An algorithm-based statistical arbitrage high frequency trading system to forecast natural gas futures prices

Saulius Masteika, Ph.D.
Department of Informatics
Kaunas Faculty of Humanities
Vilnius University,
Muitines str. 8, Kaunas, LT–44280
Lithuania
E-mail: saulius.masteika@khf.vu.lt

Saulius Masteika is an associated prof. at Vilnius University Kaunas Faculty of Humanities (Lithuania). He teaches courses related to economic modelling, intellectual systems and technical analysis in financial markets. His research interests relate to quantitative finance, behavioural finance, high frequency trading, crypto currencies, commodity markets and derivatives.

Mantas Vaitonis, Student
Department of Informatics
Kaunas Faculty of Humanities
Vilnius University,
Muitines str. 8, Kaunas, LT–44280
Lithuania
E-mail: mantas.vaitonis@khf.stud.vu.lt

Mantas Vaitonis is a master degree student at Vilnius University Kaunas Faculty of Humanities (Lithuania). He is interested in computational finance, high frequency trading and technologies of high performance computing.

Virgilijus Sakalauskas, Ph.D.
Department of Informatics
Kaunas Faculty of Humanities
Vilnius University,
Muitines str. 8, Kaunas, LT–44280
Lithuania
E-mail: virgilijus.sakalauskas@khf.vu.lt

Virgilijus Sakalauskas is a professor of Informatics at Vilnius University (Lithuania). His current research interests include application of artificial intelligence methods in research of the financial and currency markets, e-banking, e-government, e-learning, evaluation of risk and management of knowledge exchange processes. Virgilijus Sakalauskas has experience as the program committee co-chair and organizer of the widely known scientific workshops and conferences.

Kęstutis Driaunys, Ph.D.
Department of Informatics
Kaunas Faculty of Humanities
Vilnius University,
Muitines str. 8, Kaunas, LT–44280
Lithuania
E-mail: kestutis.driaunys@khf.vu.lt

Kęstutis Driaunys is an associated prof. at Vilnius University Kaunas Faculty of Humanities (Lithuania). His research interests relate to artificial intelligence, algorithmic trading, cyber security, data protection and speech recognition. He teaches courses related to computational intelligence at Vilnius University.

An algorithm-based statistical arbitrage high frequency trading system to forecast natural gas futures prices

Abstract:
Professional fund managers, investment banks and regulatory authorities raise the question about the impact of algorithmic trading to trading businesses, economy and market efficiency. The questions arise if high frequency trading (HFT) provides more efficient, liquid markets and is economically beneficial. In this paper an algorithm based on statistical arbitrage is tested. Statistical arbitrage is a well-known trading strategy where profit arises from pricing inefficiencies between correlated financial instruments. An algorithm of statistical arbitrage tries to find a pair of correlated instruments that move together and take long/short positions when they diverge abnormally, hoping that the prices will converge in the near future. Recent computational expansion in financial modeling and the ever increasing demand for order execution speed is moving market making and price discovery strategies into high frequency trading (HFT) or the milliseconds realm, where HFT already generates nearly 2/3 of the overall trading volume. In this report we apply high frequency data from NYMEX exchange to test a trading system based on statistical arbitrage in one of the most liquid futures market, i.e. natural gas futures. The overall results suggest that statistical arbitrage in HFT environment significantly outperforms traditional trading strategies, provides liquidity to the markets and denies the efficient market hypothesis.

Key words: efficient market hypothesis, high frequency trading, statistical arbitrage, pairs trading, futures market, algorithmic trading.

JEL classification: G14, G17, C63, B26

Statistinio arbitražo algoritminė didelio dažnio prekybos sistema gamtinių dujų ateities sandorių kainos pokyčio prognozavimui

SANTRAUKA

Profesionalūs fondų valdytojai, investiciniai bankai bei vertybinių popierių biržų priežiūros institucijos siekia išsiaiškinti kokį poveikį verslui, ekonomikai ir rinkų efektyvumui daro algoritminė prekyba finansų rinkose. Keliamas klausimas ar didelio dažnio prekybos sistemos (HFT) prisideda prie likvidesnių bei ekonomikai naudingų rinkų formavimosi. Šiame darbe yra tiriamas prekybos algoritmas pagrįstas statistiniu arbitražu. Trumpai apibūdinant galima teigti, kad statistinio arbitražo algoritmas identifikuoja koreliuojančias vertybinių popierių poras bei ieško nukrypimų tarp šių vertybinių popierių kainų pokyčių. Identifikavus nukrypimus atidaromos priešingos viena kitai pozicijos, t.y. pirkimo ir pardavimo trumpam sandoriai, tikintis, kad atstumas tarp kainų grįš į ankstesnį, istoriškai nusistovėjusį koreliacinį lygį. Pastarojo meto kompiuterinių technologijų plėtra biržinėje prekyboje bei nuolatinė konkurencija dėl spartesnio sandorio įvykdymo greičio, taikant statistinio arbitražo strategijas, sąlygojo didelio dažnio prekybos sistemų (HFT) kūrimąsi bei milisekundinių sandorių plėtrą. Šiuo metu HFT prekyba generuoja beveik du trečdalius visos rinkos apyvartos. Visa tai lėmė šio tyrimo aktualumą. Tyrime naudoti HFT gamtinių dujų ateities sandorių, vienų iš likvidžiausių finansinių instrumentų, istoriniai duomenys. Duomenys imti iš Niujorko prekių biržos- NYMEX. Apibendrinant atliktus tyrimus galima teigti, kad statistinio arbitražo strategija HFT aplinkoje leidžia pasiekti teigiamų investicinių rezultatų, paneigia efektyvios rinkos hipotezę bei duoda pridėtinę vertę ekonomikoje.

Raktiniai žodžiai: efektyvios rinkos hipotezė, statistinis arbitražas, koreliuojančių porų prekyba, ateities sandorių rinka, automatinės prekybos sistemos.

An algorithm-based statistical arbitrage high frequency trading system to forecast natural gas futures prices

Introduction

One of the least discussed methods on the science base of financial engineering is the high frequency trading (HFT). HFT has taken place after the U.S. Securities and Exchange Commission (SEC) authorized electronic exchange in 1998 but the subject was brought to deeper analysis just round 2010. There is no clear definition of HFT but in the most recent review prepared by the staff of SEC, HFT is attributed as using extraordinarily high speed and sophisticated programs for generating orders and ending the trading day in a flat position (that is, not carrying unhedged positions overnight) (SEC, 2014). Statistical information reveals that high frequency trading covers more than a half of the overall trading volume in the markets, and market share of HFT is still further growing (Zubulake&Lee, 2011; SEC, 2014). Due to the fact that in order to work with HFT, one has to be able to respond to market changes in milli- or even microseconds, HFT is exclusively done by algorithmic robots and also computers. Previous empirical researches suggest that HFT and increased low-latency activity improve traditional market quality, like decreasing spreads, increasing displayed depth in the limit order book, and lowering short-term volatility (Hasbrouck&Saar, 2013). A recent report by the HFT leader in the market reveals great importance of algorithmic trading, as filed with the Securities and Exchange Commission, Virtu Financial, Inc. was able to perform only with losing one day out of 1238 days using electronic trading strategies (Cifu, 2014). Therefore, following the trend of HFT expansion, SEC's reports, and also considering that HFT and other forms of low latency or algorithmic trading strategies continue to be in the focus of foreign regulators, Central banks or Securities and Exchange Commissions (CFTC&SEC, 2011; Brogaard *et al.*, 2013), a deeper look at HFT from academic perspective was chosen, which will hopefully bring some more clarification on the issue as well.

Algorithmic trading incorporates various trading strategies related to statistical arbitrage and pair trading, and is applied in different markets (Nath, 2003; Perlin, 2009; Sakalauskas&Kriksciuniene, 2012; Carrion, 2013; Hendershott&Riordan, 2012; Miao, 2014; Cartea&Penalva, 2012; Menkveld, 2013). Nath implemented simple pairs trading strategy with automatic extreme risk control using the entire universe of securities in the highly liquid secondary market for U.S. government debt (Nath, 2003). The main objective of Perlin research was to verify the performance and risk of pairs trading in the Brazilian financial market for different frequencies of the database (Perlin, 2009). Sakalauskas and Kriksciuniene investigate if the changes of financial time series can be explained by increase of intensity of flow of market news. The paper by Carrion provides evidence regarding successful HFT trading performance using a sample of NASDAQ trades and quotes that directly identifies HFT participation (Carrion, 2013). Hendershott and Riordan examine the role of algorithmic traders in liquidity supply and demand in the 30 Deutscher Aktien Index stocks on the Deutsche Boerse. It was found that algorithmic traders consume liquidity when the bid-ask quotes are narrow and supply liquidity when it is wide (Hendershott&Riordan, 2012). Miao proposed a high frequency and dynamic pairs trading system based on a market-neutral statistical arbitrage strategy using a two-stage correlation and co-integration approach. The proposed pairs trading system was applied to equity trading in U.S. equity markets. It was concluded by Miao that, overall, the system exceeded the S&P500 index performance; it is also worth noticing that it performed well during the two months in which the S&P 500 index had negative returns (Miao, 2014). Cartea and Penalva analyzed the impact of HFT in financial markets. They found that the price impact of liquidity trades is higher in the presence of the HFT and is increasing with the size of the trade. Some negative side effects of HFT are mentioned in Cartea and Penalva paper arguing that professional traders lose revenue in every trade intermediated by HFT and, because of HFT, volume of trades increases (Cartea&Penalva, 2012). Menkveld in his paper characterized the trading strategy of a large high-frequency trader and concluded that HFT trader incurs a loss on its inventory but earns a profit on the bid-ask spread. The study showed that performance of the strategy is very sensitive to the cost of capital assumptions. In the research HFT employs a cross-market strategy where a half of its trades materialize on a large incumbent market and the other half on a small, high-growth entrant market (Menkveld, 2013). This review of HFT strategies has revealed that dominant HFT strategies contribute to market liquidity, i.e. market making strategies (Menkveld, 2013; Gomber et al., 2011; Hendershott&Riordan, 2012; Zubulake&Lee, 2011), and to price discovery and market efficiency, i.e. arbitrage strategies (Miao, 2014; Aldridge, 2013; Nath, 2003; Perlin, 2009; Hanson&Hall, 2012). The shortage of hard scientific evidences about the profitability of the strategies mentioned above in HFT environment was a driving force for this study. In summary, HFT experimental study in futures market was chosen for the following reasons:

1. Concentration on one commodity gives ensured correlation between a trading pair, i.e. contracts with different expiration dates of the same underlying commodity are likely respond to the new upcoming information
2. High liquidity of high frequency environment and no overnight positions reducing the risks of market sudden spikes when the markets are closed
3. The most recent data in milliseconds used for trading what makes the system more reliable, since the speed of making a trade in statistical arbitrage is one of the most essential factors to be profitable
4. Exchanges with electronic order handling structure where the biggest part of trading volume is generated in HFT by algorithmic machines;

In this paper, the system related to market efficiency or price discovery and statistical arbitrage was chosen for a deeper analysis. Contribution of this paper lies in testing statistical arbitrage strategy in HFT environment with the Natural gas futures contracts, according to the statistics of futures industry, one of the most liquid instruments in global energy market (Acworth, 2011).

1. Statistical arbitrage and pairs trading

The roots of statistical arbitrage can be traced back to the first hedge funds, i.e. around 1950, running statistical arbitrage strategies using mathematical models to find pricing inefficiencies, where long and short positions helped to reduce market risks (Ferguson&Laster, 2007). The hedged portfolios as well as statistical arbitrage are profitable when long position earns more or loses less than the short position. It is worth mentioning that before advancements in computational science arbitrageurs managing large positions had risk-free arbitrage opportunities based on actual pricing flaws or price change delays between different correlated markets. Today the technological aspect is far more important than just the ability to trade big positions and have parallel access to correlated markets. The markets are electronic, trade in milliseconds; and more market participants have real-time access to market data feeds. In summary, technological advances moved statistical arbitrage into high frequency trading managed purely by high-tech computers and robotic algorithms.

Pair's trading is one of the most common strategies of statistical arbitrage and has been widely used by professional traders, institutional investors, and hedge fund managers since 1980 (Vidyamurthy, 2004; Dunis *et al.*, 2010; Gatev *et al.*, 2006; Hogan *et al.,* 2004). Historically, pair's trading is a trading strategy taking advantage of market inefficiencies based on a pair of stocks. The perception is to identify two stocks that move together and to take long and short positions simultaneously when they diverge abnormally (Miao, 2014; Elliot *et al.*, 2005). It is expected that the prices of the two stocks will converge to a mean in the future (Caldeira&Moura, 2013; Perlin, 2009). An algorithm of statistical arbitrage in futures market also needs to take a pair of correlated assets. The ones with the highest correlation are futures contracts of different delivery months (Masteika&Rutkauskas, 2012). It was decided to focus on statistical arbitrage or pairs trading strategy based on M. S. Perlin previous research (Perlin, 2009). Similar researches have been tested and also generated some promising results in Nath and Miao studies (Nath, 2003; Miao, 2014). Before strategies were applied they were adapted to fit the author's needs; we did the same adapting the strategy to futures market and HFT environment in order to work out better market trading results and adapt to technological advances. The following chapter presents an algorithm for statistical arbitrage in more detail.

2. Statistical arbitrage in HFT

An algorithm for statistical arbitrage finds a pair of correlated financial instruments that throughout history have followed each other, and then wait till the spread between them widens. When the widening happens an algorithm bets for the financial instruments to return to their historical spread. In this research it was chosen to trade future's contracts of the same underlying asset, i.e. natural gas. Natural gas future's contracts with different expiration dates are highly correlated and liquid to trade, giving a sound reason to choose them for this research.

3. Methodology of the automated trading system

An algorithm is constructed from the following steps:

1. High frequency data normalization
2. Selection of a correlated pair
3. Defining a moving window for trading and data normalization
4. Setting triggers for a long/short positions
5. Performance assessment of the trading system

In the following chapters all the steps are discussed in more detail.

3.1 High frequency data normalization

An algorithmic trading system needs HFT data to be normalized. The main problem with HFT data arises due to the discrepancies between time stamps of correlated contracts. The time stamp problem can be seen in *Table 1* and *Table 2*. The tables are provided below.

Table 1 HFT data example for NGV12 contract

Futures Contract	*Day*	*Time stamp*	*Operation*	*Trade Price*	*Trade Size*	*Bid Price*	*Bid Size*	*Ask Price*	*Ask Size*
NGV12	09/05/2012	18:00:00.128	Quote					2.8	5
NGV12	09/05/2012	18:00:00.128	Quote			2.799	13		
NGV12	09/05/2012	18:00:00.128	Quote					2.8	6
NGV12	09/05/2012	18:00:00.161	Quote					2.8	7

NGV12	09/05/2012	18:00:00.634	Trade	2.799	1				
NGV12	09/05/2012	18:00:00.634	Trade	2.799	1				
NGV12	09/05/2012	18:00:00.634	Trade	2.799	2				
NGV12	09/05/2012	18:00:00.634	Quote			2.799	9		
NGV12	09/05/2012	18:00:00.634	Quote					2.8	8
NGV12	09/05/2012	18:00:00.634	Trade	2.799	3				
NGV12	09/05/2012	18:00:00.634	Trade	2.799	1				
NGV12	09/05/2012	18:00:00.634	Trade	2.799	1				
NGV12	09/05/2012	18:00:00.634	Quote			2.799	4		
NGV12	09/05/2012	18:00:00.634	Quote					2.8	11

Table 2 HFT data example for NGX12 contract

Futures Contract	*Day*	*Time stamp*	*Operation*	*Trade Price*	*Trade Size*	*Bid Price*	*Bid Size*	*Ask Price*	*Ask Size*
NGX12	09/05/2012	18:00:00.129	Quote					2.942	17
NGX12	09/05/2012	18:00:00.129	Quote			2.94	18		
NGX12	09/05/2012	18:00:00.138	Quote					2.942	18
NGX12	09/05/2012	18:00:00.150	Quote					2.942	19
NGX12	09/05/2012	18:00:00.494	Quote			2.94	19		
NGX12	09/05/2012	18:00:00.612	Quote			2.94	15		
NGX12	09/05/2012	18:00:00.612	Quote					2.942	20
NGX12	09/05/2012	18:00:00.622	Quote			2.94	14		
NGX12	09/05/2012	18:00:00.622	Quote			2.94	9		
NGX12	09/05/2012	18:00:00.622	Quote					2.942	23
NGX12	09/05/2012	18:00:00.622	Quote			2.94	7		
NGX12	09/05/2012	18:00:00.622	Quote					2.942	24
NGX12	09/05/2012	18:00:00.622	Quote			2.94	6		
NGX12	09/05/2012	18:00:00.622	Quote					2.942	25

Table 1 and *Table 2* show that *Time stamps* for trades or bid/ask changes during the same trading second differ, e.g. in *Table 1* there is a *Time stamp* of *Quote* for NGV12 contract change at *18:00:00.161* while the other contract (NGX12) has *Quote* changes just at milliseconds *18:00:00.150* and *18:00:00.494*, and no activity at millisecond *18:00:00.161*. NGV12 is a natural gas futures contract for October, 2012 and NGX12 is November, 2012 contract (Masteika et al., 2012). Therefore, a comparison of time stamp sequences of correlated financial instruments is needed. In case time stamps are different, the contract which does not have the same time stamp as the other is filled in with missing time stamp and the previous bid/ask or trade prices. Such a fill brings time stamps of correlated contracts together, at the same keeping the prices accurate.

High frequency data normalization is also important for recalculating the prices of correlated contracts to a particular unit in order to remove the noise of price changes and compare the changes more qualitatively. The normalization was done as follows: for each contract i, price $P(i,t)$, empirical mean $\mu(i,t)$ and standard deviation $\sigma(i,t)$ were calculated, and then the equation (1) was applied. The method was also applied by M.Perlin (Perlin, 2009) in his research related to pairs trading in Brazilian financial market.

$$p(i,t) = \frac{P(i,t)-\mu(i,t)}{\sigma(i,t)}, \qquad (1)$$

where the value $p(i,t)$ represents the normalized price of asset i at time t. Normalized prices were calculated within a particular moving window of the time series which is going to be discussed in more detail in Chapter 3.3. All prices of correlated contracts are transformed into normalized units when using the equation (1).

3.2 Selection of the correlated pair

The most liquid, close to expiration and therefore correlated contracts of natural gas futures were selected for an algorithmic testing. Natural gas futures contract is a standardized contract between two parties to exchange natural gas of standardized quantity and quality for the price agreed today - the future's price, but with delivery occurring at a specified future date - delivery date. At any stage, the market consists of a number of contracts, or "delivery months", that have expiration dates stretching out into the future (Masteika et al., 2013). In this research natural gas October, 2012 and November, 2012 contracts were chosen. Successive contracts of the same underlying asset were selected to construct a correlative pair of a commodity.

3.3 Defining moving window for trading and data normalization

Moving window for trading and data normalization is necessary to be defined in order for an algorithm to monitor the changes in behavior of the correlated contracts. The size of moving window is a matter of empirical research and is rather sensitive because it can cause the algorithm to over-trade or miss the abnormal behavior of the correlated contracts. The size of moving window depends on what sensitivity of the system is needed. For example, longer moving window causes lower sensitivity meaning that small abnormal behavior of the changes in price will not even be noticed. However, if shorter moving window is chosen, even the smallest noise will be noticed and trading signals triggered.

3.4 Setting triggers for long/short positions

Trading signal is going to be triggered every time the value of the spread of the normalized pair's prices is higher than a given predefined threshold d. The value of d is arbitrary and works as a filter for creating trading signals. One has to choose d very carefully because it cannot be too high, in which case only a few trading signals would be created, and cannot be too low, because too many trading signals would be created, which would result in high transaction costs.

According to the strategy introduced by M. Perlin it is necessary to establish which position for which asset should be taken (long or short). Let's say one side of a pair price is $A(i,t)$ and the other $B(i,t)$. If the value of $A(i,t)$ is abnormally higher than $B(i,t)$, then a short position is kept for stock $A(i,t)$, and a long position is made for the pair of stock $B(i,t)$, else do the opposite. Such position is kept till the value of the spread becomes smaller than the threshold d (Perlin, 2009). A fragment of an algorithm to trigger a trade is presented below:

while d > spread
if $A(i,t)$>$B(i.t)$ then short $A(i,t)$ and long $B(i,t)$
else long $A(i,t)$ and short $B(i,t)$

However, in this experimental research only bid prices of one contract and ask prices of the other were applied. Thus, whenever the distance of the given d is reached, the contract with bid prices was shorted and long position of the contract with the ask price was taken. Market orders were taken in order to avoid price slippage discrepancies in experimental research.

The logic behind such an algorithm is that when the distance between correlated contracts composing a pair is too large, there is a good probability that such prices are going to converge in the future, and this logic can be explored for profit purposes.

3.5 Performance assessment of the trading system

One of the main objectives of this paper is to test and assess the performance of the trading system based on statistical arbitrage in high frequency trading environment. The trading system's returns were assessed by calculating the price differences, while trading correlated commodity contracts during the given period of time. After the trading period came to an end, the array with the prices of opening and closing positions were used to calculate the profit/loss of that period. Profits from long (2) and short (3) positions were calculated by applying the following equations:

$$PL = \sum_{i=1}^{K}(ls(i) - lb(i) - C \qquad (2)$$

$$PS = \sum_{i=1}^{K}(ss(i) - sb(i) - C \qquad (3)$$

The variable *PL* represents a profit from a long position and *PS* - a profit from a short position. Variable *i* identifies the trade for which the profit or loss is calculated. It means that the profit from long position equals to the difference between asset i sell - *ls*, and buy - *lb* values, multiplied by *K* (number of contracts), minus *C* (transaction cost); see equation (2). The profit from short position equals to the difference of assets i sell – *ss*, and buy - *sb*, multiplied by *K* (number of contracts), minus C (transaction cost); see equation (3). The total profit was calculated by applying equation (4):

$$TP = PL + PS \qquad (4)$$

All the calculations were implemented in numerical computing language *Matlab*, software developed by MathWorks, Inc.

4. Database of the research

High frequency trading quotes and trade prices for natural gas October and November futures' contracts were taken from NYMEX Mercantile exchange. High frequency data from 2012-09-04 till 2012-09-20 was applied for the study. Time records were in milliseconds. Hence, statistical arbitrage strategy was back tested with high frequency data unlike in previous studies, using daily, hourly or minutely data sets. An average number of normalized records were approx. 350000 per day.

Trading *volume*, *Bid* size and *Ask* size were recorded as a number of contracts. High frequency data matrix was generated from bid/ask prices and time records.

In order to trade HFT statistical arbitrage in a liquid market, the most active trading period during the day was calculated. An average of trades per every 15 minutes was calculated during a period of one month. The most active trading periods for NGV and NGX contracts during the day are presented in Figure 1.

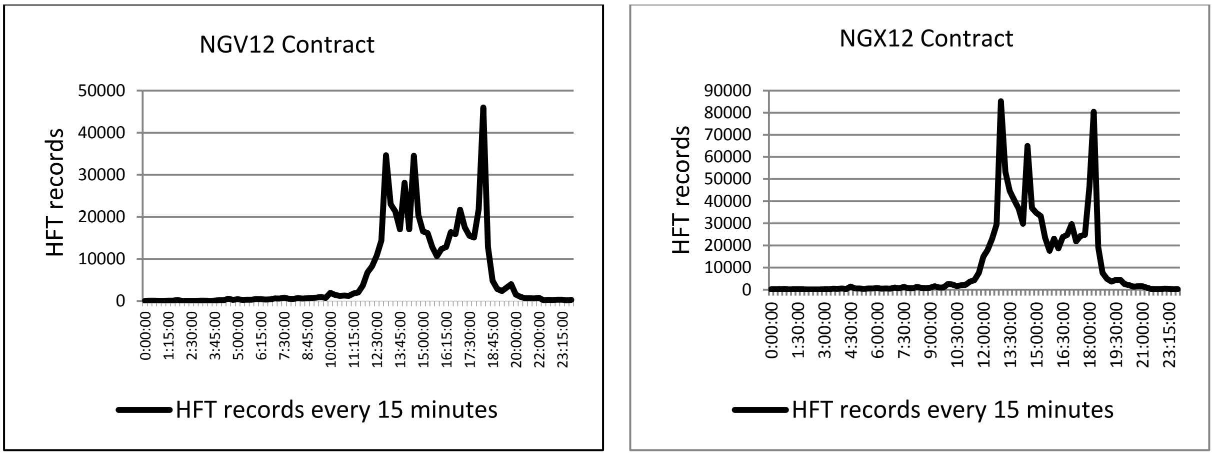

Figure 1. NGV and NGX trading activity during a day (every 15 minutes)

X axis shows time periods in quarters of an hour and *Y* axis shows the number of trades every 15 minutes. Figure 1 reveals that the most active and therefore best fitted period for high frequency trading during a day is the period between 11:15:00 and 18:45:00. Historical HF data from these periods were taken for the further data normalization which is more thoroughly explained above in Chapter 3.1.

5. Experimental results

A number of experiments were carried out in order to find the best parameters of the trading system. The following tables represent the best cases applying different parameters. *Table 3* shows the results of the system with the following parameters: moving window *(MW)* for trading and data normalization is set to 100, a trigger coefficient or the spread of the normalized pair's prices *d* is set to 5, maximal period for keeping an open position *(MP)* is 1000, position size per trade is set to 10 contracts. There is no clear pricing for high frequency trading since HFT is often considered as a market making or liquidity providing strategy run by firms acquiring rebates and discounts or even premiums from exchanges and market brokers for liquidity provision. Therefore, for experimental purposes transaction cost is set to zero.

Table 3. Experimental results when *MW*=100; *d*=5; and *MP*=1000

Trading day	*Records*	*Trades*	*Profit*	*Standard deviation*
2012-09-04	429757	1702	0.13	0.01
2012-09-05	396571	2702	0.09	0.01
2012-09-06	371582	1794	0.13	0.02
2012-09-07	253560	1870	0.3	0.03
2012-09-09	4824	24	0.01	0.03
2012-09-10	461334	3324	0.41	0.03
2012-09-11	475755	2852	0.25	0.02
2012-09-12	505672	1852	0.14	0.02
2012-09-13	474112	1594	0.37	0.03
2012-09-14	339556	1372	0.15	0.03
2012-09-16	4550	22	0	0
2012-09-17	416413	1556	0.22	0.02
2012-09-18	192279	800	-0.29	0.18
2012-09-19	323962	506	-0.02	0.09
2012-09-20	393060	949	0.24	0.02

Table 3 shows that throughout the period of experimental research, the HFT system based on statistical arbitrage, where *MW*=100; *d*=5; and *MP*=1000 generated negative results on 18th and 19th of September. *Table 4* shows the results when *MW* is shortened and set to 50. All other parameters were left unchanged.

Table 4. Experimental results when *MW*=50; *d*=5; and *MP*=1000

Trading day	*Records*	*Trades*	*Profit*	*Standard deviation*
2012-09-04	429757	3162	0.12	0.02
2012-09-05	396571	5340	0.37	0.01
2012-09-06	371582	4050	0.17	0.01
2012-09-07	253560	3424	0.19	0.02
2012-09-09	4824	54	0.03	0.03
2012-09-10	461334	7200	0.41	0.02
2012-09-11	475755	6342	0.39	0.02
2012-09-12	505672	4690	0.2	0.01
2012-09-13	474112	4654	0.25	0.01
2012-09-14	339556	3788	0.12	0.02
2012-09-16	4550	44	0.01	0.06
2012-09-17	416413	2332	0.26	0.01
2012-09-18	192279	1822	0.07	0.01
2012-09-19	323962	2670	-0.05	0.04
2012-09-20	393060	4556	0.25	0.01

Table 4 reveals that the system with shortened *MW* parameter still generates some negative results, i.e. -0.05 loss on 19th of September. Additional experimental studies with the parameters revealed that positive results during all of the period were generated with the following parameters: *MW*=50; *d*=6.5; *MP*=1000, see *Table 5*.

Table 4. Experimental results when *MW*=50; *d*=6.5; and *MP*=1000

Trading day	*Records*	*Trades*	*Profit*	*Standard deviation*
2012-09-04	429757	2146	0.16	0.02
2012-09-05	396571	4008	0.34	0.01
2012-09-06	371582	3302	0.17	0.01
2012-09-07	253560	2302	0.23	0.02
2012-09-09	4824	12	0	0
2012-09-10	461334	4784	0.35	0.02
2012-09-11	475755	4892	0.18	0.01
2012-09-12	505672	2122	0.10	0.01
2012-09-13	474112	2422	0.12	0.01
2012-09-14	339556	3274	0.05	0.01
2012-09-16	4550	16	0.01	0.03
2012-09-17	416413	4298	0.21	0.01
2012-09-18	192279	1400	0.06	0.01
2012-09-19	323962	2142	0.09	0.01
2012-09-20	393060	3560	0.17	0.01

Table 5 above shows the daily profits from HFT system and confirms the results revealed in IPO prospectus by High Frequency Trading market leader Virtu Financial, Inc, where only one losing trading day out of 1237 days was generated (Cifu, 2014). The chart in *Figure 2* illustrates daily results of an algorithm-based on a statistical arbitrage HFT system.

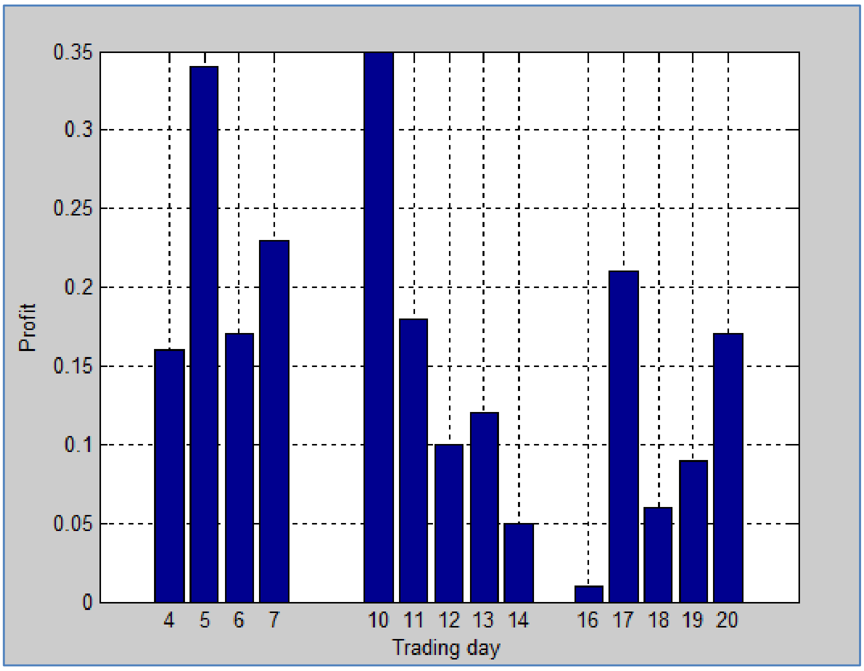

Figure 2. Daily HFT system results

The best results were obtained with the lower values of a moving window for trading and data normalization *(MW)* and higher values of trigger coefficient *d*. The following charts in *Figure 3* illustrate an algorithmic HFT performance during the day. Charts above represent the most profitable days of the research period, and the ones below - the least profitable days.

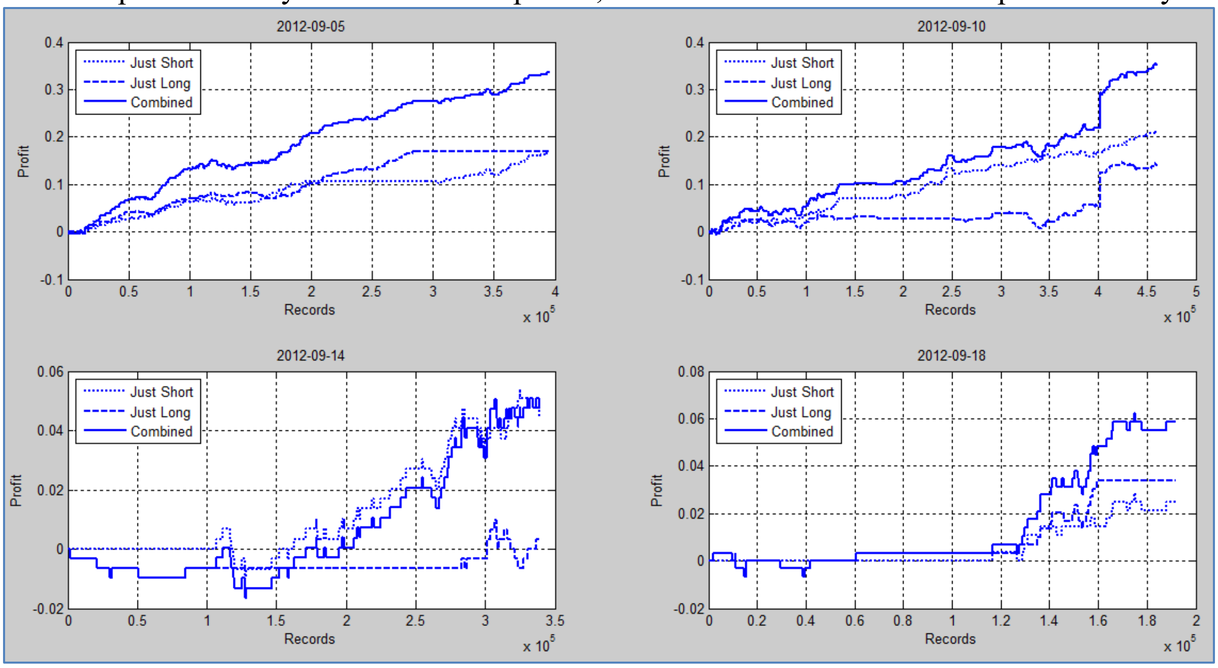

Figure 3. Intraday results of the most and least profitable days

Figure 3 gives us a broader view on how the system works within a trading day. The charts show rather small fluctuations of the profit having low standard deviation without respect to whether a particular day was more appropriate for the system or not. The charts reveal that less profitable days occur because of fewer trades rather than wide fluctuations or a series of unproductive trades, which increases the value of the system since low fluctuations in profit changes, is the most important goal of all trading systems.

In summary, the results are promising but differ from the ones obtained by M. Perlin, who suggest that the use of lower *d* gives higher profitability (Perlin, 2009). We think that the reason HFT system based on statistical arbitrage gives different results if compared with intraday or end of day trading is because of high frequency data specifics, where a considerably increased number of records are encountered during the same period of time. Such a huge number of records in HFT raise a possibility to trade more frequently and quicker if compared to other market strategies, which also determines noisier environment and demand for a less sensitive triggering of the trades, i.e. a higher value of the coefficient *d*. The higher values of *d* generate fewer trades and filter noisy triggers making the system more stable. Another reason higher value of *d* works better is because of the predefined and fixed trading pair, where no swaps between various contracts are implemented as in previous researches. High frequency data in milliseconds generates enough trading opportunities for statistical arbitrage, and no pair changes were implemented. In the previous research M. Perlin was also able to apply bidirectional trading changing short and long positions on different pairs. However, in this research a fixed pair trading was implemented using *bid* prices of one contract and *asks* of the other.

Conclusions

Recent technological advances have made trading in the markets electronic, fast, and algorithmic. Instead of humans, computers replicate the role of market makers, specialists or liquidity providers but at a much higher rate of speed. The number of derived financial instruments and an increased interconnectedness between the markets has caused increased opportunities for profits arising from pricing inefficiencies or price move delays between securities, or in other words for strategies based on statistical arbitrage. These factors were driving forces to test the system based on statistical arbitrage in HFT as a leading environment of modern investing.

One of the most liquid markets, i.e. Natural gas futures, was taken as a research object. Historical HFT data from NYMEX exchange was applied in experimental calculations. In the aim of the research to investigate the efficiency of the proposed HFT system based on an algorithmic statistical arbitrage, normalization of the data of two correlated futures contracts of the same underlying asset was tested. The proposed solution did fill the mismatched lines of the data streams allowing measurement of spread changes between the contracts.

The main conclusion of the paper is that statistical arbitrage has positive results when applied to HFT in natural gas futures market. As well as in *end of day* and *intraday* trading, statistical arbitrage can also be applied in HFT environment. Experimental studies revealed that HFT generates noisier data streams and hence demand for a less sensitive triggering of the trades, i.e. a higher value of the coefficient *d*. The experiment has also showed that a fixed pair of financial instruments can be applied in statistical arbitrage and generate stable profits.

In summary, any trading system will be challenged against the efficient market hypothesis claiming that market prices already incorporate all the information needed and that beating the market is nearly impossible. However, this study is one more denial of the hypothesis and has proved the efficiency of trading algorithms and systems, at least the ones employing extremely short holding periods, algorithmic computation and electronic market infrastructure. Nevertheless, the proposed HFT system needs additional analysis in real time HFT environment before it can be applied as a trading tool. Therefore, the results can be a starting point for further research among businesses or financial institutions being able to collect the additional funds necessary for real time HFT data feeds and microstructure.

Further development of the system could involve same strategy studies with a larger sample of values of parameters *d*, *MW*, and *MP*. In order to increase the performance of the system, additional studies are also recommended with different sectors of futures market, like grains, metals, financials, currencies; bidirectional arbitrage testing; pair selection, and costs calculation for liquidity providing and HFT infrastructure. The conclusion can be made that statistical arbitrage and high frequency trading gives positive results; generates aditional liquidity to the markets and economy and can be attractive to market infrastructure developers and profesional market participants who implement the technologies of quantitative finance and financial modeling in their business.

References

Acworth, W. (2011) “Record Volume 2010 (Annual Volume Survey)”, Futures Industry, March 2011, pp. 12-29.

Aldridge, I. (2013) *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. 2 edition, Wiley.

Brogaard, J., Hendershott, T., Riordan R. (2013) “High frequency trading and price discovery”, ECB Lamfalussy fellowship programme/ Working paper series, No. 1602, European central bank Press.

Caldeira, J. F., Moura, G. V. (2013) “Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy”, Revista Brasileira de Financas, Vol.11, No. 1, pp. 49–80.

Carrion, A. (2013) “Very fast money: High – frequency trading on the NASDAQ”, Journal of Financial Markets, Vokume 16, Issue 4, pp. 680 – 711.

Cartea, A., Penalva, J. (2012) “Where is the value in high frequency trading?”, Quarterly Journal of Finance, Vol. 2, No. 3, pp. 1–46.

CFTC, SEC (2011). Recommendations regarding regulatory responses to the market events of May 6, 2010: Summary Report of the Joint CFTC-SEC Advisory Committee on Emerging Regulatory Issues. https://www.sec.gov/spotlight/sec-cftcjointcommittee/021811-report.pdf (accessed September 20, 2013)

Cifu, D.A. (2014). Virtu Financial, Inc. registration statement, Form S-1 (Registration statement under the securities act of 1933). US securities and exchange commission. https://www.sec.gov/Archives/edgar/data/1592386/000104746914002070/a2218589zs-1.htm (accessed August 05, 2014)

Dunis, C., Giorgioni, G., Laws, J., Rudy, J. (2010) “Statistical arbitrage and high-frequency data with an application to Eurostoxx 50 equities”, Alternative investment analyst review, in series: Trading strategies, pp.35-57.

Elliot, R.J., Hoek, J.V., Malcolm, W.P. (2005), “Pairs Trading”, Quantitative Finance, Vol. 5, Issue 3, pp. 271 – 276.

Ferguson, R., Laster, D. (2007) “Hedge Funds and systemic risk“ in *Hedge funds, credit risk transfer and financial stability* by Cole, R.T., Feldberg, G., Lynch D., Financial Stability Review- Special issue on hedge funds, Banque de France, No. 10, pp. 45-54.

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.

Gomber P., Arndt B., Lutat M., Uhle T. (2011). *High-Frequency Trading. Goethe University*, Deutsche Börse Group, p.86.

Hanson, T.A., Hall, J.R. (2012) “Statistical arbitrage trading strategies and high frequency trading”, Midwest Finance Association 2013 Annual Meeting Paper. Available at SSRN: http://ssrn.com/abstract=2147012

Hasbrouck, J., Sarr, G. (2013) “Low-latency trading”, Journal of Financial Markets, Vol. 16, Issue 4, pp. 646 – 679.

Hendershott, T., Riordan, R. (2012). Algorithmic Trading and the Market for Liquidity. Journal of Financial and Quantitative Analysis, Vol. 48, No.4, pp.1001-1024.

Hogan, S., Jarrow, R., Teo, M., Warachka, M. (2004), “Testing market efficiency using statistical arbitrage with applications to momentum and value strategies”, Journal of Financial Economics, Vol. 73, pp. 525 – 565.

Masteika S., Driaunys K., Rutkauskas A.V. (2013) “Historical data formation for back test and technical analysis in North American futures market”, Transformations in Business and Economics, Vol. 12, No. 1A, pp. 473-488.

Masteika, S., Rutkauskas, A.V. (2012) „Research on futures trend trading strategy based on short term chart pattern”, Journal of Business Economics and Management, Vol. 13, No.5, pp. 915-930.

Masteika, S., Rutkauskas, A.V., Alexander, J.A. (2012) “Continuous futures data series for back testing and technical analysis”, IACSIT Conf. proceedings, Singapore, Vol. 29, IACSIT Press, pp.265-269.

Menkveld, A.J. (2013) “High frequency trading and the new market makers”, Journal of Financial Markets, Vol. 16, No.4, pp. 712-740.

Miao, J.G. (2014) “High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach”, International Journal of Economics and Finance, Vol.6, No. 3, pp. 96-110.

Nath, P. (2003) "High Frequency Pairs Trading with US Treasury Securities: Risks and Rewards for Hedge Funds", London Business School. *Working paper series*. Available at SSRN: http://ssrn.com/abstract=565441

Perlin, M. (2009) “Evaluation of Pairs Trading Strategy at Brazilian Financial Market", Journal of Derivatives & Hedge Funds, Vol. 15, pp. 122-136. Available at SSRN: http://ssrn.com/abstract=952242

Sakalauskas V., Kriksciuniene D. (2012) “The impact of public information to stock market fluctuations”, Transformations in Business & Economics, vol. 11, no. 1(25), p. 84-99.

SEC, (2014), *Equity Market Structure Literature Review, Part II: High Frequency Trading*, U.S. Securities and Exchange Commission, Staff of the Division of Trading and Markets. Available on the SEC’s website: http://www.sec.gov/marketstructure/research/hft_lit_review_march_2014.pdf

Vidyamurthy, G. (2004), *Pairs Trading – Quantitative Methods and Analysis*, John Wiley & Sons, Inc., New Jersey, p. 210.

Zubulake P., Lee S. (2011). *The High frequency game changer: how automated trading strategies have revolutionized the markets.* Aite group. Wiley trading, p. 162.