MASTEROPPGAVE

**Pairs Trading With Vanilla Options
BE331E Finance And Investment**

Aleksander Rinaldo

ABSTRACT

The idea behind the strategy of pairs trading is to exploit the relationship between two securities by finding two securities that move similarly over time. The theory is that by buying one stock long while selling another stock short within the same sector, when the securities diverge from equilibrium a profit will be yielded when the two securities eventually converge back to equilibrium.

The investment strategy of 'Pairs Trading' was pioneered in the 1980's and is a market neutral strategy enabling traders to profit from almost any market condition. Although this strategy was initially only available to hedge funds and investment banks, it has more recently become a viable option for small-scale investors thanks to the increasing development of ICT tools and internet-based brokers.

For this paper I have attempted to follow the methodologies used in the earlier studies conducted by Gatev, Rouwenhorst and Goetzmann (2006) and Do & Faff (2009). However in this paper I have chosen to apply vanilla options instead of going long and short in the underlying. The options are set with a strike at the money.

I chose use the companies on the S&P 500 and used a co-integration approach to find all subsets of two companies. One group of securities was significant at 1% level, one at a 5% level and the last at 10% level. The indicator to trade was calculated using the mispricing expressed in standard deviations using a 50-day simple moving average.

FOREWORD

This thesis was developed as a final study in the Master of Science in Business, specialisation of Finance and Business at Universitet I Nordland, Bodø Campus. The period of study extends over the entire spring semester 2014 and has a study loading of 30 credits. I wish to thank my supervisor, Thomas Leirvik for the help, encouragement and feedback during the formulation of this paper. I would also like to thank my wife, Gemma for her ongoing support and feedback during this semester.

Any error or omissions found throughout this paper are my (Aleksander Rinaldo) responsibility

Contents

Figures

1. Introduction

For many years, practitioners and academics of investment management have tried to beat the market by applying fundamental analysis and technical analysis. In the mid-1980s Nunzio Tartaglia gathered a team of physicists, mathematicians and computer scientist to discover arbitrage opportunities in the equity market. The team used sophisticated statistical methods to develop highly technical trading programs. One of these programs identified pairs of securities whose prices tended to move together. In 1987 they reportedly made a $50 million profit for the firm Bass (1999). The team was disbanded in 1989 after low performance. However, quantitative trading strategies, and in particular pairs trading have since become an increasingly popular strategy as the methodology has gradually been easier to access for the average trader due to technological advances. Therefore, in more recent years this kind of trading, often-noted *algorithm trading*, has grown significantly.

Pairs trading involves finding two different assets, or pairs, that are co-integrated and are close economic substitutes. By identifying the equilibrium between the stock pairs, one can exploit any situation where the price of one of the pairs breaks out of the equilibrium. Avellaneda and Lee (2010, p. 761) State that by investing frequently in small positions over different stocks the portfolio will be diversified, which would produce a low volatility investment strategy that has low covariance with the market. The trading strategy relies on the fact that it is likely that pairs will revert back to their long-term equilibrium. An accepted explanation for this equilibrium is the law of one price, which states that similar products should sell for similar price It is also worth noting that the trading strategy involves high frequency trading, which would encompass an increased amount of trading cost.

A common example of co-integrated pairs is Coca-Cola (KO) and Pepsi (PEP). Both companies registered under the same industry classification, producing carbonated soda drinks. **Feil! Finner ikke referansekilden.** shows how both companies have shared their

highs and lows. Because of the companies shared similarities one would expect the securities to be affected by the same market and share similar trends.

Figur 1-1 A time-series plot of the prices of Pepsi and Coca-Cola

By looking at these two companies we can see that in year 2000 the shares are moving closer together, we would therefore go long in PEP (red) and short in KO (blue). By 2002 the securities have moved away from equilibrium and we would therefore expect reversion back to equilibrium. The trader would therefore take a short position in Pepsi and a long position in coca cola.

David Shaw, a computer science professor recognised as one of the most successful quantitative traders on Wall Street, and an early Tartaglia's protégé, implies that the success of his firm D.E. Shaw relied on an early entry into the market. In an interview, put together, by New York Times journalist Joseph Kahn, he says he believes that when his secret algorithms 'spot a buy' it only gives him a marginally better chance of generating profit than if he were to flip a coin.

If the market is operating efficiently one would expect that a trading strategy revolving around past price dynamics and simple contrarian principle would not be able to generate money. Therefore the risk adjusted return should not be positive. As pairs trading has been subject to rigorous research by several academics including Vidyamurthy (2004, Gatev, et al. (2006, Do and Faff (2010, Do and Faff (2012), it seems that after accounting for all costs regarding trading the excess return has diminished gradually and is practically gone by 2008.

One common factor seen in these studies is taking opposite positions by going long and short. The cost of taking the wrong position where the security does not return back to its expected equilibrium can thus be prime reason for not generating excess return. This paper seeks to reduce this loss by buying options instead of going long and short.

Engle and Granger (1987) expanded the results proved in the paper by Granger (1981) regarding long run dynamics. Long-run multiplier, error correcting model and partial adjustment models were used to estimate the long and short run dynamics. Yet, it is the discovery of modelling two non-stationary variables upon each other where the error term is stationary that has made it possible to do pairs trading. Pairs trading is therefore trading on the expectation of the long run dynamics.

Lee specified three main factors that statistical arbitrage encompasses:
I.   Trading signals are rule based.
II.   The trading strategy is market neutral, due to zero betas with the market and
III.   The mechanism for generating excess return is statistical.
Avellaneda and Lee (2010)

By applying options to pairs trading I seek to answer the question; 'does the profitability increase or decrease during the period in regards to prior studies'. I also seek to ascertain if the cost losses due to non-convergence is decreased.

2.   Market efficiency

In the stock market, prices are under constant change to equilibrate the supply and demand. The stock market is there to provide capital to businesses and at the same time allocate ownership to investors. The efficiency of this reallocation of resources has undertaken rigorous research. Most neoliberal and neoconservative economic theories are founded on the belief that markets are efficient. The political philosophy neoliberalism has been a driver for free trade on open deregulated markets.

This chapter will focus on research done on market efficiency; I focus on the central contributor of testing the efficiency Fama (1965, 1970, 1991, 1998). Later research is also provided questioning the early research done by Fama and Malkiel.

Earlier research led by mathematician Bhacheliers and his study on the development of prices during the 1900s, pointed towards a change of which was following a random walk. "His " fundamental principle" for the behaviour of prices was that speculation should be a 'fair game'; in particular, the expected profits to the speculator should be zero. With the benefits of the modern theory of stochastic processes, we know now that the process implied by this fundamental principle is a martingale" Fama (1970, p. 389)

For a time series to be classified as a random walk it requires independence. "In statistical terms independence means that the probability distribution for the price change during time t is independent of the sequence of price changes during previous time periods" Fama (1965, p. 35)

A martingale is the probability theory, simply stating that there is no possibility of predicting the mean of the future winning behaviour based on past and current information. More specifically the recognised martingale definition is the process satisfying:

$$E(X_{n+1}|X_1, \ldots, X_n) = X_n$$

$$E(X_{n+1}|X_1, \ldots, X_n) - X_n = 0.$$

Stating that the average "winnings" from observation n to observation n+1 are zero. If the equity market were to behave as a martingale it would be impossible to gain excess return in the long term.

If there is no possibility of predicting future price based on previous price there would have to be independence between the observed values. Fama explains this independence: "If there are many sophisticated traders who are extremely good at estimating intrinsic value, they would be able to recognize situations where the price of a common stock is beginning to run up above its intrinsic value. Since they expect the price to move eventually back towards its intrinsic value, they have an incentive to sell this security or

to sell it short… Thus their action will neutralize the dependence in the noise-generating-process, and successive price changes will be independent" Fama (1965, p. 38)

The equity market would be efficient if the prices of all the securities fully reflected all available information. As Fama (1970) writes, this statement is to general to be testable. However, the conditions of market equilibrium were assumed to be stated in terms of expected returns.

Fama (1970) Describes the expected return theory as:

$$E(\tilde{p}_{j,t+1}|\ \Phi_t) = [1 + E(\tilde{r}_{j,t+1}|\ \Phi_t]p_{jt}$$

Simply stating that the expected upcoming price given all available information is equal to current price plus the expected return given all available information times the current price.

Where;

E is the expected value operator.

p_{jt} is the price of security j at time t.

$\tilde{p}_{j,t+1}$ is the price of the security at t+1 (with reinvestment of any intermediate cash income from the security).

$\tilde{r}_{j,t+1}$ is the one-period percentage return $(\tilde{p}_{j,t+1} - p_{jt})/\ p_{jt}$.

Φ_t is a general symbol for whatever set of information is assumed to be 'fully reflected' in the price at t.

Tildes indicate that $\tilde{p}_{j,t+1}$ and $\tilde{r}_{j,t+1}$ are random variables at t. The value of the equilibrium expected return $E(\tilde{p}_{j,t+1}|\ \Phi_t)$ is determined the information that Φ_t 'is fully reflected' in the formation of price p_{jt}.

$x_{j,t+1}$ is the difference between the observed price and the expected value of the price that was projected at *t* on the basis of the information Φ_t.

$$x_{j,t+1} = p_{j,t+1} - E(p_{j,t+1}| \Phi_t)$$

$$E(x_{j,t+1}| \Phi_t) = 0$$

With a return set equal to zero the model is referred to as the ' fair game' efficient market model.

Fama (1970) Points to the fair game model where by any trading system based on historical information would have expected return equal to zero. Pairs trading assume that the securities in the long run would return back to its equilibrium. The expectation of this behavior is only based on past information, and by the fair game model it would therefore not be possible to generate excess return when accounted for risk.

The submartingale model states that the expected value of t+1 of any security is equal or greater than the current price. This assumption: "Implies that such trading rules based only on the information in Φ_t cannot have greater expected profits than a policy of always buying-and-holding the security during future period in question"Fama (1970, p. 386)

2.1   The random walk model

A combination of two assumptions brought together contributed to the random walk model. First the assumption that all available information was fully reflected in the price, and second that the return was identically distributed. Fama (1970, p. 386) wrote this model as:

$$f(r_{j,t+1}| \Phi_t) = f(r_{j,t+1})$$

Here the conditional and marginal probability distribution is the same as the non conditional. Also the density function has to be the same for all t-values.

This is an example of a random walk generated; the interesting thing with this is that from a technical point of view you would expect the security to be in an upward trend, but clearly in this case it would not mean anything. It might look systematic, but these values are following a random walk.

Fama (1970) Elaborates on this by stating that if the expected return on a security j is constant over time, then we have

$$E(\tilde{r}_{j,t+1}|\,\Phi_t] = E(\tilde{r}_{j,t+1})$$

From this formula it is easy to see that the expected return is independent from the information available at time t. Fama (1970) targues that the random walk is an extension of the fair game model by expressing that "the

Figur 2-1 Random Walk

mean of the distribution of $\tilde{r}_{j,t+1}$ is independent of the information available at t, Φ_t, whereas the random walk model in addition says that the entire distribution is independent of Φ_t".
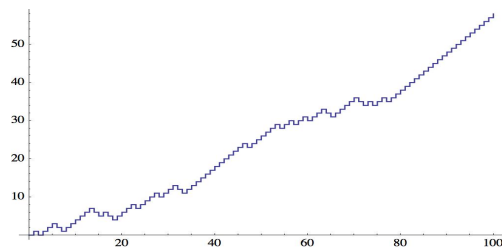
Fama (1970) not only argues that the random walk is an extension of the fair game model but that the fair game model falls short because it does not take into account the environment in which investor's tastes and preferences create new information that is used to generate return equilibrium in which return distribution repeats itself over time.

Due to the complexity of the 'real world' inference are often made with assumptions. They can be justified or not justified. Market efficiency theory contains three assumptions:

I.   There are no transaction costs in trading securities.
II.   All available information is costless-ly available to all market participants.
III. All agree on the implications of current information for the current price and distribution of future prices of each security. In such a market the current price of a security obviously 'fully reflects' all available information.

Fama (1970) argues that this is not unrealistic, that even when the transaction costs are high there is no causality in the fact that the price is not fully reflected by the available

information at time t. He also argues that if enough investors have a sufficient amount of information it would create an efficient market, and therefore even if not every investor had sufficient information the market would still be efficient.

The weak form of market efficiency focuses on historical information, where any trading strategy that was based on current or historical information would not be able to generate excess return. This is due to the fact that all information is reflected in the price. Focus here is on the fair game model, where return pointed towards a sub martingale expected return model. Depending on which order the research was conducted it was often referred to as a 'random walk'. However, Fama (1970) states that even if the random walk has been cited as a good approximation, the researcher would continue to point towards the 'fair game' model. "What resulted was a theory of efficient markets stated in terms of random walks, but usually implying some more general form of the 'fair game' model."Fama (1970, p. 409)

Market efficiency took a step further when researchers began analyzing how efficiently new information became reflected in the market price. The semi strong market efficiency still concludes that all prior information is reflected in the price. However, it now also looks at how quickly new information become evaluated and transferred into the securities. Information that is taken into account is the fundamental data, which could be accounting practices, stock splits, annual reports etc. Fama (1970) looks at how the securities react after these fundamental releases, he also concludes that to a large extent the market is efficient. "We shall conclude that, with but a few exceptions, the efficient market model stands up well" Fama (1970, p. 383)

Conversely, (Bodie et al 2009) states that there are fundamental ratios such as the PE ratio that are able to 'predict' unexplained risk adjusted returns. Another strategy is the buying of stocks just after the release of positive quarterly reports. This is a result which is in large a contrast to the semi efficient market model, where Fama states: "The available semi-strong form evidence on the effect of various sorts of public

announcements on common stock returns is all consistent with the efficient market model." Fama (1970, p. 409)

The strong form of market efficiency is concerned with the monopoly effect, determining whether or not there are any investors/hedge funds in particular that would have monopolistic access to information that would be relevant for the development of prices, and whether this information would be reflected in the market. Many regulations have been implemented to prevent investors from benefiting from inside information. For some this would be enough to conclude that the market does not operate under a strong form of market efficiency.

However, there have been a lot of criticisms since the Global Financial Crisis (GFC). In addition to these criticisms, large sales were reported ahead of the GFC.

Fama (1970) indicates that there are specialists at the stock exchange with inside information that would be able to benefit from restricted information. However, apart from those few the remaining investors would have to manage on the information given in the public, and therefore would be subject to strong market efficiency.

The weak form of market efficiency where markets are in equilibrium and where temporary disequilibrium is self-correcting has been disputed. If the prices on the securities are in equilibrium where arbitrage profits have been completely removed is it possible that a competitive economy would always be in equilibrium? Grossman and Stiqlitz argue that is it not by stating, "Clearly not, for then those who arbitrage make no (private) return from their (privately) costly activity. Hence the assumption that all markets, including those for information, are always in equilibrium and always perfectly arbitraged are inconsistent when arbitrage is costly" Grossman and Stiglitz (1980, p. 393)

The semi strong form of market efficiency regarding how quickly new information is reflected in the price has also been challenged by several researchers.

Income has a significant effect on security prices, and one would therefore expect that the market would to some extent "price in" their expected future income of a given security. Ball and Brown (1968) created a study by developing two alternative models, both looking at the expectation prior to an income release(market expectation) and then studying the market behavior when expectations were not realized. "suggests that the market begins to anticipate forecast errors early in the 12 months preceding the report" Ball and Brown (1968, p. 171) Although the preciseness of anticipation or expectation of error would increase the closer one gets to the release of the report, it does raise the question of the markets efficiency. Ball and Brown (1968) among others[1] documented that there was a predictability in abnormal returns.

Bernard and Thomas (1990, Bernard and Thomas (1989) explain that it is difficult to understand why the market does not react fully on releases as publically available as earnings. They go on to give an idea of why the market is behaving in such a way. "Given that a firm announces positive (negatively) unexpected earning for quarter t, the market tends to be positively (negatively) surprised in the days surrounding the announcement for quarter t+1… This evidence is consistent with a market that "fails to adequately revise its expectations for quarter t+1 earnings upon receipt of the news for quarter t" Bernard and Thomas (1990, p. 306).

The publication by Shiller (2003) "From efficient markets theory to behavioral finance", argues that the efficient market hypothesis is outdated and that the research has taken a wider perspective including psychology and sociology. In this paper Shriller presents the efficient market model $P_t = E_t P_t^*$

Where,

P_t is the current price of any security

and $E_t P_t^*$ is the expected future price.

---

[1] Joy, Litzenberg and McEnally(1977), Watts(1978), Rendleman, Jones and Latane(1982)

From here it follows that future price is equal to the current price plus some forecast error U_t. Shiller (2003) points out that a forecast which obtains autocorrelation in its residuals is violating the assumption that the error term should be random, and that this often is an indication of omitted variables in the equation, or in other words, not all information is taken into account. If there were a pattern in U_t it would be contradictory to the market efficiency hypothesis where all information is fully reflected in the price.

Shiller expands on this by stating, "the fundamental principle of optimal forecasting is that the forecast must be less variable then the variable forecasted. Any forecaster whose forecast consistently varies through time more than the variable forecasted is making a serious error, because then high forecasts would themselves tend to indicate forecast positive errors, and low forecast indicate negative errors." Shiller (2003, p. 85)

Shiller (2003) highlights that we have to take into account human nature and the herding behavior of the investors, stating that financial markets are not always functioning well and that there is clear evidence of periods where markets do not reflect all the available information.

Under the strong form of market efficiency it is claimed that if you want to exceed the stock market return, you would have to take more risk than the stock market as a whole. Maximum possible return would therefore be a function of how much risk the participant would be willing to accept.

To test the validity of the strong market efficiency several alternatives have been proposed. Yet, "efficient market theorists specify two methods to test tenet number three (strong form of market efficiency). The first method is statistical inference. This involves, calculating serial correlation coefficients of stock price changes. If the serial correlation coefficient is zero or close to zero, this supports assuming serial independence in the price data. Therefore, one can infer that stock-trading rules cannot work. The second method requires using a stock trading rule, based solely on past prices – where expected

profits are greater and risk lower than they would be under a naïve buy and hold policy"
Prentis (2012, p. 24)

Prentis (2012) provides a systematic and clear overview of the inherent pitfalls with
research in favor of the market efficiency.

He provides three main reasons supported by earlier research:

   I.   Using the wrong data

   II.   Using the wrong method to analyze data and

   III. Jumping to conclusions without sufficient evidence.

Prentis (2012) also provides quantitative results contradicting Fama (1970)'s statement
that beating the stock market over long period of time should be impossible while at the
same time have lower risk than a buy and hold benchmark portfolio.

"The relative maxima and minima stock trading rule makes substantially more money at
significantly less risk than the naïve buy-and-hold policy. Efficient market theorists say
this thorough beating of the US stock market should be impossible to achieve using only
a stock trading rule, based solely on past prices. Thus, tenet number three and the weak
form of the theory of efficient markets is invalid during this early stock market period,
prior to US government financial markets regulation. This calls the theory of efficient
markets into question." Prentis (2012, p. 28)

Nevertheless, Easton and Kerin (2010) points out that it makes more sense to ask to what
extent the market is efficient or not. They also stipulate how important it is to separate the
efficiency the market has in regards to pricing securities, and the efficiency of the pricing
of the market as a whole. In other words a separation is needed between micro and macro
efficiency. While there is a clear separation between micro and macro I will note that the
market prospects do have an effect on the expected value of a single security.

In regards to pairs trading it has been shown by Gatev, et al. (2006) that profits were generated during a sample period of several years, this also questions Fama (1970)'s theory of strong market efficiency.

While this paper does not try to provide evidence either for or against market efficiency, it attempts to try and provide a wide overview of the studies conducted on market efficiency. Furthermore, it should be noted that this is not a conclusive list on the published research on market efficiency.

3.   Arbitrage

Arbitrage is defined as "the simultaneous purchase and sale of the same, or essentially similar, security in two different markets for advantageously different prices"Shleifer and Vishny (1997, p. 35) Arbitrage is a broad concept with many combinations. To give an example, suppose a given exchange rate (say between Australian Dollars and Norwegian Kroner) two different quotes are given by two different traders, at the same time. One would be able to buy the kroner in the cheaper market and sell them in the more expensive.

The textbook-typical example of arbitrage consists of simultaneously buying a product in one market and selling it in another in order to avoid exposure to the market. The gains made from this transaction then become the arbitrage profit. It is argued that while opportunities such as this may exist in the financial market at certain times, they tend to disappear fairly quickly and that arbitrage opportunities are more likely to exist in new and underdeveloped markets.

However, keeping the market in equilibrium is an ongoing process and this would mean that new opportunities for arbitrage would exist every time there was a divergence from equilibrium. Traders who use pair trading attempt to profit on this strategy, which is to transform the price to its equilibrium. In the paper by Gatev, et al. (2006) it is mentioned there would have to be a profit to undertake this action, and that the excess return is the payment for transferring the market into equilibrium.

In theory arbitrage requires no capital and entails no risk. In reality one would have to have capital and to some extent endure risk. Shleifer and Vishny (1997)

Pairs trading is said to be an arbitrage strategy though arbitrage as a concept is often referred to as being risk-free. However, even if one were able to generate a market neutral portfolio it would not mean that it is risk free. In other words pairs trading is a risk arbitrage.

Different theories have been developed for the explanation of why pairs diverge from each other. One is the fundamental change in one or both of the co-integrated pairs. However, if the change were the due cause of fundamental change there would be no reason why the pairs should remain co-integrated.

3.1   Pure arbitrage

A pure arbitrage opportunity is defined as *"a zero-cost trading strategy that offers the possibility of a gain with no possibility of a loss* Bondarenko (2003, p. 875).

While in earlier times, before the late technological advances, well-resourced arbitrageurs could have taken advantage of such 'pure arbitrage opportunities', today markets operate on a much more level playing field. This decrease in 'pure arbitrage opportunities has led to an increase in market efficiency through more developed markets.

According to Damodaran (2014) at New York University, there are 2 main reasons why finding pure arbitrage in modern markets will be a rare occurrence. The first is that identical assets are uncommon in real-life situations and the second is that even if you were to find two identical assets, it would raise the question as to why these pricing differences were allowed to exist.

One example of such a strategy would be to exploit the mispricing of options through put call parity, yet this would acquire the trader to be able to identify mispricing if it exists.

Chen (1996) names pure arbitrage as "the strongest form of arbitrage", and states that in

order for a pure arbitrage pricing formula to exist "a complete market is necessary", which requires enough related assets. Despite the fact that such an arbitrage is not usually found, theoretically it can exist. In theory, you are able to walk away with sure profits having invested no money and taking no risk, which seems too good to be true. It is the actual nature of pure arbitrage that makes such an opportunity likely to be short lived. Even the smallest pricing differences will be noticed and acted upon in a market scrutinized by investors looking for riskless profits. Once noted, these price differences will be immediately taken advantage of and in the process, disappear.

Yet, there exist studies claiming to show opportunities of pure arbitrage. Sawicki and Hric (2001) examine 84 Czech stocks that trade on the two Czech exchanges – the Prague Stock Exchange (PSE) and the Registration Places System (RMS)- and find that prices adjust slowly across the two markets, and that arbitrage opportunities exist (at least on paper) –the prices in the two markets differ by about 2%. These arbitrage opportunities seem to increase for less liquid stocks. While the authors consider transactions cost, they do not consider the price impact that trading itself would have on these stocks and whether the arbitrage profits would survive the trading.

Due to this quick exploitation of pricing differences, in order for success at pure arbitrage to occur two requirements must be fulfilled; they are "access to real-time prices and instantaneous execution" Damodaran (2014, p. 33). In addition to these requirements one must first have access to substantial debt at favorable interest rates. Because the pricing differences making pure arbitrage theoretically possibly will more than likely be very small, such access to substantial debt can prove helpful since it can "magnify the small pricing differences" Damodaran (2014, p. 33)

3.2   Relative-value arbitrage

Pairs trading classifies as relative value arbitrage. This involves investing in assets that can be argued to have relative similar value. However, since this is not pure arbitrage there is the possibility of losses and it therefore involves risk. This 'risky arbitrage strategy' pairs trading, is said to be market neutral. Meaning that the covariance, beta, is equal to zero.

Many securities can be used in relative-value arbitrage. And the combinations are endless, in pairs trading with options we look at securities on the S&P 500. It is therefore important to identify the securities that share historical trends under equal industry classifications. By using sub industry classifications one hope to reduce the chances of finding cointegrated pairs on the basis of luck. The likelihood of finding pairs without an economic relationship is dependent in the significance level chosen for the test. Dough, the significance would never be able to completely eliminate the chance completely. Originally pairs trading involved taking opposite positions (short and long) in cointegrated securities.

 "When the prices of the two securities diverge, the relative-vale arbitrageur buys one security and shorts the other, then when the prices converge again, the relative-value arbitrageur closes the trade" Barclayhedge (2014). Due to the risks associated with relative-value arbitrage, the major players include large institutional investors such as hedge funds and investment banks.

Jones and Izabella (2010) state that according to the National Bureau of economic research, since the collapse of Lehman Brothers in 2008 the US markets have been inundated by "astonishing" pricing abnormalities – such opportunities on which relative-value arbitrage strategies depend upon. Furthermore they are not the only markets to see such discrepancies. Pricing discrepancies opened up on the UK's sovereign debt market in 2009 and due to a massive program by the Bank of England to buy back gilts, coupled with the UK treasury issuing large amounts of bonds are the same time leading to a "distortion in pricing of UK government bonds of different maturities" Jones and Izabella (2010). This allowed relative-value arbitrage opportunities to occur as some hedge funds took advantage of the bond market.

In fact relative-value arbitrage, or pairs trading opportunities can be both common and successful. A paper Gatev, et al. (2006) found that the investment strategy of pairs trading was tested against daily data from 1962-2002, the robustness of the excess returns

indicated that pairs trading profited from the temporary mispricing of close substitutes. Another important part of this test is the time delay on their test meaning that they applied the test years after first discovering its profitability. By having a strategy that performs just as good in sample as out of sample is an indicator of a successful strategy since it does not rely on over fitting parameters to 'in sample' data.

Though, pairs trading have been classified under relative arbitrage, it also have been categorized under speculative or pseudo arbitrage, which could be considered the counterpart to pure arbitrage. Where pure arbitrage presents the ideal situation of 'riskless' arbitrage, arbitrage in the real world involves at least some risk, and these more risky arbitrage situations are referred to as pseudo or speculative arbitrage.

According to Damadoran at New York University, pseudo arbitrage is not really considered arbitrage in the pure sense of the word. He states that in pseudo arbitrage investors take advantage of what they see as similar (though not identical) "assets that are mispriced, either relative to their fundamentals or relative to their historical pricing; you then buy the cheaper asset and sell the more expensive one and hope to make money on the convergence" Damodaran (p. 14) If they are correct in their assumptions, the difference should narrow over time, yielding profits Damodaran (2012).

The more an investment strategy moves away from pure 'riskless' arbitrage and the more it exposes investors to significant risk, it becomes categorized as pseudo or speculative arbitrage strategy Damodaran (2014). However, to be successful when deciding to move toward pseudo-arbitrage it is important that these risks are kept under control and that the "financial leverage in your strategy" Damodaran (2012) is reduced accordingly. Furthermore, investors should recognize that size can be both wield both positive and negative results and as more investment funds are obtained and execution costs reduced, getting "into and out of positions quickly, and without a price impact" Damodaran (2012, p. 471) will prove more difficult .

It is worth noticing that it is often within the field of pseudo arbitrage we see hedge funds

in their numerous forms.

4.   RISK

4.1   Execution risk

Execution risk; "the chance that a desirable transaction cannot be executed within the context of recent market prices or within limits proposed by an investor" Gastineau and Kritzman (1999, p. 130) Investors face execution risk in almost all financial instruments.

Due to the nature of execution within financial instruments it is generally impossible to close several transactions at exactly the same time. This means there is a possibility that when one part of the deal is initiated, for whatever reason, the second part of the deal cannot be closed (or executed) at a profitable price, resulting in an execution that is worse than expected. This is the most common form of execution risk and is called slippage.

4.2   Convergence risk

In the context of arbitrage strategies, the fact that mean reversion affects conditional volatility, and therefore risk, is of particular importance. Convergence trades "assume explicitly that the spread between two positions, a long and a short, is mean reverting, and if the mean reversion is strong, then the long horizon risk is smaller than the square root volatility" Allen, et al. (2009, p. 72). Those who manage risk, may often need to assess the risk involved in a particular trading strategy with a different view of risk, and such convergence trades may create such a difference Allen, et al. (2009). In these instances it is common for those managing risk to keep a null hypothesis of market efficiency.

The risk of a convergence trade is that the expected convergence does not occur, or that it takes too long, even diverging before converging. The danger of price divergence is that due to the synthetic nature of leveraged, convergence trades and the fact that they involve a short position, a trader may run out of capital before the trade can make money in the long term. In relation to pairs trading the problem would be if the pairs do not converge

quick enough and therefore the expected mean reversion does not occur before the option is expired.

While many, if not most economists are of the opinion that markets are self-stabilizing in the long run, there is a large body of published research to suggest that destabilizing dynamics can exist in markets. "Convergence trading typically absorbs shocks, but an unusually large shock can be amplified when traders close prematurely" according to Kambhu (2006, p. 1). Xiong (2001) found that convergence traders with logarithmic utility functions could trade in a way that would amplify the affect of market shocks, if the shock were large enough (to delete their capital), although they usually traded in a way that was market stabilizing. As a result, when such traders suffered such a capital loss they would try to reverse their convergence trade positions, driving prices along the same line as the initial shock Xiong (2001).

4.3   Liquidity risk

Liquidity risk is the risk that stems from the lack of marketability of an investment that cannot be bought or sold quickly enough to prevent or minimize a loss Nikolaou (2009). The higher the probability of liquidity, the higher the associated liquidity risk becomes, and when this possibility becomes certainty the liquidity risk reaches its highest threshold and illiquidity materialises. According to the European Central Bank "there is an inverse relationship between illiquidity and liquidity risk, given that the higher the liquidity risk, the higher the probability of becoming illiquid, and therefore, the lower the liquidity" Nikolaou (2009, p. 16)

There are two main types of liquidity risk. They are; the risks associated with market liquidity and the risks associated with funding liquidity. Market liquidity risk refers to an asset that cannot be sold due to a lack of liquidity in the market. This could be due to the widening the spread on an offer or making explicit liquidity reserves, for example. On the other hand, funding liquidity refers to liability that either; can only be met for a price that is not financially viable, cannot be met when due or that presents a systemic risk SWIFT (2011)

A range of internal and external factors can cause liquidity risk. These include, but are not limited to; high off balance sheet exposure, highly sensitive financial market and depositors, heavy reliance on corporate deposits and sudden economy shock. However the European Central bank stresses that the real roots of liquidity risk lie in "information asymmetries and the existence of incomplete markets" Nikolaou (2009)

Since the GFC the importance of being able to adequately manage liquidity risk has become increasingly important. Financial institutions are being driven to strengthen and improve their liquidity risk management strategies due to post-GFC increases in the cost of liquidity, larger funding spreads, higher volatility and reduced market confidence SWIFT (2011). According to SWIFT's 2011 in-depth market survey among cash, liquidity and liquidity risk managers at financial institutions around the world "82% of respondents recognise a lack of ability to manage and report the liquidity position at a firm-wide level on a daily basis" and "91% indicated that they have a lack of ready-made liquidity risk analytics and business intelligence" SWIFT (2011, p. 6). However, according to the Principles for Sound Liquidity Risk Management and Supervision (2008) "a bank should actively manage its intraday liquidity positions and risks to meet payment and settlement obligations on a timely basis under both normal and stressed conditions" BIS (2008, p. 20) It is argued that issues surrounding liquidity risk management relate to a lack of common standards and industry practices SWIFT (2011). It is further argued that collaborative solutions are needed to resolve such issues in the future.

4.4   Short squeeze

Short squeeze is observed when "the price of a stock rises significantly and speculators are forced to cover their positions to limit their losses "Obienugh (2010, p. 347). When a short squeeze occurs, it generally means that short sellers are being edged out of their short positions, usually at a loss. Short squeeze has the potential to occur when short selling and may occur in an automated way if the stop-loss orders of short-sellers were in place with their brokers Obienugh (2010).

The benefit of predicting a short squeeze means that one can take advantage of the situation in which panicked short-sellers are causing a further rise in price due to short term demand Obienugh (2010). An experienced short-squeezer would buy the stock which it was on the rise and sell it at its peak, being able to "swoop in" at precisely the right time Gobel (2013). There are several predictors of short squeeze and they include short interest percentage, short interest ratio, and daily moving average charts Gobel (2013). In theory by interpreting the daily moving average charts, a trader is suppose to be able to predict short squeezes by calculating the short interest percentage and short interest ratio.

Employing a short squeeze strategy is not without risk. If the stock has peaked, it has the potential to fall and the success of the strategy will depend on whether the stock can be successfully sold during its peak Gobel (2013).

4.5   Credit risk

Credit risk is most simply defined as "the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms" BFIS (2000). The assessment of credit risk involves the borrowers ability to repay investors, who lend their capital, and when a perceived credit risk is high one will demand a higher rate of interest. (HKIB) (2010)

Credit risk can be classified into several groups; credit default risk, concentration risk, country risk etc. Credit default risk may impact all credit-sensitive transactions, including; loans, securities and derivatives. Concentration risk has the "potential to produce large enough losses to threaten a banks core operations" UniCredit (2012) and Country risk involves the loss associated with a sovereign state defaulting on its obligations or freezing foreign currency payments UniCredit (2012).

There are many sources of credit risk for banks that can exist throughout their everyday activities, however probably the largest and most obvious are that of loans. Other sources

of risk include acceptances, interbank transactions, foreign exchange transactions and more BFIS (2000). Although, through the GFC, it is clear that even for traders there is a risk of having counterparties not being able to pay back there depth, resulting in bankruptcy. For a trader the chances of experience trading risk in such form mentioned above is very low, as the financial bank is standing as the traders counterparty and not the trader of the opposite position as you.

While 'credit risk' refers to the probability of loss due to a borrower or counterparty not being able to make repayments on their debt; 'credit management' is "the practice of mitigating those losses by understanding the adequacy of both a bank's capital and loan loss reserves at any given time" Sas (2014). And according to the Basel Committee (2000) the effective management of credit risk and an understanding of the relationships between credit risk and other types of risk are essential components to the long-term success of any banking institutions BFIS (2000). Since the GFC and the corresponding "credit crunch" the importance of such management has been highlighted, including a higher demand for transparency and more strict regulations such as those outlined in the Basel III Sas (2014).

4.6   Long-Term Capital Management

Long-Term Capital Management (LTCM) was a hedge fund management firm based in Greenwich, Connecticut that utilized absolute-return trading strategies combined with high financial leverage. In 1998 they narrowly avoided bankruptcy when "a group of its major creditors worked out a reconstructing deal that recapitalized the firm" Haubrich (2007, p. 1). However some say the crisis almost "blew up the worlds financial system" Jorion (2000, p. 24). It is now known that LTCM severely underestimated its risk due to its reliance on short-term history and risk concentration Jorion (2000). As such, the LTCM crisis provides an excellent example of poor risk management in the extreme.

5.   PAIRS TRADING

5.1   Stationarity of variables

Pairs trading relies heavily on econometrical tools like OLS. The assumptions made for OLS are well known and described in almost every Econometric book, instead some assumptions will be skipped in order to focus on what is particularly important for this study, that is stationarity and non stationarity.

Stationary variables should be reverting back to its mean, in other words mean reverting. "Formally, a time series y_t is stationary if its mean and variance are constant over time, and if the covariance between two values from the series depends only on the length of time separating the two values, and not on the times at which the variables are observed."Hill, et al. (2008, p. 476)

Figur 5-1 Price data

$E(y_t) = \mu$ (Constant mean)

$var(y_t) = \sigma^2$ (constant variance)

$cov(y_t, y_{t+s}) = cov(y_t, y_{t-s}) = \gamma_s$ (Covariance depends on s and not t)

Ingersoll (1987)

Figur 5-2 First difference

The reason for needing to know if a time series is stationary or not is due to the danger of retrieving significant values from a regression on variables that is unrelated. Running OLS on non-stationary variables can create this problem and analyses like this are called spurious regressions. To the right is an image of the stock prices of Google, as you can see it is not mean reverting, it looks like it has both cycles and is trending.

By performing analysis on financial time series like equity's, it will be evident that the prices are non-stationary. Hill, et al. (2008, p. 477) state that one of the reasons that stock prices are not stationary is the chance of bankruptcy. To handle the implication of non-stationary data advice is often given to take the first difference of the time series.

Resulting in a price change and not the change itself. Mathematically the first difference is often shown as $\Delta y = y_t - y_{t-1}$.

From the difference in Google prices, it can be noted that the fluctuations seemed to be centered on zero. This is an indication that the time series is now stationary. Another indicator if a time series is stationary or not can be found by looking at the autocorrelation of a time series.

Below in figure 5 you can see the autocorrelation of Google with its 60 lags for the no stationary time series, and the autocorrelation of the same company done on the first difference in figure 6. As you can see the autocorrelation does not look like it's dying off on the price data, but on the first difference it looks too be stationary. For the autocorrelation plot (figure 6) of the difference it also seems to be mean reverting, where its mean is zero.
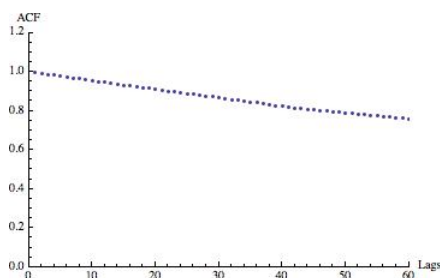
Figur 5-3 Autocorrelation Function Price Data
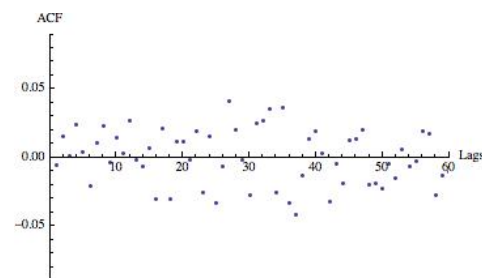
Figur 5-4 Autocorrelation Function First Difference

5.2   Co-integration

The Economic theory whereby time series propose dynamism and the practice of modeling relationships between economic variables has long been existent. It was discovered at an early stage that economic variables in many cases are non-stationary. Accrued knowledge regarding analysis of non-stationary variables has suggested that one should take the first difference transforming the data to become stationary.

However, it was Granger who suggested in 1981 that a relationship between co-integration and error correction models existed. Although it was first coined by Engle and Granger (1987) providing representation, estimation and testing. They discovered that if two time series are I(1)- in other words non stationary, one would expect the difference

between the two variables to be non stationary - I(I). More precisely, one would expect that $y_t - \beta x_t$, where both series are I(1) would produce another time series of I(1). However, in some cases $e_t = y_t - \beta x_t$, is I(0). In this case e_t is said to be the long run equilibrium and the deviation from this equilibrium would therefor be stationary with constant variance. Another noticeable point is that y_t and x_t share similar stochastic trends. Since the discovery of this relationship, this theory has been applied to many different areas, especially within macroeconomics, in aspects such as consumption and long and short-term interest rate, among others. It was through this relationship that pairs trading was developed.

In fact, the two economists Engle and Granger (1987) that made this discovery received the Nobel Prize in economics.

Post the publication a number of papers were dedicated to the discovery, in particularly Johansen (1991) provided a study focusing on the likelihood ratio test of cointegration rank. "Conducting inference on the number of cointegrating relations as well as the structure of these without imposing a priori structural relations. This is accomplished by fitting the general imposing Vector Autoregressive Model(VAR), which is used to describe the variation of the data, and then formulating questions concerning structural economic relations as hypotheses on parameters of the VAR model."Johansen (1991, p. 3). The paper provided an step by step procedure for addressing problems finding cointegrating relationships in nonstationary data, problems regarding estimating of this relationship and last testing economic hypothesis about the structure.

5.3   Method
One of the pioneers of pairs trading, Nunzio Tartaglia, states the reason why the method works is due to humans having a tendency to trade on stocks that are going up and not stocks that are going down, and that it is this psychology that has made the strategy so popular. Avellaneda and Lee (2010) believe the strategy works due to undisciplined investors who have overreacted. The theory of overreaction has been extensively

reviewed[2], although not necessarily in regards to pairs trading but more towards the stock market in general.

Another argument for the success of pairs trading is the market neutral strategy. By taking many positions both short and long the portfolio does not co-vary with the market, and the beta with the portfolio and the market portfolio is zero. Having a market neutral strategy would also be important when buying options. Because one call and one put is being bought it would be expected that the covariance with the market should be zero, and that the market neutrality would still be valid.

5.4   Trading rules

There are countless different signals used when setting up a trading system. How the operator sets up the trading algorithm will depend on his or her preferences. Some operators are concerned with long-term movements while others are concerned with short-term movements; the range can vary from years to seconds.

A very simple method used by Gatev, et al. (2006) involves extracting the standard deviation of the spread of the normalized price series during the preface when identifying co-integrated pairs. The deviation would be used as a signal during the trading period. When the spread increases above a certain number of standard deviations the trader goes short in the lowest selling stock and long in the highest. Interestingly with such a simple strategy, a profit was provided after the transaction cost.

Another method is to create a moving average from the shares and adding a standard deviation above and below the stocks. The moving average would operate as the equilibrium and a break through two standard deviations above or below the mean would indicate that the stocks have moved to far apart and are likely to return back to equilibrium.

2
  Kiseok Nam, Chong Soo Pyun, and Stephen L Avard, "Asymmetric Reverting Behavior of Short-Horizon Stock Returns: An Evidence of Stock Market Overreaction," *Journal of Banking & Finance* 25, no. 4 (2001).

5.5   Some clarifications

In theory, pairs trading is a rather simple concept. It attempts to exploit the relationship where the law of one price is violated. Within academia the focus has been towards two main forms of pairs trading. The first is statistical arbitrage, where the purpose is to discover mispricing by analyzing a time series of price information. The second is risk arbitrage, which according to Hill, et al. (2008) is about the strategy in regards to fusion between two companies. There has often been confusion surrounding the statement that statistical arbitrage is risky. This is due to the aforementioned theory that arbitrage is suppose to be risk free. Yet, this is not the case when it comes to pair trading. Another common confusion is the mix of terms, leading to the assumption that risk arbitrage and statistical arbitrage are the same, which is not the case.

Pairs trading includes a range of different strategies and investment programs according to Gatev, et al. (2006), where the similarity between them can be summarised as:

I.   Signals developed for entering or exiting the market are systematically and rule based, unlike fundamental analyses.
II.   Unequal positions means that collecting the trades together gives a beta of zero in regards to market and is therefore market neutral.
III. Mechanism to generate excess return is statistical.

The spread between the co-integrated pairs are said to be stationary, which would mean as fore mentioned, that it is an average reverting process with constant variance. It should be the case that the average of the spread is the equilibrium within that pair. It could be argued that it would therefore be easier to estimate when the process is starting to diverge than it would be to estimate when it starts to converge. The problem with this assumption is the cost of transaction; a small deviation from the equilibrium would not necessarily be large enough to cover the transaction cost you, which could result in many trades not providing any profit.

Though the cost of options is used in this paper, a similar problem occurs. To avoid the problem of entering many positions that do not cover transaction costs, the trade signal will be made on convergence. This is due to the fact that when looking at a normal distribution, the further away one moves from its mean the more 'extreme' are the events, and one could expect fewer observations in this 'extreme' area. However when observing a value very close to the mean it does not provide increasingly high probability that the next value would be even further away from the mean. Therefore, entering on divergence can impose higher cost because of the lack of a good indicator; consequently this would mean that one could take many positions, which would not diverge further from equilibrium.

5.6   PAIRS TRADING AND OPTIONS

5.6.1   Problems when using options

When applying options to pairs trading there are adjustments that need to be made. First and foremost is the premium. The cost of the premium means that if the stocks do not diverge too far from each other it would result in a loss, even if the stocks do as predicted and converge back to equilibrium one might end up with a loss if the spread is not large enough when entering the market.

The timeframe for the option is also crucial. When going short and long in two stocks the time period does not affect the cost in the same way as it would do with options. If the timeframe was increased it may result in a cost of the option being higher than the return from the trade. The other scenario that would generate problems would be if the timeframe of the option were too short, resulting in the payment of the premium for an option, which has not reached the stage of converging back to the equilibrium.

5.6.2   OBSERVATIONS

There are several ways in which the stocks are able to move together and it is important to clarify how this can occur and what the implication of it would be. I am not going to outline all possibilities, since there could be many, but I shall detail some. Of course the easiest and most optimal way is that the two time series within a pair move from each

other and then convert back to a mean without any form of trending. This would mean that the equilibrium would be at the same price level at convergence as it was when they started to diverge from each other.

The other option would be that the stocks converge from each other but do trend. In this case there could be several options. One option could be that stock number one, with the highest value, increases in price and never diverges back, but stock two, with the low price, follows the upward trend at a higher rate then stock one. In this case the time it takes for the stocks to converge depends on how large the difference of increase is between the two stocks. This implies that one stock is tracking the other stock.

When looking at two stocks that have equal value it can be hard to establish cause and affect relationships. However, if looking at an index vs. a stock theory, simple reasoning would make you see that clearly the stock of one company does not move the value of a large index. The relationship would in fact be the other way around.

6.   METHODOLOGY

6.1   TOOLS:
I chose to use Mathematica because of its versatility, options and simplicity when it comes to writing code, and in particular because there are such a large number of books that provide information about the program.

6.2   DATASET:
The sample was obtained from all shares under the S&P 500 from the period January 2000 to August 2013 through backward looking. The date was extracted as daily close prices from The Centre for Research in Securities Prices (CRSP) and contained every share within the index or market. CRSP. is one of the 12 Research and Learning Centres at Chicago Booth. They aim to bridge theory and practice with trusted data solutions.

CRSP data is used by investment practitioners to backtest strategies, as well as to benchmark investment performance. CRSP. has been responsible for advancing the body of knowledge in finance, economics and other related fields since 1960. They have

achieved this by providing research-quality data to both scholarly researchers and investment practitioners alike.

Implied volatility and historical volatility was extracted for Optionmetrics, which provides historical option price data, historical volatility implied volatility etc. With over 300 institutional subscribers it is regarded as a high quality data provider. OptionMetrics (2014)

Number of observed days was 3264 days covering a total of 13 years. The sample period includes noteworthy events like the Dot Com Crash (DCC) and the Global Financial Crisis (GFC). The purpose of identifying them is so that we can compare the return generated during these events with the period's return that does not contain extreme events. Data extracted from Thomson Reuters are 3 month US Treasury bill yield.

Because of the high liquidity requirements for securities added to the S&P 500 index there was no need to initially exclude any securities from the sample. The committee selecting the securities has requirements of market capitalization that is greater than or equal to US$ 4.0 Billion among others.

The shares are divided into Global industrial classification sub industries to avoid the problems pointed out by Do and Faff (2012, p. 264), who state that:
Generating pairs over the entire portfolio can create pairs that are not closely related to each other and therefore would violate the assumption of the law of one price, due to the fact that they are not close economic substitutes.

When the pairs are matched only on how close they move together there is a possibility of neglecting pairs that are co-integrated and have a large reversal in the price spread. It is important to note that the spread and reversal effect is what generates the return. Generating pairs that are moving so closely together that the opportunity of divergence does not come into play would provide no opportunity to gain profit at all.

When pairs move closely together there is a chance of opening a trade where the reversal effect is not large enough to cover the bid-ask bounce and transaction cost even though the stocks end up converging back to equilibrium.

The test period for co-integrating is based on Do and Faff (2012, Do and Faff (2010, Gatev, et al. (2006). All three papers used a 12-month period to test and extract pairs, following 6 months of trading. This would mean that at first, extract 252 trading days, and then if the pair is co-integrated during this period, trade on that pair. After a 6 month period go back one year from the last date to test again to see if the pairs are co-integrated. As long as the pairs are co-integrated this process continues. The number of cointegrated period's were 16 735, while number of cointegrated pairs were 1199.

Sub Industry Classification	Number of pairs	Sub Industry Classification	Number of pairs
Life Sciences Tools & Services	6	Fertilizers & Agricultural Chemicals	3
Technology Hardware, Storage & Peripherals	21	Air Freight & Logistics	6
Health Care Distributors	6	Household Products	6
Health Care Equipment	78	Cable & Satellite	6
Property & Casualty Insurance	21	Specialized Finance	3
IT Consulting & Other Services	3	Restaurants	10
Application Software	10	Apparel, Accessories & Luxury Goods	13
Semiconductors	65	Hypermarkets & Super Centers	1
Data Processing & Outsourced Services	55	Communications Equipment	15
Multi-Utilities	91	Railroads	6
Electric Utilities	78	Diversified Support Services	1
Independent Power Producers & Energy Traders	1	Integrated Telecommunication Services	10
Managed Health Care	10	Drug Retail	1
Life & Health Insurance	21	Integrated Oil & Gas	6
Pharmaceuticals	36	Airlines	1
Multi-line Insurance	10	Diversified Chemicals	6
Residential REITs	3	General Merchandise Stores	6
Internet Software & Services	10	Health Care Services	6
Biotechnology	21	Homebuilding	3
Semiconductor Equipment	3	Industrial Conglomerates	3
Electrical Components & Equipment	10	Movies & Entertainment	3
Asset Management & Custody Banks	36	Research & Consulting Services	3
Specialized REITs	10	Oil & Gas Drilling	15
Internet Retail	10	Industrial Machinery	36
Automotive Retail	6	Specialty Chemicals	10
Insurance Brokers	1	Oil & Gas Exploration &	1

		Production	
Oil & Gas Exploration & Production	119	Investment Banking & Brokerage	6
Industrial Gases	3	Automobile Manufacturers	1
Electronic Components	1	Trading Companies & Distributors	1
Steel	3	Construction & Engineering	3
Personal Products	1	Retail REITs	6
Paper Packaging	3	Apparel Retail	6
Consumer Finance	6	Consumer Electronics	1
Aerospace & Defense	55	Leisure Products	1
Diversified Banks	15	Health Care REITs	1
Regional Banks	36	Home Improvement Retail	1
Computer & Electronics Retail	1	Casinos & Gaming	1
Oil & Gas Equipment & Services	15	Advertising	1
Metal & Glass Containers	1	Electronic Manufacturing Services	1
Semiconductors	1	Department Stores	3
		Oil & Gas Storage &	
Auto Parts & Equipment	1	Transportation	6
Systems Software	10	Food Retail	2
Packaged Foods & Meats	45	Home Furnishings	1
Construction Machinery & Heavy Trucks	5	Tobacco	6
Broadcasting	3	Oil & Gas Refining & Marketing	3
		Environmental & Facilities	
Soft Drinks	7	Services	3
Hotels, Resorts & Cruise Lines	6	Specialty Stores	6

6.3   Sub sample

The datasets are usually split into separate samples in the periods where extreme
events are hitting the market. This is to separate the abnormalities from the general
behavior of the stock market.

Do and Faff (2012, p. 264) reported a change in the profitability during the two bear
markets (DCC & GFC) and even if the profitability of pairs trading on average
declined from 2000 to 2009 there was an increase in the profitability during the
DCC and the GFC. It is therefore important to note these periods when looking at the
result of the trading strategy.

Sub sample period	Commencing date	Ending date9
DCC(2000-2002)	Nasdaq reaches all time high closing price(end of the day) at 27.03.2000, whit a closing price of 4704.72	January 2002 Amazon announces first quarter profit.

Interval (2003-2008)	Normal trading period starts 01.01.2003	01.08.2008
GFC (2008-2009)	September 14.th 2008 Lehman Brothers Collapse. And the freezing of wholesale money markets. "Ted Spread"	31/12/2009 Arbitrary date to mark end of turbulent 15-month period.

By looking at an example we can demonstrate how to divide the subsamples. The "TED spreads" are used as measures of the credit risk for inter-bank lending. From the graph it can be seen that the TED spread is the difference between the Libor and the Treasury bill. A higher TED Spread is an indication that the counterparties see each other as more risky. The sharp increase occurred around September 14.th 2008. This is used to indicate the start of the GFC.
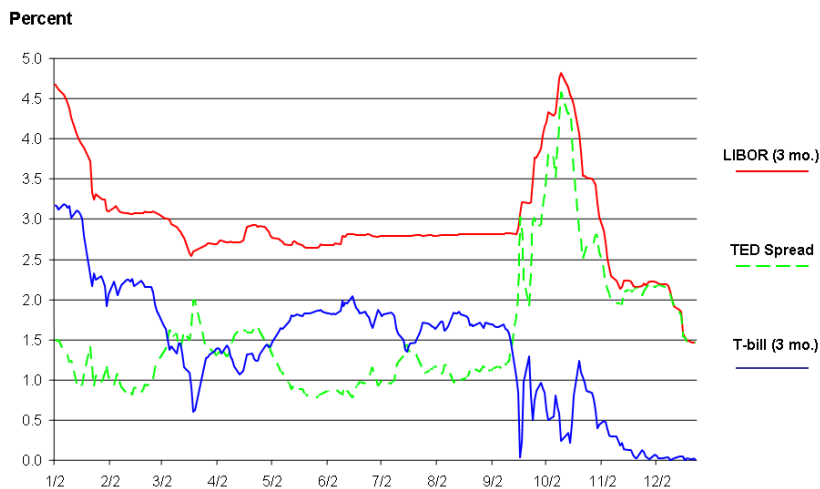
Figur 6-1 TED Spread
"Global Financial Crisis of 2008â   "2009 | the Ted Spread Â   " an Indicator of Credit Risk Â   "
Increased Dramatically During September 2008" 2013)

6.4   Selecting pairs

When picking the pairs we start by generating all subsets of two under each industry classification group. Regressing one security upon the other gives us the residuals, which is needed for further analysis. The deriving of the coefficients and the residuals is

explained in nearly every econometric textbook[3], therefore the formulas will only be provided after the derivation.

$$b = (X'X)^{-1}X'y$$

$$\hat{y} = X(X'X)^{-1}X'y$$

$$P_x = X(X'X)^{-1}X'$$

$$\hat{e} = (I - P_x)y$$

Where,

$$(X'X) = \begin{matrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{matrix}$$

X is an nx2 matrix.
b is the coeffecient. In this case b is a 2x1 matrix because we are including the intercept.
P_x is the projection matrix
$\hat{e} = (I - P_x)y$ is the estimated residuals.

This is the straightforward process for simple OLS. Where X is a Nx2 matrix and the transpose of it provides the 2xN matrix. In column one are the 1 values and in column two are the X_n values. P_x is the projection matrix and the residuals(u) is as seen in the last formula derived by subtracting the observed y values from the estimated $\hat{y}$ variables.

Unit root test for stationarity is then conducted to test whether the time series is stationary or not. It is known as a augmented Dickey-Fuller (ADF) test. "Since we cannot observe e_t, we test the stationarity of the least square residuals, $\hat{e}_t = y_t - b_1 - bx_t$, using a Dickey-Fuller test."Do and Faff (2010).

$$\Delta \hat{e}_t = \alpha + \gamma \hat{e}_{t-1} + \sum_{s=1}^{m} a_s \Delta \hat{e}_{t-s} + v_t$$

Where;

---

[3] William H. Greene, *Econometric Analysis*, 7th ed. (Boston: Prentice Hall, 2012); R Carter Hill, William E Griffiths, and Guay C Lim, *Principles of Econometrics*, vol. 5 (Wiley Hoboken, NJ, 2008); Marno Verbeek, *A Guide to Modern Econometrics*, 4th ed. (Chichester: Wiley, 2012).

$$\Delta \, \hat{e}_t = \hat{e}_t - \hat{e}_{t-1}$$

$\gamma = (p - 1)$ is the first coefficient

$\Delta \hat{e}_{t-1} =$ first difference $(\hat{e}_{t-1} - \hat{e}_{t-2})$

$a_s =$ the second coefficient

α is the intercept

v_t is the estimated residuals

$$H_0 : \gamma = 0, \qquad H_1 : \gamma = 1$$

If H_0 is not rejected it would be concluded that the series is nonstationary. This can easily be seen from the model due to the fact that a random error term would be left if γ equals zero. By rejecting H_0 it is concluded that the time series is stationary. The summation of the lags of the first difference is calculated to make sure that the autocorrelation is taken out of the residuals so that one is left with v_t, which is random. The autocorrelation that is persistent can be seen in the earlier plot of the autocorrelation function. The number of lags can be increased until it is certain that the v_t does not contain an autocorrelation.

Some econometricians have discussed the problem of using an ADF test on the residuals. "Because the OLS estimator 'chooses' the residuals in a cointegrating regression to have as small a sample variance as possible, even if the variables are not cointegrated, the OLS estimator will make the residuals 'look' as stationary as possible. "Hill, et al. (2008, p. 489). This means that by using a simple test such as the ADF test it may result in getting a type one error, where the hypothesis of non-stationary too often is being rejected.

We can overcome this problem by using Phillips and Ouliaris (1990) (PP) test. The test is very similar to the Dicky Fuller test, but there are a couple of advantages to using PP. As with the ADF test, the PP test corrects for autocorrelation, but at the same time a lag length does not need to be specified. The other advantage is that by modifying the Dickey Fuller test statistic, the test is robust in relation to the general form of heteroskedastisity in the error term.

To simply describe the test I have extracted the setup from the manual:STATA"Http://Www.Stata.Com/Manuals13/Tspperron.Pdf - Google Search" 2014)

$$y_i = \alpha + \rho y_{i-1} + e_i$$

$$z_p = n(\hat{\rho}_n - 1) - \frac{1}{2} * \frac{n^2 \sigma^2}{2 s_n^2}(\hat{\lambda}_n^2 - \hat{\mathcal{V}}_{0,n})$$

$$Z_t = \sqrt{\frac{\hat{\mathcal{V}}_{0,n}}{\hat{\lambda}_n^2}} * \frac{\hat{\rho}_n - 1}{\sigma^2} - \frac{1}{2} * (\hat{\lambda}_n^2 - \hat{\mathcal{V}}_{0,n}) * \frac{1}{\hat{\lambda}_n} * \frac{n * \hat{\sigma}}{s_n}$$

$$\hat{\mathcal{V}}_{j,n} = \frac{1}{2} \sum_{i=j+1}^{n} \hat{u}_i \hat{u}_{i-j}$$

$$\hat{\lambda}_n^2 = \hat{\mathcal{V}}_{0,n} + 2 \sum_{j=1}^{q} (1 - \frac{j}{q+1}) \hat{\mathcal{V}}_{0,n}$$

$$s_n^2 = \frac{1}{n-k} \sum_{i=1}^{n} \hat{u}_i^2$$

An alternative approach is to account for serial correlation in the regression residuals, and by doing so one could use Newey-West methodology. In essence, this is the Phillips Perron test but then applied to the regression residuals. This has been programmed into mathematica, where the result has been back tested with Eviews and Stata to make sure that the programming is giving the correct result.

6.5   The ordinary least square and its correction to Newey West (HAC. standard erros).

$$w_j = 1 - \frac{j}{H}$$

$$H = \sqrt[4]{T}$$

$$\bar{V}(b) = (\sum_{t=1}^{T} x_t(x_t)\grave{})^{-1} TS(\sum_{t=1}^{T} x_t(x_t)\grave{})^{-1}$$

$$S^* = \frac{1}{2} * \sum_{t=1}^{T} e_t^2 x_t x + \frac{1}{T} * \sum_{j=1}^{H-1} w_j * \sum_{s=j+1}^{T} e_s e_{s-j}(x_s x\grave{}_{s-j} + x_s x\grave{}_{s-j})$$

The square root of the diagonal of matrix $\hat{V}^*$ is the newey west standard errors.

The last formula is giving the HAC standard errors. Since this is a very straightforward concept covered in almost all econometric books, no further explanation will be entered into. For reference, the specific formulas and a deeper explanation for this is provided in Verbeek (2008, p. 126)

6.6   Trading strategy

The trading indicator used reflects Reverre (2001, p. 476),

where

$$m = \frac{\Delta - MA_{50}}{\sigma_{50}(\Delta)}$$

Where:

Δ, is the difference between the two shares.

MA_{50}, is the moving average over the last 50 days on the difference between the shares.

$\sigma_{50}(\Delta)$, is the standard deviation of the difference the last 50 days.

The prices used for calculating trading signals are close price adjusted for dividends. Since we divide by the standard deviation we end up with a distribution where the standard deviation is equal to one. Indicator m is therefore representing the mispricing in terms of standard deviations. When m is equal or less than -2 we buy a put on stock 1 and a call on stock 2.

Time to expiry is set to from 0 to 60 days, exiting on the third Friday in any month, if this Friday does not exist due to holidays etc. the closest date before is chosen. Since one year of significant co-integration provides a trading period of 6 months, a limit has been set such that it is not possible to enter a trade after the third Friday in the last month of the 6-month trading period. For companies entering the S&P500 prior to the start of the trading these months would be June and December. However if the company does enter the S&P500 after 2000, the 6th month would be dependent on when the date the company entered the S&P500.

After entering a trade, a new option price is calculated every day, where only the strike is fixed at the same level as when entering the trade. If the value of the option is increased, generating an annual return of 20 or more percent the paper is sold. If during the period this does not occur the option is held to expiry. It is important to notice that the all options entered are at the money(ATM).

6.7   The Black –Scholes –Merton Model

The Black-Scholes-Merton-Model was developed in the 1970's. The development of the model was based on many of the previous assumptions that were applied when trying to estimate option prices. However Fischer Black, Myron Scholes and Robert Merton had a breakthrough - and that was choosing the correct discount rate to use for the payoffs. The payoffs had already been correctly calculated, but by using the capital asset pricing model Black, Scholes and Merton were able to show the relationship between the market-required return on the option and the required return on the stock.

"In general, it seems clear that the higher the price of the stock, the greater the value of the option. When the stock price is much greater than the exercise price, the option is almost sure to be exercised. The current value of the option will thus be approximately equal to the price of the stock minus the price of a pure discount bond that matures on the same date as the option, with a face value equal to the striking price of the option."Black and Scholes (1973, p. 638)

When calculating the options price we need to make assumptions. A probability distribution for the stock price development is required.

The Black-Scholes-Merton models underlying the assumptions derived by Black, Scholes and Merton are as follows:

I.   Stock price behavior corresponds to the lognormal model with μ and σ constant.

II.   There are no transaction costs or taxes. All securities are perfectly divisible.

III. There are no dividends on the stock during the life of the option.

IV.   There are no riskless arbitrage opportunities.

V.   Security trading is continuous.

VI.   Investors can borrow or lend at the same risk free rate of interest.

VII. The short-term risk free rate of interest, r, is constant

Black and Scholes (1973, p. 640)

μ: Expected return on the stock

σ: Volatility of the stock price

Thus we assume:

$$\frac{\Delta S}{S} \sim \phi(\mu \, \Delta t \, , \sigma^2 \, \Delta t)$$

Where ΔS is the change in stock price in time Δt. As $\phi \, (m, v)$ a mean and variance can be seen as well as the denotation of normal distribution. Hull (2008) expands upon this by stating that due to this assumption, the stock price of any further time would be expected to have a lognormal distribution.

The fact that the lognormal distribution only operates as zero or above is expected since a company with a stock price at zero would be bankrupt. If we plot the Google stock prices in a histogram we can see how the distribution looks to be an approximate for the lognormal distribution.

The expected value (or mean) of S_t is $S_0 e^{\mu T}$ and its variance is $S_0^2 e^{2\mu T}(e^{\sigma^2 T} - 1)$
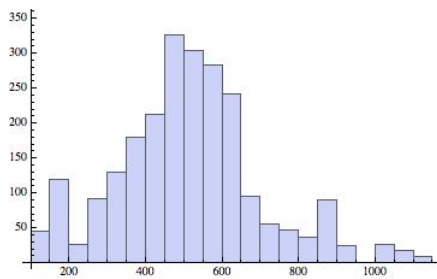
Figur 6-2 Google Stock Price

When a variable has the lognormal distribution, then its natural logarithm is normally distributed. Looking at the return of the security in figure 9, it is apparent that the return also simulates a normal distribution.
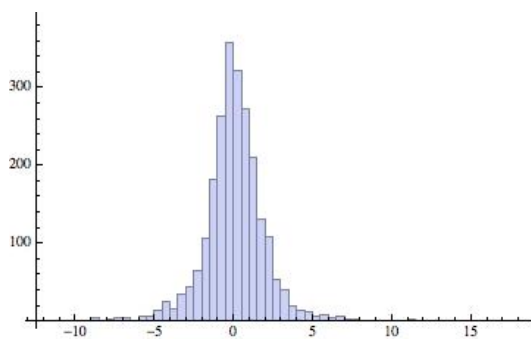
Figur 6-3 Google Stock Price

It is worth noting with the stock return, that the "fat" tails are representing, for example, the GFC, as well as other such extreme events.

6.8   Expected return.

It is to be expected that when an investor is looking at different securities they would demand higher return from the securities that impose more risk than from the securities with a lower risk. Therefore a lower risk implies lower return and a higher risk implies higher return. The interest rate also affects the stock return, the higher the interest rate the higher the expected return for any stock, because the risk that would result from taking money out of the bank and investing it in a security would need to be paid for.

Hull (2008) Specifies that the stock price return is

$$S_T = S_0 e^{RT} \qquad \text{-->} \qquad R = \frac{1}{T} Ln \frac{S_T}{S_0},$$

"R the continuously compounded return actually realized over a period time of length T years."Hull (2008)

When attempting to estimate the price of the options, the general Black-Scholes-Merton Model will be used. In order to estimate the price it is crucial to be able to apply the correct volatility.

6.9   Volatility

Volatility in the market originates from new information reaching the market. "New information causes people to revise their opinions about the value of the stock" Hull (2008, p. 302). This reevaluation of the stock, in some cases, would change the price of the stock, causing volatility in the market.

By having historical prices for a stock it is easy to estimate historical volatility with standard deviations. The stock prices used in these calculations come in intervals, and the intervals used through this paper are in days.

Definitions:

$n + 1$: Number of observations

S_i: Stock price at the end of ith interval, where i=0,1…n

τ: Length of time interval in a year

$$u_i = \ln \left(\frac{S_i}{S_{i-1}}\right)$$

To estimate the standard deviation of u_i:

$$s = \sqrt[2]{\frac{1}{n-1} \sum_{i=1}^{n} (u_i - \bar{u})^2}$$

or

$$s = \sqrt[2]{\frac{1}{n-1}\sum_{i=1}^{n}u_i{}^2 - \frac{1}{n(n-1)}\left(\sum_{i=1}^{n}u_i\right)^2}$$

u_i = Where $\overline{u}$ is the mean of the u_i

Hull (2008, p. 306) states that "difficulties arises when choosing the value for n". The suggestion is that a larger number of n would provide a more accurate value. However, the larger the timeframe, the more likely it is that the old values would have no relationship to the next period's volatility. It is exactly this point that makes it difficult to choose the value of n.

Take note that there is an importance when it comes to trading day vs. calendar day. The volatility per annum is calculated from the volatility per trading day using the formula:
Volatility per annum = Volatility per trading day
$* \sqrt{Number\ of\ trading\ days\ per\ annnum}$

There are many different recommendations when it comes to picking the time frame that should be used. Hull (2008), suggests using the same time interval as period to expiry when estimating volatility. For example if you are estimating an option with a year to expiry you could use one-year historical volatility.

Although the estimated volatility has been extracted from OptionMetrics, the use of timeframe has been chosen on the basis of Hull (2008) recommendations. Therefore since the longest timeframe a option can have is set to two months, I have chosen to extract 60 days volatility from dataset provided from OptionMetrics (2014). How realistic this volatility is certainly up for discussion as stochastic volatility is an extremely large and challenging area within finance. One could try to test with different GARCH models however in the industry one would expect hedge funds to develop specific models for each security.

6.10 The Black-Scholes-Merton Pricing Formulas

The presentation here is the developed through the studies of Merton (1971) ,Black and Scholes (1973) and Merton (1976)

$$c = S_0 e^{-qT} N(d_1) - K e^{-rT} N(d_2)$$

$$p = K e^{-rT} N(-d_2) - S_0 e^{-qT} N(-d_1)$$

where ,

$$d_1 = \frac{\ln(S_0/K) + (r - q + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(S_0/K) + (r - q - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

Black and Scholes (1973, p. 644)

Please note that the call and the put in these formulas have the q value in them, which represent dividends, this would mean that we are able to take into account the dividends when there are stocks that pay dividends.

Where:

N(x) → cumulative probability function for a standardized normal variable.

S_0 → is the stock price

K → is the strike price

r → is the risk free interest rate

T → is the time to expiration and

σ → is the volatility of the stock price

N(d_1[4]) and N(d_2)

---

[4] The limit of the natural logarithm of x as x approaches one is zero, by entering at the money the d1 and d2 as in the formula gets reduced.

"The $N(d_2)$ is the probability that a call option will be exercised in a risk neutral world. And $S_0N(d_1)e^{rT}$ is the expected stock price at time T in a risk neutral world when stock price less than the strike price are counted as zero. " Hull (2008)

7.   Result

7.1   Estimation of option prices

The dataset from OptionMetrics provided only the bid ask spread. A mean was therefore chosen for the options with high liquidity, and then the implied, and historical volatility was used to estimate the prices to see if one were able to get an estimated option price similar to the historical price from OptionMetrics. This is an example of the real price seen in the market and the historical price.
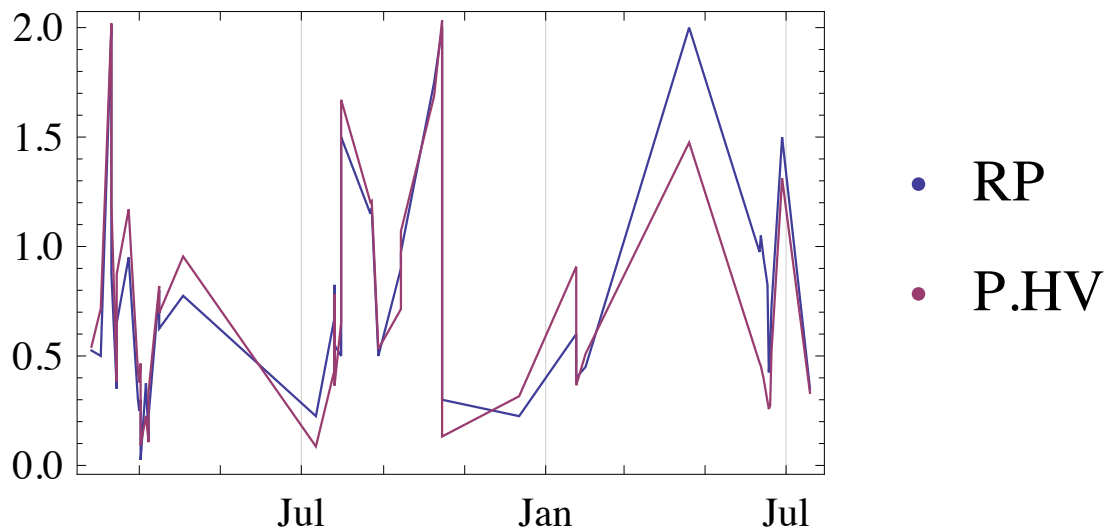
Figur 7-1 Comparison between real historical price and estimated option price.

	ABC--Real Price
Mean	0.744565
Median	0.6125
Max	2
Min	0.025
Variance	0.250414
Std.Dev	0.500414
Skewness	1.1237
Kurtosis	3.67129
Coefficient Variation	0.672089
T.Stat.Jarque	12.7022
PValue	0.0262362
Sum	34.25
Sum Sq. Dev	11.2686
Observation	46

	ABC--Price Hist.V
Mean	0.707429
Median	0.524613
Max	2.03268
Min	0.0865404
Variance	0.249823
Std.Dev	0.499823
Skewness	1.06958
Kurtosis	3.42119
Coefficient Variation	0.706534
T.Stat.Jarque	10.7759
PValue	0.0337196
Sum	32.5417
Sum Sq. Dev	11.242
Observation	46

As we can see from this result the BSM model does a fairly good job estimating the option price. However this was not the case for all options. A larger output was generated and can be found in Appendix A.

7.2   CALCULATION OF RESULT

Because of the extreme result I see it as necessary to provide an example of how the return is calculated. Firstly, take a look at the shortenings made:

D= Days to expiry,

PA= Price on the underlying security A

PB= Price on the underlying security B

TI= Trading Indicator

VA=Volatility on the underlying security A

VB= Volatility on the underlying security B

DA= Dividend payed by security A

DB= Dividend payed by security B

Call A= Equals daily Option price with same strike as set when entering; in this case I entered 25 feb. 2002.

RP= Return on period

D	Date	PA	PB	TI	VA	VB	DA	DB	I	DDA	DDB	Call A	RP
18	25feb2002	27.9	44.895	-2.455555969	0.287691	0.204184	0.095		0.072280463		0.099819462	0.927153551	
17	26feb2002	27.67	46.025	-3.725668515	0.28808	0.213576			0.072280463	0	0.099839506	0.780879437	
16	27feb2002	27.68	46.09	-3.288944313	0.284411	0.210828			0.072280463	0	0.099859553	0.747250899	
15	28feb2002	27.2	46.77	-3.774775738	0.283558	0.211043			0.072280463	0	0.099879605	0.503463954	

14	01mar2002	31	47.255	-0.620171434	0.434333	0.211771		0.072702019	0	0.099899076	3.431563031	2.701180917
11	04mar2002	32.71	48.605	0.259558743	0.457281	0.220597		0.072702019	0	0.099959618	4.951306313	
10	05mar2002	31.95	47.525	0.050994054	0.461161	0.229153		0.072702019	0	0.099979807	4.209330214	
9	06mar2002	32.92	46.525	1.818603651	0.460397	0.232047	0.1	0.072702019	0	0	5.119144282	
8	07mar2002	32.78	45.665	2.290313544	0.455113	0.234269		0.072702019	0	0	4.964213132	
7	08mar2002	34.06	46.85	2.204388221	0.463251	0.241964		0.072702019	0	0	6.219656172	
4	11mar2002	33.7	46.81	1.839932127	0.468752	0.246703		0.072702019	0	0	5.832491899	
3	12mar2002	33.04	46.91	1.200370208	0.472755	0.244862		0.072702019	0	0	5.164340061	
2	13mar2002	32.85	46.35	1.432093089	0.467451	0.244057		0.072702019	0	0	4.966105606	
1	14mar2002	32.46	46.025	1.333418404	0.463338	0.241879		0.072702019	0	0	4.568048007	
0	15mar2002	33.08	46.885	1.114483413	0.464836	0.245706		0.072702019	0	0	0	

In the sample you can see that the price changes from 0.927153551 to 3.431563031 producing a return over the period 25feb2002 to 01mar2002 of 2.701180917. This shows how an adjustment in price can make a huge change in volatility, which in turn produces a large change in option price. Return that provided over 1000% annual return was removed from the sample since this would be unrealistic and would be the case of companies with poor liquidity, therefore the values would be unobservable in the market.

7.3   RESULT YEARLY

The result was quite extreme and so I would like to point out that the need for further studies to be done on historical option prices to see if the strategy would be profitable.

Year	Return	Sharp Ratio
2001	21.64%	4.578
2002	29.93%	5.11406
2003	27.18%	1.20492
2004	22.65%	4.12026
2005	17.62%	3.4747
2006	29.99%	2.11654
2007	27.67%	0.952911
2008	21.85%	1.73623
2009	27.33%	1.82158
2010	18.22%	1.59783

2011 30.85%   2.48927
2012 19.30%   1.79839

It is immediately clear that this is not a realistic; there are several reasons for this. However, it is mostly because of the fact that the liquidity is not taken into account. On top of that, the dividend provides a problem in the cases where the dividend is almost as large as the security price. Subtracting such a large value from an option means that the option is valued at almost zero. When an option the days after dividend is payed again is estimated and is in the money the return on the option is extreme. It is therefore made an adjustment, removing options that costs less then five cent. The sharp ratio and the return is correlated at approximately 0.92 from the Pearson Correlation test. A high correlation would be expected since return is used to calculated the sharp ratio. The sharp Ratio is a risk measure that is often used however criticisms of its use have been sine in later papers within mathematics for example Bailey, et al. (2013) Pseudo-Mathematics and Financial Charlatanism, because of the violation when using it on returns that are not normally distributed.

7.3.1   RETURN 2002
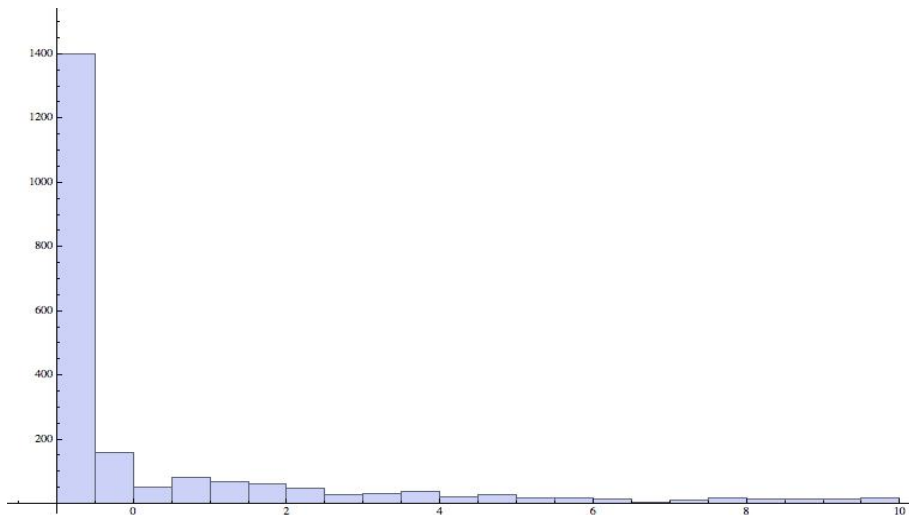
Figur 7-2 Return Plot 2002

Figure 11 shows the histogram of the return of 2002. It is clear that most of the transactions end up being out of the money and therefore produce 100% negative return.

The annual return is driven by the options that are providing extreme profit. The probability of experiencing these huge returns when trading on real world data on would expect to be extremely low. In fact any of these returns are so high that one can almost guarantee that they would not be traded in the real market.

Using a portfolio and weighting the investment based on significance level of the shares produced the same return plus minus 2% per year. There were no indication of particularly strong correlation between return and significance level. The portfolio was subject to days with huge return but also days with huge drawdowns. These drawdowns came on days where almost all of the options produced 100% negative return.

Looking at the return over the entire period shown in figure 12 provide the same result, annual return are driven by extreme profit but most of the return is negative.

Figur 7-3 Return Plot for 2002-2013

8.   WELL KNOWN BIASES

8.1   Survivorship bias

Many losing funds are closed and merged into other funds to hide poor performance, and as such, 'mutual fund attrition' can cause problems for researchers Elton, et al. (1996)

The exemption of such funds from performance studies occurs because these funds no longer exist. The causes of the attrition is either because "their performance is very poor over a period of time or because their total market value is sufficiently small so that management judges that it not longer pays to maintain the fund" Elton, et al. (1996, p. 1098). However, to only study funds that survive will in turn overstate the measured performance of the funds and thus creates survivorship bias.

In Elton, Gruber & Blake's 1996 study, they showed that survivorship bias appeared larger in the small-fund sector than in large mutual funds. Although now, most of the more recent performance studies have strived to eliminate such bias by "incorporating all available data on fund returns" Carpenter and Lynch (1999, p. 339). However several problems are still met by researchers when trying to eliminate survivorship bias. Carpenter and Lynch (1999) state that some level of survivorship bias is inevitable if complete data sets on dissolving funds are unavailable. They also state that even when samples include all available data "some methodologies introduce survivorship bias by imposing minimum survival requirements" Carpenter and Lynch (1999, p. 339).

Correction for attrition and its corresponding bias is important on many levels. Samples that do not correct for attrition will "overstate the return that mutual funds earn for their investors" Elton, et al. (1996, p. 1098). Because funds with differing objectives may have differing rates of attrition, it may also differentially impact the return reported for mutual funds with different goals and for these reasons attempts should be made to avoid it (Elton, Gruber, Blake).

8.2   Look ahead bias

Look ahead bias is found in studies that includes information or data that would not have been available or subsequently known during the time period being analysed Baquero, et al. (2005); Daniel, et al. (2009). Many researchers have found results that show look-ahead bias methodologies will materially bias statistics, by having requirements that mean funds must survive a minimum period of time after a ranking period Carpenter and Lynch (1999); Baquero, et al. (2005).

The impact of look-ahead bias is "exacerbated for hedge funds due to their greater level of total risk" Baquero, et al. (2005, p. 2). In fact, Baquero suggests that look-ahead bias poses a serious problem. In fact, without correcting for look-ahead bias, average returns of poorly performing funds may be overestimated by as much as 3.8% per year Baquero, et al. (2005)

It is considered very difficult, if not impossible, to conduct a simulation of the performance of investment strategies completely void of look-ahead bias; and one can never be 100% sure that the bias has in fact been removed. Daniel, et al. (2009). However, Daniel etc suggests that one way to address these biases "requires first recognising their existence, and then to model how survival probabilities depend on historical returns, funding age and aggregate economy-wide shocks" Daniel, et al. (2009, p. 9) .Because the hedge fund industry is highly unregulated, a careful analysis is required in order to eliminate look-ahead bias as datasets can be subject to backfilling biases. Baquero, et al. (2005).

8.3   Data Snooping

Data snooping bias is the result of using data mining to uncover relationships in data. The process of data mining involves automatically testing huge numbers of hypotheses against a single data set by extensively searching for combinations of variables that may show a correlation. Due to the fact that typically new data sets on which to test hypotheses cannot be create independently of the data, data snooping bias becomes an issue (Sullivan). Sullivan states that "the common practice of using the same data set to formulate and test hypotheses introduces data-snooping biases, which if not accounted for, invalidate the assumptions underlying classical statistical inference" Timmermann, et al. (1998, p. 98).

This problem has been practically unavoidable in analysis of time-series data and the main issue in combating data snooping lies in the lack of sufficiently simple and practical methods capable of assessing the potential dangers of data snooping in a given situation

White (2000). However, a 2000 paper by White published in Econometrica suggests that their new procedure, the 'Reality Check' provides "simple and straightforward procedures for testing the null that the best model encountered in a specification search has no predictive superiority over a given benchmark model, permitting account to be taken of the effects of data snooping" White (2000, p. 1115)

9.   Summary

9.1   Criticisms

Many research studies involve bias in one-way or another. To what extent the bias affects the results of the study varies and thus so the quality does also. In this study there are a couple of things the reader should be aware of, however, through careful reading of this paper, it may have already been noticed.

First is the 'backward ' looking on the S&P 500. This imposes survivorship bias, because you are running a test on data where only the companies that have survived through the in sample period (ISP), have been tested. If this was conducted by going long and short, it would have imposed bigger problems since there would be longer periods in trade.

Yet, buying options with the max length of two months to expiry combined with a low number of companies going bankrupt in the ISP it is expected that this bias would not have any significant effect on the results presented. Nevertheless, there are examples of companies that have been quickly removed from the index through the trading period, for example Goldman Sachs. Which could impose different results.

Second is the trading signal used. There are two things that make the trading signal questionable, the first being the squaring of the values. Squaring is a well-known way of dealing with negative values, but it provides an obvious problem, the large values are weighted more. One simple way of overcoming this obstacle, especially in a simple calculation of standard deviation is to use absolute values instead of squaring.
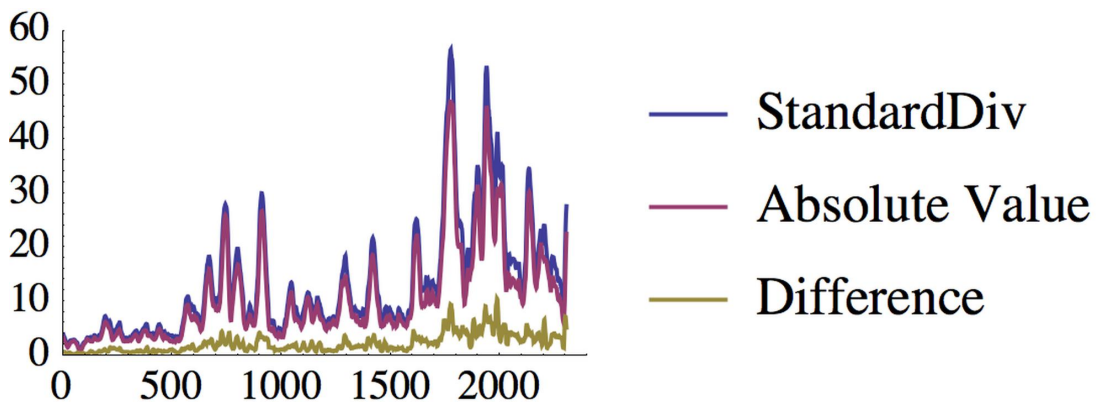
Figur 9-1 Return Plot for 2002-2013

You can see from the diagram that the difference is clearly increasing with larger values.
The reason for using standard deviation instead of absolute value is simply because it is
the same indicator used with previous research on pairs trading which would make the
research more comparable. The major problem with using this indicator is the underlying
assumption of normal distribution. I therefore tested for normal distribution between two
cointegrated shares to see if they were normally distributed. Among the 11 distribution
tests I used were Anderson Darling, Cramer-von and Mises, Jarque -Bera ALM, in the
small sample I tested none of the differences were normally distributed at a 5%
significant level.

Yet, trading strategies that rely on standard deviation is not unusual. One can also find
arguments of the central limit theory, however this is clearly misguiding since the central
limit theorem is based on a large sample without extreme values, and the stock market is
subject of fat tail distribution imposing extreme values.

Third are the adjustments of the BSM model. Even though the adjustment of dividend is
taken into account the taxation of the dividend payed out and the effect it has on the
underlying is not. This imposes a problem because the change in the underlying when
dividend is payed out differs among the companies.

Forth is the choice of volatility. Estimation of stochastic volatility often implies specific
models to specific companies due to the fact that the strategy is applied on a large scale it

was not considered possible to estimate the volatility for each company. However, could try to apply the general GARCH 1.1 model to estimate the volatility seeing to what extend the result would be altered.

Fifth is the liquidity. Although only a quick comparison was made with the bid ask price of historical option prices it was easy to see that most of the option taken place theoretically did not exist in the market. Therefore it would be interesting to see how many estimated options were found and traded in the historical one. The quick look suggested a very low percentage.

Another limitation is that the calculations have also been coded, all the way down to the matrix calculations. Before any of the coding was applied it was compared with eviews and Stata to make sure that the formulas/functions applied into mathematica were producing the correct result. The computations of different distributions were done for the normal distribution, but due to the fact that the code was extremely slow I was forced to use the code already built into mathematica.

In addition to this, the matrix calculations were shown to be significantly faster using the function like least square or linear solve in Mathematica.

One of the challenges with using a system like Mathematica is the lack of built in packages for econometrics, however the limitations of the system is only the users mathematical knowledge. Although no problems occurred when programing least square or newey west/white heteroskedasticity consistent standard errors, I did have a problem with computing Johanson Cointegration test and the timeframe I had to work within forced me to abandon it.

The bias that might occur due to this is the potential for a problem with the loss of quality of the Johansen test – which is the vector error correction model (VECM) allowing for the testing of cointegrating where it is possible to have shifts in the distribution and

where one is able to include for a short and a long term relationship. This in addition with the usage of maximum likelihood allowed the quality of cointegration to really stand out.

If given a larger timeframe I would have spent time making sure that the Johansen test were computed correctly so that I would be able to do the test.

9.2   CONCLUSION AND FURTHER RESEARCH

This paper shows how an extreme gap can occur between real world trading and estimations. Aside from this, there is little else to comment on other to state the need for further studies using real historical prices to see if the strategy is profitable.

REFERENCES

(HKIB), Hong Kong Institute of Bankers. *Credit Risk Management*. Chichester: John Wiley & Sons Ltd : distributor John Wiley and Sons Ltd : distributor John Wiley & Sons Australia Ltd : distributor John Wiley & Sons Australia Ltd : distributor John Wiley & Sons Cape Town : distributor John Wiley & Sons Inc, 2010.

Allen, Linda, Jacob Boudoukh, and Anthony Saunders. *Understanding Market, Credit, and Operational Risk: The Value at Risk Approach*. John Wiley & Sons, 2009.

Avellaneda, Marco, and Jeong-Hyun Lee. "Statistical Arbitrage in the Us Equities Market." *Quantitative Finance* 10, no. 7 (2010): 761-82.

Bailey, David H, Jonathan M Borwein, M Lopez de Prado, and Qiji Zhu. "Pseudo Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on out-of-Sample Performance." *Submitted Notices of the AMS, September* (2013).

Ball, Ray, and Philip Brown. "An Empirical Evaluation of Accounting Income Numbers." *Journal of accounting research* 6, no. 2 (1968): 159-78.

Baquero, Guillermo, Jenke Ter Horst, and Marno Verbeek. "Survival, Look-Ahead Bias, and Persistence in Hedge Fund Performance." *Journal of Financial and Quantitative Analysis* 40, no. 03 (2005): 493-517.

Barclayhedge. "Hedge Fund Relative Value Arbitrage Strategy | Hedge Fund Education." http://www.barclayhedge.com/research/educational-articles/hedge-fund-strategy-definition/hedge-fund-strategy-relative-value-arbitrage.html.

Bass, Thomas A. *The Predictors: How a Band of Maverick Physicists Used Chaos Theory to Trade Their Way to a Fortune on Wall Street*. Henry Holt & Company, 1999.

Bernard, Victor L, and Jacob K Thomas. "Evidence That Stock Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings." *Journal of Accounting and Economics* 13, no. 4 (1990): 305-40.

———. "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium." *Journal of Accounting research* 27, no. 3 (1989): 1-36.

BFIS, Bank of international settlements. "Principles for the Management of Credit Risk - Final Document." (2000).

BIS, Bank for international settlements. "Principles for Sound Liquidity Risk Management and Supervision - Bcbs144.Pdf." (2008).

Black, Fischer, and Myron Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81, no. 3 (1973): 637-54.

Bondarenko, Oleg. "Statistical Arbitrage and Securities Prices." *Review of Financial Studies* 16, no. 3 (2003): 875-919.

Carpenter, Jennifer N, and Anthony W Lynch. "Survivorship Bias and Attrition Effects in Measures of Performance Persistence." *Journal of financial economics* 54, no. 3 (1999): 337-74.

Chen, Ren-Raw. *Understanding and Managing Interest Rate Risks*. Vol. 1: World scientific, 1996.

Damodaran, Aswath. *Investment Philosophies: Successful Strategies and the Investors Who Made Them Work*. Vol. 665: Wiley. com, 2012.

———. "Not Riskless, Not Even Close..Pseudo or Speculative Arbitrage." http://people.stern.nyu.edu/adamodar/pdfiles/invphilslides/session29.pdf.

———. "Too Good to Be True? The Dream of Arbitrage."
        http://www.stern.nyu.edu/~adamodar/pdfiles/invphiloh/arbitrage.pdf.

Daniel, Gilles, Didier Sornette, and Peter Woehrmann. "Look-Ahead Benchmark Bias in
        Portfolio Performance Evaluation." *The Journal of Portfolio Management* 36, no.
        1 (2009): 121-30.

Do, Binh, and Robert Faff. "Are Pairs Trading Profits Robust to Trading Costs?". *Journal
        of Financial Research* 35, no. 2 (2012): 261-87.

Do, Binh, and Robert W Faff. "Does Simple Pairs Trading Still Work?". *Financial
        Analysts Journal* 66, no. 4 (2010): 83-95.

Easton, Steve, and Paul Kerin. "Market Efficiency and the Global Financial Crisis."
        Australian Economic Review 43, no. 4 (2010): 464-68.

Elton, Edwin J, Martin Jay Gruber, and Christopher R Blake. "Survivor Bias and Mutual
        Fund Performance." *Review of Financial Studies* 9, no. 4 (1996): 1097-120.

Engle, Robert F, and Clive WJ Granger. "Co-Integration and Error Correction:
        Representation, Estimation, and Testing." *Econometrica: journal of the
        Econometric Society* (1987): 251-76.

Fama, Eugene F. "The Behavior of Stock-Market Prices." *The journal of Business* 38, no.
        1 (1965): 34-105.

———. "Efficient Capital Markets: A Review of Theory and Empirical Work*." *The
        journal of Finance* 25, no. 2 (1970): 383-417.

Gastineau, Gary L, and Mark P Kritzman. *Dictionary of Financial Risk Management*.
        Vol. 52: John Wiley & Sons, 1999.

Gatev, Evan, William N Goetzmann, and K Geert Rouwenhorst. "Pairs Trading:
        Performance of a Relative-Value Arbitrage Rule." *Review of Financial Studies*
        19, no. 3 (2006): 797-827.

"Global Financial Crisis of 2008â "2009 | the Ted Spread Â " an Indicator of Credit
        Risk Â " Increased Dramatically During September 2008."
        http://www.alarmdetector.com/knowledge/global-financial-crisis-img4.html.

Gobel, Reyna. "The Short Squeeze Method."
        http://www.investopedia.com/articles/stocks/08/short-squeeze-profits.asp.

Greene, William H. *Econometric Analysis*. 7th ed. Boston: Prentice Hall, 2012.

Grossman, Sanford J, and Joseph E Stiglitz. "On the Impossibility of Informationally
        Efficient Markets." *The American economic review* (1980): 393-408.

Haubrich, Joseph G. "Some Lessons on the Rescue of Long-Term Capital Management."
        (2007).

Hill, R Carter, William E Griffiths, and Guay C Lim. *Principles of Econometrics*. Vol. 5:
        Wiley Hoboken, NJ, 2008.

"Http://Www.Stata.Com/Manuals13/Tspperron.Pdf - Google Search."
        https://http://www.google.no/?gws_rd=cr&ei=G_0EU9ydEKH-ygOE2oC4Ag -
        q=http:%2F%2Fwww.stata.com%2Fmanuals13%2Ftspperron.pdf.

Hull, John. *Fundamentals of Futures and Options Markets*. Pearson Prentice Hall, 2008.

Ingersoll, Jonathan E. *Theory of Financial Decision Making*. Vol. 3: Rowman &
        Littlefield, 1987.

Johansen, Soren. "Estimation and Hypothesis Testing of Cointegration Vectors in
        Gaussian Vector Autoregressive Models." [In English]. *Econometrica* 59, no. 6
   (Nov 1991 1991): 1551.

Jones, Sam, and Kaminska Izabella. "Hedge Funds Reap Rewards from Treasuries." *Financial Times*, September 15, 2010 2010.

Jorion, Philippe. "Risk Management Lessons from Long‐Term Capital Management." *European financial management* 6, no. 3 (2000): 277-300.

Kambhu, John. "Trading Risk, Market Liquidity, and Convergence Trading in the Interest Rate Swap Spread - 0605kamb.Pdf." http://www.newyorkfed.org/research/epr/06v12n1/0605kamb.pdf.

Merton, Robert C. "Option Pricing When Underlying Stock Returns Are Discontinuous." *Journal of financial economics* 3, no. 1 (1976): 125-44.

———. "Theory of Rational Option Pricing." (1971).

Nam, Kiseok, Chong Soo Pyun, and Stephen L Avard. "Asymmetric Reverting Behavior of Short-Horizon Stock Returns: An Evidence of Stock Market Overreaction." *Journal of Banking & Finance* 25, no. 4 (2001): 807-24.

Nikolaou, Kleopatra. "Liquidity (Risk) Concepts: Definitions and Interactions - Ecbwp1008.Pdf." *Working Paper Series* (2009).

Obienugh, JP. *Jonbull's Stock Guide*. Trafford Publishing, 2010.

OptionMetrics. "Optionmetrics - the Industry Standard for Options Research." http://www.optionmetrics.com/.

Phillips, Peter CB, and Sam Ouliaris. "Asymptotic Properties of Residual Based Tests for Cointegration." *Econometrica: Journal of the Econometric Society* (1990): 165-93.

Prentis, Eric L. "Early Evidence on Us Stock Market Efficiency:" Market Vs. State" Debate and Deregulation Implications." *Economics & Finance Review* 2, no. 8 (2012).

Reverre, Stephane. *The Complete Arbitrage Deskbook*. McGraw-Hill, 2001.

Sas. "Credit Risk Management | Sas." (2014).

Sawicki, Julia, and Jurac Hric. "Arbitrage Opportunities in Parallel Markets: The Case of the Czech Republic." (2001).

Shiller, Robert J. "From Efficient Markets Theory to Behavioral Finance." *Journal of Economic Perspectives* 17, no. 1 (Winter2003 2003): 83-104.

Shleifer, Andrei, and Robert W Vishny. "The Limits of Arbitrage." *The Journal of Finance* 52, no. 1 (1997): 35-55.

SWIFT. "Managing Liquidity Risk Collaborative Solutions to Improve Position Management and Analytics." (2011).

Timmermann, Allan, Halbert White, and Ryan Sullivan. "The Dangers of Data-Driven Inference: The Case of Calendar Effects in Stock Returns." Financial Markets Group, 1998.

UniCredit. "Credit Risk." (2012).

Verbeek, Marno. *A Guide to Modern Econometrics*. Wiley. com, 2008.

———. *A Guide to Modern Econometrics*. 4th ed. Chichester: Wiley, 2012.

Vidyamurthy, Ganapathy. *Pairs Trading: Quantitative Methods and Analysis*. Vol. 217: Wiley. com, 2004.

White, Halbert. "A Reality Check for Data Snooping." *Econometrica* 68, no. 5 (2000): 1097-126.

Xiong, Wei. "Convergence Trading with Wealth Effects: An Amplification Mechanism in Financial Markets." *Journal of Financial Economics* 62, no. 2 (2001): 247-92.

APPENDIX A

	LO--Real Price
Mean	2.47588
Median	1.075
Max	46.15
Min	0.04
Variance	24.2807
Std.Dev	4.92755
Skewness	5.65369
Kurtosis	42.0935
Coefficient Variation	1.99022
T.Stat.Jarque	20559.9
PValue	0
Sum	700.675
Sum Sq. Dev	6847.16
Observation	283
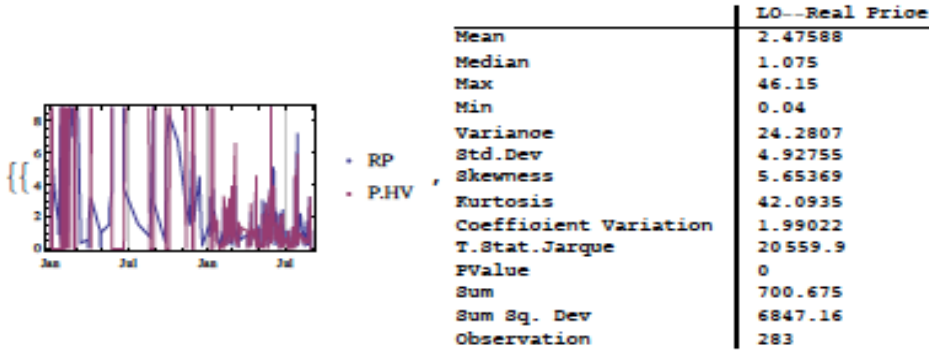
	LO--Price Hist.V
Mean	13.4268
Median	0.813892
Max	86.8284
Min	0.
Variance	756.758
Std.Dev	27.5092
Skewness	1.78076
Kurtosis	4.27379
Coefficient Variation	2.04883
T.Stat.Jarque	173.616
PValue	0
Sum	3799.78
Sum Sq. Dev	213406.
Observation	283

	LO--Real Price
Mean	2.47588
Median	1.075
Max	46.15
Min	0.04
Variance	24.2807
Std.Dev	4.92755
Skewness	5.65369
Kurtosis	42.0935
Coefficient Variation	1.99022
T.Stat.Jarque	20559.9
PValue	0
Sum	700.675
Sum Sq. Dev	6847.16
Observation	283
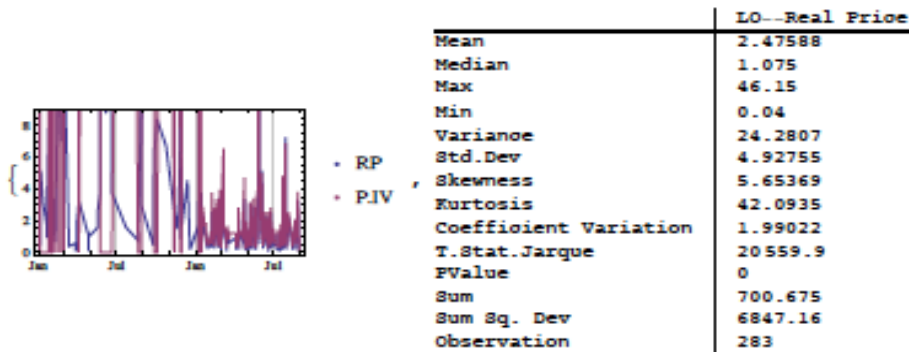
	LO--Price Im.Vol
Mean	13.7209
Median	1.22956
Max	86.8284
Min	0.
Variance	749.718
Std.Dev	27.381
Skewness	1.77974
Kurtosis	4.27349
Coefficient Variation	1.99557
T.Stat.Jarque	173.431
PValue	0
Sum	3883.02
Sum Sq. Dev	211421.
Observation	283

$$\left\{\left\{\left\{\right.\right.\right.$$

$\cdot$ RP

$\cdot$ P.HV

,

	LOW--Real Price
Mean	1.28843
Median	0.805
Max	19.325
Min	0.03
Variance	3.89811
Std.Dev	1.97436
Skewness	5.00946
Kurtosis	33.7279
Coefficient Variation	1.53238
T.Stat.Jarque	41417.4
PValue	0
Sum	1207.25
Sum Sq. Dev	3648.63
Observation	937

,
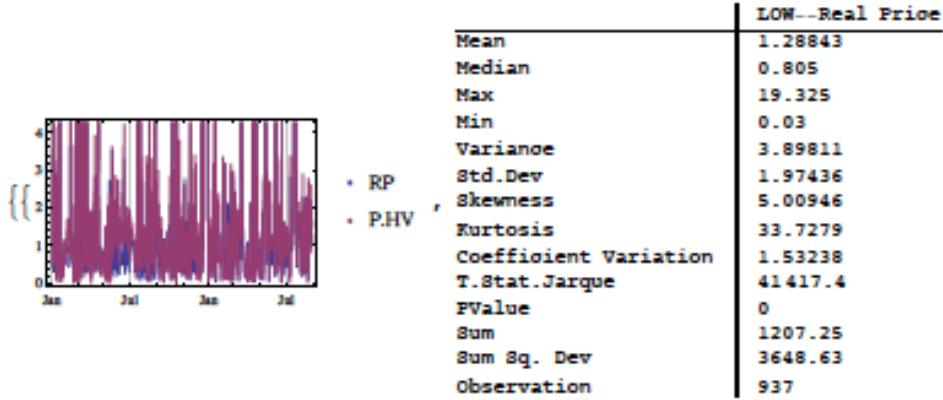
	LOW--Price Hist.V
Mean	1.48331
Median	1.01023
Max	19.3012
Min	3.38894×10^{-6}
Variance	4.1734
Std.Dev	2.04289
Skewness	4.57067
Kurtosis	29.4216
Coefficient Variation	1.37725
T.Stat.Jarque	30991.3
PValue	0
Sum	1389.86
Sum Sq. Dev	3906.3
Observation	937

$\Big\}$,

$\cdot$ RP

$\cdot$ P.IV

,

	LOW--Real Price
Mean	1.28843
Median	0.805
Max	19.325
Min	0.03
Variance	3.89811
Std.Dev	1.97436
Skewness	5.00946
Kurtosis	33.7279
Coefficient Variation	1.53238
T.Stat.Jarque	41417.4
PValue	0
Sum	1207.25
Sum Sq. Dev	3648.63
Observation	937

,
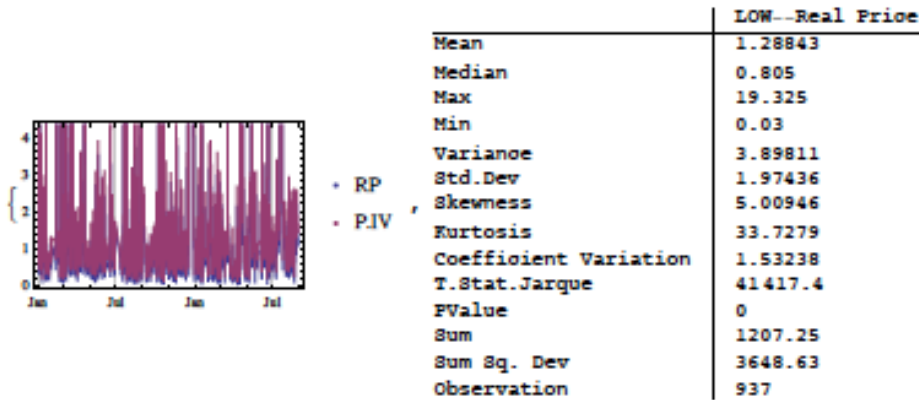
	LOW--Price Im.Vol
Mean	1.59427
Median	1.08739
Max	19.8747
Min	0.101198
Variance	4.33714
Std.Dev	2.08258
Skewness	4.70307
Kurtosis	30.42
Coefficient Variation	1.30629
T.Stat.Jarque	33317.4
PValue	0
Sum	1493.83
Sum Sq. Dev	4059.56
Observation	937

$\Big\}\Big\}$,

	LRCX--Real Price
Mean	1.48429
Median	0.9625
Max	12.65
Min	0.05
Variance	3.39061
Std.Dev	1.84136
Skewness	3.83303
Kurtosis	21.5092
Coefficient Variation	1.24057
T.Stat.Jarque	3757.79
PValue	0
Sum	311.7
Sum Sq. Dev	708.638
Observation	210
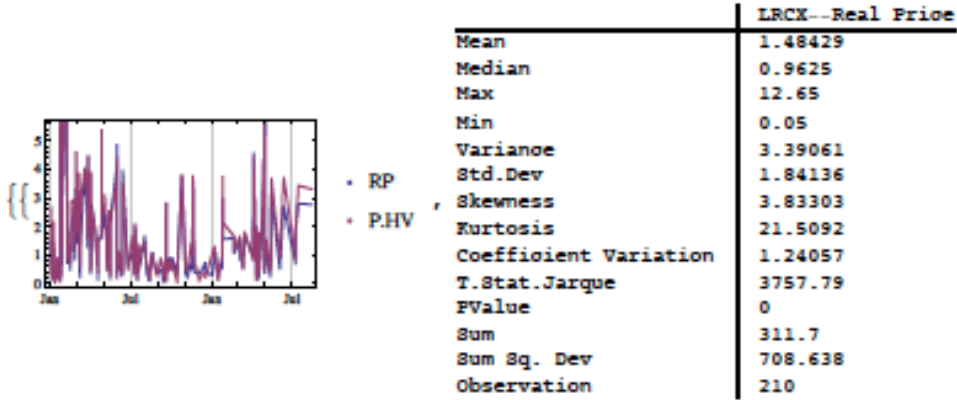
	LRCX--Price Hist.V
Mean	1.62359
Median	1.01318
Max	13.1268
Min	0.012267
Variance	3.8018
Std.Dev	1.94982
Skewness	3.59321
Kurtosis	19.799
Coefficient Variation	1.20093
T.Stat.Jarque	3125.4
PValue	0
Sum	340.954
Sum Sq. Dev	794.577
Observation	210

	LRCX--Real Price
Mean	1.48429
Median	0.9625
Max	12.65
Min	0.05
Variance	3.39061
Std.Dev	1.84136
Skewness	3.83303
Kurtosis	21.5092
Coefficient Variation	1.24057
T.Stat.Jarque	3757.79
PValue	0
Sum	311.7
Sum Sq. Dev	708.638
Observation	210
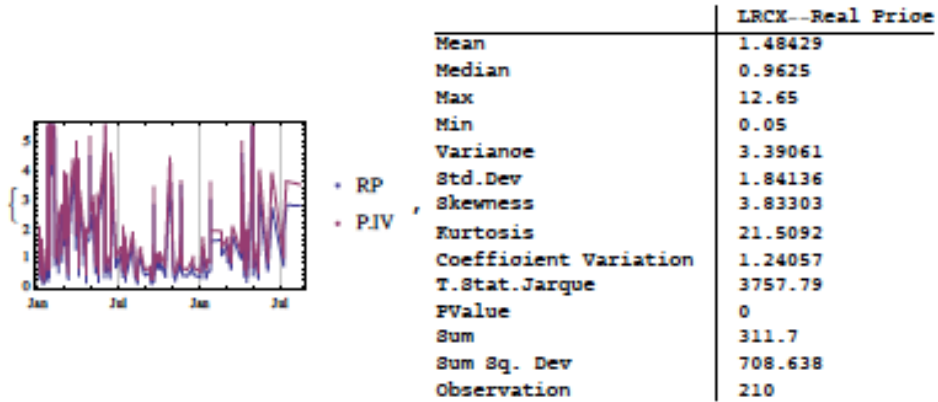
	LRCX--Price Im.Vol
Mean	1.84755
Median	1.34454
Max	13.0272
Min	0.158088
Variance	3.61385
Std.Dev	1.90101
Skewness	3.51538
Kurtosis	19.0673
Coefficient Variation	1.02894
T.Stat.Jarque	2879.15
PValue	0
Sum	387.985
Sum Sq. Dev	755.294
Observation	210

	LSI--Real Price
Mean	0.447845
Median	0.3625
Max	4.8
Min	0.075
Variance	0.240691
Std.Dev	0.490603
Skewness	6.29417
Kurtosis	54.8905
Coefficient Variation	1.09547
T.Stat.Jarque	15634.5
PValue	0
Sum	51.95
Sum Sq. Dev	27.6795
Observation	116
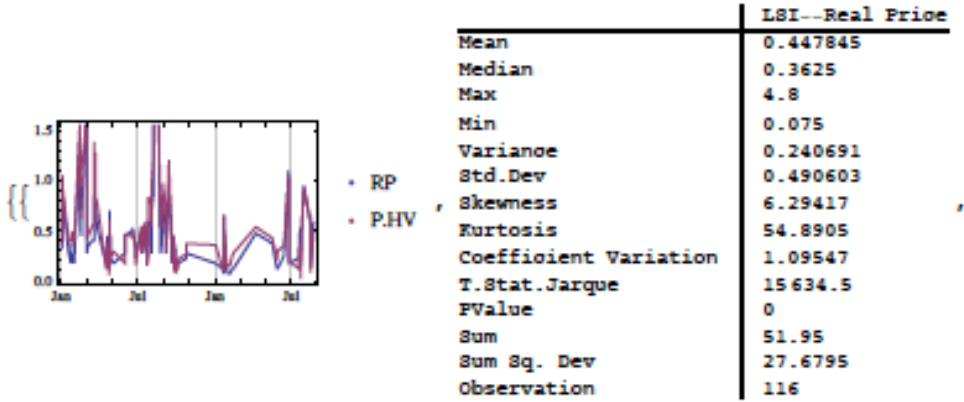
	LSI--Price Hist.V
Mean	0.529841
Median	0.424533
Max	4.71473
Min	0.0192154
Variance	0.264599
Std.Dev	0.514392
Skewness	4.9305
Kurtosis	39.0896
Coefficient Variation	0.970842
T.Stat.Jarque	7673.6
PValue	0
Sum	61.4615
Sum Sq. Dev	30.4289
Observation	116

	LSI--Real Price
Mean	0.447845
Median	0.3625
Max	4.8
Min	0.075
Variance	0.240691
Std.Dev	0.490603
Skewness	6.29417
Kurtosis	54.8905
Coefficient Variation	1.09547
T.Stat.Jarque	15634.5
PValue	0
Sum	51.95
Sum Sq. Dev	27.6795
Observation	116
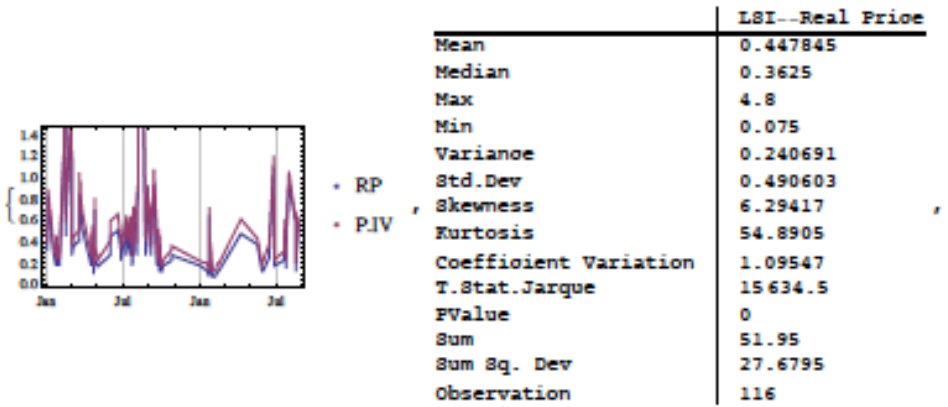
	LSI--Price Im.Vol
Mean	0.54347
Median	0.450522
Max	4.95692
Min	0.12686
Variance	0.250679
Std.Dev	0.500679
Skewness	6.14564
Kurtosis	53.4625
Coefficient Variation	0.921263
T.Stat.Jarque	14792.8
PValue	0
Sum	63.0425
Sum Sq. Dev	28.8281
Observation	116

{{

	LUK--Real Price
Mean	0.58125
Median	0.55
Max	1.6
Min	0.05
Variance	0.166946
Std.Dev	0.40859
Skewness	1.12637
Kurtosis	4.32431
Coefficient Variation	0.702951
T.Stat.Jarque	9.42575
PValue	0.0424276
Sum	6.975
Sum Sq. Dev	1.83641
Observation	12

,
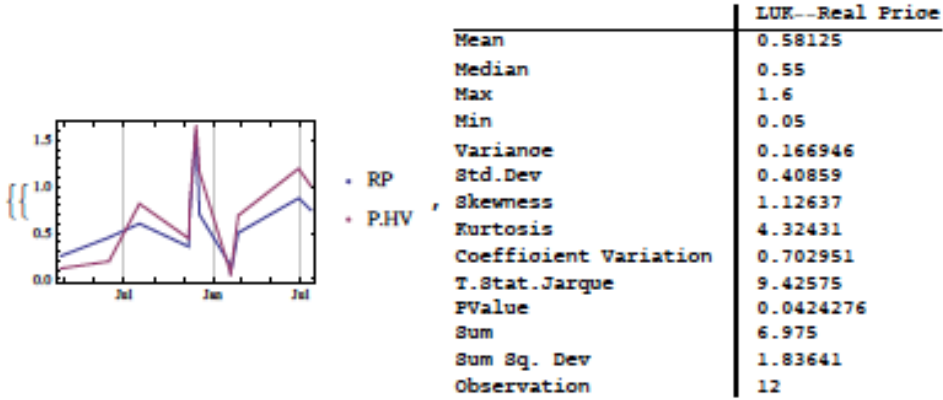
	LUK--Price Hist.V
Mean	0.680832
Median	0.738233
Max	1.67121
Min	0.031127
Variance	0.275055
Std.Dev	0.524457
Skewness	0.289368
Kurtosis	2.06758
Coefficient Variation	0.770317
T.Stat.Jarque	0.640786
PValue	0.696662
Sum	8.16999
Sum Sq. Dev	3.0256
Observation	12

},

	LUK--Real Price
Mean	0.58125
Median	0.55
Max	1.6
Min	0.05
Variance	0.166946
Std.Dev	0.40859
Skewness	1.12637
Kurtosis	4.32431
Coefficient Variation	0.702951
T.Stat.Jarque	9.42575
PValue	0.0424276
Sum	6.975
Sum Sq. Dev	1.83641
Observation	12

,
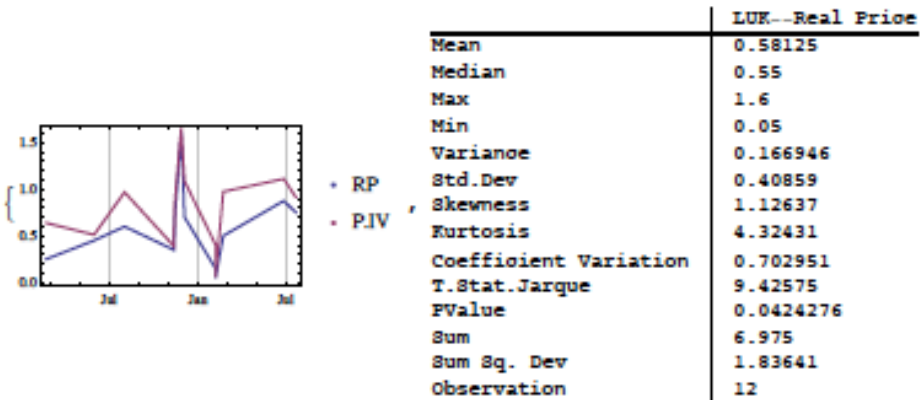
	LUK--Price Im.Vol
Mean	0.797748
Median	0.889949
Max	1.65176
Min	0.0594576
Variance	0.177896
Std.Dev	0.421777
Skewness	0.166937
Kurtosis	2.83688
Coefficient Variation	0.528709
T.Stat.Jarque	0.23863
PValue	0.89288
Sum	9.57298
Sum Sq. Dev	1.95685
Observation	12

}},

	LUV--Real Price
Mean	0.611111
Median	0.3875
Max	2.225
Min	0.05
Variance	0.294662
Std.Dev	0.542828
Skewness	1.32074
Kurtosis	3.62434
Coefficient Variation	0.888264
T.Stat.Jarque	19.1196
PValue	0.007931
Sum	33.
Sum Sq. Dev	15.6171
Observation	54

{{ , • RP • P.HV ,
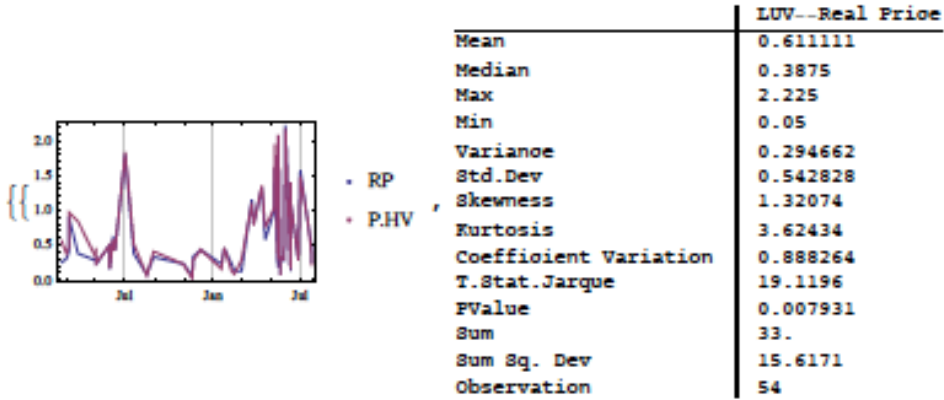
	LUV--Price Hist.V
Mean	0.659262
Median	0.450709
Max	2.17429
Min	0.00520304
Variance	0.338907
Std.Dev	0.582157
Skewness	1.31325
Kurtosis	3.65078
Coefficient Variation	0.883043
T.Stat.Jarque	19.0382
PValue	0.00799422
Sum	35.6002
Sum Sq. Dev	17.9621
Observation	54

},

	LUV--Real Price
Mean	0.611111
Median	0.3875
Max	2.225
Min	0.05
Variance	0.294662
Std.Dev	0.542828
Skewness	1.32074
Kurtosis	3.62434
Coefficient Variation	0.888264
T.Stat.Jarque	19.1196
PValue	0.007931
Sum	33.
Sum Sq. Dev	15.6171
Observation	54

{ , • RP • P.IV ,
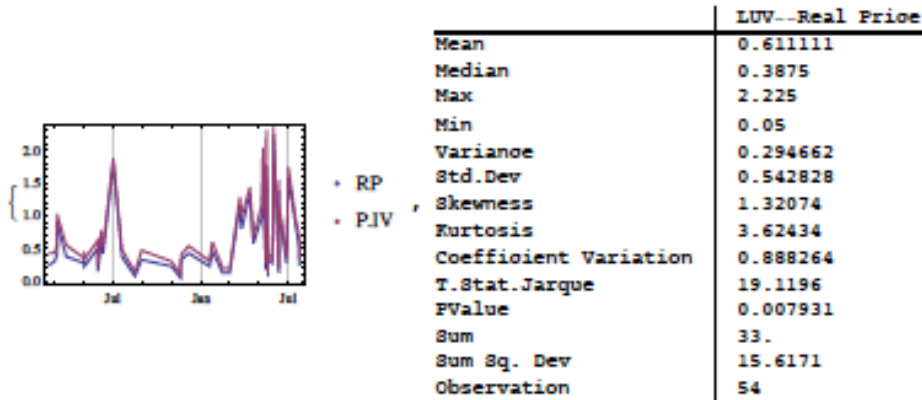
	LUV--Price Im.Vol
Mean	0.747272
Median	0.527756
Max	2.35189
Min	0.0991742
Variance	0.367197
Std.Dev	0.605968
Skewness	1.33878
Kurtosis	3.67907
Coefficient Variation	0.810907
T.Stat.Jarque	19.8482
PValue	0.00736522
Sum	40.3527
Sum Sq. Dev	19.4614
Observation	54
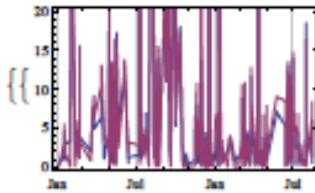
}},

	MA--Real Price
Mean	4.98093
Median	2.13
Max	94.45
Min	0.055
Variance	73.2798
Std.Dev	8.56036
Skewness	5.40943
Kurtosis	45.0465
Coefficient Variation	1.71863
T.Stat.Jarque	29530.
PValue	0
Sum	1798.11
Sum Sq. Dev	26380.7
Observation	361

	MA--Price Hist.V
Mean	5.41265
Median	2.68302
Max	92.14
Min	0.
Variance	84.2815
Std.Dev	9.1805
Skewness	4.65155
Kurtosis	33.4605
Coefficient Variation	1.69612
T.Stat.Jarque	15886.6
PValue	0
Sum	1953.97
Sum Sq. Dev	30341.3
Observation	361
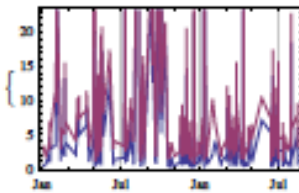
	MA--Real Price
Mean	4.98093
Median	2.13
Max	94.45
Min	0.055
Variance	73.2798
Std.Dev	8.56036
Skewness	5.40943
Kurtosis	45.0465
Coefficient Variation	1.71863
T.Stat.Jarque	29530.
PValue	0
Sum	1798.11
Sum Sq. Dev	26380.7
Observation	361

	MA--Price Im.Vol
Mean	7.68075
Median	4.83804
Max	97.6628
Min	0.60471
Variance	84.1066
Std.Dev	9.17096
Skewness	4.65471
Kurtosis	35.5448
Coefficient Variation	1.19402
T.Stat.Jarque	17948.4
PValue	0
Sum	2772.75
Sum Sq. Dev	30278.4
Observation	361

	NOV--Real Price
Mean	2.22196
Median	1.41
Max	23.45
Min	0.035
Variance	8.55263
Std.Dev	2.92449
Skewness	4.22821
Kurtosis	24.8687
Coefficient Variation	1.31618
T.Stat.Jarque	15380.5
PValue	0
Sum	1459.82
Sum Sq. Dev	5610.53
Observation	657

$\{\{$ · RP · P.HV $,$
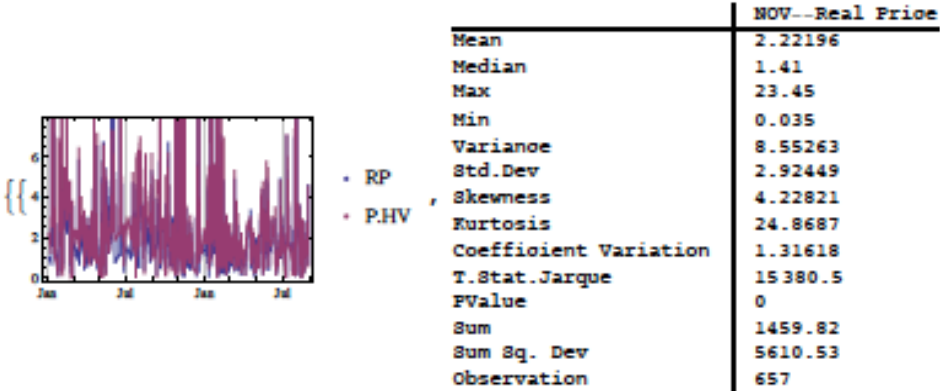
	NOV--Price Hist.V
Mean	2.53902
Median	1.75221
Max	22.6577
Min	1.96456×10^{-6}
Variance	9.23289
Std.Dev	3.03857
Skewness	3.7211
Kurtosis	20.5438
Coefficient Variation	1.19675
T.Stat.Jarque	10159.
PValue	0
Sum	1668.14
Sum Sq. Dev	6056.78
Observation	657

$\}$,

	NOV--Real Price
Mean	2.22196
Median	1.41
Max	23.45
Min	0.035
Variance	8.55263
Std.Dev	2.92449
Skewness	4.22821
Kurtosis	24.8687
Coefficient Variation	1.31618
T.Stat.Jarque	15380.5
PValue	0
Sum	1459.82
Sum Sq. Dev	5610.53
Observation	657

$\{$ · RP · P.IV $,$
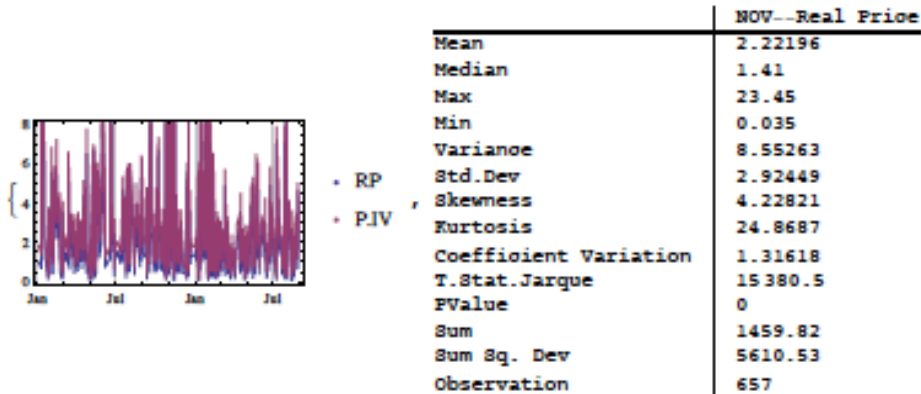
	NOV--Price Im.Vol
Mean	2.84308
Median	2.02943
Max	24.6803
Min	0.124914
Variance	9.24857
Std.Dev	3.04115
Skewness	3.88726
Kurtosis	22.3581
Coefficient Variation	1.06967
T.Stat.Jarque	12174.7
PValue	0
Sum	1867.91
Sum Sq. Dev	6067.06
Observation	657

$\}\}$,

	NSC--Real Price
Mean	1.61948
Median	1.0875
Max	20.75
Min	0.075
Variance	5.29174
Std.Dev	2.30038
Skewness	5.87318
Kurtosis	45.7645
Coefficient Variation	1.42045
T.Stat.Jarque	15342.4
PValue	0
Sum	278.55
Sum Sq. Dev	904.887
Observation	172
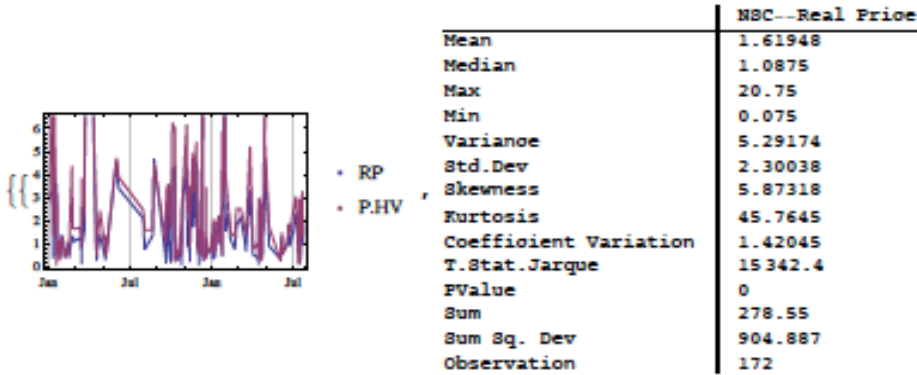
	NSC--Price Hist.V
Mean	2.28772
Median	1.6506
Max	21.8409
Min	0.0745359
Variance	6.85309
Std.Dev	2.61784
Skewness	4.7474
Kurtosis	33.6968
Coefficient Variation	1.1443
T.Stat.Jarque	8051.31
PValue	0
Sum	393.487
Sum Sq. Dev	1171.88
Observation	172

	NSC--Real Price
Mean	1.61948
Median	1.0875
Max	20.75
Min	0.075
Variance	5.29174
Std.Dev	2.30038
Skewness	5.87318
Kurtosis	45.7645
Coefficient Variation	1.42045
T.Stat.Jarque	15342.4
PValue	0
Sum	278.55
Sum Sq. Dev	904.887
Observation	172
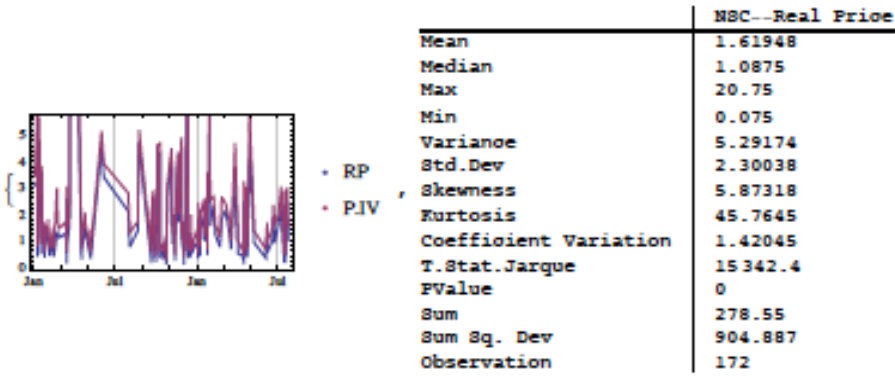
	NSC--Price Im.Vol
Mean	2.10224
Median	1.56035
Max	22.5096
Min	0.261397
Variance	6.55842
Std.Dev	2.56094
Skewness	5.81518
Kurtosis	45.0221
Coefficient Variation	1.2182
T.Stat.Jarque	14829.9
PValue	0
Sum	361.584
Sum Sq. Dev	1121.49
Observation	172

	NVDA--Real Price
Mean	0.574688
Median	0.405
Max	5.325
Min	0.04
Variance	0.303526
Std.Dev	0.550932
Skewness	3.04556
Kurtosis	17.3
Coefficient Variation	0.958662
T.Stat.Jarque	13211.1
PValue	0
Sum	745.945
Sum Sq. Dev	393.673
Observation	1298

· RP
· P.HV

	NVDA--Price Hist.V
Mean	0.561315
Median	0.383406
Max	5.42418
Min	0.0000577832
Variance	0.406405
Std.Dev	0.637499
Skewness	3.16282
Kurtosis	16.9781
Coefficient Variation	1.13572
T.Stat.Jarque	12871.1
PValue	0
Sum	728.587
Sum Sq. Dev	527.108
Observation	1298

	NVDA--Real Price
Mean	0.574688
Median	0.405
Max	5.325
Min	0.04
Variance	0.303526
Std.Dev	0.550932
Skewness	3.04556
Kurtosis	17.3
Coefficient Variation	0.958662
T.Stat.Jarque	13211.1
PValue	0
Sum	745.945
Sum Sq. Dev	393.673
Observation	1298

· RP
· P.IV

	NVDA--Price Im.Vol
Mean	0.740146
Median	0.556139
Max	5.90514
Min	0.0957084
Variance	0.378234
Std.Dev	0.615007
Skewness	2.95003
Kurtosis	16.304
Coefficient Variation	0.830927
T.Stat.Jarque	11581.9
PValue	0
Sum	960.709
Sum Sq. Dev	490.57
Observation	1298

	PCLN--Real Price
Mean	7.17248
Median	4.35
Max	68.25
Min	0.075
Variance	62.3283
Std.Dev	7.89483
Skewness	2.08589
Kurtosis	9.03806
Coefficient Variation	1.10071
T.Stat.Jarque	2611.23
PValue	0
Sum	8248.35
Sum Sq. Dev	71615.2
Observation	1150
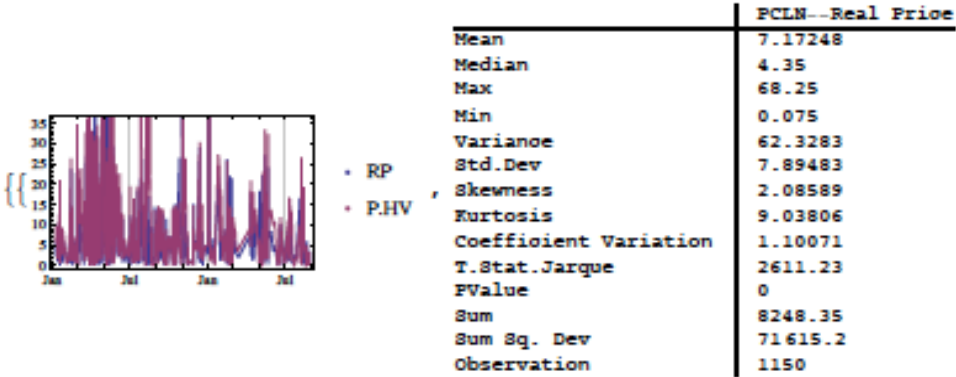
	PCLN--Price Hist.V
Mean	9.97351
Median	6.59555
Max	74.376
Min	0.
Variance	109.596
Std.Dev	10.4688
Skewness	1.84595
Kurtosis	7.4699
Coefficient Variation	1.04966
T.Stat.Jarque	1628.72
PValue	0
Sum	11469.5
Sum Sq. Dev	125926.
Observation	1150

	PCLN--Real Price
Mean	7.17248
Median	4.35
Max	68.25
Min	0.075
Variance	62.3283
Std.Dev	7.89483
Skewness	2.08589
Kurtosis	9.03806
Coefficient Variation	1.10071
T.Stat.Jarque	2611.23
PValue	0
Sum	8248.35
Sum Sq. Dev	71615.2
Observation	1150
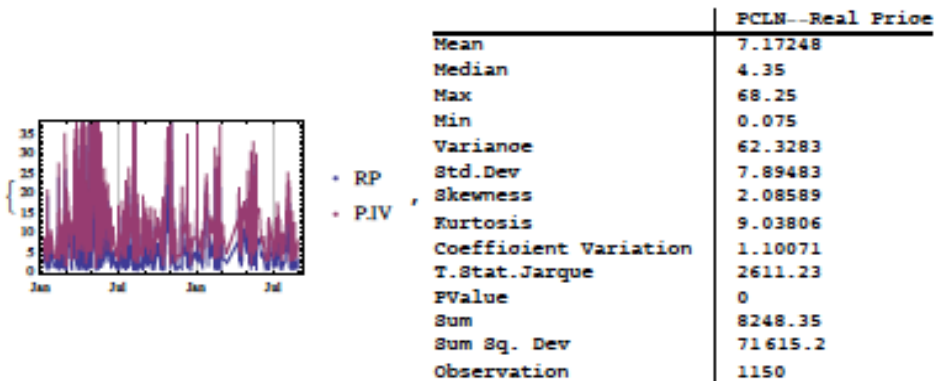
	PCLN--Price Im.Vol
Mean	12.4079
Median	9.57514
Max	76.3081
Min	0.945711
Variance	93.5696
Std.Dev	9.67314
Skewness	2.01593
Kurtosis	8.41818
Coefficient Variation	0.779594
T.Stat.Jarque	2210.88
PValue	0
Sum	14269.1
Sum Sq. Dev	107511.
Observation	1150

APPENDIX B

Joining price and vol

```
ClearAll["Global`*"]
```

Aleksander Rinaldo

```
In[1]:=  << AleksanderRinaldo`
```

Data Import

```
sp = Import["/Users/aleksanderrinaldo/Desktop/Dataset
        adjusted by Thomas/crsp clean s&p500.csv"];

vol = Import["/Users/aleksanderrinaldo/Desktop/Dataset
        adjusted by Thomas/60 day vol.csv"];

int = Flatten[
    Import["/Users/aleksanderrinaldo/Desktop/Dataset adjusted by Thomas/correct
        interest.xlsx"], 1];

class = Flatten[Import[
    "/Users/aleksanderrinaldo/Desktop/Dataset adjusted by Thomas/cl.xlsx"], 1];
```

```
In[21]:=  mulig = Import["/Users/aleksanderrinaldo/muligCorrectEx.mx"] /.
      {x_, s_, a_} :> {x, s, a, 0, 0, 0.};
```

Volatility

```
pick = Cases[Intersection[vol[[All, 2]], sp[[All, 2]]], Except["ticker"]];

join[data1_, data2_, i_] := Block[{sn = i, d1 = data1, d2 = data2, v,
    sel, d1P, d2P, keep, d1P2, iS, p, pr, date, dString}, sel[sn] = True;
  d1P = Select[d1, sel[#[[2]]] &];
  d2P = Select[d2, sel[#[[2]]] &];
  iS = Intersection[d1P[[All, 1]], d2P[[All, 1]]];
  date = Sort[DateList[#] & /@ iS];
  dString =
    (ToLowerCase@DateString[#, {"Day", "MonthNameShort", "Year"}]) & /@ date;
  p = {#, ___} & /@ dString;
  v = Flatten[Cases[d2P, #, 1, 1] & /@ p, 1];
  pr = Flatten[Cases[d1P, #, 1, 1] & /@ p, 1];
  insertColumn[pr, v[[All, 3]], -1]
 ]

volP = join[sp, vol, #] & /@ pick;
```

Interest

```
intRule = Cases[int, {{y_, m_, x_, s___}, i_} :> {y, m, v___} → i];
```

```
intRange = DateRange[{2000}, {2014}];

joint = Transpose[{intRange, intRange /. intRule}];

converg = Cases[joint,
    {x_, y_} :> {ToLowerCase@DateString[x, {"Day", "MonthNameShort", "Year"}], y}];

interestRule = Cases[converg, {x_, y_} :> x -> y];

aInterst[x_, rul_] := Block[{d = x, r = rul},
    insertColumn[d, d[[All, 1]] /. r, -1]]

addingInterest = aInterst[#, interestRule] & /@ volP;
```

Sorting Into groups

```
subIndustries = {n_, __, #, ___} :> {n, #} & /@ Union[class[[3 ;;, 3]]];

pp = Cases[class, #] & /@ subIndustries;

scom = Subsets[#, {2}] & /@ pp;

jPair = Flatten[Cases[#, {{x_, __}, {v_, __}} :> {x, v}] & /@ scom, 1];

ts[xdata_, xpair_, xts_] :=
 Block[{data = xdata, jP = xpair, en, two, pEn, pTo, bes, a, b, cmp, tf = xts,
    df, ma, stDi, diff, ts, date, aP, bP, dA, dB, vA, vB, i, ise, kov, ei},
   en = jP[[1]];
   two = jP[[2]];
   pEn = Flatten[Select[data, MatchQ[#[[1, 2]], en] &], 1];
   pTo = Flatten[Select[data, MatchQ[#[[1, 2]], two] &], 1];
   ise = Intersection[pEn[[All, 1]], pTo[[All, 1]]];
   If[Length[ise] > 60,
    Scan[bes[#] = True; &, ise];
    a = Select[pEn, bes[#[[1]]] &];
    b = Select[pTo, bes[#[[1]]] &];
    cmp = MapThread[Join[#1, #2] &, {a, b}];
    df = cmp[[All, 3]] - cmp[[All, 9]];
    diff = Drop[df, tf - 1];
    ma = MovingAverage[df, tf];
    stDi =
     Table[StandardDeviation[df[[1 + n ;; tf + n]]], {n, 0, Length[diff] - 1}];
    ts = PadLeft[((diff - ma) / stDi), Length[cmp]];
    date = cmp[[All, 1]]; aP = cmp[[All, 3]]; dA = cmp[[All, 4]];
    bP = cmp[[All, 9]]; dB = cmp[[All, 10]]; vA = cmp[[All, 5]];
    vB = cmp[[All, 11]]; i = cmp[[All, 12]];
    kov = Transpose[{date, aP, bP, ts, vA, vB, dA, dB, i}];
    ei =
     insertRow[kov, {"Date", "PA", "PB", "TS", "VA", "VB", "DA", "DB", "I"}, 1];
    insertRow[ei, {en, two, "", "", "", "", "", "", ""}, 1]]]


addTSignal = DeleteCases[ts[addingInterest, #, 50] & /@ jPair, Null];
```

Testing for cointegration

```
s[sig_] := Block[{d = sig, ms, range, eR, f, vB, v, ru, rul, nw, model1, model,
     rull, val, residuals, ℋ, menu, pValue, lengthPValue, lf, fE, dob},
   ms = SplitBy[d[[3 ;;]], DateList[#[[1]]][[2]] &];
   vB = Cases[#, {g_, a_, b_, v___} :> {g, Log[a] // N, Log[b] // N, v}] & /@ ms;
   range = Table[List[1 + n, 18 + n], {n, 0, 500 000, 6}];
   eR = DeleteCases[If[#[[2]] ≤ Length[ms[[All]]], #] & /@ range, Null];
   f = vB[[#[[1]] ;; #[[2]]]] & /@ eR;
   fE = ms[[#[[1]] ;; #[[2]]]] & /@ eR;
   v = Position[f, DirectedInfinity[h_]];
   model = If[Length[v] < 1,
      LinearModelFit[Flatten[#[[1 ;; 12]], 1][[All, 2 ;; 3]], b, b] & /@ f,
      ru = Cases[v, {s___, x_} :> {s, x + 1}];
      val = Extract[f, #] & /@ ru;
      rul = Thread[v[[All]] → val];
      nw = ReplacePart[f, rul];
      LinearModelFit[Flatten[#[[1 ;; 12]], 1][[All, 2 ;; 3]], b, b] & /@ nw];
   residuals = #["FitResiduals"] & /@ model;
   ℋ = UnitRootTest[#[[All]], Automatic, "HypothesisTestData"] & /@ residuals;
   pValue = Partition[Flatten[#["TestData", All] & /@ ℋ, 1][[All, 2]], 4];
   menu = Join[{"UnitRootTest:", "Dickey-Fuller F",
        "Dickey-Fuller T", "Phillips-Perron F", "Phillips-Perron T"},
      pValue[[#]]] & /@ Range[Length[pValue]];
   lengthPValue = Table[Length[Select[pValue[[t]], # < 0.10 &]],
     {t, 1, Length[pValue]}];
   lf = Range[Length[f]];
   dob = MapThread[
     If[# === 4, Prepend[Join[d[[1 ;; 2]], Flatten[fE[[#2]][[13 ;;]], 1]], #3]] &,
     {lengthPValue, lf, menu}]]

enComb = s[#] & /@ addTSignal;

pick2 = Cases[enComb, {{"UnitRootTest:", __}, __}, -1];
```

Trading System

Pairs trading with vanila options

Data

In[2]:= ```pick2 = Import["/Users/aleksanderrinaldo/pick2 2.mx"];```

In[1458]:= ```ivi = ((1.20) ^ (1 / 250)) - 1;```

In[1459]:= ```ivi```

Out[1459]= 0.000729552

Code

```
In[4]:= fA[px_, dts_, int_] := Block[{s = px, dove = dts, nDD, diff, i = int, interest, p},
          nDD = dove[[#]] & @@ {1 ;; # & @@ Position[dove, s[[1]]][[1]]};
          interest = i[[#]] & @@ {1 ;; # & @@ Position[dove, s[[1]]][[1]]};
          diff = Length[DayRange[s[[1]], #]] - 1 & /@ nDD;
          p = PadRight[
            MapThread[s[[7]] * e ^ (-#2 * (#1 / 360)) &, {diff, interest}], Length[dts]];
          p /. (Select[Reverse[p], # > 0 &, 1][[1]] → 0)
         ]

In[5]:= fB[px_, dts_, int_] := Block[{s = px, dove = dts, nDD, diff, i = int, interest, p},
          nDD = dove[[#]] & @@ {1 ;; # & @@ Position[dove, s[[1]]][[1]]};
          interest = i[[#]] & @@ {1 ;; # & @@ Position[dove, s[[1]]][[1]]};
          diff = Length[DayRange[s[[1]], #]] - 1 & /@ nDD;
          p = PadRight[
            MapThread[s[[8]] * e ^ (-#2 * (#1 / 360)) &, {diff, interest}], Length[dts]];
          p /. (Select[Reverse[p], # > 0 &, 1][[1]] → 0)
         ]

In[6]:= dcc[corrected_] := Module[{x = corrected, dts, aP, a, upA, bP, upB},
          dts = DateList[#] & /@ x[[All, 1]];
          (*A*)
          aP = Cases[x, {d_, h___, f_, y_, r_} /; f > 0 :> {DateList[d], h, f, y, r}];
          upA = If[Length[aP] > 0,
            insertColumn[x, Total[#] & /@ (fA[#, dts, x[[All, -1]]] & /@ aP^T), -1],
            insertColumn[x, Table[0, {Length[dts]}], -1]];
          (*b*)
          bP = Cases[x, {d_, h___, f_, y_, r_} /; y > 0 :> {DateList[d], h, f, y, r}];
          upB = If[Length[bP] > 0,
            insertColumn[upA, Total[#] & /@ (fB[#, dts, x[[All, -1]]] & /@ bP^T), -1],
            insertColumn[upA, Table[0, {Length[dts]}], -1]];
          insertColumn[upB, Length[DateRange[#, DateList[x[[-1, 1]]]]] - 1 & /@ dts, 1]]

In[104]:= plus[data_, in_, call_, put_] :=
          Block[{s = data, p = put, c = call, strikePut, strikeCall, puts,
            aaReturn, bbReturn, dta, dtb, interests = in, calls, p1, c1},
           strikePut = s[[1, 3]] * p;
           strikeCall = s[[1, 4]] * c;
           p1 = PadRight[putA[#, strikePut] & /@ s[[1 ;; -2]], Length[s]];
           puts = ReplacePart[p1, h_ :> (0.20 * s[[1, 3]]) /; h == 1 && p1[[1]] == 0];
           c1 = PadRight[callB[#, strikeCall] & /@ s[[1 ;; -2]], Length[s]];
           calls = ReplacePart[c1, h_ :> (0.20 * s[[1, 4]]) /; h == 1 && c1[[1]] == 0];
           dta = insertColumn[s, puts, -1];
           dtb = insertColumn[s, calls, -1];
           {aaReturn[dta, interests, strikePut], bbReturn[dtb, interests, strikeCall]}]
```

In[1558]:=
```
aaReturn[dta_, interests_, strikePut_] :=
  Block[{v = dta, r = interests, on, k, strike = strikePut, return, returnDisc},
   return = 1 + (# - v[[1, -1]]) / (v[[1, -1]]) & /@ v[[All, -1]];
   returnDisc = MapThread[# / (1 + r) ^ (v[[1, 1]] - #2) &, {return, v[[All, 1]]}] - 1;
   If[Length[on = Flatten[Position[returnDisc, h_ /; h > r, 1, 1]]] > 0,
    {v[[#, 2]] & @@ on, "P'a", v[[1, -1]],
     (return[[#]] & @@ on) - 1, v[[1, 2]], v[[#, -1]] & @@ on},
    {v[[-1, 2]], "P'a", v[[1, -1]], If[0 < (strike - v[[-1, 3]]),
      (strike - (v[[-1, 3]] + v[[1, -1]])) / (v[[-1, 3]] + v[[1, -1]]),
      -1], v[[1, 2]], "HTE"}]]
```

In[1557]:=
```
bbReturn[dta_, interests_, strikeCall_] :=
  Block[{v = dta, r = interests, on, k, strike = strikeCall, return, returnDisc},
   return = 1 + (# - v[[1, -1]]) / (v[[1, -1]]) & /@ v[[All, -1]];
   returnDisc = MapThread[# / (1 + r) ^ (v[[1, 1]] - #2) &, {return, v[[All, 1]]}] - 1;
   If[Length[on = Flatten[Position[returnDisc, h_ /; h > r, 1, 1]]] > 0,
    {v[[#, 2]] & @@ on, "C'b", v[[1, -1]],
     (return[[#]] & @@ on) - 1, v[[1, 2]], v[[#, -1]] & @@ on},
    {v[[-1, 2]], "C'b", v[[1, -1]], If[0 < (v[[-1, 4]] - strike),
      (v[[-1, 4]] - (strike + v[[1, -1]])) / (strike + v[[1, -1]]),
      -1], v[[1, 2]], "HTE"}]]
```

In[1559]:=
```
minus[data_, in_, call_, put_] :=
  Block[{s = data, p = put, c = call, strikePut, strikeCall, puts, aaReturn,
    bbReturn, dta, dtb, interests = in, calls, maaReturn, mbbReturn, p1, c1},
   strikePut = s[[1, 4]] * p;
   strikeCall = s[[1, 3]] * c;
   p1 = PadRight[putB[#, strikePut] & /@ s[[1 ;; -2]], Length[s]];
   puts = ReplacePart[p1, h_ :> (0.20 * s[[1, 4]]) /; h == 1 && p1[[1]] == 0];
   c1 = PadRight[callA[#, strikeCall] & /@ s[[1 ;; -2]], Length[s]];
   calls = ReplacePart[c1, h_ :> (0.20 * s[[1, 3]]) /; h == 1 && c1[[1]] == 0];
   dtb = insertColumn[s, puts, -1];
   dta = insertColumn[s, calls, -1];
   {maaReturn[dta, interests, strikeCall], mbbReturn[dtb, interests, strikePut]}]
```

In[1561]:=
```
maaReturn[dta_, interests_, strikeCall_] :=
  Block[{v = dta, r = interests, on, k, strike = strikeCall, return, returnDisc},
   return = 1 + (# - v[[1, -1]]) / (v[[1, -1]]) & /@ v[[All, -1]];
   returnDisc = MapThread[# / (1 + r) ^ (v[[1, 1]] - #2) &, {return, v[[All, 1]]}] - 1;
   If[Length[on = Flatten[Position[returnDisc, h_ /; h > r, 1, 1]]] > 0,
    {v[[#, 2]] & @@ on, "C'a", v[[1, -1]],
     (return[[#]] & @@ on) - 1, v[[1, 2]], v[[#, -1]] & @@ on},
    {v[[-1, 2]], "C'a", v[[1, -1]], If[0 < (v[[-1, 3]] - strike),
      (v[[-1, 3]] - (strike + v[[1, -1]])) / (strike + v[[1, -1]]),
      -1], v[[1, 2]], "HTE"}]]
```

In[1562]:=
```
mbbReturn[dta_, interests_, strikePut_] :=
  Block[{v = dta, r = interests, on, k, strike = strikePut, return, returnDisc},
   return = 1 + (# - v[[1, -1]]) / (v[[1, -1]]) & /@ v[[All, -1]];
   returnDisc = MapThread[# / (1 + r) ^ (v[[1, 1]] - #2) &, {return, v[[All, 1]]}] - 1;
   If[Length[on = Flatten[Position[returnDisc, h_ /; h > r, 1, 1]]] > 0,
    {v[[#, 2]] & @@ on, "P'b", v[[1, -1]],
     (return[[#]] & @@ on) - 1, v[[1, 2]], v[[#, -1]] & @@ on},
    {v[[-1, 2]], "P'b", v[[1, -1]], If[0 < (strike - v[[-1, 4]]),
      (strike - (v[[-1, 4]] + v[[1, -1]])) / (v[[-1, 4]] + v[[1, -1]]),
      -1], v[[1, 2]], "HTE"}]]
```

In[13]:=
```
compsN[neededx_, data_, interestyD_, call_, put_] := Block[
    {s, ndX = neededx, fix = data, pk, st, nD, ptrn, enD, rng, trD, nter, trd, ex, r,
     exr, rose, exD, tradDat, correct, returnTrade, pos, d, dat, vt, lastMonth, vv},
    dat = Cases[fix, {sa_, da_, ka_, ja_, va_, sj_, fj_, oj_, q_}];
    s = dat[[4 ;;, 4]];
    pk = Select[s, # ≥ 2 || # ≤ -2 &, 1];
    st = Position[s, #] & @@ pk;
    d = DateList[#] & /@ Cases[dat[[4 ;;, 1]], _String];
    ptrn = Flatten[d[[#]]] & @@ st /. {v_, p_, i_, j___} :>
        {{v, p, w_, j} /; w ≥ i, {v, p + 1, _, j} /; MemberQ[d, {v, p + 1, _, j}],
         {v + 1, 1, _, j} /; MemberQ[d, {v + 1, 1, _, j}] && p == 12};
    exr = Flatten[Cases[ndX, #] & /@ ptrn /. {t_, n_, o_, b_, i_, j_} :>
        Identity[{t, n, v_, b, i, j} /; v ≤ o], 1];
    vv = DeleteCases[Cases[d, #] & /@ exr, {}];
    exD = Last[Last[vv]];
    tradDat = dat[[(#1 + 3 ;; #2 + 3) & @@ Flatten[{st, Position[d, exD]}]]];
    correct = dcc[tradDat];
    returnTrade = If[correct[[1, 5]] > 0, plus[correct, interestyD, call, put],
       minus[correct, interestyD, call, put]];
    vt = Join[#, {0, 0}] & /@ {returnTrade};
    pos = Position[DateList[#] & /@ d,
        Last[Sort[{DateList[#[[1]]], DateList[#2[[1]]]} & @@ returnTrade]]] + 4;
    Join[dat[[1 ;; 3]], dat[[# ;;]] & @@ pos[[1]], vt]]
```

In[14]:=
```
putA[ld_, strike_] :=
  Block[{s = ld, k = strike}, FinancialDerivative[{"European", "Put"},
    {"StrikePrice" → k, "Expiration" → s[[1]] / 252}, {"InterestRate" → s[[-3]],
     "Volatility" → s[[6]], "CurrentPrice" → (s[[3]] - s[[-2]])}]]
putB[ld_, strike_] := Block[{s = ld, k = strike},
  FinancialDerivative[{"European", "Put"},
    {"StrikePrice" → k, "Expiration" → s[[1]] / 252}, {"InterestRate" → s[[-3]],
     "Volatility" → s[[7]], "CurrentPrice" → (s[[4]] - s[[-1]])}]]
callA[ld_, strike_] := Block[{s = ld, k = strike},
  FinancialDerivative[{"European", "Call"},
    {"StrikePrice" → k, "Expiration" → s[[1]] / 252}, {"InterestRate" → s[[-3]],
     "Volatility" → s[[6]], "CurrentPrice" → (s[[3]] - s[[-2]])}]]
callB[ld_, strike_] := Block[{s = ld, k = strike},
  FinancialDerivative[{"European", "Call"},
    {"StrikePrice" → k, "Expiration" → s[[1]] / 252}, {"InterestRate" → s[[-3]],
     "Volatility" → s[[7]], "CurrentPrice" → (s[[4]] - s[[-1]])}]]
```

```
In[18]:= returnN[data_, exInterest_, anToDInterest_, call_, put_] :=
        Module[{z = data, x = exInterest, dT, led, lP, tL, rUl,
          nw, nededEx, exPos, signal, joinReturn, rr, nEwRule, rule,
          te, nest, return, nE, a, a1, a2, a3, i = anToDInterest},
         dT = DateList[#] & /@ z[[4 ;;, 1]];
         led = Last[Intersection @@ {dT, x}];
         lP = (Position[dT, led][[1, 1]]) + 3;
         tL = Length@dT + 3;
         a = z[[lP ;; tL, 4]];
         rule = Identity[{fo___, #, fa___} :> {fo, 1, fa}] & /@ a;
         nw = z /. rule;
         nE = Range[dT[[1, 1]], dT[[-1, 1]] - 1];
         nededEx = Union@Flatten[Cases[x, #] & /@
             Join[{#, s___} & /@ nE, {{dT[[-1, 1]], s_, p__} /; s ≤ dT[[-1, 2]]}], 1];
         nest = FixedPointList[If[Length[Select[#,
                 Function[{f}, f[[4]] ≥ 2 || f[[4]] ≤ -2], 1]] > 0,
             compsN[nededEx, #, i, call, put], {{1, 1, 1, 1, 1}}] &, nw];
         return = Union@Cases[Flatten[nest, 1], {zo_, za_, zi_, zu_} :> {zo, za}];
         Join[z[[1 ;; 3]], return]]
```