Norges Handelshøyskole

FIN11 Trading and Market Microstructure

FIN11 Term paper

Predefined Information

Start date:	2021-11-02 09:00 AM	**Term:**	2021 12
End date:	2021-12-13 02:15 PM	**Assessment type:**	Norwegian 6-point scale (A-F)
Examination type:	Term paper (70%)		
Flow code:	191 FIN11 1 O3 2021 12		
Internal assessor:	(Anonymised)		

Participant

Candidate No.:	1058

Information from participant

The address from which I take this exam is: *:	Badstustredet 6B

I am aware of what constitutes cheating and the consequences of cheating on an exam. I hereby declare that I will NOT cheat on this exam. *: Yes

Group

Group name:	(Anonymised)
Group number:	9
Other members:	1021, 1029, 1049

FIN11 – Trading and Market Microstructure

Pairs trading with Russian oil

A hidden arbitrage opportunity?



Candidate Number: 1029, 1021, 1049, 1058

Pages: 38

Word count: 6265

(Excluding words in table of content, tables, references and code)

Table of Contents

1. Introduction 3

2. Arbitrage explained 4

2.1 Basic concept of arbitrage 4

2.2 The Arbitrage Pricing Theory Model; The mathematics behind arbitrage 5

2.3 Our chosen method of arbitrage 6

2.3.2 Speculative Arbitrage 6

2.3.3 Pairs trading - Our selected method 7

3. Developing a Pairs Trading Strategy 8

3.1 Identify a trading pair 8

3.1.1 Qualitative analysis 9

3.1.2 Quantitative Analysis 10

3.2 Overview of the selected Pair 11

3.2.1 Rosneft company snapshot 11

3.2.2 Lukoil company snapshot 12

3.3 Defining the trading strategy 12

3.4 Optimizing the trading strategy 14

3.4.1 Arbitrage Bounds 14

3.4.2 Implementing a Stop-Loss 16

4. Measuring performance: 17

5. Performance Analysis, 2017 – September 2021: 18

5.1 Performance analysis under different market conditions: 18

5.2 Market neutrality 20

5.3 Strategy Weaknesses 20

5.3.1 Is the strategy representative to evaluate pairs trading strategy? 20

5.3.2 Data restrictions 21

5.3.3 Capital cost 21

5.3.4 Attracting investors 21

5.4 Revisiting the choice of arbitrage bounds 21

5.5 Further research possibilities 22

6. Conclusion 22

References: 24

Appendix: R code 25

1. Introduction

Long/Short- Equity strategies have risen in popularity among professional investors ever since the dotcom crash in 2000, thanks to their market neutral characteristics (Dunis & Ho, 2005). These equity strategies are essentially based on the concept of an arbitrage opportunity called Pairs Trading, which was allegedly first introduced by Morgan Stanley back in the 80´s (Gatev et. al. 2006). Market neutral strategies are considered a great hedge for otherwise equity and fixed income-based portfolios, as they are uncorrelated to the market and therefore unaffected by sudden downturns. Such strategies do not necessarily outperform the equity or fixed Income markets, but offer substantial diversification benefits. Our fascination for this topic led us to the goal of creating our own Pairs Trading Strategy and see if we can achieve a profit or not. We were also curious what kind of impact different market conditions such as Liquidity, Volatility and Market Return would have on such a strategy especially with regard to the Covid-19 sparked flash crash last year. We summarized this in the following research question:

"Can we design a profitable pairs trading strategy and how is it affected by changing market conditions"

The paper proceeds with a theoretical overview of arbitrage trading and Pairs Trading in particular. This chapter is than followed by a description of our strategy design process. We divided the group into two teams. One of those was responsible for designing the strategy based on data from 2012 to 2016, while the other critically analyzed and reflected the performance over the implementation period from 2017 onwards, specifically focusing on the impact of changing market conditions. In the end we discuss the weaknesses and shortcomings of our study and provide an outlook how we could continue from there.

2. Arbitrage explained

2.1 Basic concept of arbitrage

Arbitrageurs trade instruments whose prices are correlated. Correlated instruments are characterized by simultaneous rise and fall. Instruments that are dependent on common fundamental factors normally have correlated prices and are the basis of arbitrage trading. Arbitrageurs seek opportunities where the prices diverge from their normal relation. Consequently, arbitrageurs are exploiting inefficiency in the marked (Arbitrage, 2021). By going long in the relatively cheap instrument and short in the relatively expensive instrument, arbitrageurs profit when/if the prices converge again. These short and long positions are called the legs and construct the hedge portfolio arbitrageurs seek to optimize. When the value of an instrument falls, gains in the short legs will offset the losses in the long legs and vice versa. Figure 1 shows a visual representation of the concept. The intersection between short and long position is where profit is earned. Analyzing different markets containing the same instruments are generally the practice to seek out arbitrage opportunities.

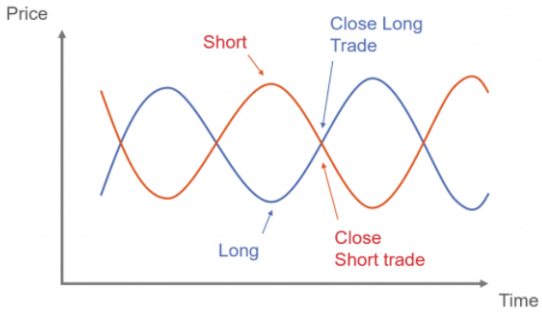


Figure 1: Visualization of pairs trading

To better understand the hedge portfolio of an arbitrage trader we can view the trade as an instrument in itself. When buying the portfolio, you buy the short position and sell the long position. Thereby, when you sell the portfolio, you sell the short position and buy the long position. Consequently, an arbitrage trader is no other than a trader using various strategies to trade their arbitrage portfolio.

Arbitrage trading has different effects on the market. Arbitrages connect buyers and sellers, thereby making the market more efficient. By having legs in different markets, arbitrageurs move and provide liquidity thereby ensuring prices reflect their relative values. Arbitrageurs compete with dealers in providing liquidity. The dealer usually carries the losses on this rivalry because arbitragers generally act on extra information. In the following sections will the mathematical background and different types of arbitrage.

2.2 The Arbitrage Pricing Theory Model; The mathematics behind arbitrage

In order to price assets and identify arbitrage opportunities, the arbitrage pricing theory model (APT) is sometimes used. The APT model can be regarded as a substitution of the capital asset pricing model (CAPM). Both models assume a linear relationship between assets' expected returns and their covariance with other factors. In CAPM, but also commonly in the APT model, the beta 'β' is used as a fitting covariance. This covariance reflects the changes in the expected return of the asset in relation to changes in the expected return of the market. Based on a regression performed using historical returns, the expected return of an asset can be plotted as a formula containing the standard expected return of the asset and the influence of other factors, such as beta. An asset can have one influencing factor, but realistically this is often not practical given the large number of factors that might influence an asset's returns. In this linear formula, the effect a factor has on the expected return of the asset can be regarded as the risk premium of this factor. ATP differs from CAPM, in that CAPM prices a combination between the market portfolio and a risk free asset (i.e. T-bills) and under ATP, the expected return of any well diversified portfolio can be plotted. It is to be noted that under ATP, it is required that any plotted asset is a diversified portfolio such that any idiosyncratic risk is diversified away and only systematic risk is left (Huberman, 2005).

As an example, we can assume that the two formulas for the expected returns of the well diversified portfolios A & B are as such:

- $E(RA) = 10{,}00\ \% + 1{,}5\% * \beta$
- $E(RB) = 8{,}00\ \% + 1{,}5\ \% * \beta$

Given these formulas, there is an arbitrage opportunity. Regardless of the value of β, we can gain a 2 % premium without any risk involved. The value of β does not have to be identical to create arbitrage opportunities, since we can also combine one of our portfolios with a risk-free asset (i.e. T-bills) in order to realize an identical exposure to β.

Following this, it would be interesting to apply ATP in the real market to try to create and realize pure arbitrage trading opportunities. Realistically though, there are several factors which would complicate this. First of all, ATP assumes that assets are priced as if markets offer no arbitrage opportunity (Huberman, 2005). If there were, investors would be quick to take advantage. If we look at our example above, then investors would be quick to buy A & sell B. This would increase the price of A and decrease the price of B, thus decreasing the

expected return of A and increasing the expected return of B until the expected returns of A & B become identical and there are no longer any arbitrage opportunities. To profit from arbitrage opportunities, a trader would have to be extremely fast in order to trade before the opportunity disappears. In other words, arbitrage opportunities would quickly disappear.

Secondly, in order to realize arbitrage trading opportunities, a trader would need to be able to successfully identify all factors which affect the expected return of an asset. If any of the applying factors are wrong or missing, then any found arbitrage opportunities are invalid. Given the infinitely large number of factors that might influence the return of an asset, this might prove difficult (Huberman, 2005).

Lastly, the exposure of factors applied to an asset might very well vary over time. As soon as this changes, any previously found arbitrage opportunities might no longer apply. Based on the following, applying the ATP model to try to create pure arbitrage trading opportunities might prove to be quite difficult. It might still be possible however, to create a profitable strategy by trying to capture statistical arbitrage. One way to achieve this is by pairs trading, which we will explain in the next section.

2.3 Our chosen method of arbitrage

There are several methods for arbitrage trading. They all involve exploiting prices deviating from the assumed fundamental value. Where they differ is on the basis of risk that arbitrageurs assume. Pure arbitrage exploits the convergence of instruments that tend to return to their average value. Theoretically pure arbitrage always converges to mean average making it risk free. Historically, this has not always been the case. A more realistic way of exploiting arbitrage is speculative arbitrage which we will discuss next.

2.3.2 Speculative Arbitrage

Speculative arbitrage is a more common, but also riskier type of arbitrage. Speculative arbitrage involves a hedge portfolio where the values are nonstationary. A value is nonstationary if all future values are expected to be higher or lower than current level. In contrast to mean reversing values, nonstationary have a tendency to wander around without returning to a particular value. Hence, the name “speculative arbitrage”, but also referred to as risk arbitrage.

Speculative arbitrageurs seek opportunities where they believe there is a short-term tendency toward mean reversion. The instruments still follow a nonstationary path due to instrument-specific valuation factors. The mean reversion may be a result of different scenarios, like inconsistent pricing or mispricing of the common factors. By discovering the inconsistent- or mispricing you can anticipate a mean reversion and profit. Keep in mind that speculative arbitrage can be quite risky because there is no guarantee for mean reversion and the fact that the market can stay irregular longer than you can stay liquid.

2.3.3 Pairs trading - Our selected method

Pairs trading is a market neutral strategy made from trading just two securities. In the following we will explain a variant of pairs trading known as statistical arbitrage pairs trading.

The central idea behind the strategy is that two securities with similar characteristics must have prices that move together. This does not mean that the price for the two securities is the same, but that the relationship between the two stocks should stay relatively alike. Think of it as one stock price divided by the other stock price. The resulting relationship-number is the spread. Market inefficiencies can cause one of the securities prices to lag or outpace the other, increasing the spread, but with time the prices should move parallel and naturally the spread should close. This is something that an arbitrage trader can profit from. When it is observed that the spread is particularly large, one can short the overpriced security and go long in the underpriced security. With time, the spread should narrow, and the arbitrageur closes his/her position having earned the excess spread over the mean spread.

Pairs trading is profitable and possible due to mispricing of instrument-specific factors, or inconsistency in the pricing of the two common fundamental valuation factors (Harris, 2002). However, new information or changes in relative prices may change the instrument-specific factors and thereby move the fundamental price. Such permanent price movements are detrimental to a pair’s trader position.

A pairs trading strategy is a market neutral trading strategy that tries to profit by making use of inefficiencies in financial markets (Naicker, 2015, p. 2). Inefficiencies in financial markets can result in mispricing of instruments that are in relation to one another. The theory assumes

securities with similar characteristics should have correlated prices, and therefore diverging from the alleged mean will result in an equilibrium spread. By identifying pairs that are historically correlated, the pairs trader can bet on a mean reversion when the two prices diverge. By having one long leg in the relatively cheap instrument and one short leg in the expensive one, the arbitrageur profit is realized by conversion between the instruments. Consequently, closing their position as the instrument converges as illustrated in figure 1. Keep in mind mean reversion is a short-term speculation in nonstationary instruments.

3. Developing a Pairs Trading Strategy

3.1 Identify a trading pair

In order to come up with and implement a viable pairs trading strategy, it is first necessary to select a suitable pair of assets. In order to do this, we will first use a qualitative approach to identify a pool of companies that should move together. Using "Yahoo Finance" data from 2012 to 2016 (01-01-2012 to 31-12-2016), we will then calculate their actual correlations over the given time period. We would have liked to use a longer time span, but the data from earlier years on yahoo finance has large gaps, rendering it useless. Additionally, our first-choice data source – the Bloomberg terminal of the university had already exceeded its data limit for the calendar year. Thus, we had to settle with this 5-year testing and almost 5 year implementation period. Nevertheless, we should be able to identify a suitable pair for our intended strategy. If that is not the case, we will have to repeat the qualitative step and identify a new pool of securities. After selection, we will use data from 2017 to today (01-01-2017 to 15/11/2021) to test our strategy for the previously found asset pairs and to see how it performs under different market conditions. We aim to prevent 'data snooping' or 'optimization bias', which occurs when parameters are optimized only for a specific dataset with no guarantee of success using other data, by choosing different data sets for the strategy testing and implementation period (White, 2000). However, we must acknowledge that there still is a bias, because we already know that all companies in our pool still exist today, thus removing the risk of bankruptcy.

3.1.1 Qualitative analysis

There are two common ways to identify suitable pairs for a pairs trading strategy. The first one is to use a purely data driven approach where you compare the price movements of thousands of companies and match them respectively. The second one is to use qualitative metrics and fundamental analysis to identify companies that have similar business models, revenue streams, are exposed to the same markets etc. and should therefore by the law of one price be similarly valued. Since we believe the first approach would exceed our capacities, we opted for a hybrid approach. In a first step we will identify a sample of companies that should exhibit similar price movements using a heuristic approach. After that we will check their correlations with each other to identify the best pair among them.

We are looking for companies whose prices we expect to correlate in the future. Intuitively, it would make sense to first narrow our focus to companies with similar characteristics such as industry, company activities and country. The more of these characteristics our selected asset pair share, the more the prices of the two assets are likely to correlate given that the movements of their stock prices are affected by many of the same economic, political, and social factors.

After careful consideration we decided to settle with Russian oil and gas companies. Their business models are very similar encompassing parts of the value chain along the oil and gas business. What turned out to be the deciding factor for us is that all companies´ revenues and by that also stock prices are heavily reliant on the current prices of two commodities: Oil and Gas. The market prices of these are the main profit driver for companies in this sector. We wanted to settle on companies from one specific region so besides Russia we also considered Europe and the United States. However, the US companies revenues seemed very globally diversified, introducing different risks for each company. While there are several big players in the European oil and gas sector like Shell, Total Energies, British Petroleum and Equinor those are all based out of different countries, two also being outside the EU. Here we saw risks of country specific regulations as well as other dynamics that could lead to asymmetric price movements. That left Russia as a country with a large enough private sector in this industry for us to consider. Russian companies of course often struggle with being targeted by sanctions from western nations due to their country's foreign politics, making them less popular investments. However, we saw this as a great opportunity as pairs trading should be

market neutral, we would carry none of this risk. Additionally, the Russian Oil and Gas companies almost exclusively run their operations in their home country removing the source of diversity among them that discouraged us to look at the American market.

Next, we had to think about whether to pick small, medium, or large cap companies as they are affected by very different dynamics on the stock market. For example, inclusion in certain indexes, which are usually composed by market capitalization is a big price driver for large caps, especially with the growing popularity of passive investment strategies. In our mind large caps seemed the most suitable, because in smaller companies' single events that only affect one of the possible pair can have a much bigger impact. Such events include but are not limited to the loss of key personnel, clients or contracts. Large caps have a more diversified business so single incidents have a smaller impact on the bottom line.

That's how we ended up settling on Rosneft, Gazprom, Lukoil and Novatek as the companies we would run our first correlation analysis with. Their stock prices over the respective period is displayed in Figure 2.

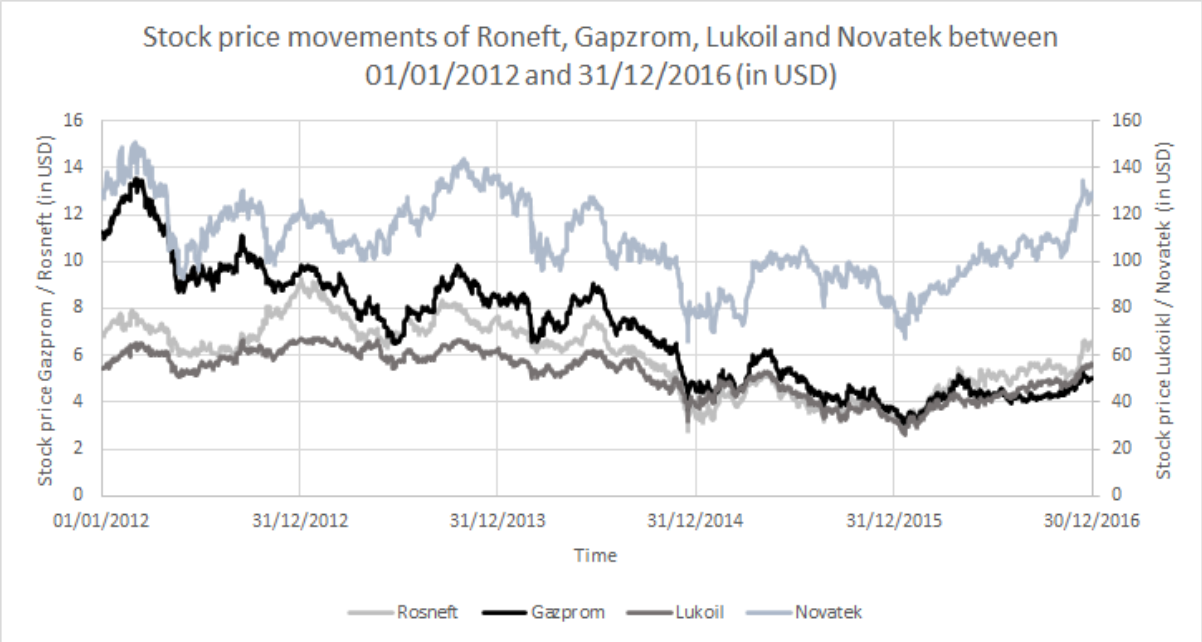


Figure 2: stock price movements of Rosneft, Gazprom, Lukoil and Novatek between 01/01/2012 and 31/12/2016 (in USD)

3.1.2 Quantitative Analysis

From the four selected companies we are now looking for a pair with a particularly high stock price correlation that is needed for a strategy based on mean reversion. Since we are looking for highly correlated assets, we should aim to find a pair with a correlation of at least 0.80.

We used closing prices from the international order book (IOB) of the London Stock Exchange where the Global Depository receipts (GDRs) of the companies are traded (Yahoo Finance 2021a, Yahoo Finance 2021b, Yahoo Finance 2021c, Yahoo Finance 2021d) We chose this option over the Moscow Stock Exchange to save us from messy calculations with fluctuating currency prices as the IOB is denominated in USD.

The formula to calculate the correlation of two stocks is:

$$Corr_{xy} = \frac{cov(x,y)}{\sigma x \times \sigma y}$$

Using the corresponding excel function yielded the following results:

	Rosneft	**Gazprom**	**Lukoil**	**Novatek**
Rosneft		0,83	0,95	0,81
Gazprom	0,83		0,85	0,78
Lukoil	0,95	0,85		0,79
Novatek	0,81	0,78	0,79	

Table 1: Correlation matrix of movements in stock prices between Rosneft, Gazprom, Lukoil and Novatek between 01-01-2012 and 31-12-2016

Our heuristic approach was successful, as even the lowest exhibited correlation of any pair with 0.78 is still considerable. However, even on this very high level the pair of Rosneft and Lukoil standout with a correlation of 0.95. Out of this sample this makes them the obvious choice to build our strategy around.

Another common mathematical technique to identify trends in a multivariate time-series is the Engle-Granger-Cointegration (Engle & Granger 1987). Unfortunately, the replication of their work with our pair surpassed our abilities. However, given the high correlation and common qualitative characteristics we assume with some confidence that there is a long-term equilibrium spread between the pairs.

3.2 Overview of the selected Pair

3.2.1 Rosneft company snapshot

To maintain the setting of our scenario the following description is based on the company's annual report of 2016 (Rosneft 2016).

Rosneft is a Russian energy company with headquarters in Moscow. The majority of their revenues totaling 4,887B Rubles in 2016 come from equal parts crude oil and refined oil sales. They also produce Gas, but it is a much smaller portion of their business. The company was privatized exactly 10 years ago in 2006 and the Russian government still indirectly owns 50% + 1 share of Rosneft. Rosneft operates along the entire oil production value chain from the exploration of new oil fields, their exploitation, subsequent refining and even distribution to end customers through for example their own gas station chain. Thereby a particular emphasis is placed on the exploration of new fields with the company striving to uncover new oil reserves at the same rate they are exploiting known ones. While the majority of their operations is located in Russia Rosneft exports their technology and knowledge in 23 other countries worldwide and operates three major oil refineries in central Europe.

3.2.2 Lukoil company snapshot

Lukoil has slightly higher sales (5,227B Rubles) and Net Income (206B Rubles) (Lukoil 2016). Contrary to Rosneft they are not state owned, but management held around 35% of equity capital in 2016. Still they are presumed to stand under heavy influence by the Russian government. Similarly, to Rosneft they produce both oil and gas, but the gas business is miniscule compared to the oil business. The focus of their business activities is western Siberia, but they also run smaller operations in 30 countries across the globe. The company even used to run a gas station network in the US that was however discontinued with the start of the Crimea crisis. Lukoil has integrated all vertical activities along the oil industry value chain and thus runs almost the same business model as Rosneft, just focusing on different oil and gas fields within Russia.

3.3 Defining the trading strategy

Our strategy is a simple variation of the general method of pairs trading (Gatev et. al 2006; Krauss 2017). We define a measure of spread for the pair as well as arbitrage bounds (from here on just bounds).

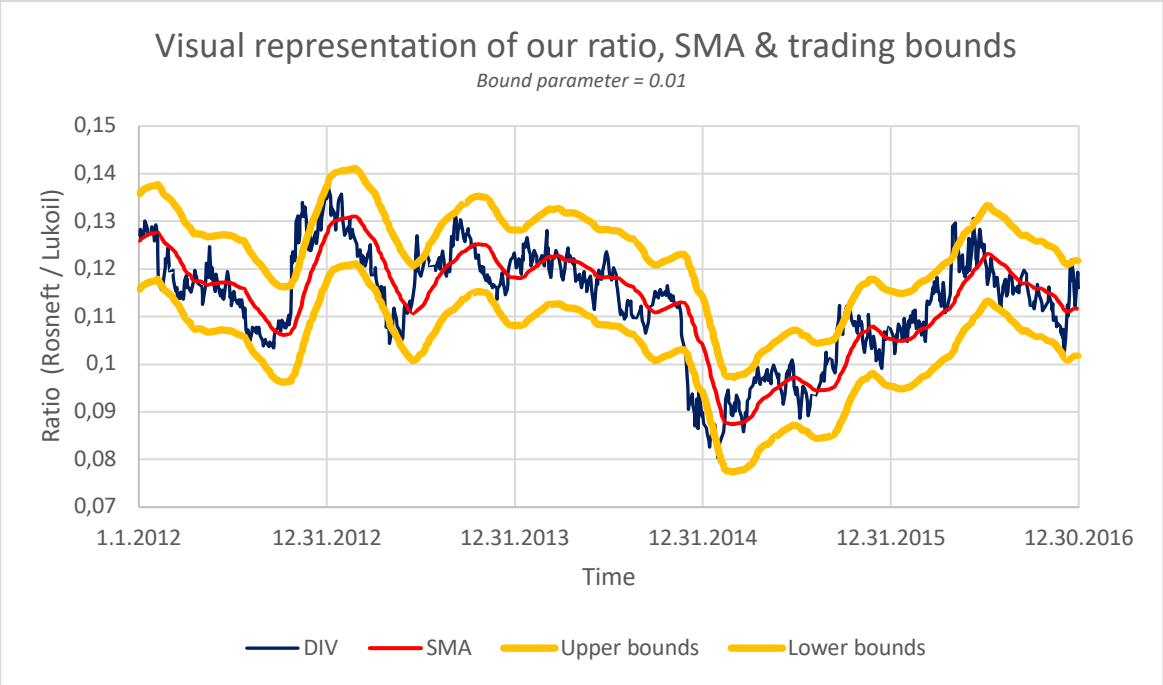


Figure 3: Visual representation of our trading metrics, using a random parameter value of 0.01

The dark blue line called 'DIV' calculated by dividing the stock price of Rosneft by the price of Lukoil. Pairs Trading assumes that the stock prices of the pair move in unison. Therefore, the ratio should stay constant over the long-term and a change in this ratio signal the over-or underpricing of one or both securities. Of course, since the stock prices are not perfectly correlated, we should allow for some margin for the DIV to move. To do so, we will use a 50-day SMA of the DIV line and apply bounds to the 50-day SMA. We are limited in our computational capacity and can only use one SMA. We find the 50-day SMA to be approximately in between medium and long term. The upper bound is calculated as SMA[t] + parameter value and the lower bound is calculated as SMA[t] - parameter value. Whenever the DIV line crosses the lower bound, we open a short position in Rosneft and a long position in Lukoil. This makes our strategy almost capital neutral as the acquisition of the long position can be financed by the sale of the short position. However, borrowing and broker fees will lead to some capital requirements. Whenever, the DIV line crosses the upper bound, we open a long position in Rosneft and a short position in Lukoil. Since multiple signals to open a trade might be generated in a short notice, as the DIV line jumps around the lower or upper bound, we will limit the number of open trades to one at a time. The position will be closed, if at any time the DIV line crosses the 50-day SMA. Therefore, a small parameter value should lead to a significantly higher number of trades than a large value, as the bounds are narrower and smaller price movements are sufficient to execute a trade.

3.4 Optimizing the trading strategy

3.4.1 Arbitrage Bounds

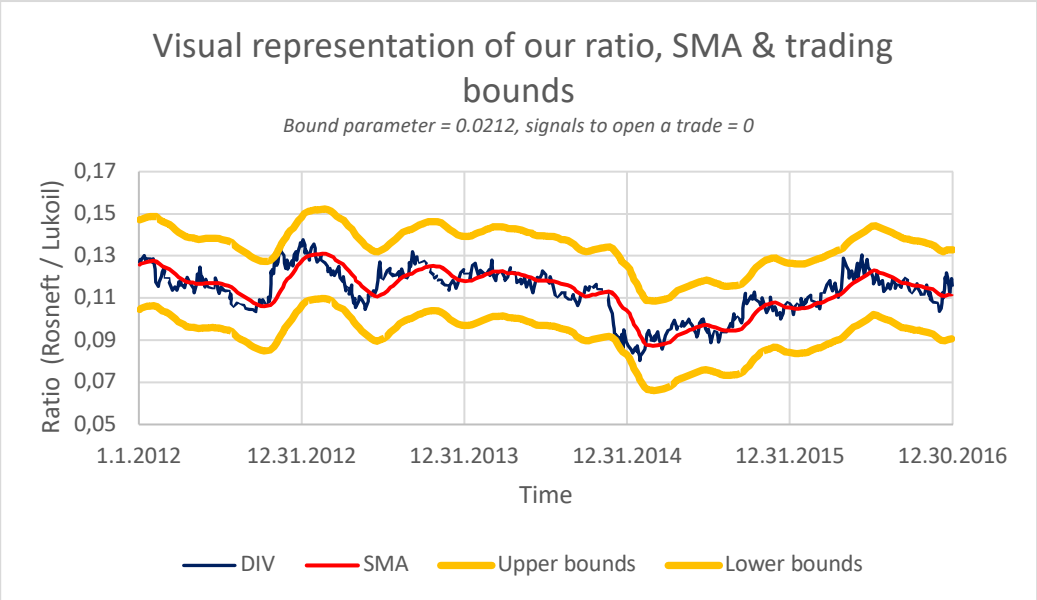


Figure 4: Visual representation of our trading metrics, where the bound parameters are such that the DIV does not cross the upper or lower bounds, such that no trades are triggered

To achieve the highest possible profit, we wrote an algorithm in R to simulate our trades (Figure 5). This would allow us to identify the profit maximizing parameter. First, we must determine for which values to run the algorithm. Using solver in Excel, we find that p = 0.0212 is the first value for which zero trades are triggered, meaning that we only have to analyze lower values.

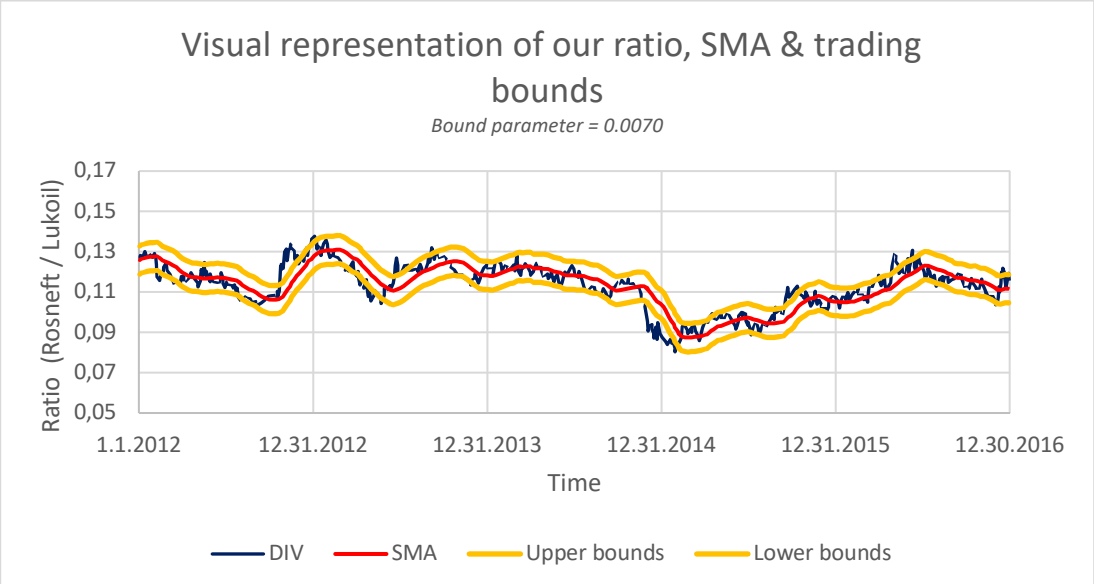


Figure 5: Visual representation of our trading metrics at the parameter value which maximizes our profit

Applying our algorithm (using values for p between 0 and 0.0212 and maximizing for P&L) yielded the following results:

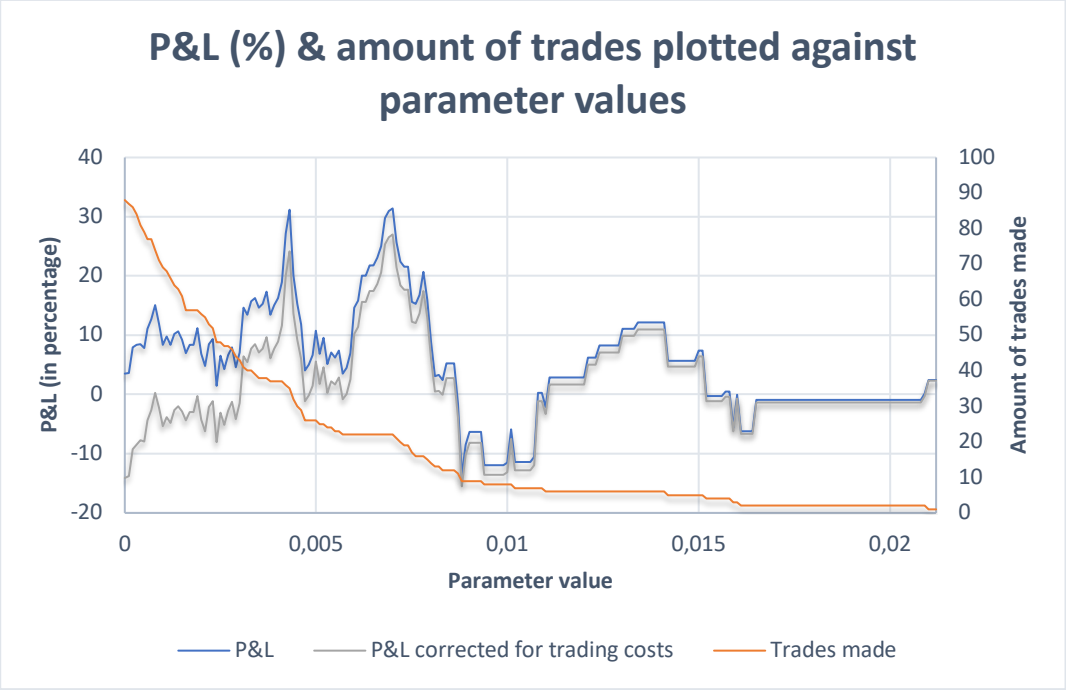


Figure 6: P&L (in %) & the amount of trades for every value of p

As expected, we observe that the number of trades and the value for p are negatively correlated. The highest realized 4-year returns are 31.40 % at p = 0.0070 and 31.18 % at p = 0.0043. Whilst the return at p = 0.0070 is already higher, the difference between the two widens even further as transaction costs are taken into consideration. Transaction costs are critically important when considering an optimal parameter value, given that the amount of trades made fluctuates significantly given different parameter values. Using a parameter value of 0.0070 results in 37.15% less transactions. Based on this, we will use p = 0.0070 as our parameter value. Considering transaction costs of 0.05% per trade based on rates provided by Interactive Brokers (2021), our realized P&L using p = 0.0070 amounts to 27 %.

Another interesting observation is that the highest observed loss for our strategy amounts to only 13.7 %. This implies that our strategy is actually creating value (i.e. reversing our strategy does not yield better results).

3.4.2 Implementing a Stop-Loss

Trade type	Amount (per asset traded)		Combination	Amount (per two assets traded)
Higher loss than 5%	9		One loss, one profitable	17
Between - 5% & + 5%	23		Both loss	0
Higher profit than 5%	12		Both profit	5

Table 2: Trade information using p = 0.0070

When analyzing our trades, we found a substantial number of trades with substantial losses. As expected in pairs trading, a significant amount of trades resulted in a profit on one asset and a loss on the other. It's also worth noting that not a single trade resulted in a loss on both assets. The first observation prompted us to test adding a Stop-loss to the strategy. Of course, the stop-loss had to be applied to the entire portfolio, not a single position. If the total profit of our long position in asset 1 and our short position in asset 2 exceeds the stoploss threshold, then the position is closed, and the loss realized. Again, we ran an analysis to identify the optimal stop-loss value, given our previously determined Bollinger Bands with our algorithm. Sadly, we did not manage to optimize both the parameter and the stop-loss at the same time.

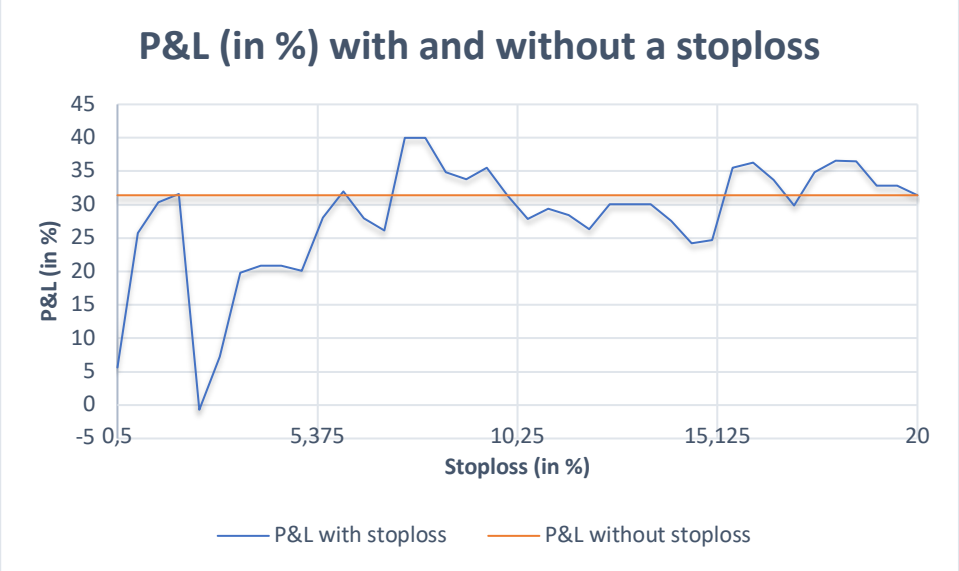


Figure 7: Rate of return when introducing stoploss

Conclusion: stoploss of 7.5% and 8% both increase our P&L from 31.40% to 40% (not considering transaction costs). We'll go with 8% because of the larger margin.

4. Measuring performance:

Risk-free rate (RFR):

Although our portfolio is supposed to be market-neutral, it is as we acknowledge, not completely risk free. We will look further into exactly how risky the portfolio is, but in any case, we want to measure our performance up against a risk-free proxy. In this case the two stocks are traded on the London stock exchange. Therefore, it makes sense to use a risk-free asset from the UK. We've landed on the London Interbank Offered Rate (Libor). This is the go-to metric for all banks in the UK. We'll use the 12-month yield from January 2017 as our risk-free rate throughout the whole period. Using historical data, we've found that the rate was at 1.71% as of January 2017.

Benchmark:

Following the pairs trading strategy, our portfolio should in theory be disconnected from the market. As a benchmark for "the market" we've decided to use FTSE all-share. This is an index comprising of the 600 largest firms listed on the London stock exchange.

Formulas:

To measure our performance relative to the benchmark and the risk-free rate, over the period 2017- September 2021, we'll utilize a few different formulas:

P&L: Total profits made or lost after each trade. With and without transaction costs.

Return on investment. Percentage gain on the initial sum invested. This will be calculated over the entire period, as well as for each year.

$$ROI: \frac{End\ value\ of\ investment - Cost\ of\ investment}{Cost\ of\ investment}$$

Excess returns. Returns achieved above the risk-free rate:

$$Excess\ return: Return - RFR$$

Beta: Measures the portfolios volatility in relation to the overall market (in our case the FTSE all-share):

$$Beta = \frac{Covariance\ portfolio\ \&\ market}{Variance\ portfolio}$$

5. Performance Analysis, 2017 – September 2021:

How did our strategy perform? Over the nearly five-year period from January 2017 to November 2021, we experienced years with losses and years with profits. Overall, our strategy netted us a return on investment of 18.5%. This is substantially better than our benchmark, the FTSE all-share index, and netted us an excess return of 10.8%. To help the reader put this into context, we have calculated our P&L for a scenario where the initial investment is $ 1000 (Table 3). The return-on-investment amounts to $ 184.5, assuming trading costs of 0.05% flat for each trade.

Year	P&L	P&L with TC	ROI	FTSE	Excess return
2017	$ 0.19	$ -7.92	-0.69%	7.24%	-7.93%
2018	$ -101.68	$ -108.17	-10.72%	-7.54%	-3.18%
2019	$ 139.54	$ 131.44	13.24%	6.06%	7.18%
2020	$ 116.24	$ 102.64	10.36%	-10.24%	20.60%
YTD	$ 71.03	$ 66.56	6.76%	14.02%	-7.26%
Result	$ 225.31	$ 184.54	18.45%	7.62%	10.83%

Table 3: P&L with initial sum of 1000 pounds invested

5.1 Performance analysis under different market conditions:

To answer our thesis question; how pairs trading performs under different market conditions we can look at our portfolios performance when markets are liquid/illiquid, and when the markets are volatile/steady. In our period of testing, there is particularly one year with large volatility.

In the covid-shock-year 2020, volatility went through the roof. By looking at table 4 we see that the measured standard deviation for the FTSE is double that of any other year. We see that reflected in our portfolio as well. When looking at the correlation between excess returns and volatility we can deduct that the volatility is a major driver for returns. For the five years, we calculated the correlation between deviation and excess returns to 0.82. In 2020 the spread between the two stocks was twice that of any normal year. In the high volatility covid-year, we managed an ROI of 10.4% which netted us an excess return

Year	Standard Deviation portfolio	Standard deviation FTSE	Beta
2017	0.0047	0.0049	-
2018	0.0035	0.0076	-
2019	0.0036	0.0070	-
2020	0.0063	0.0180	-
YTD	0.0018	0.0072	-
Whole period	0.0129	0.0101	-0.0080

Table 4: Yearly standard deviation of our portfolio and the FTSE all-share. Beta for the five year period

of 20.6%. This makes sense as a pairs trading strategy is most profitable when the stocks go haywire, and the spread enlarges. The covid crisis made for extreme scenarios where the market was at times very inefficient. In this case, Lukoil stock price seemingly recovered faster than Rosneft's (figure 8). This made for a large spread, which resulted in very profitable trades.

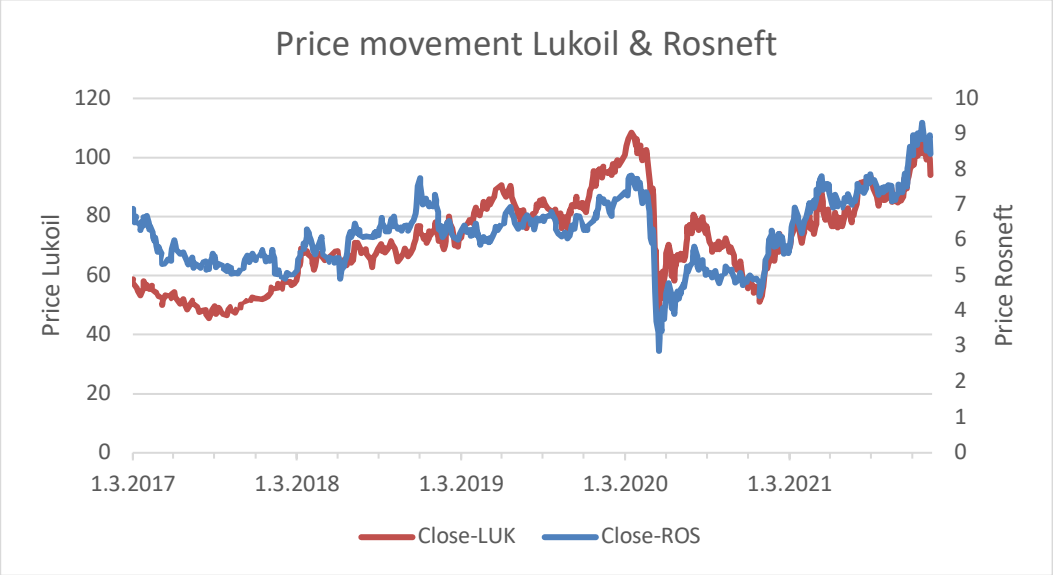


Figure 8: Price movement of the stocks throughout our active period

Interestingly, by P&L and ROI, this was not our top year. In 2019 we managed an ROI of 13.2% and an excess return of 7.18%. Our best year in terms of ROI. By looking at the trades, this was not the year where we had the greatest number of trades, nor the year with the longest total holding period, but it was the year with the longest individual trades. Meaning that each trade was held for a relatively long period of time, 53 days in average. If we look at figure 8, we can tell that there were two periods of large spreads. We don't know exactly what caused the spread, but we do know that they were excessive, causing our "algorithm" to trade.

Year	Number of trades	Total holding period days	Average holding period days
2017	4	85	21.25
2018	4	164	41
2019	3	159	53
2020	5	107	21.4
YTD	1	28	28
Result	17	543	32.93

Table 5: Number of trades and holding period

We also took a look at our portfolios performance in comparison to how liquid the market was. We see that two of our worst years, 2016 and 2017 were also the most liquid years. This makes sense as a liquid market is an efficient market. An illiquid market is an inefficient

market (Kapur, 2019). Pairs trading is a method of exploiting market inefficiencies, and when the market is very efficient, the method should be less profitable.

5.2 Market neutrality

Pairs trading is in theory supposed to be a market neutral trading strategy. We wanted to compare our portfolios movements with that of the benchmark, to reveal whether this was really the case in our period of testing. Using Beta, we can compare the two. The beta between our portfolio and the FTSE all-share turned out to be -0.008. This is a very low correlation and would indicate that our portfolio was in fact market neutral. However, we want to nuance this conclusion a little bit. Although the size of the spread does not follow the market as a whole, as we have seen, there is some correlation between volatile markets and our portfolios performance. Especially during 2020, the covid-year.

5.3 Strategy Weaknesses

To be able to construct and execute a pairs trading strategy in such a short timespan a few weaknesses are inevitable. Some of the most crucial weaknesses will be discussed in the following sections.

5.3.1 Is the strategy representative to evaluate pairs trading strategy?

Our strategy traded on a single relationship between two stocks. Thereby, we are limiting our risk to the development in the oil market and other factors related to the respective companies. It is difficult to conclude weather our strategy is convertible to other markets or a single struck of luck. An alternative method would be to create a portfolio with the same strategy in different markets. Such an approach would have given better insight regarding how robust our strategy is between markets with different conditions. Nevertheless, we are adequate with the result and see it as a great foundation for later expansions into other markets.

5.3.2 Data restrictions

Another restrain in our strategy is the amount of data that we traded on. Our dataset consisted of closing prices only. Every short or long position was a result of the prices at the end of the day, which usually differs from opening prices. Due to school restrictions on data accumulation, we were not able to acquire hourly prices on our stocks.

5.3.3 Capital cost

In theory out strategy could almost be capital neutral as the acquisition of the long position can be financed by the sale of the short position. Nevertheless, our strategy would require some additional funding. Thereby, borrowing and broker fees would lead to some capital requirements. Interest on these capital requirements is not taken into account and will make our strategy less profitable.

5.3.4 Attracting investors

Attracting investors is a crucial part when wanting to scale our strategy. However, some investors would likely be worried by the volatility and early years with losses and there is a chance they would have withdrawn their money. Consequently, we would have had to explain why we were losing money and how it would change in the future, but such a pitch is no long-term substitute for absent performance.

5.4 Revisiting the choice of arbitrage bounds

When optimizing our trading bounds in the beginning, we found two outstanding values (0.007 and 0.0043) of which we decided to pick the one with less trades to minimize transaction cost. Seeing as in some periods, like 2021 we spent a lot of time outside the market, we wanted to see what the other value with more narrow bounds would have yielded. We found that with a P&L of 33.23% (before TC) this parameter significantly outperformed our strategy, also totaling 35 trades instead of our 22.

Subsequently we ran a full-scale analysis for many parameter values.

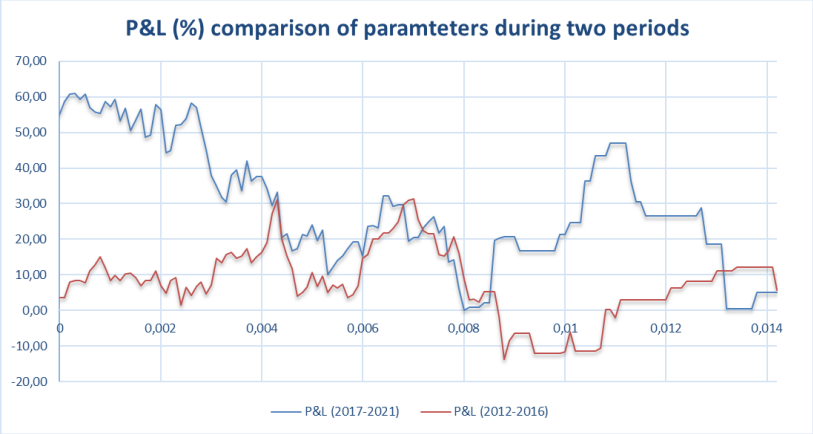


Figure 9: Comparison of parameter values

Interestingly, in our implementation period small values yielded significantly better returns than bigger ones and the strategy performed better on average over the parameters.

5.5 Further research possibilities

If one wanted to continue this project and elaborate on our current findings, we would suggest building a portfolio of many pairs trading strategies so that results are more representative for pairs trading in general and not only one single strategy. Furthermore, a portfolio whose value is tracked on a minutely basis, will allow for a more comprehensive study of the relationship between pairs trading and market conditions than our annual numbers.

6. Conclusion

The purpose of this paper has been to explore the conditions profitability of pairs trading. We started with presenting the theoretical background for arbitrage. Mathematical explanation and pairs trading were introduced in depth. Furthermore, we selected suitable pairs based on defined criteria. A pairs trading strategy was constructed based on data from 2012 until 2016. We then implemented our strategy for the period 2016 until 2021. Lastly, we evaluated the performance of the strategy.

To answer our thesis question; can we design a profitable pairs trading strategy and how is it affected by changing market conditions; we can divide the question into two. First looking at the profitability. The pairs trading strategy experienced losses the two first years. For the next

3 years the strategy was profitable and gave returns better than that of our benchmark, FTSE. The strategy performed well overall and netted us a total ROI of 18,45%. The second part of our thesis question is harder to answer precisely. We primarily defined market conditions as changes in volatility and liquidity. The covid-crisis made for a good period of testing, with extreme market conditions. A pairs trading strategy is in theory, supposed to be market neutral, but is most profitable when the markets are inefficient. We found that our portfolio had a very low correlation with the movements of the benchmark, but we were most profitable when the market went haywire. In this context, our strategy was positively affected by the market inefficiencies caused by the pandemic.

Nevertheless, the strategy does contain several weaknesses. Some of the most central weaknesses were that our strategy was based on companies that have survived. The market conditions have been beneficial for our testing period. The cost of capital is not included in our strategy. And lastly, our data is greatly restricted making us only react to closing prices. However, with limited time and resources we feel to have presented adequate results and believe to have a strong foundation for further developing a functional pairs trading strategy.

References:

Alliance Bernstein. (2012, January 01). Market-Neutral Strategies: An All-Weather Investment Option. Retrieved from Alliance Bernstein: https://www.alliancebernstein.com/CmsObjectABD/PDF/Research_WhitePaper/MarketNeutralWhitepaper06.08.12.pdf

Arbitrage (2021). Retrieved December 12, 2021, from Investopedia website: https://www.investopedia.com/terms/a/arbitrage.asp

Dunis, C., Ho, R. Cointegration portfolios of European equities for index tracking and market neutral strategies. *Journal of Asset Management.,* **6,** 33–52 (2005)

Engle, R. F. and Granger, C. W. J. (1987) 'Co-integration and Error Correction: Representation, Estimation and Testing', Econometrica, 55(2), 251–76.

Gatev, E., Goetzmann, W., & Rouwenhorst, G. (2006, Februrary 13). Pairs Trading: Performance of a Relative-Value Arbitrage Rule., *The Review of Financial Studies*, Volume 19, Issue 3, Fall 2006, Pages 797–827

Harris, L. (2002). Trading and Exchanges: Market Microstructure for Practitioners. New York, USA: Oxford University Press Inc.

Huberman, G. (2005). Arbitrage pricing theory (No. 216). Staff Report.

Interactive Brokers. *Commissions - Tiered | Interactive Brokers U.K. Limited*. (2020, January 11). IBKR. Retrieved November 25, 2021, from https://www.interactivebrokers.co.uk/en/index.php?f=39753&p=stocks2

Kapur, S. (2019), *Secrets of Sand Hill Road.* San Francisco, USA: Portfolio Publishing.

Krauss, C., Statistical Arbitrage Pairs Traiding Strategies: review and Outlook. Journal of Economic Surveys (2017) Vol. 31, No. 2, pp. 513–545.

Lukoil., (2016), Lukoil Annual Report 2016, https://www.lukoil.com/FileSystem/9/289049.pdf?dl=1. accessed: 20.11.2021

Naicker, S. (2015), *Evaluation of the performance of a pairs trading strategy of JSE listed firms*. Johannesburg: Faculty of Commerce. Retrieved [02.11.2021] from:

https://wiredspace.wits.ac.za/bitstream/handle/10539/21506/Thesis%20-%20Pairs%20Trading%20-%20Shreelin%20Naicker.pdf?isAllowed=y&sequence=1

Pairs Trading - A Real-World Guide - AlgoTrading101 Blog. (2020, September 24). Retrieved November 2, 2021, from Quantitative Trading Ideas and Guides - AlgoTrading101 Blog website: https://algotrading101.com/learn/pairs-trading-guide/

Rosneft., (2016), Rosneft Annual Report 2016. https://www.rosneft.com/upload/site2/document_file/a_report_2016_eng.pdf. accessed: 20.11.2021

White, H. (2000). A reality check for data snooping. *Econometrica*, *68*(5), 1097-1126.

Yahoo Finance (2021a), Rosneft Stock price history, available at: https://finance.yahoo.com/quote/ROSN.IL/history?p=ROSN.IL. accessed 17.11.2021

Yahoo Finance (2021b), Lukoil Stock price history, available at: https://finance.yahoo.com/quote/LKOD.IL?p=LKOD.IL&.tsrc=fin-srch. accessed 17.11.2021

Yahoo Finance (2021c), Gazprom stock price history, available at: https://finance.yahoo.com/quote/OGZD.IL?p=OGZD.IL&.tsrc=fin-srch. accessed 17.11.2021

Yahoo Finance (2021d), Novatek stock price history, available at: https://finance.yahoo.com/quote/NVTK.IL?p=NVTK.IL&.tsrc=fin-srch. accessed 17.11.2021

Appendix: R code

Script used to find the optimal p-value

```
backtest <- function(parameter, amountpertrade){


library(readxl)

TRADE <- read_excel("TRADE.xlsx")

#TRADE <- as.numeric(TRADE)
```

```
n <- as.numeric(nrow(TRADE))

open_trades <- 0

trades_made <- 0

total_PL <- 0


# to keep track of our trades, we're creating an empty dataframe

results1 = data.frame(matrix(

  vector(), 0, 12, dimnames=list(c(),
c("Trade_number","Date_opened","Date_closed","P1_Lukoil","P2_Lukoil","POS_Lukoil","
PL_Lukoil","P1_Rosneft","P2_Rosneft","POS_Rosneft","PL_Rosneft","Total_PL_trade"))),

  stringsAsFactors=F)


# based on our parameter value, add columns containing:

# 1: SMA + parameter

# 2: SMA - parameter


for(i in 1:n){

  TRADE$SMA_UP[i] <- (TRADE[i,5]+parameter)

  TRADE$SMA_DOWN[i] <- (TRADE[i,5]-parameter)

}

i <- 0

# now add another column containing a signal to open a trade, with:

# 1 when DIV > SMA_UP (meaning: short rosneft, buy lukoil)

# 2 when DIV < SMA_DOWN (meaning: buy rosneft, short lukoil)


for(i in 1:n){

  if(TRADE[i,4]>TRADE[i,6]){
```

```
    TRADE$SIGNAL[i] <- 1
  }
  else if(TRADE[i,4]<TRADE[i,7]){
    TRADE$SIGNAL[i] <- 2
  }
  else{
    TRADE$SIGNAL[i] <- 0
  }
}
i <- 0

# now let's generate signals to close our position
# we want to do this whenever DIV crosses the moving average

# first we create a column telling us whether DIV is below or above the MA

for(i in 1:n){
  if(TRADE[i,4] > TRADE[i,5]){
    TRADE$UPORDOWN[i] <- "ABOVE"
  }
  else if(TRADE[i,4] < TRADE[i,5]){
    TRADE$UPORDOWN[i] <- "DOWN"
  }
}
i <- 0
```

```
# then we use this column to generate a signal to sell our position
# everytime the value of this column changes

for(i in 1:n){
  if(i < 2){
    j <- 1
  }
  else{
    j <- i - 1
  }
  if(TRADE[i,9] != TRADE[j,9]){
    TRADE$CLOSEPOS[i] <- 1
  }
  else if(TRADE[i,9] == TRADE[j,9]){
    TRADE$CLOSEPOS[i] <- 0
  }
}
i <- 1

# now that our data is all set, we can simulate the actual trading

for(i in 1:n){
  if(TRADE[i,8] == 1 & open_trades == 0){
    trade_opened <- i
    pos_ros <- -1 #short
    pos_luk <- 1 #long
```

```
    open_trades <- 1 }
  if(TRADE[i,8] == 2 & open_trades == 0){
    trade_opened <- i # to get the relevant price later
    pos_ros <- 1 #long
    pos_luk <- -1 #short
    open_trades <- 1}
  if(TRADE[i,10] == 1 & open_trades == 1){
    trade_closed <- i # to get the relevant price later
    trades_made <- trades_made + 1 # just to keep track


    # adding the trade to our results table
    results1[trades_made,1] <- trades_made
    results1[trades_made,2] <- TRADE[trade_opened,1]
    results1[trades_made,3] <- TRADE[trade_closed,1]
    results1[trades_made,4] <- TRADE[trade_opened,3]
    results1[trades_made,5] <- TRADE[trade_closed,3]
    results1[trades_made,6] <- pos_luk
    results1[trades_made,7] <- as.numeric(pos_luk * ((TRADE[trade_closed,3]-
TRADE[trade_opened,3])/TRADE[trade_opened,3]) * amountpertrade)
    results1[trades_made,8] <- TRADE[trade_opened,2]
    results1[trades_made,9] <- TRADE[trade_closed,2]
    results1[trades_made,10] <- pos_ros
    results1[trades_made,11] <- as.numeric(pos_ros * ((TRADE[trade_closed,2]-
TRADE[trade_opened,2])/TRADE[trade_opened,2]) * amountpertrade)
    results1[trades_made,12] <- results1[trades_made,7] + results1[trades_made,11]
    # since we've now closed our position, we're resetting open_trades to 0
    open_trades <- 0
```

```
    }
}
a = as.numeric(nrow(results1))
for(b in 1:a){
results1$perc_gain[b] <- round((results1[b,12]/10000),4)
}
sum <- 0
for(b in 1:a){
  sum <- sum + results1[b,12]
}
# create a list to be able to return multiple outputs
output = data.frame(matrix(
  vector(), 0, 2, dimnames=list(c(), c("SUM","Trades_made"))),
  stringsAsFactors=F)

output[1,1] <- sum
output[1,2] <- trades_made

return(output)
}

# now that this is all working, time to search for the optimal bounds
# parameter values between 0 & 0.0212
# we're creating a table for our results
analysis = data.frame(matrix(
  vector(), 0, 3, dimnames=list(c(), c("Parameter_value","P&L","Trades_made"))),
```

```
  stringsAsFactors=F)


p = 0.0212
i = 1
while(p>0){
analysis[i,1] <- p
c <- backtest(p,100)
analysis[i,2] <- c[1,1]
analysis[i,3] <- c[1,2]
p <- p-0.0001
i <- i + 1
print(p)
}


##### MAKING GRAPHS ####


# graph 1: P&L plotted against trading bounds
library(ggplot2)


ggplot(analysis, aes(x=Parameter_value, y=P.L)) +
  geom_line(color="darkblue", linetype="solid") + ggtitle("4-year P&L (in %) for different
trading bound values") +
  xlab("Interval value for trading bounds") + ylab("4-year P&L (in %)") +
  theme(plot.title = element_text(color="red", size=14, face="bold.italic"),
  axis.title.x = element_text(color="blue", size=14, face="bold"),
  axis.title.y = element_text(color="#993333", size=14, face="bold")) +
  scale_size_manual(values=c(4, 6)) + theme_bw()
```

```
# graph 2: P&L (left), trades(right), plotted against trading bounds
ggplot(analysis, aes(x=Parameter_value)) +
  geom_line(aes(y = P.L), color = "darkred") +
  geom_line(aes(y = Trades_made), color="steelblue") +
  theme(plot.title = element_text(color="red", size=14, face="bold.italic"),
  axis.title.x = element_text(color="blue", size=14, face="bold"),
  axis.title.y = element_text(color="#993333", size=14, face="bold")) +
  scale_size_manual(values=c(4, 6)) + theme_bw() + ggtitle("P&L (%) & amount of trades made")
  + labels = c("P&L", "Amount of trades")
library("writexl")
write_xlsx(results,"C:/Users/stefa/OneDrive/Desktop/Trading & Market Microstructure//testB.xlsx")

results = data.frame(matrix(
  vector(), 0, 2, dimnames=list(c(), c("SL_Value","PL_Value"))),
  stringsAsFactors=F)
c <- 0
d <- 1
while(c<20.5){
```

Script used to simulate stop losses

```
library(readxl)
TRADE <- read_excel("TRADE.xlsx")
#TRADE <- as.numeric(TRADE)

n <- as.numeric(nrow(TRADE))
```

```
parameter <- 0.0070

amountpertrade <- 100

open_trades <- 0

trades_made <- 0

total_PL <- 0

SL_HIT <- 0

SL <- (-1)*c

# to keep track of our trades, we're creating an empty dataframe

results1 = data.frame(matrix(

  vector(), 0, 12, dimnames=list(c(),
c("Trade_number","Date_opened","Date_closed","P1_Lukoil","P2_Lukoil","POS_Lukoil","
PL_Lukoil","P1_Rosneft","P2_Rosneft","POS_Rosneft","PL_Rosneft","Total_PL_trade"))),

  stringsAsFactors=F)


# based on our parameter value, add columns containing:

# 1: SMA + parameter

# 2: SMA - parameter


for(i in 1:n){

  TRADE$SMA_UP[i] <- (TRADE[i,5]+parameter)

  TRADE$SMA_DOWN[i] <- (TRADE[i,5]-parameter)

}

i <- 0

# now add another column containing a signal to open a trade, with:

# 1 when DIV > SMA_UP (meaning: short rosneft, buy lukoil)

# 2 when DIV < SMA_DOWN (meaning: buy rosneft, short lukoil)
```

```
for(i in 1:n){
  if(TRADE[i,4]>TRADE[i,6]){
    TRADE$SIGNAL[i] <- 1
  }
  else if(TRADE[i,4]<TRADE[i,7]){
    TRADE$SIGNAL[i] <- 2
  }
  else{
    TRADE$SIGNAL[i] <- 0
  }
}
i <- 0

# now let's generate signals to close our position
# we want to do this whenever DIV crosses the moving average

# first we create a column telling us whether DIV is below or above the MA

for(i in 1:n){
  if(TRADE[i,4] > TRADE[i,5]){
    TRADE$UPORDOWN[i] <- "ABOVE"
  }
  else if(TRADE[i,4] < TRADE[i,5]){
    TRADE$UPORDOWN[i] <- "DOWN"
  }
}
```

```
i <- 0

# then we use this column to generate a signal to sell our position
# everytime the value of this column changes

for(i in 1:n){
  if(i < 2){
    j <- 1
  }
  else{
    j <- i - 1
  }
  if(TRADE[i,9] != TRADE[j,9]){
    TRADE$CLOSEPOS[i] <- 1
  }
  else if(TRADE[i,9] == TRADE[j,9]){
    TRADE$CLOSEPOS[i] <- 0
  }
}
i <- 0

# now that our data is all set, we can simulate the actual trading

for(i in 1:n){
  if(TRADE[i,8] == 1 & open_trades == 0){
    trade_opened <- i
```

```
    pos_ros <- -1 #short
    pos_luk <- 1 #long
    open_trades <- 1 }
  if(TRADE[i,8] == 2 & open_trades == 0){
    trade_opened <- i # to get the relevant price later
    pos_ros <- 1 #long
    pos_luk <- -1 #short
    open_trades <- 1}
  if(open_trades == 1 & SL_HIT == 0){
    trade_closed <- i
    PL_tempLUK <- as.numeric(pos_luk * ((TRADE[trade_closed,3]-
TRADE[trade_opened,3])/TRADE[trade_opened,3]) * amountpertrade)
    PL_tempROS <- as.numeric(pos_ros * ((TRADE[trade_closed,2]-
TRADE[trade_opened,2])/TRADE[trade_opened,2]) * amountpertrade)
    PL_tot <- PL_tempLUK + PL_tempROS
    if(PL_tot < SL){
      SL_HIT <- 1
      print("executing SL")
      print(TRADE[trade_opened,1])
      print(TRADE[i,1])
      print(trades_made + 1)
    }
  }
  # to close the trade
  if(open_trades == 1 & TRADE[i,10] == 1 | SL_HIT == 1){
    trade_closed <- i # to get the relevant price later
    trades_made <- trades_made + 1 # just to keep track
```

```
    # adding the trade to our results table
    results1[trades_made,1] <- trades_made
    results1[trades_made,2] <- TRADE[trade_opened,1]
    results1[trades_made,3] <- TRADE[trade_closed,1]
    results1[trades_made,4] <- TRADE[trade_opened,3]
    results1[trades_made,5] <- TRADE[trade_closed,3]
    results1[trades_made,6] <- pos_luk
    results1[trades_made,7] <- as.numeric(pos_luk * ((TRADE[trade_closed,3]-
TRADE[trade_opened,3])/TRADE[trade_opened,3]) * amountpertrade)
    results1[trades_made,8] <- TRADE[trade_opened,2]
    results1[trades_made,9] <- TRADE[trade_closed,2]
    results1[trades_made,10] <- pos_ros
    results1[trades_made,11] <- as.numeric(pos_ros * ((TRADE[trade_closed,2]-
TRADE[trade_opened,2])/TRADE[trade_opened,2]) * amountpertrade)
    results1[trades_made,12] <- results1[trades_made,7] + results1[trades_made,11]
    # since we've now closed our position, we're resetting open_trades to 0
    open_trades <- 0
    SL_HIT <- 0
   }
}
a <- as.numeric(nrow(results1))
sum <- 0
for(b in 1:a){
  sum <- sum + results1[b,12]}
results[d,1] <- SL
results[d,2] <- sum
```

```
print(c)

c <- c + 0.5

d <- d + 1

}
```