in results. This means that the model could be further developed in a way that addresses its shortcomings and would produce comparable results to the existing research on PT approaches. Also when examining the massive returns and direction predicting simultaneously, the model does seem to have promising predicting capabilities for pairs trading stocks.

5. Conclusions

In this paper PT was examined, approaches in literature were reviewed and an experimental combined forecasts methodology was applied to daily OSE data. The methodology consists of making daily pairwise forecasts of spreads for all possible (77*76)/2=2926 pairs in OSE with Elman Neural Networks using the first 5 years of data for training and forecasting spreads for the next 5 years. Stocks are ranked with TOPSIS, which combines our daily forecasts into simple rankings for each stock for each day in the trading period. Then a trading simulation is made, where we applied portfolios from 1 to 38 pairs consisting of the expected top performers for each day, in other words top ranked stocks are bought and bottom ranked stocks are shorted. The portfolio number is also the number of pairs included in the portfolio. Positions are opened daily and closed on the next day, new positions opened immediately after closing the previous according to the rankings and continuing until the end of our 5 year trading period using the top ranked stocks for each day.

The goal of this paper was to review Pairs Trading and continue Mikkelsen & Kjaerlands (2018) work on PT with ML methods. The same market is used with daily data and the trading period in this thesis exceeds that of Mikkelsen & Kjaerlands (2018). An experimental methodology first introduced by Huck (2009) is applied in this thesis. Both Huck (2009) and this study has promising results, which solidify grounds for further research and development of the methodology. In the next part we will answer the research questions of this paper.

5.1. Answers to research questions

Sub-question: What is known about pairs trading in academic literature?

The academic research covers a wide variety of approaches to PT as presented in chapter 2 Literature Review. The traditional approaches as defined by Krauss (2017) consist of distance approach, cointegration approach, time-series approach, stochastic control approach and other approaches. All methodologies seem to have possibilities for further research, especially in the field of high-frequency data. The shortcomings of research is sometimes in applying the statistical method, usually not testing them adequately. Sometimes, especially in this paper and some others, the shortcoming is not taking into account the real-life applicability of the trades, in other words trading volumes in high-frequency ticks. This paper adds to the body of research in PT and further solidifies Huck's (2009 & 2010) methodology as a viable option for further academic research, especially if HF data could be included, and maybe even for practitioners to apply.

Sub-question: How does Huck's (2009 & 2010) method compare with other approaches in Pairs Trading?

Combining forecasts is an approach that is applied when conventional models either are not accurate enough, or when we could produce better results with combining forecasts. Huck (2009 & 2010) and this paper's methodology assumes mean reversions, like when using the Distance Approach or if you would use linear regressions. The co-integration method on the other hand tests for mean-reversions and trades are done accordingly. Combining forecasts could use multiple forecasting methods and combine them or it can be done as we have – by using all possible pairwise forecasts and combining them to an outranking system and trading accordingly. Krauss (2017) criticized that especially co-integration approach and distance approach can sometimes get nearly identical results because they use similar trading rules. Maybe the combined forecasts approach could have a solution to this problem, such as using the forecasts to also make trading rules.

Main Research Question: How does Huck's (2009 & 2010) combined forecasts methodology work with OSE daily data 6/2006-6/2016?

The methodology works with the data and results are very promising. Direction tests fell below 50% with bigger portfolios, but on the other hand when examining returns and standard deviations the methodology does seem very promising. We can say that as the methodology is now, it shows promising results but further research is needed that would address the shortcomings of this model. The main shortcomings are over-optimistic assumptions on opening and closing the trades and not addressing trading volumes.

5.2. Lessons learned

A take home message of this research is that there is a lot of potential, but still some problems to address in PT research. High-frequency data is the new trend in PT research, it is a step towards more frequent trading. Stübinger & Bredthauer (2017) note that intraday trading research using stock prices in PT started in 2010. While addressing the problems of the model used in this study doesn’t necessarily require intraday trading, it could be applied to intraday as well and it could provide solutions for signals of trades. Intraday has had promising results with different kinds of methodologies and if you are interested in PT as a researcher or practitioner, it is suggested to review literature.

Examining stock data with ML is a double-edged sword. A lot of potential could be uncovered, but it is also risky and uncertain as the models don’t directly rely on statistically proving co-movement or other statistical relations between the stocks, but are rather just an application to tap on known relations with a different methodology. ML shows a lot of potential, as these methods can find different kind of connections than the methodologies limited to proving co-movement or other relations between stocks. While the application in this paper should still be developed to validly compare results, it shows great promise. When applying ML one should understand the known relations and possible shortcomings of each applicable methodology.

5.3. Future research directions

Pairs Trading seems like a promising area of research in finance, it’s relatively new and various statistical models and approaches can be applied and developed to examine it. In such a new field authors like Krauss (2017) and Stübinger & Bredthauer (2017) have made interesting thorough reviews with critique that helps researchers to easier understand the concepts and methodologies in PT and point out what future research could focus on and what is relevant to develop. The following questions are good to note for all researchers in PT, but especially when applying a combined forecasts method such as the one in this paper.

Further research should address the questions of volumes and real-world applicability. Could these trades be done in real life trading? Could the volumes traded in tick-by-tick trading provide enough volumes for this strategy to be profitable in practice? How could this strategy be applied to tick-by-tick data?

For the methodology used in this paper, the first problem is the opening trade. Could we use morning ticks for signal to open positions and afternoon ticks for actually opening positions? Or could we find another way to apply data to this methodology to bring the results more relevant in real-world trading?

Different forecasting methods are being tested on equities by many authors, such as Hucks (2019) follow-up research. There is also a wide body of forecasting methods to use with a combined forecasts approach, which can be applied as Huck (2019) is doing with his follow-up research. If the problems for real-life trading would be addressed, these approaches could be more validly tested with traditional portfolio performance metrics and compared to the body of research done with traditional approaches to PT.

5.5 Critique

This thesis examines an experimental approach to PT. Much research is done on statistical arbitrage, and a small part of it has a specific interest in PT. This papers' main shortcomings are reviewed in the following paragraphs and the biggest critique for this paper is not addressing them. If I would start this paper again, I would focus on these shortcomings. But for now they are left for future research to develop solutions for.

Effectively we train (77*76)/2=2926 neural networks which are not validated or tested, instead they are ran through a TOPSIS ranking algorithm and then traded accordingly. Traditional approaches to PT should be done in the spirit of supervised ML and data science in general, meaning that they should be tested and validated in statistical terms. All models have an economical, statistical and econometrical theory basis because they are more or less based on the mean-reversion of equities. The methodology we apply here can use the same arguments than the traditional methods, however this methodology is more like an unsupervised ML developed and applied in the context of PT. The methodology applied in this paper doesn't test or validate for mean-reversions, it is theoretically proven that they exist and the model's goal is to capture short-term profits through the ranking of combined forecasts. Combined forecasts are used in general when making good fits with statistical models is hard or when we can improve accuracy with combining forecasts.

The trades are done using the same days data – in essence we are using the closing value of the current day and forecasting spreads for the next one. The problem is, that we are assuming we could make the buys with the same prices as we are using for forecasting. Solutions for this could be to use high-frequency data, use only morning data for the last day used in the model and make the trades using high-frequency stock values of the afternoon. A similar albeit smaller problem is with the selling or closing of the trades – we are assuming we can close the trades in the closing price of the day instead of checking high-frequency tick by tick data. Also much of the research in PT might not take into account trading volumes, or only partly addresses them by removing the most illiquid stocks and assumes that the trades in their trading simulations could be done in the real world. Therefore the results, especially measured in portfolio returns should be seen rather as a proof of concept and very theoretical rather than practical.

These shortcomings are not only for this thesis – they are present in some of the research in PT and statistical arbitrage in general. Research has to start at some point, and this is the starting point for this particular methodology. Better programming practices could also be used, in this paper the programming was done to get this base methodology up and running. Data science has other forecasting approaches to try, and the rules for trading should be examined with a financial viewpoint. Practitioners are probably ahead in testing different kinds of trading signals or rulesets, but researchers could also move focus towards them.

REFERENCES

Avellaneda, M. and Lee, J.-H. (2010) Statistical arbitrage in the US equities market. Quantitative Finance. 10, 7, 761–782.

Basheer, I.A., & Hajmeer, M. (2000) Artificial neural networks: fundamentals, computing, design, and application. Journal of Microbiological Methods 43, 3–31.

Bowen, D.A., Hutchinson, M.C. & O'Sullivan, N. (2010) High frequency equity pairs trading: transaction costs, speed of execution and patterns in returns. The Journal of Trading 5, 3, 31–38.

Brooks, C. (2008) Introductory Econometrics for Finance. Second edition. Cambridge University Press.

Chong, E. K. P. & Zak, S. H. (2001) An Introduction to Optimization. Second edition. A Wiley-Interscience Publication, John Wiley & Sons, Inc.

Christoffersen, P. & Diebold, X. (2006) Financial asset returns, direction-of-change forecasting, and volatility dynamics. Management Science 52, 8, 1273–1287.

Cummins, M. & Bucca, A. (2012) Quantitative spread trading on crude oil and refined products markets. Quantitative Finance 12, 12, 1857–1875.

Do, B. & Faff, R. (2010) Does simple pairs trading still work? Financial Analysts Journal 66(4): 83–95.

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

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

Ferreira, L. (2008) New tools for spread trading. Futures: News, Analysis & Strategies for Futures, Options & Derivatives Traders 37, 12, 38–41.

Gatev, E., Goetzmann, W. N. & Rouwenhorst, K., G. (2006) Pairs Trading: Performance of a Relative-Value Arbitrage Rule. The Review of Financial Studies, Vol. 19, No. 3, pp. 797-827.

Hagan, M.T., Demuth, H. B., Beale, M. H. & De Jesus, O. (2014) Neural Network Design. Second Edition. eBook.

Huck, N. (2009) Pairs selection and outranking: An application to the S&P 100 index. European Journal of Operational Research 196, p. 819–825.

Huck, N. (2010) Pairs trading and outranking: The multi-step-ahead forecasting case. European Journal of Operational Research 207, p. 1702-1716.

Huck, N. (2013) The high sensitivity of pairs trading returns. Applied Economics Letters, 20, 14, 1301–1304.

Krauss, C. (2017) Statistical Arbitrage Pairs Trading Strategies: Review And Outlook. Journal of Economic Surveys, 31, 2, 513-545.

Krauss, C. and Stübinger, J. (2017) Nonlinear dependence modeling with bivariate copulas: statistical arbitrage pairs trading on the S&P 100. Applied Economics, 49, 52, 5352–5369.

Lei, Y. & Xu, J. (2015) Costly arbitrage through pairs trading. Journal of Economic Dynamics and Control, 56, 1–19.

Leitch, G. & Tanner, J. E. (1991) Economic Forecast Evaluation: Profit Versus the Conventional Error Measures, American Economic Review 81, 3, 580-90.

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

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

Mikkelsen, A. (2018) Pairs trading: the case of Norwegian seafood companies. Applied Economics, 50, 3, 303–318.

Mikkelsen, A. & Kjaerland F. (2018) High-frequency pairs trading on a small stock exchange. International Journal of Economics and Financial Issues, 8, 4, 78–88.

McCullock, J. Mnemosyne Studio. [Webpage, Accessed 25.8.2020] [http://mnemstudio.org/neural-networks-elman.htm

Puspaningrum, H., Lin, Y.-X. & Gulati, C.M. (2010) Finding the optimal pre-set boundaries for pairs trading strategy based on cointegration technique. Journal of Statistical Theory and Practice 4, 3, 391–419.

Quek, C., Pasquier, M. & Kumar, N. (2007) A novel recurrent neural network-based prediction system for option trading and hedging. Applied Intelligence 29, 138–151.

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

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

TITLON database. [Webpage, Accessed 20.8.2020] http://titlon.uit.no/

Vidyamurthy, G., 2004. Pairs Trading: Quantitative Methods and Analysis. Hoboken, N.J: John Wiley & Sons.

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

Zell, A. et al. (1998), 'SNNS Stuttgart Neural Network Simulator User Manual, Version 4.2', IPVR, University of Stuttgart and WSI, University of Tübingen. Available: http://www.ra.cs.uni-tuebingen.de/SNNS/welcome.html

APPENDICES

Appendix 1. Elman Network with RSNNS-package, trainlen -variable is the length of the training period

```
for (a in 1:ncol(tickers2)){
  set.seed(0)
  inputs[1:trainlen,1]=returns[1:trainlen,tickers2[1,a]]
  inputs[1:trainlen,2]=returns[2:(trainlen+1),tickers2[1,a]]
  inputs[1:trainlen,3]=returns[3:(trainlen+2),tickers2[1,a]]
  inputs[1:trainlen,4]=returns[1:trainlen,tickers2[2,a]]
  inputs[1:trainlen,5]=returns[2:(trainlen+1),tickers2[2,a]]
  inputs[1:trainlen,6]=returns[3:(trainlen+2),tickers2[2,a]]
  outputs[1:trainlen,1]= returns[4:(trainlen+3),tickers2[1,a]]
  outputs[1:trainlen,2]= returns[4:(trainlen+3),tickers2[2,a]]
  model <- elman(inputs, outputs, size=12, learnFuncParams=c(0.1), maxit=1000)
  i=1
  if (a>1){
    if (tickers2[1,a]!=tickers2[1,(a-1)]){
      r=r+1
      b=1
    }
  }
  for (i in 1:tradelen){
    inputs[1:trainlen,1]=returns[(i+1):(trainlen+i),tickers2[1,a]]
    inputs[1:trainlen,2]=returns[(i+2):(trainlen+i+1),tickers2[1,a]]
    inputs[1:trainlen,3]=returns[(i+3):(trainlen+i+2),tickers2[1,a]]
    inputs[1:trainlen,4]=returns[(i+1):(trainlen+i),tickers2[2,a]]
    inputs[1:trainlen,5]=returns[(i+2):(trainlen+i+1),tickers2[2,a]]
    inputs[1:trainlen,6]=returns[(i+3):(trainlen+i+2),tickers2[2,a]]
    pred=predict(model,inputs)
    forecastup[a,i]=pred[trainlen,1]
    forecastlow[a,i]=pred[trainlen,2]
    spreads[r,b,i]=pred[trainlen,1]-pred[trainlen,2]
  }
  b=b+1
}
```

Appendix 2. TOPSIS with topsis-package. Tradelen -variable is the length of the trading period. Rankings is the matrix where daily rankings are saved for all alternatives (stocks).

```
#weights w & impacts i for the model
w=1
i=1
a=1
#prevectorize w & i
for (a in 1:ncol(returns)-1){
  i[a]=("+")
  w[a]=1
}
#prevectorize rankings
rankings=matrix(0, nrow=tradelen,ncol=ncol(returns))
#Loop runs TOPSIS for every day and saves results to matrix
for (a in 1:tradelen){
  mcdm=topsis(spreads[,,a],w,i)
  rankings[a,]=mcdm[,3]
}
```

Appendix 3. Trading sim with direction statistics, tradret -variable is a subset of returns of the trading period

```
for (a in 2:nrow(portfolios)){
  for (i in 1:(ncol(portfolios)-1)){
    shorts[1:i]=1-tradret[a-1,names(tail(sort(rankings[a-1,]),i))]
    longs[1:i]=1+tradret[a-1,names(sort(rankings[a-1,])[1:i])]
    portfolios[a,i]=sum(((portfolios[a-1,i]/(i*2))*longs[1:i])
                        +((portfolios[a-1,i]/(i*2))*(shorts[1:i])))
    directions[a-1,i]=(sum(longs[1:i]>1)+sum(shorts[1:i]>1))/(i*2)
  }
  portfolios[a,39]=portfolios[a-1,39]*(sum((1/77)*(1+tradret[a-1,])))
}
```

Appendix 4. Trading sim with trading costs

```
for (a in 2:nrow(portfolios)){
  for (i in 1:(ncol(portfolios)-1)){
    shorts[1:i]=1-tradret[a-1,names(tail(sort(rankings[a-1,]),i))]
    longs[1:i]=1+tradret[a-1,names(sort(rankings[a-1,])[1:i])]
    tcportfolios[a,i]=(sum(((0.99961*tcportfolios[a-1,i]/(i*2))*longs[1:i])+
                         ((0.99961*tcportfolios[a-1,i]/(i*2))*
                            (shorts[1:i]))))*0.99961
  }
  tcportfolios[a,39]=tcportfolios[a-1,39]*(sum((1/77)*(1+tradret[a-1,])))
}
```