



UPPSALA
UNIVERSITET

Pairs Trading: an Extension to the Cointegration Approach

Can a cointegration approach based on low frequency data trading still beat the market in contemporary years?

Abstract: *This paper examines the (in contemporary literature inconclusive) usefulness of cointegration between stock prices as basis for a trading strategy. The primary contribution to previously used frameworks of the paper is the implementation and use of error correction models for selection of stocks to trade on. Evaluation is done through simulated results running the algorithm on the sectors of the Standard and Poor's 500 index in the years 2005 through 2014. Results indicate that trading strategies of this nature may be very successful even in recent years given that the universe of tradeable stocks within a sector is sufficiently large. The application of error correction models improve average returns, though in a way not originally anticipated.*

Authors: Isak Aggeborn, Olof Hansson
Supervisor: Lars Forsberg
Uppsala University
Department of Statistics
Spring term, 2017

Thanks to: Johan Vegelius, Uppsala University Department of Statistics

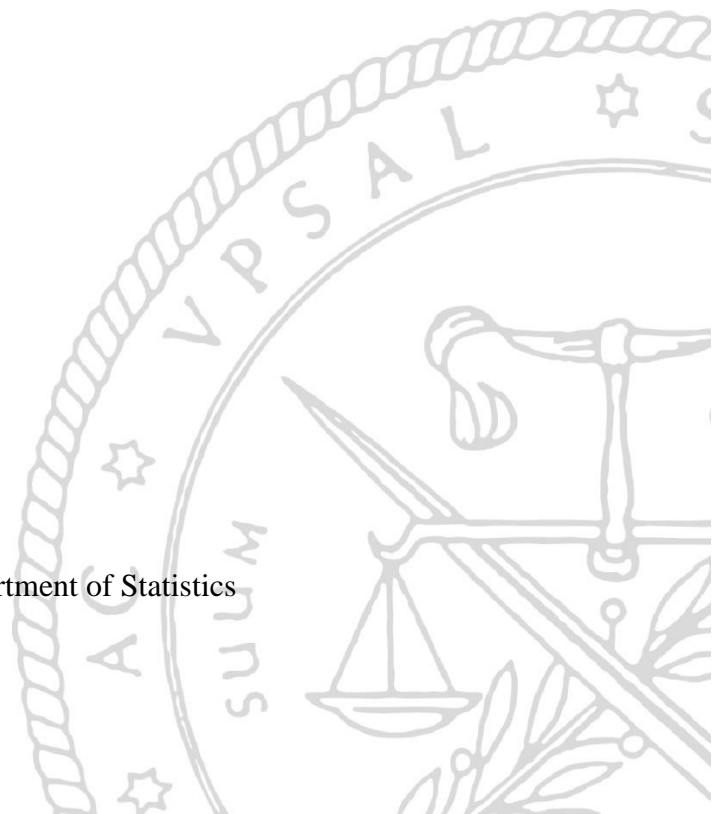


Table of Contents

1. Introduction.....	p 3
Cointegration Based Trading – a Crash Course.....	p 3
Purpose and Research Question.....	p 5
2. Data.....	p 5
3. Method.....	p 5
3.1 Theoretical Frameworks.....	p 5
Augmented Dickey-Fuller Tests.....	p 5
Applications.....	p 6
Error Correction Models.....	p 8
Applications.....	p 9
3.2 Practical Implementation.....	p 10
Sufficient Level of Divergence, Convergence and Stop Loss.....	p 10
Alternative Methods.....	p 13
Perceived Advantages of this Implementation.....	p 14
Returns and Equity Calculations.....	p 14
4. Results.....	p 16
The Big Picture – does Cointegration Based Trading Still Work?	p 16
A Sneak Peek: Visual representations.....	p 17
Error Correction Model – a Surprising Find.....	p 19
The Largest Sector: Consumption.....	p 19
Financials – Barely Large Enough.....	p 21
When there Just aren't Enough Trades – Information.....	p 23
5. Summary.....	p 25
Further Research – Where to Go Now?	p 25
6. References.....	p 26
7. Technical Appendix: Code Used to Perform Trading Analysis (R).....	p 27

Introduction

The use of cointegration between stock prices as a basis for a trading strategy is hardly a new concept. However, it is still a topic of contemporary interest in that it's up for debate to which extent these models are still profitable in today's markets. Contrary to the early days when trading models based on these properties was relatively unknown, it is far less certain that whatever statistical arbitrages are available have not already been exploited by other actors on the market.

In recent literature, among others the following authors have found trading based on models of this nature net excess returns: Gatev, Goetzman & Rouwenhorst, "Pairs Trading: Performance of a Relative Value Arbitrage Rule" (1998, 2006), George Miao, "High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach" (2014), Engelberg, Gao & Jagannathan, "An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity" (2009).

However, the following authors (once again, amongst others) do not find similar success: Bowen, Hutchinsinon & O'Sullivan, "High Frequency Equity Pairs Trading: Transaction Costs, Speed of Execution and Patterns in Returns" (2010), Do & Faff, "Does Simple Pairs Trading Still work?" (2010), Hoel, "Statistical Arbitrage Pairs: Can Cointegration Capture Market Neutral Profits?" (2013).

In other words, it's still very much an unresolved question whether these models are still profitable in a more contemporary setting. It is in this light that this paper takes its onset: We shall here make certain extensions to the cointegration based approach to pairs trading and subsequently evaluate whether given said extensions the strategy can still provide returns in excess of the market. For the results of this paper to be as relevant as possible, we will apply and evaluate our strategy on the stocks on the Standard and Poor's 500 index.

To make results as comparable to previous literature as possible (and most likely as a detriment to the results arrived at), we will use low frequency (daily) data in the when making the decisions on when to trade.

Cointegration Based Trading – a Crash Course

Pairs trading has existed in multiple forms for a long time. Our study focuses on one specific form of pairs trading - using cointegration of time series as method to identify suitable pairs to trade in. The core concept of pairs trading is to identify a pair of stocks where the difference between the prices of the two stocks is, for some reason, believed to have mean-reverting properties.

Essentially, the strategy will always revolve around identifying periods where the discrepancy between the price of the two stocks is larger than it ought to be and expected to shrink at a later point. One would therefore be able to make a profit from taking a long position in the

"undervalued" stock and a short position in the "overvalued" stock at this point in time and closing the position when the price discrepancy has returned to some expected level.

In this paper, we will use the cointegration method for identifying pairs of stocks suitable for trading in. As hinted at by the name, the method relies on identifying pairs of shares (or in a more complex implementation, a set of shares) that given the information in the in sample period is likely to have a cointegrating relationship.

If the time series of prices of two stocks have a cointegrating relationship, it implies that while the prices would appear driven by a stochastic (random) process individually (which would be seen examined with say a unit root test), these stochastic components are related for the two time series in such a way that there exists a linear combination between the two time series where the stochastic components cancel.

In other words, the core of the strategy essentially revolves around identifying pairs of stock prices that in some linear combination form a stationary time series. The usefulness of this is rather self-evident: should the series be stationary, a significant divergence would presumably be temporary and expected to revert in the future time periods. This is of course where a strategy of this nature would initiate a trade, taking one long and one short position and cancelling them later when the price discrepancy has returned to a level specified by the strategy.

Purpose and Research Question

The aim of this paper is to investigate whether a trading method based on the cointegrating property of stock prices can be profitable in recent years. To get there, one extension and one optimization compared to previously used frameworks will be implemented:

1. Error correction models will be estimated for pairs of stocks in consideration for trade, and the properties of said models used to screen for pairs most suitable for trading.
2. A somewhat more complex and likely suitable stop loss function compared to what's seen in much of the literature will be employed.

The research question this paper aims to contribute with an answer to is: "Can a trading strategy utilizing the aforementioned properties based on the cointegration of stock prices provide returns in the excess of the market? If so, in which settings?"

Data

Our primary subject of study will be stocks on Standard and Poor's 500 (hereafter referred to as S&P 500). This chosen trading universe contains the largest stock companies on the US market. Because of this, the items on S&P500 will of course change from on a year by year basis. As shall be seen in later section, our method will require the presence of the stock on S&P 500 throughout both the in- and out of sample period. For this reason, a sum of 500 stocks will never be considered at one point in time. Rather, the subset that is present throughout all years considered in the current implementation of the trading strategy will be used. If for example the in sample period is three years (2000-2002) and the out of sample period was one year (2003), only stocks present through 2000-2003 will be considered.

As per standard in literature, a division into industry sectors is used. The reason for doing this is the likely assumption that stocks belonging to the same industry sector would be more likely to co integrate over time, compared to stocks from different industry sectors. Therefore, the share of actual co integrating relationships found relative to the share of pairs identified due to spurious co integration across the in sample period would be improved on. We make use of the Global Industry Classification Standard (GICS).

When implementing a strategy like this, one of the first questions that arises is the frequency of data to be considered. We opt to use data on a daily frequency rather than high frequency for on advice from the supervisor of this thesis and to avoid the question of whether results are driven by the strategy of choice or the chosen method of imputing values where no bids/asking prices are given at one tick in time.

In all cases data will (and has already been in the case of S&P 500) be retrieved from the Eikon database.

Method

Theoretical Frameworks

Augmented Dickey Fuller Tests

Of key importance in the application of a trading algorithm based in the cointegration of stock prices is being able to test for stationarity in a time series. The Augmented Dickey-Fuller Test (ADF) provides a mean of doing so through testing for the presence of a unit root in the time series. If a unit root is present, the time series is non-stationary – this forms the null hypothesis of the test.

The model on which the test is performed is the following:

$$\nabla Y_t = c + aY_{t-1} + (\phi_1 \nabla Y_{t-1} + \phi_2 \nabla Y_{t-2} + \dots + \phi_k \nabla Y_{t-k}) + \varepsilon_t$$

The intuition behind this test is rather clear-cut: If the lagged level of the series (Y_{t-1}) provides no information regarding the change in Y_t , then the series shares the stochastic trend nature of the random walk. Hence the null is accepted. If on the other hand stationarity is present, a negative value of a is expected to be estimated. We thus have the following hypotheses:

$$H_0 = a \geq 0 \text{ Unit root is present}$$

$$H_a = a < 0 \text{ Unit root is not present}$$

The test function takes the following form (following a non-standard distribution):

$$ADF_{obs} = \frac{\hat{a} - 1}{\hat{\sigma}_{\hat{a}}}$$

In accordance with the reasoning above, a sufficiently negative value leads to the rejection of the null hypothesis and acceptance of the alternative hypothesis of stationarity.

Applications

Identifying pairs of stocks whose prices forms a cointegrating relationship, which forms the basis of the trading strategy perused in this article, requires a series of tests for stationarity.

Naturally, one necessary component is that the prices of the individual stocks are not stationary. While such a finding would be very surprising, it needs to be ensured. Denoting the price of the stock Y_t , the procedure outlined above may be applied to ensure this property. Should the null hypothesis be rejected, implying that the stock is $I(0)$, it would not be considered for trading. In this practical application, a p-value of 5 % was used as cut-off threshold.

A second property required to be present for a stock to be of interest is that the first difference of the price of the stock is stationary. This may be tested in similar fashion, using the following model specification:

$$\nabla^2 Y_t = c + a\nabla Y_{t-1} + (\phi_1 \nabla^2 Y_{t-1} + \phi_2 \nabla^2 Y_{t-2} + \dots + \phi_9 \nabla^2 Y_{t-9}) + \varepsilon_t$$

The hypotheses to be tested are the same as in previous test, namely:

$$H_0 = a \geq 0 \text{ Unit root is present}$$

$$H_a = a < 0 \text{ Unit root is not present}$$

Similarly, the test statistic is of the same form (see below), and a cut-off p-value of 5 % is elected for this application:

$$ADF_{obs} = \frac{\hat{a} - 1}{\hat{\sigma}_{\hat{a}}}$$

In this case, it is required that the alternative hypothesis is accepted. Hence, should the test statistic be such that the alternative hypothesis is discarded in favour of the null hypothesis at this stage, such a stock will not be considered for trade.

Any stocks within the considered trading universe (which in practice will refer to a specific sector of the S&P 500 index) that exhibit these properties may be considered as candidates to potentially have the required cointegrating property. At this point, it is clear that neither of the stocks price is stationary individually, i.e. it contains a stochastic trend. For two stocks to cointegrate, it must be the case that the stochastic trends in the two time series are driven by the same factors, and in some linear combination may cancel out.

Should this be the case, the changes in one stock's price would be explainable by the changes in the other stocks price. Referring to the two stocks as Y and X, finding the ratio in which X explains Y is a simple linear estimation of Y on X:

$$Y_t = c + \beta X_t + \varepsilon_t$$

This regression may of course be run for any two stocks Y and X – however it needs not (in fact most likely is not) be the case that the stock prices vary such that the stochastic trends cancel out. Fortunately, this may easily be tested by examining if the residual ε_t is stationary. Should ε_t be stationary, it implies that the specific combination $Y_t - \beta X_t$ must be stationary as is (if not already) evident by rewriting the equation above:

$$Y_t - \beta X_t = c + \varepsilon_t$$

At this point, one may carry out an ADF test for the stationarity of the time series of the residual ε_t , the model would then take the following form:

$$\nabla \varepsilon_t = c + a\varepsilon_{t-1} + (\phi_1 \nabla \varepsilon_{t-1} + \phi_2 \nabla \varepsilon_{t-2} + \dots + \phi_k \nabla \varepsilon_{t-k}) + u_t$$

Naturally, the null-, alternative hypothesis and test statistic takes the same form as in previous applications:

$$H_0 = a \geq 0 \text{ Unit root is present}$$

$$H_a = a < 0 \text{ Unit root is not present}$$

$$ADF_{obs} = \frac{\hat{a} - 1}{\hat{\sigma}_{\hat{a}}}$$

In this application, rejection of H_0 and thereby acceptance of the alternative hypothesis of stationarity implies that the stocks Y and X cointegrate in the relation $Y - \beta X$.

However, it is not clear that Y should be estimated on X and not vice versa. Hence, the following specification needs also be tested:

$$X_t = \beta' Y_t + c' + \varepsilon'_t$$

The corresponding model specification required to test for the stationarity of this residual would be:

$$\nabla \varepsilon'_t = c + a \varepsilon'_{t-1} + (\phi_1 \nabla \varepsilon'_{t-1} + \phi_2 \nabla \varepsilon'_{t-2} + \dots + \phi_k \nabla \varepsilon'_{t-k}) + u_t$$

This test will be identical to the one outlined above. We elect to only consider a pair of stocks Y and X for trading should both these tests for stationarity (on ε_t and ε'_t) reject the null hypothesis of a unit root. The p-value will in general not be the same for the two tests. The natural choice is to trade on the relationship where the associated test statistic provides the lowest p-value – i.e. most strongly rejects the null hypothesis of non-stationarity. This practice is hence implemented.

As response to previous literature identifying breaks from cointegrating structure as a major source of losses in trading models of this nature, we choose to utilize a cut-off p-value of 2 % when performing tests for stationarity in cointegrating relationships, so as to reduce the number of false positives accepted.

Error Correction Models

Error correction models are a specific type of VAR representation of the relationship between two variables that are of particular use given that the two variables exhibit certain characteristics. Generally speaking, an error correction model between two variables Y_t and X_t may take the following form:

$$\nabla Y_t = c + \alpha \nabla X_t + \beta (Y_{t-1} - c' - \beta' X_{t-1}) + \epsilon_t \quad (1)$$

This representation will be of use when Y_t and X_t are such that:

1. Both Y_t and X_t are non-stationary. That is, neither variable is $I(0)$.
2. Both Y_t and X_t are $I(1)$ - the first difference of both variables is stationary.
3. There exists a cointegrating relationship between Y_t and X_t .

If two variables exhibit these variables, then all parts of (1) will be stationary ($I(0)$), as $Y_{t-1} - c' - \beta' X_{t-1}$ is the stationary residual from the cointegrating relationship between Y_t and X_t . Meaningful estimates of the parameters of the equation may therefore be estimated under the conditions that the variables Y_t and X_t have these properties.

Of particular interest to the applications in this paper is the parameter β . This parameter is an estimate of the degree to which the variable Y_t responds to divergence from the constant expected value of the cointegrating relationship between Y_t and X_t . The parameter is expected to be negative should such a relationship exist, as a negative parameter indicates that Y_t responds to a divergence in such a way as to diminish the difference to the constant expected value. The larger the parameter β , the more the time series of Y_t develops in such a way as to offset a divergence.

Applications

The first two characteristics required for the error correction representation of the two variables Y_t and X_t to be of interest is a prerequisite for stock prices to be considered for trading in the first place, for reasons outlined above in the section regarding the Augmented Dickey-Fuller test. The third criterion, that there exists a cointegrating relationship between Y_t and X_t , is not fulfilled for most pairs of stock prices Y_t and X_t . However, for the pairs in which trading is conducted, such a property is believed to be present.

Therefore, an error correction model may be meaningfully estimated for all pairs of stocks Y_t and X_t considered for trading. The benefit of undertaking this practice is rather straightforward: Ponder that Y_t and X_t are cointegrating time series of stock prices. Estimating an error correction model of the form of (1) will thus provide an estimate of how quickly the price of stock Y responds to a divergence from the constant value of the cointegrating relationship. *Ceteris paribus*, trading on a pair of stocks Y and X will be preferable to a pair of stocks Z and K if the estimated parameter $\hat{\beta}$ is larger for the pair Y and X than it is for the pair Z and K .

This result stems from the fact that the larger estimate of $\hat{\beta}$ indicates that the price of Y responds quicker to revert to the constant value of the cointegrating relationship between Y and X than does correspondingly the price of Z . Hence, one would expect quicker reversal back to the constant expected value in the pair Y and X compared to the pair Z and K . Other things equal, this implies that the same arbitrage could be obtained in a shorter time frame - a clear advantage.

However, trading on the pair of stocks involves trading in both assets. Therefore, the response of both stock prices traded on to a divergence from the constant expected value of the cointegrating relationship is of interest. Hence, the following estimation must also be carried out for stock X :

$$\nabla X_t = c' + \alpha' \nabla Y_t + \gamma(Y_{t-1} - c' - \beta' X_{t-1}) + \epsilon'_t \quad (2)$$

In this representation, γ will give the rate at which X responds to a divergence. Note that γ in (2), contrary to β in (1), is expected to take positive values. The relevant question in the setting of this paper is of course how greatly Y and X responds jointly to a divergence from the constant value of the cointegrating relationship.

The cointegrating relationship between Y and X that presumably has been found before the implementation of the error correction model was found to be given by $Y_t = c' + \beta' X_t + u_t$. The observant reader will realize that in fact, $(Y_{t-1} - c' - \beta' X_{t-1}) = u_{t-1}$. For such a pair of stocks, positions will be taken in the ratio of longing Y by one unit and short selling X by β' units. Thus, the total holding will be $k(Y - \beta' X)$ where k is some scalar constant. Let's assume for now that $k = 1$. Hence, the total response rate of the shares held to a divergence

from the constant expected value will be given by $\beta - \beta'\gamma$. Both terms are expected to be negative, and the larger the absolute value of the expression, the more rapidly the two variables are expected to respond to revert to the constant expected value of $Y - \beta'X$.

Practical Implementation

The method used to identify pairs for trading based on exhibiting the property of cointegration was outlined in the theoretical section above. However, the trading algorithm is of course somewhat more complex than so in its implementation. This section will outline the practicals of the implementation of the trading strategy apart from what was covered in the theoretical section.

Sufficient Level of Divergence, Convergence and Stop Loss

Once a pair of shares on which trading is to be undertaken given fortuitous enough circumstances are present has been isolated, the question remains as to what the actual decision rules regarding trading will be. Essentially, it boils down to answering three questions: At what level of divergence from the constant expected value should a trade be opened and positioned be taken? At what level of divergence should said positions be sold off? When should a trade be shut down, positions sold off and any losses accepted if things go wrong?

The most frequent approach to answering the first question within the literature is to require that the level of divergence exceeds two standard deviations relative to the in-sample period. Say shares are traded in the relationship $Y - 2X$. That is, some multiple of one share of Y is bought and the same multiple of 2 shares of X are sold short. Assume that the standard deviation of the value of $Y - 2X$ is 10 during the in-sample period. This would imply that a deviation of 20 from the average during the in sample period is sufficient to initiate a trade in the out-sample period.

Some authors elect to use a more aggressive strategy, taking positions at one standard deviation instead. We choose to go with two standard deviations, partly because it makes results more directly comparable to previous studies and partly because avoiding trades of poor quality has been identified as a key point of interest in said studies.

Further, we apply the requirement that convergence must have started before we take a position. That is, it's not sufficient that $Y - 2X$ diverged by 20 in the example used above. Instead, we require that this has occurred, and subsequently the divergence has decreased to 1.75 standard deviations. In the example used, a trade will be opened if $Y - 2X$ takes a value of 20 or larger and subsequently decreases to 17.5 or lower during the out-sample period, relative to the in-sample mean value. For the reader that finds it hard to keep track of the

numbers, refer to the first section of the results section where a visual representation is provided.

The second question, regarding when a position is to be sold off, is simpler. Two strategies are possible: either hold the assets until complete convergence has occurred, or sell off the assets once sufficient level of convergence has occurred. It should be pointed out here that the latter strategy may be superior to the former strategy even if complete convergence would occur at a later point, as it allows a trade to be completed in a shorter time frame. We elect to settle for 0.5 standard deviations as sufficient level of convergence. Framed in the example above, if a trade has been opened in the shares Y and X in the relationship $Y = 2X$, then these will be sold off once the divergence from the in-sample mean decreases to 5, given that the standard deviation during the in-sample period is 10.

Perhaps the most complicated question is the third question, that is how the stop loss should reasonably be formulated. The literature sports many choices of stop loss, most seemingly rather arbitrary (set levels, or a given number of time periods). While perhaps sufficient for academic purposes, the choice was made to implement a more complex but likely more suitable stop loss function for the algorithm reported on in this paper. It consists of three decision rules:

1. Given that only a short time has elapsed, a more “forgiving” stop loss is applied: Should the level of divergence exceed 2 standard deviations (which would be 0.25 standard deviations greater than where positions were taken), the trade will be cancelled and losses accepted at current levels.
2. Past this point, but still within a reasonable time frame, a trade will be cancelled should divergence return to as large a value as it was at the point where positions were taken. Should the divergence be on the interval greater than 1.75 s.d when transition is made between the two stop loss functions, the assets are sold off and losses accepted at current levels.
3. If a trade is open and continuously at levels of positive return but does not reach the level of convergence required for positions to be sold off (0.5 standard deviations), an end point is defined where the assets are sold off and profits taken at current levels.

The thought process behind implementing these decision rules is rather simple. Based on the cointegrating properties of the stocks bought and sold short during the in-sample period, it is expected that the divergence will return to the in-sample mean. However, some “noise” is to be expected in the movement of the stock prices – hence a greater degree of freedom must be allowed during an initial period such that positions are not immediately sold off as soon as the level of divergence increases slightly.

However, after a period of time has elapsed, it is expected that divergence should be closer to the in-sample mean than it was when positions were taken. This is the second component of the stop-loss. Finally, if too much time were to pass by where divergence is closer to the in-sample mean than where positions were taken but never sufficiently close to the required level

(0.5 standard deviations), it makes sense to cancel the trade and take whatever profits are made at this point for two reasons: Firstly, it may be that something has occurred that breaks the cointegrating relationship identified in the in-sample period. Secondly, the trade is open for a long period of time, effectively occupying capital that could be committed to a more profitable trade.

Naturally, the question remaining is at where the cutoffs between the different decision rules of the stop loss should be implemented. It appears rather obvious that a set number of days would be suboptimal – as demonstrated the by ECM regressions, one should not expect convergence at the same rate for all cointegrating pairs of shares. Instead, we elect to do it by relating to the average number of days to convergence for the individual set of shares during the in-sample period. The first (more forgiving, set at 2 s.d.) decision rule of the stoploss is implemented at a time frame smaller than 25 % of the average time to convergence in-sample. The second (1.75 s.d.) is then correspondingly implemented should the time from entry into the trade exceed 25 % of the average time to convergence in-sample. Lastly, should the trade still be open when the time from entry into the trade hits 120 % of the average time to convergence in-sample, the trade is cancelled.

The observant reader will realize that there is further a complication to this matter. Ponder we're trading in the two stocks X and Y, and the spread from the in-sample mean at day 5 is 20 while on day 6 it's 17. Let's assume our algorithm tells us that in this setting, we should purchase Y and sell X short in some ratio if the spread falls to 17.5. That is, during day 6, the price crossed 17.5 at some point in time. Two possibilities are possible: the model can either be specified to trade intra-day, in which case the purchase would be made when the spread hit 17.5, or it could be specified to trade only at the closing of the day (purchasing at a spread of 17). Naturally, the corresponding situation occurs when the stocks are later sold off.

In order to keep the model somewhat relevant in a modern framework, the choice was made to allow for intra-day trading. Evaluating the results of a model that trades only at one point in time each bank day simply seems irrelevant if one is to compare it to contemporary trading strategies. Notice that this is not the same as taking advantages of high frequency data. Opportunities that present themselves intraday but are not available at the closing of the day will not be capitalized on since the algorithm uses day frequency data.

While the theoretical nature of the implementation of the ECM model has been outlined in previous section above, its practical implementation deserves a note. Initially, only pairs of stocks with larger than average absolute value for the error-correction parameter were accepted. For reasons which will be made clear in the results section, this criterion was flipped. Pairs of stocks with a value larger than 120 % of the average value were disregarded in the practical implementation of the algorithm. This was coupled with a very basic screening process, requiring (1) that the pair of stocks had diverged sufficiently and subsequently converged in the in-sample period and (2) that the average time to do so was less than 125 bank days (approximately half a year).

The above description is a complete depiction of the trading algorithm. However, of perhaps just as significant importance is the choice of in- and out-sample periods. The trading algorithm operates on bank days, with a year consisting on average of slightly more than 250 days. As standard within this strand of literature, we employ a rolling window. Our in-sample consists of 800 bank days, the out-sample is 50 bank days long. Hence, the first trading window has an in-sample of bank days 1 through 800, and an out-sample of days 801 through 850. In the subsequent window, the in-sample period ends in 850 – this is the rolling window, the out-sample in one period is part of the in-sample of the following period.

The reader will probably notice that the out-sample is substantially longer than the in-sample. There are two reasons for electing this type of specification. Firstly, we notice an improvement in the quality of the trades undertaken when using a longer in-sample period, especially beyond 300 bank days. We attribute the majority of this result to the fact that a longer in-sample requires cointegration across entire years – hence avoiding potential issues if a pair of stocks was to cointegrate only during some seasons. We also notice a slight improvement with shorter out-sample periods. It is the belief of the authors that this is primarily due to the fact that a longer out-sample period implies a larger time frame where cointegrating structures within the in-sample period may break up.

Alternative Methods

There are alternative methods that may be perused when studying pairs trading. Apart from electing to use a different test for cointegration, an approach which has received quite some exploration in the literature is the use of a sum square distance measure as a “proxy” for cointegration. Ponder two stocks X and Y are considered for trade. This measure would then essentially be:

$$S^2 = \sum (Y - X)^2$$

Stocks are then selected based on minimizing the S^2 measure. The profitability of a strategy of this nature depends on identifying pairs with the property of cointegration, though as evident it does not test for this property. Authors of previous literature has argued that the majority of pairs of stocks identified in this way should also have the property of cointegration (Gatev et al., 2006).

While this method may or may not be superior to performing actual tests for cointegration in a practical implementation, it was dismissed in this paper if for no other reason because it is not very convincing from a theoretical point of view.

A second alternative method which could have been used is the stochastic measure of spread between two shares. This method models the spread between two stocks as a function of a latent variable, typically taken to follow a Brownian motion. It (the latent variable that is) is then presumed to be equal to the spread plus a term of white noise – this imposes and allows to

capture mean-reverting properties. Also this method is disregarded in this paper, based on the reason that it would not be as clear whether potential results were solely driven by cointegrating properties in the stocks traded on.

Perceived Advantages of this Implementation

At this point, the in dismissible and highly relevant question that most readers probably ask is “Why should I believe this paper would reach any results not already previously derived?”. For one, the implementation of the ECM framework has potentially large effects on returns. The rate at which trades of sufficiently good quality can be concluded is just as important as the quality of said trades. The importance of the number of opportunities identified will be made obvious in the subsequent section concerning the results of the paper.

Secondly, the stop loss function used and outlined above is surprisingly more advanced than most we’ve found in previous literature. The discrepancy in choices is large, and most seemingly taken ad hoc. To illustrate this with some contemporary examples, Engelberg, Gao & Jagannathan (2009) use a stop-loss of 10 days, Mark Whistler recommends a stop loss of 10 % of divergence on which trade is entered in his book “Capturing profits and hedging risk with statistical arbitrage strategies”, while George Miao (2014) uses a stop loss of 100 % of divergence level at which trade is entered. We believe that approach taken in this paper should hold a distinct advantage to alternatives of this nature.

Returns and Equity Calculations

When evaluating the results of the trading strategy, a pivotal point is of course the way returns and equity are calculated. Starting with returns calculations, this issue may not be as straight forward as it would appear. The reason is that the trading method applied in this paper short sells potentially as much or more stocks than it purchases. Illustrating with an example: let’s say we trade in shares X and Y. Our algorithm provides that we should purchase stocks in the ratio $Y - 2X$. That is, for every stock of Y we purchase, we should short sell 2 stocks of X.

Assuming the price of Y to be 2 and the price of X to be 1, this means that if we were to purchase $1 \times (Y - 2X)$, we would hold 200 worth of stocks in Y and -200 worth of stocks in X. Ponder that the trade is successful and 20 days later positions are sold off with a net profit of 50. What would be the return to this trade? Clearly 50 was made, but we held a net of 0 value in stocks to achieve this return. Additionally, different regulations may apply when selling stocks short, requiring the trader to commit additional capital.

With the aim to be as transparent and keep things as simple as possible, we elected to calculate the return in the following manner:

$$\text{Return} = \frac{\text{Net result of trade}}{\text{Purchasing value of stocks held long} + \text{purchasing value of stocks held short}}$$

Hence, the return measure is a relative one, capturing the profit/losses of the trade relative to the size of the investments made (be that positive or negative). In the illustrative example above, the return would be $\frac{50}{200+200} = 12.5 \%$.

Return measures tells something about the quality of the trades identified. However, they do not capture the profitability of the trading strategy. Even a return as high as the fictive one used in the illustrative example would be terrible if the average time taken to achieve said return was say five years. If the same return could be obtained within a month, it'd be absurdly profitable.

In order to evaluate the performance of the trading algorithm, we therefore constructed a capital allocation model to which the algorithm adheres. It's rather simple, and takes the following properties:

- Whenever a trade opportunity is identified, a certain amount of capital is allocated to this trade and deducted from the amount of capital available.
- Whenever a trade is cancelled (convergence, stop loss or end of out-sample), the amount of capital committed to this trade plus/minus earnings/losses from the trade are refunded to the amount of capital available and may be reinvested.

To be more precise, denoting the amount of capital available by CA, investment into a trade opportunity is made by the size $10 + (CA-10)/3$. Naturally, a condition that $CA > 10$ is imposed such that no trades are taken if $CA < 10$.

This function provides an easy means of keeping track of the relative equity at the end of an out-sample period to the equity at the onset of the trading period. Graphs displaying the equity curves of the different sectors are presented in the subsequent results section. The observant reader will notice that the function specified is not very aggressive in the way that it allocates the available capital at a point in time to identified available trading opportunities. Likely, results achieved could be improved with an approach that more actively attempts to keep all available capital occupied in trades at any point in time.

However, as a significant part of the paper consists of evaluating the viability of cointegration-based trading, we feel it's more important to undertake more trades (committing less to each) to give a more complete picture of the quality of the trades identified by an algorithm of this type.

Results

The Big Picture – does Cointegration Based Trading Still Work?

There are two requirements that must be met for the trading strategy perused in this paper to be successful: (1) the trades identified must on average be sufficiently profitable, (2) they must be closable quickly enough and numerous enough such that capital is effectively turned over. We shall explore these in greater detail later on, but the quick answers to these two requirements are:

- The trades identified are on average of sufficiently good quality.
- The number of opportunities identified is only large enough to form a successful strategy if the sector considered is large enough.

It is standard procedure in trading strategies based on cointegration to consider stocks within the same industry. The reason is that there is then some economic justification making it plausible that stocks would be cointegrated. Since we've followed this standard, the results are presented on a section basis.

The largest sectors where most trades are identified are in order: Consumption, financials, information. The remaining sectors are small, with few opportunities identified. For this reason, there is little to report on smaller sectors apart from that not much happens. We will therefore leave these out of the discussion as they add very little, the reader interested may either obtain these by running the code provided in the appendix on the appropriate data or request it.¹

With that said, overall performance measures of the strategy for the 3 sectors discussed above are presented in the table below:

Table 1: Performance of the trading strategy in the consumption, financials and information sectors.

Sector	Average return*	Equity growth
Consumption	0.689706 %	138.2887 %
Financials	0.379063 %	51.8362 %
Information	0.684242 %	20.0621 %

*Average returns refers to the following measure: Each out-sample period, the relative average return of the trades conducted in the period is calculated. The average return reported in the table is the mean of all these values for each time period where trading occurs within a given sector.

Table 1 provides two insights: the trades identified are on average sufficiently good, but it appears that the amount of trading opportunities identified is more critical than the quality of the identified opportunities. Naturally, this result is somewhat dependent on the way which the algorithm allocates available capital to a trade. Undoubtedly, an algorithm allowed to

¹ For a complete set of results, the authors may be reached on the following email: olofhan2235@gmail.com

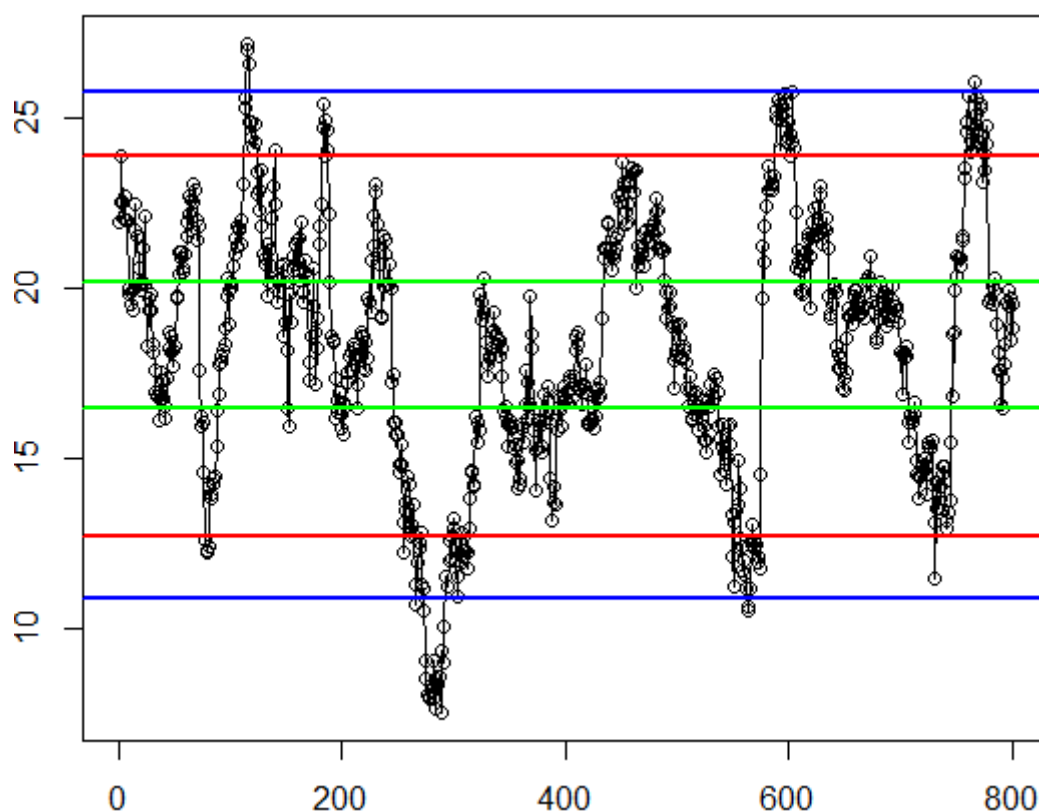
allocate more or less all available capital to a trade opportunity if they are scarce would likely perform better in the smaller sectors.

As a point of reference, the S&P index grew by approximately 55 % during the time period trading was conducted on. Summing up the big picture, the results indicate that cointegration based trading can produce returns well in the excess of market returns given that the sector in which it is implemented is sufficiently large that enough opportunities to enter trades may be identified.

A Sneak Peek: Visual representations

Before looking into the specifics of the results, it may be illustrative to consider an example of a pair of stocks that are traded on. The example below is taken for two stocks in the consumption sector. The first graph displayed below displays the observed price of the difference between the two stocks in the cointegrating relationship in the in-sample period.

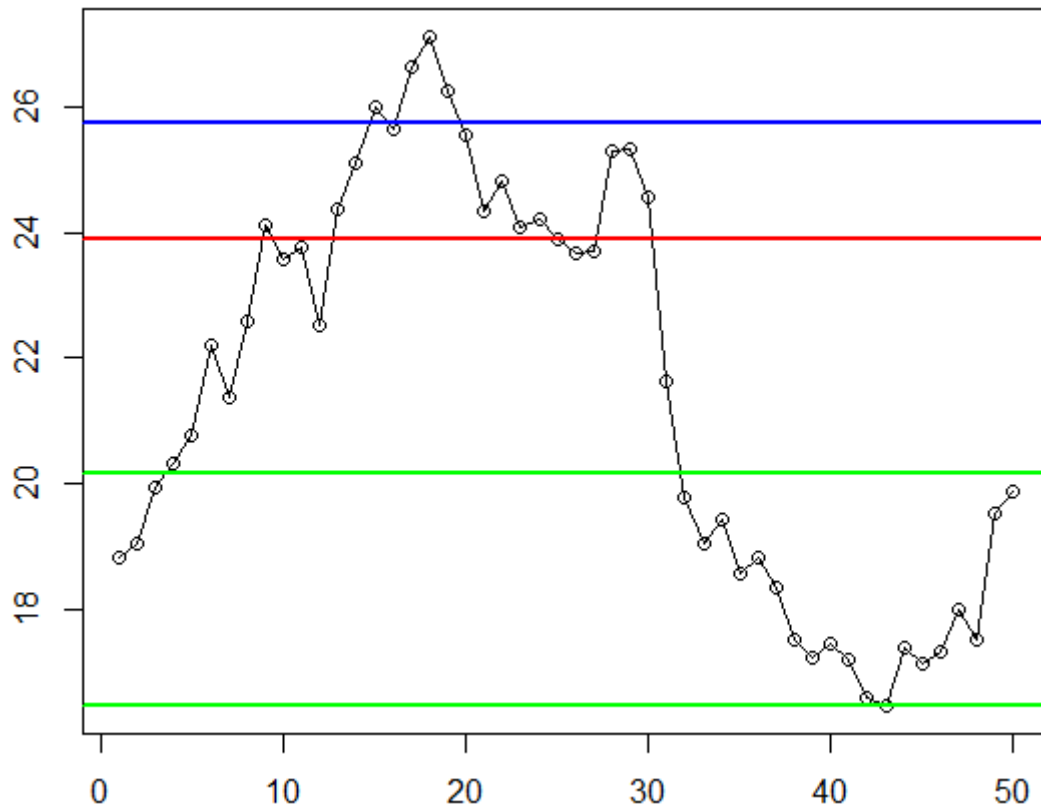
Graph 1: The difference between two stock prices in the cointegrating relationship during an in-sample estimation.



The blue lines indicate the level of divergence which must have been exceeded before a trade would be considered, had it been a trading period. The red lines indicate the level of spread at which a trade would have been opened if sufficient divergence had occurred previously. Finally, the green lines indicate the level at which positions would have been sold off.

Graph two shows the relationship between the same shares in the corresponding out-sample period.

Graph 2: The difference between the same two stock prices in the cointegrating relationship during the out-sample period.



Positions were taken in the two stocks when the graph crossed the red line, and sold off when the graph crossed the green line. Notice that the reason the stocks are not sold off as the spread increases again after positions have been taken is that the first, more forgiving phase of the stop-loss is still active. Had the spread remained at these higher levels exceeding the red line for too long, the second stop loss would have taken effect and positions would have been sold off netting losses. This example demonstrates the importance of a useful stop loss function.

Error Correction Model – a Surprising Find

As outlined in the method section, we expected to find that stocks with a larger absolute value of λ and γ (see expressions below for reference) in the error correction model would perform better.

$$\begin{aligned}\nabla Y_t &= c + \alpha \nabla X_t + \beta(Y_{t-1} - c' - \beta' X_{t-1}) + \epsilon_t \\ \nabla X_t &= c' + \alpha' \nabla Y_t + \gamma(Y_{t-1} - c' - \beta' X_{t-1}) + \epsilon'_t\end{aligned}$$

However, to our perplexion empirical results clearly showed the opposite. Run across numerous time periods, we consistently noted that stocks with larger absolute value of β and γ did close out trades entered faster (as predicted), but with worse results! This latter seemed very counterintuitive, as the poor results seemed to be driven by hitting the stop loss function more frequently – therefore likely representing a break from the cointegrating relationship identified during the in-sample period. This is surprising as a large absolute value of β and γ indicates larger responses in Y_t and X_t to return from deviations from the cointegrating relationship in $t - 1$.

We credit the supervisor for this thesis, Lars Forsberg, with finally identifying the likely cause behind this result. The equation for the error correction model is estimated assuming there is some cointegrating relationship between stocks Y and X . However, the largest downfall to a trading strategy of this sort is when stocks are found to appear cointegrated in the in-sample period, but in fact are not. Said relationship will then of course most frequently brake up in the out sample period, and losses are likely.

Pairs of stocks where a larger absolute value of β and γ is estimated may be more likely to pass the test for cointegration spuriously as they exhibit a greater degree of some mean reverting property that is in fact not true cointegration. We believe this to be the reason that we find higher quality trades among pairs of stocks where β and γ take a small magnitudes, and fail to find any other conceivable explanation behind the results consistently indicating this.

For this reason, all models were respecified to elect not the stocks with the largest absolute value of β and γ , but rather stocks that had comparatively smaller absolute values of β and γ . More specifically, trading was conducted on stocks that had no greater than 120 % of the average weighted (see theory section) value for β and γ .

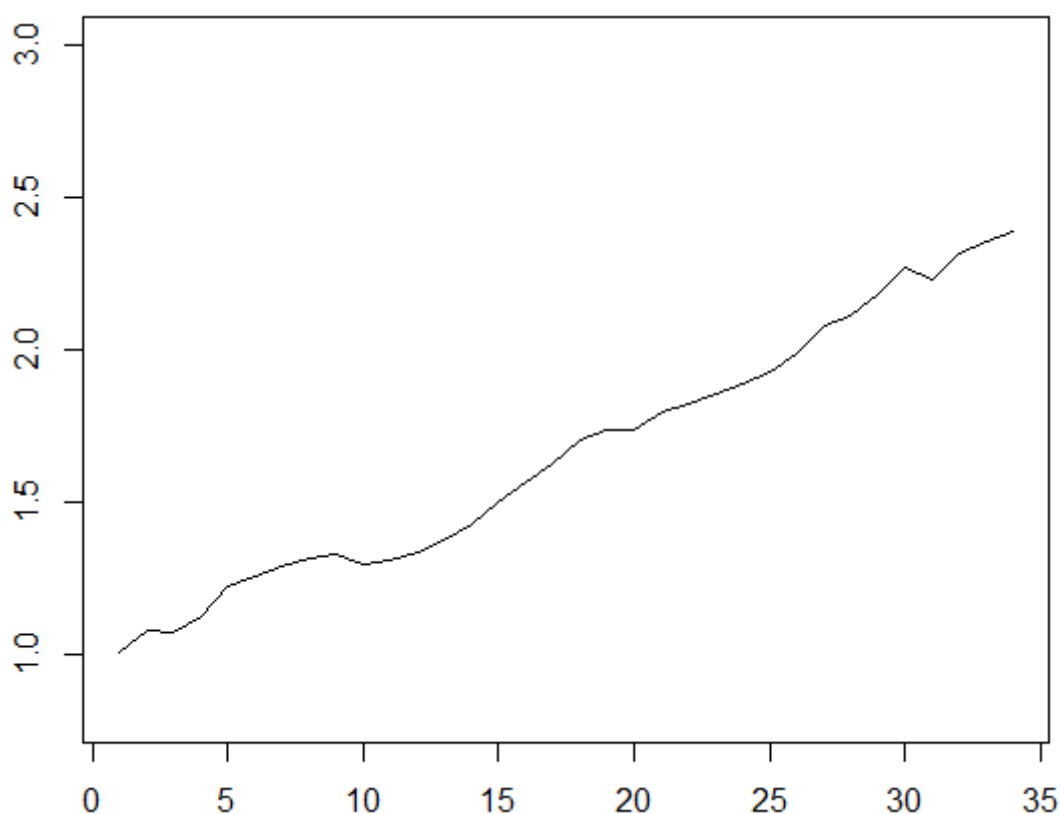
The Largest Sector: Consumption

The largest sector of the S & P 500 index is the consumption index. In the time period considered, 99 stocks were listed within the consumption section throughout all of the years studied. This is almost twice the number of any other sector. Keep in mind that the number of potentially cointegrating relationships grows with the square of the size of the trading

universe considered. Hence, twice the number of stocks implies four times as many potentially cointegrating relationships.

Our results indicate that a trading strategy based on the principle of cointegration can be very successful in this type of a framework. The graph below shows the development of the equity of the trading portfolio in the time period traded on (March 2008 through November 2014):

Graph 3: Equity of the portfolio trading on stocks in the consumption sector



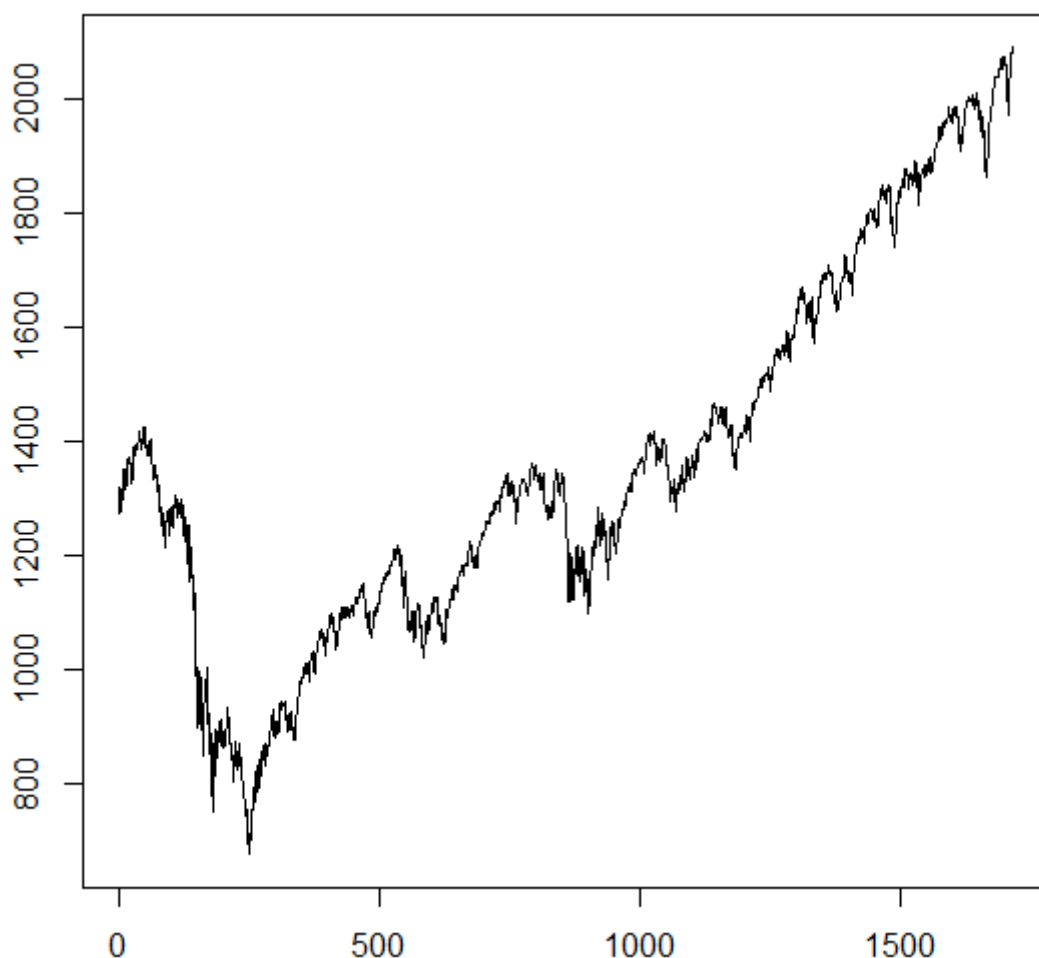
The vertical axis is an index starting in 1 at the beginning of the time period. The horizontal axis shows the number of trading periods undertaken cumulatively, each trading period being 50 bank days long.

For reference, the S & P 500 index grew with approximately 55 % within this time frame. The equity of the portfolio grew with 132.29 %. Hence, our trading strategy provided an increase in value by more than twice the market increase in this time period. As discussed previously, this result would likely be larger with a strategy that more aggressively committed capital to trades to keep it “busy” in trades at all times as much as possible – but we don’t consider this the main aim of the paper.

Perhaps more importantly, the quality of the trades identified are good across the entire time period. This showcases the very desirable property of a strategy of this type: the results are not dependent on the performance of the market. It is as likely to make profit in a scenario where all stocks are declining in value as it is in a scenario where all stocks are improving.

To see the difference, it may be illustrative to compare to the development of the market index during this time period:

Graph 4: The S&P 500 index, March 2008 through November 2014



The vertical axis displays the index, the horizontal axis gives the ordered bank days in the time period traded on.

Simple inspection by eye shows that the cointegration based approach to trading yielded far more stable returns than the market did.

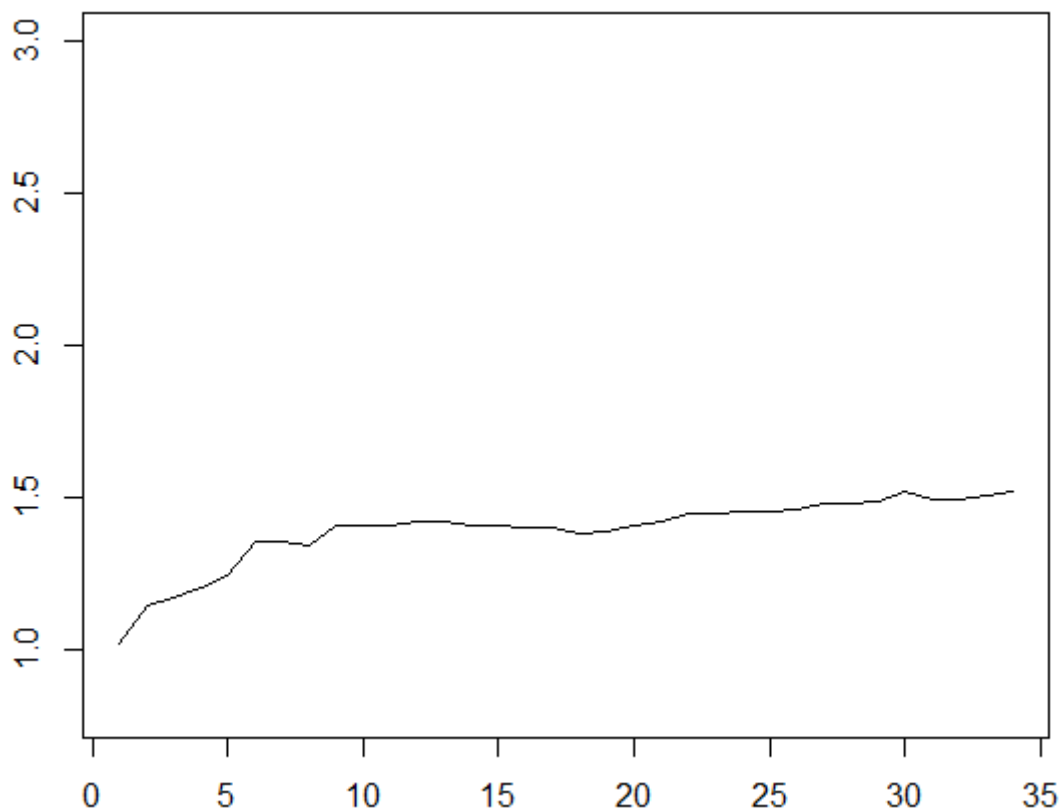
Financials – Barely Large Enough

The sector in which the algorithm identified the second most opportunities is the financial sector. In general, the trades undertaken in the financial sector were of lower quality than the trades in the consumption sector. The relative earning of the average trade is approximately 55 % of the relative earning of the average trade in the consumption sector. A speculation as to why this may be the case is that it could be that identified cointegrating relationships have a larger chance of breaking up subsequently in the financial sector during the years of the great crash around 2008 and the recovery thereafter. It could also be that cointegration based trading simply performs worse on these stocks, either due to some property of the stocks themselves or due to arbitrages being competed for more fiercely.

Notice that had it been possible to keep capital active in trades to the same extent as the consumption sector, this margin is still good enough to earn approximately 50 % more than the market index. However, the sector is smaller compared to the consumption sector and not as many opportunities are identified, and as a result the strategy appreciates equity by approximately 52 %. This is slightly less than the increase in value of the market during the time period. A trading strategy investing more aggressively would most likely beat the market. The method employed in this paper is very ineffective at binding up capital in trades during periods where only one or two trading opportunities are presented, which happens multiple times in the financial sector.

A graphical representation provides a rather good overview of the results for this sector:

Graph 5: Equity of the portfolio trading on stocks in the financial sector



The uneven rate of change visible in the graph corresponds very well to the amount of identified trade opportunities within the sector. By far the most trades per trading window are identified and capitalized in during the first time periods. Little happens with few identified opportunities for trades during the years in the middle of the time period, but a few periods of more trades occur towards the end of the time period.

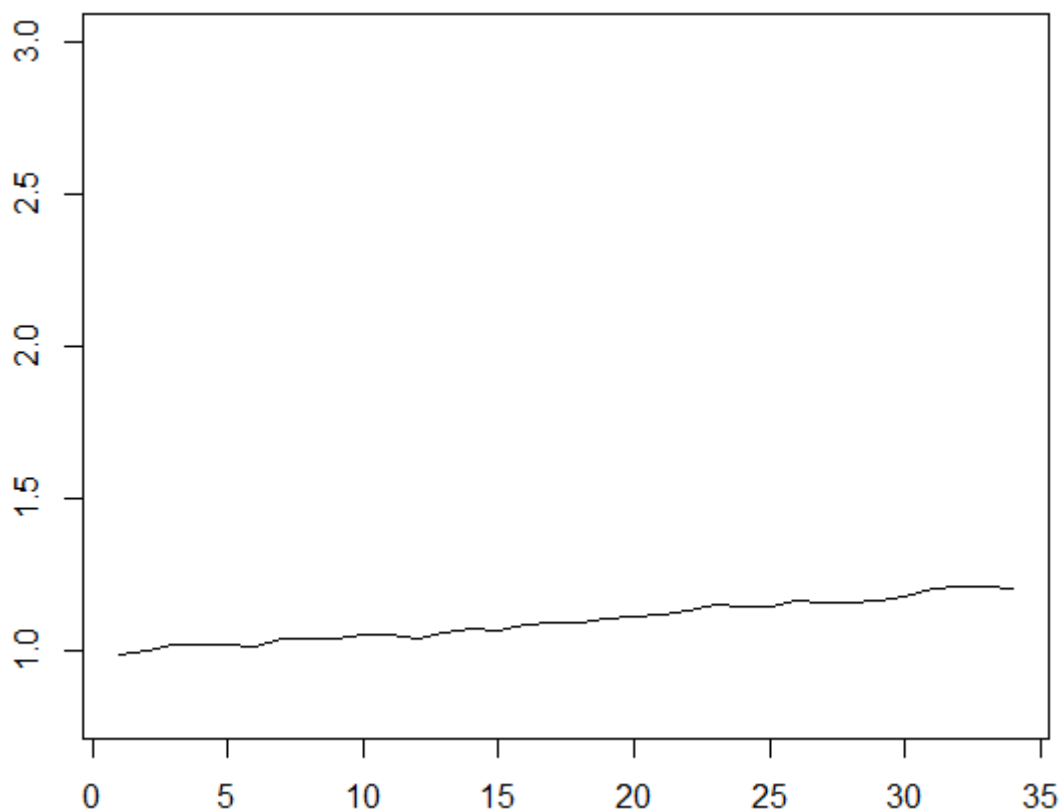
Though the graph hasn't got the smooth shape of the equity curve for the consumption sector, notice that this is not due to large losses in equity, but rather relative lack of activity leading to

relative lack of equity improvement. The method still provides a very stable returns pattern with small losses compared to the market.

When there Just aren't Enough Trades – Information

The third largest sector in terms of identified trades is the information sector. The results of the information sector further highlights the observations made in the financial sector. Trades undertaken in this sector are on average of almost exactly the same quality as trades in the consumption sector, yet equity growth is not even one sixth of what it is in the consumption sector. The reason behind this is of course that the number of trading opportunities is very limited. No point in reiterating already stated points, the graph over the equity curve for the information sector looks as follows:

Graph 6: Equity of the portfolio trading on stocks in the information sector



Once again, we see a very stable (but insufficient) development in equity across the time period.

Summary

Our study shows that, given the extensions that have been made in this study, the cointegration approach to pairs trading may still be a very successful strategy for pairs trading even in as contemporary of a time frame as 2005-2014 (trading during 2008 through 2014). The most successful implementation of the strategy yielded an equity improvement of more than twice the growth of the market index.

However, our results indicate that for this type of a strategy to be successful, it is key that the industry examined is sufficiently large such that invested capital may be “kept busy” to large enough an extent. Smaller industries are not even close to presenting enough opportunities for an implementation of the sort outlined in this paper to be of use.

To little surprise and therefore of questionable use, it also further confirms that the results to a trading strategy of this nature are very stable with few losses compared to net investments on the market.

On the contrary, a very surprising find was made regarding the implementation and use of ECM models for the traded pairs. The theoretically supported and initially presumed use for the ECM model was to identify pairs which did not only cointegrate, diverge and subsequently converged during the in-sample period, but also had the properties to quickly revert to the constant expected value of the cointegrating relationship.

While it was found that pairs with a high estimated tendency to revert in said manner did indeed reach convergence quicker, they were also found to hit stop loss functions significantly more often, indicating that the cointegrating relationship between the involved stocks appeared more likely to break in the out-sample period. The proposed explanation for this is that stocks with high estimated absolute values for β and γ in the following equation are more likely to have mean reverting properties leading to spurious cointegration during in-sample periods:

$$\begin{aligned}\nabla Y_t &= c + \alpha \nabla X_t + \beta(Y_{t-1} - c' - \beta' X_{t-1}) + \epsilon_t \\ \nabla X_t &= c' + \alpha' \nabla Y_t + \gamma(Y_{t-1} - c' - \beta' X_{t-1}) + \epsilon'_t\end{aligned}$$

For this reason, the opposite implementation of what was theoretically predicted proved useful: keeping only the stocks that did not have large absolute values of β and γ yielded trades with higher average returns.

Further Research – Where to Go Now?

Given the results of this paper, we feel that the most promising outset for further research would be to investigate methods aimed at identifying more opportunities to act in divergences in cointegrating relationships between financial assets. Two obvious candidates are forfeiting

the standard industry borders in favour of including more potentially related assets and the application of a Johansens method so as to allow for cointegration between more than two assets.

A third possibility worth examining would be to implement a similar method on high-frequency data. Of course, nothing stands in the way of combining this with one or both of the aforementioned prospects of interesting topics.

References

“An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity”, Engelberg, Gao & Jagannathan, (2009), Third Singapore International Conference on Finance 2009

“High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two–Stage Correlation and Cointegration Approach” George Miao, (2014), International Journal of Economics and Finance; Vol. 6, No. 3; 2014

“Practical Methods of Financial Engineering and Risk Management: Tools for Modern Financial Professionals”, 1.st ed, 2014, Rupak Chatterjee

“Pairs Trading: Performance of a Relative-Value Arbitrage Rule” (2006), Evan Gatev, William N. Goetzmann and K. Geert Rouwenhorst, The Review of Financial Studies Vol. 19, No. 3 pp. 797-827

“Pairs Trading: Performance of a Relative-Value Arbitrage Rule” (2010), Evan Gatev, William N. Goetzmann and K. Geert Rouwenhorst

“High Frequency Equity Pairs Trading: Transaction Costs, Speed of Execution and Patterns in Returns” (2010), Bowen, Hutchinson & O’Sullivan, Journal of Trading, Summer 2010, Vol. 5, No. 3, 31-38

“Does Simple Pairs Trading Still work?” (2010) Do & Faff, Financial Analyst Journal 66, no. 4, 83-95

“Statistical Arbitrage Pairs: Can Cointegration Capture Market Neutral Profits?” (2013), Hoel C H, <https://brage.bibsys.no/xmlui/handle/11250/169897>

"Capturing profits and hedging risk with statistical arbitrage strategies" (2004), Whistler Mark

Technical Appendix: Code Used to Perform Trading Analysis (R)

```
library("tseries")
```

```
library("ecm")
```

```
# read csv file
```

```
Cons = read.csv2("Information 2005-2015.csv")
```

```
Cons2 <- Cons
```

```
#Set in-sample!
```

```
Cons3 <- Cons2
```

```
Cons3 <- data.frame(Cons3)
```

```
Cons3 <- Cons3[1600:2400,]
```

```
Cons2 <- Cons3
```

```
Assetlist <- list()
```

```
ADFstat1 <- 1
```

```
ADFstat2 <- 1
```

```
Tradinglist <- list()
```

```
Assetlistl1 <- list()
```

```

#Drop stocks I(0) and check that stocks are I(1)
for (i in 2:ncol(Cons2)){
  ADFStatX <- adf.test(Cons3[,i])
  if (ADFStatX$p.value > 0.05){
    ADFStatDiffX <- adf.test(diff(Cons3[,i]))
    if (ADFStatDiffX$p.value < 0.05){
      Assetlist1[[length(Assetlist1)+1]] <- colnames(Cons3[i])
    }
  }
}

```

```

#create pairs
for (i in 1:length(Assetlist1)){
  for (j in i:length(Assetlist1)){if (i != j){
    M <- matrix(nrow = 1, ncol = 2)
    M[1,1] <- Assetlist1[[i]]
    M[1,2] <- Assetlist1[[j]]
    Assetlist[[length(Assetlist)+1]] <- M}}
}

```

```
HR1 <- 0
```

```
HR2 <- 0
```

```
#Create linear estimations and test for cointegration of stocks
```

```

for (i in 1:length(Assetlist)){
  Linest1 <- lm(Cons3[,Assetlist[[i]][1,1]]~Cons3[,Assetlist[[i]][1,2]])
  Linest2 <- lm(Cons3[,Assetlist[[i]][1,2]]~Cons3[,Assetlist[[i]][1,1]])
}

```

```

HR1  <- as.numeric(Linest1$coefficients[2])
HR2  <- as.numeric(Linest2$coefficients[2])
ADFstat1 <- adf.test(Linest1$residuals)
ADFstat2 <- adf.test(Linest2$residuals)

if (ADFstat1$p.value < 0.02 & ADFstat2$p.value < 0.02){if (ADFstat1$p.value < ADFstat2$p.value){M
<- matrix(nrow = 1, ncol = 3)

M[1,1] <- Assetlist[[i]][1,1]
M[1,2] <- Assetlist[[i]][1,2]
M[1,3] <- HR1
Tradinglist[[length(Tradinglist)+1]] <- M} else {M <- matrix(nrow = 1, ncol = 3)

M[1,1] <- Assetlist[[i]][1,2]
M[1,2] <- Assetlist[[i]][1,1]
M[1,3] <- HR2
Tradinglist[[length(Tradinglist)+1]] <- M}}
}

```

#We're gonna need to look at the spread in these relationships

```

Spreadlist <- list()

for (i in 1:length(Tradinglist)){

  Spread <- Cons2[,Tradinglist[[i]][1,1]] - as.numeric(Tradinglist[[i]][1,3]) * Cons2[,Tradinglist[[i]][1,2]]

  SpreadMean <- mean(Spread)

  Spread2sd <- 2*sd(Spread)

  M <- matrix(nrow = 1, ncol = 2)

  M[1,1] <- SpreadMean
  M[1,2] <- Spread2sd

  Spreadlist[[length(Spreadlist)+1]] <- M}

```

#Store values from in-sample!

```

SpreadEstimationlist <- list()

```

```

for (i in 1:length(Spreadlist)){

  SpreadEstimation_Demeaned <- Cons2[,Tradinglist[[i]][1,1]] -
as.numeric(Tradinglist[[i]][1,3])*Cons2[,Tradinglist[[i]][1,2]] - Spreadlist[[i]][1,1]

  M <- matrix(nrow = length(Cons2[,Tradinglist[[i]][1,1]]), ncol = 1)

  M[,1] <- SpreadEstimation_Demeaned

  SpreadEstimationlist[[length(SpreadEstimationlist)+1]] <- M}

#build durationmeasure for insample divergense och covergense
Durationlist <- list()

for (i in 1:length(Tradinglist)){

  M <- matrix(nrow = 1, ncol = 9)

  M[1,1] <- 0    #in trade
  M[1,2] <- 0    #convergence indicator for pair
  M[1,3] <- 0    #the day we start the trade
  M[1,4] <- 0    #Days we have hold a trade that ends in stoploss
  M[1,5] <- 0    #Days we have hold a trade that ends in covergence
  M[1,6] <- 0    #times we cloose a trade whit convergence
  M[1,7] <- 0    #times we cloose a trade whit stoploss
  M[1,8] <- 0    #Average days in trade that ends in covergence
  M[1,9] <- 0    #Average dagar i trade that ends in stoploss

  Durationlist[[length(Durationlist)+1]] <- M}

for (j in 2:length(Cons2[,1])){

  for (i in 1:length(SpreadEstimationlist)){

    if (Durationlist[[i]][1,1] == 0 & abs(SpreadEstimationlist[[i]][j-1]) >= 1*Spreadlist[[i]][1,2] &
abs(SpreadEstimationlist[[i]][j]) <= 1*Spreadlist[[i]][1,2]){Durationlist[[i]][1,2] <- 1

    }

    if (Durationlist[[i]][1,1] == 0 & Durationlist[[i]][1,2] == 1 & abs(SpreadEstimationlist[[i]][j-1]) >=
1.5/2*Spreadlist[[i]][1,2] & abs(SpreadEstimationlist[[i]][j]) <=
1.5/2*Spreadlist[[i]][1,2]){Durationlist[[i]][1,1] <- 1

    Durationlist[[i]][1,3] <- j
  }
}

```

```

Durationlist[[i]][1,2] <- 0}

if (Durationlist[[i]][1,1] == 1 & SpreadEstimationlist[[i]][j-1] > 0*Spreadlist[[i]][1,2] &
SpreadEstimationlist[[i]][j]<=0*Spreadlist[[i]][1,2]){Durationlist[[i]][1,1] <- 0

Durationlist[[i]][1,5] <- Durationlist[[i]][1,5]+j-Durationlist[[i]][1,3]

Durationlist[[i]][1,3] <- 0

Durationlist[[i]][1,6] <- Durationlist[[i]][1,6]+1}

if (Durationlist[[i]][1,1] == 1 & SpreadEstimationlist[[i]][j-1] < 0*Spreadlist[[i]][1,2] &
SpreadEstimationlist[[i]][j]>=0*Spreadlist[[i]][1,2]){Durationlist[[i]][1,1] <- 0

Durationlist[[i]][1,5] <- Durationlist[[i]][1,5]+j-Durationlist[[i]][1,3]

Durationlist[[i]][1,3] <- 0

Durationlist[[i]][1,6] <- Durationlist[[i]][1,6]+1}

if (Durationlist[[i]][1,1] == 1 & 1.45*Spreadlist[[i]][1,2] <
abs(SpreadEstimationlist[[i]][j])){Durationlist[[i]][1,1] <- 0

Durationlist[[i]][1,4] <-Durationlist[[i]][1,4]+j-Durationlist[[i]][1,3]

Durationlist[[i]][1,3] <- 0

Durationlist[[i]][1,7] <- Durationlist[[i]][1,7]+1}

if (j == length(Cons2[,1])){
  Durationlist[[i]][1,8] <- Durationlist[[i]][1,5]/Durationlist[[i]][1,6]
  Durationlist[[i]][1,9] <- Durationlist[[i]][1,4]/Durationlist[[i]][1,7]
}
}}

Durationlist2 <- list()
for (i in 1:length(Durationlist)){
  Durationlist2[[length(Durationlist2)+1]] <- Durationlist[[i]][1,8]}

Durationlist3 <- list()
for (i in 1:length(Durationlist)){

```

```
Durationlist3[[length(Durationlist3)+1]] <- Durationlist[[i]][1,6]}
```

```
Durationlist2[is.na(Durationlist2)] <- 0
```

```
Durationlist3[is.na(Durationlist3)] <- 0
```

```
for (i in 1:length(Durationlist)){
```

```
  Durationlist[[i]][1,8][is.na(Durationlist[[i]][1,8])] <- 0}
```

```
asdf <- 0
```

```
for (i in 1:length(Durationlist3)){
```

```
  asdf <- asdf + Durationlist3[[i]]
```

```
}
```

```
raknaMedelvarde <- asdf/length(Durationlist3)
```

```
RaknaDagar <- 0
```

```
for (i in 1:length(Durationlist2)){
```

```
  RaknaDagar <- RaknaDagar + Durationlist2[[i]]
```

```
}
```

```
RaknaDagarMedelvarde <- RaknaDagar/length(Durationlist2)
```

```
#### Redo tradinglist and exclude trades whit high time to convergense and no convegrences  
recorded
```

```
TradinglistProxy <- list()
```

```
for (i in 1:length(Durationlist3)){
```

```
  if (Durationlist3[[i]] > 0){
```

```
    if (Durationlist2[[i]] < 125){
```



```

TradinglistProxy[[length(TradinglistProxy)+1]] <- i}
}
}

TradinglistTemp <- list()
for (i in 1:length(TradinglistProxy)){
  TradinglistTemp[[length(TradinglistTemp)+1]] <- Tradinglist[[TradinglistProxy[[i]]]]
}
Tradinglist <- TradinglistTemp

```

#Build ECM models for screening

```

ECMlist <- list()
for (i in 1:length(Tradinglist)){
  IndepVarDiff <- diff(Cons3[,Tradinglist[[i]][1,1]])

  DepVarDiff <- diff(Cons3[,Tradinglist[[i]][1,2]])

  Linest <- lm(Cons3[,Tradinglist[[i]][1,1]]~Cons3[,Tradinglist[[i]][1,2]])
  Linest2 <- lm(Cons3[,Tradinglist[[i]][1,2]]~Cons3[,Tradinglist[[i]][1,1]])
  Resids <- as.numeric(Linest$residuals)

  Resids2 <- -1/Linest$coefficients[2]*as.numeric(Linest$residuals)
  Resids <- head(Resids,-1)
  Resids2 <- head(Resids2,-1)
  Res <- Resids

```

```
Res2 <- Resids2
```

```
ECM <- lm(IndepVarDiff ~ DepVarDiff + Res)
```

```
ECMStat <- ECM$coefficients[3]
```

```
ECM2 <- lm(DepVarDiff ~ IndepVarDiff + Res2)
```

```
ECMStat2 <- ECM2$coefficients[3]
```

```
ECMlist[[length(ECMlist)+1]] <- ECMStat + abs(as.numeric(Tradinglist[[i]][1,3]))*ECMStat2  
}
```

```
ECMcumulative <- 0
```

```
for (i in 1:length(ECMlist)){
```

```
  ECMcumulative <- ECMcumulative + ECMlist[[i]]
```

```
}
```

```
ECMavg <- ECMcumulative/length(ECMlist)
```

```
## Screen tradinglist according to ECM results
```

```
TradinglistTemp2 <- list()
```

```
for (i in 1:length(Tradinglist)){
```

```
  if (abs(ECMlist[[i]]) <= 1.2*abs(ECMavg)){TradinglistTemp2[[length(TradinglistTemp2)+1]] <-  
  Tradinglist[[i]]}
```

```
}
```

```
Tradinglist <- TradinglistTemp2
```

```
Spreadlist <- list()
```

```
for (i in 1:length(Tradinglist)){
```

```

Spread <- Cons2[,Tradinglist[[i]][1,1]] - as.numeric(Tradinglist[[i]][1,3]) * Cons2[,Tradinglist[[i]][1,2]]
SpreadMean <- mean(Spread)
Spread2sd <- 2*sd(Spread)
M <- matrix(nrow = 1, ncol = 2)
M[1,1] <- SpreadMean
M[1,2] <- Spread2sd
Spreadlist[[length(Spreadlist)+1]] <- M}

Constrading = read.csv2("Information 2005-2015.csv")

Constrading2 <- Constrading

#set outsample period
Constrading2 <- data.frame(Constrading2)
Constrading2 <- Constrading2[2401:2450,]

##Constrading2[,Tradinglist[[i]][1,1]] Asset price independent
##Constrading2[,Tradinglist[[i]][1,2]] Asset price dependent
##Constrading2[,Tradinglist[[i]][1,3]] HR

SpreadTradinglist <- list()

for (i in 1:length(Spreadlist)){
  SpreadTrading_Demeaned <- Constrading2[,Tradinglist[[i]][1,1]] -
as.numeric(Tradinglist[[i]][1,3])*Constrading2[,Tradinglist[[i]][1,2]] - Spreadlist[[i]][1,1]
  M <- matrix(nrow = length(Constrading2[,Tradinglist[[i]][1,1]]), ncol = 1)
  M[,1] <- SpreadTrading_Demeaned

```

```
SpreadTradinglist[[length(SpreadTradinglist)+1]] <- M}
```

```
#creating stuff we will need for trades
```

```
Bank <- 100
```

```
ImInFor <- 0
```

```
Return <- 0
```

```
TickerTradeProfit <- 0
```

```
TickerTradeLoss <- 0
```

```
TickerTrade <- 0
```

```
TradeDayOpen <- 0
```

```
a <- 0
```

```
Loserlist <- list()
```

```
Loserlist2 <- list()
```

```
b <- 0
```

```
c <- 0
```

```
Winnerlist <- list()
```

```
Alist <- list()
```

```
Vafan <- 0
```

```
Vafanlist <- list()
```

```
GåttNer <- 0
```

```
Minneslista <- list()
```

```
DagariVunnaTrades <- 0
```

```
TickerYay <- 0
```

```
#We will need to remember some stuff for each pair traded on
```

```

for (i in 1:length(SpreadTradinglist)){
  M <- matrix(nrow = 1, ncol = 4)
  M[1,1] <- 0    #In a trade for the pair
  M[1,2] <- 0    #convergence indicator for pair
  M[1,3] <- 0    #day we enter trade
  M[1,4] <- 0    #how much we bought
  Minneslista[[length(Minneslista)+1]] <- M
}

```

#Trading algorithm below:

#for each pair, for each day:

```

for (j in 2:length(Constrading2[,1])){
  for (i in 1:length(SpreadTradinglist)){

    #trade not open and price more than 2sd away from avreage
    if (Minneslista[[i]][1,1] == 0 & abs(SpreadTradinglist[[i]][j-1]) >= 1*Spreadlist[[i]][1,2] &
        abs(SpreadTradinglist[[i]][j]) <= 1*Spreadlist[[i]][1,2]){Minneslista[[i]][1,2] <- 1
    }

    if (Minneslista[[i]][1,1] == 0 & Minneslista[[i]][1,2] == 1 & abs(SpreadTradinglist[[i]][j-1]) >=
        1.5/2*Spreadlist[[i]][1,2] & abs(SpreadTradinglist[[i]][j]) <= 1.5/2*Spreadlist[[i]][1,2] & Bank >
        10){Minneslista[[i]][1,1] <- 1

      TickerTrade <- TickerTrade +1

      Minneslista[[i]][1,3] <- j

      Minneslista[[i]][1,2] <- 0

      Minneslista[[i]][1,4] <- (10 + (Bank-10)/3)/(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]] +
        abs(as.numeric(Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]])

      Bank <- Bank - 10 - (Bank-10)/3}

    #Trade open and we crossed mean price from above

```

```

if (Minneslista[[i]][1,1] == 1 & SpreadTradinglist[[i]][j-1] > 0.25*Spreadlist[[i]][1,2] &
SpreadTradinglist[[i]][j]<=0.25*Spreadlist[[i]][1,2]){Minneslista[[i]][1,1] <- 0

b <- b +1

TickerTradeProfit <- TickerTradeProfit + 1

Return <- Return + 1/2*Spreadlist[[i]][1,2]/(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]] +
abs(as.numeric(Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]])

Bank <- Bank + Minneslista[[i]][1,4]*(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]] +
abs(as.numeric(Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]]) +
Minneslista[[i]][1,4]*1/2*Spreadlist[[i]][1,2]

Minneslista[[i]][1,4] <- 0

if(b < 10){

M <- matrix(nrow = 1, ncol = 1)

M[1,1] <- i

Winnerlist[[length(Winnerlist) + 1]] <- M}

DagariVunnaTrades <- DagariVunnaTrades + (j-Minneslista[[i]][1,3])

TickerYay <- TickerYay +1

Minneslista[[i]][1,3] <- 0}

#Trade open and we crossed mean price from below

if (Minneslista[[i]][1,1] == 1 & SpreadTradinglist[[i]][j-1] < -0.25*Spreadlist[[i]][1,2] &
SpreadTradinglist[[i]][j]>=-0.25*Spreadlist[[i]][1,2]){Minneslista[[i]][1,1] <- 0

TickerTradeProfit <- TickerTradeProfit + 1

Return <- Return +
1/2*Spreadlist[[i]][1,2]/(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]]+abs(as.numeric(Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]])

Bank <- Bank + Minneslista[[i]][1,4]*(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]] +
abs(as.numeric(Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]]) +
Minneslista[[i]][1,4]*1/2*Spreadlist[[i]][1,2]

Minneslista[[i]][1,4] <- 0

DagariVunnaTrades <- DagariVunnaTrades + (j-Minneslista[[i]][1,3])

TickerYay <- TickerYay +1

Minneslista[[i]][1,3] <- 0}

#Build stop loss! Trade open, early days

```

```

if ((j - Minneslista[[i]][1,3]) < trunc(0.25*Durationlist[[i]][1,8]) & Minneslista[[i]][1,1] == 1 &
1*Spreadlist[[i]][1,2] < abs(SpreadTradinglist[[i]][j])){

  Return <- Return + (1.5/2*Spreadlist[[i]][1,2] -
1*Spreadlist[[i]][1,2])/(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]]+abs(as.numeric(Tradin
glist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]])

  Bank <- Bank + Minneslista[[i]][1,4]*(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]] +
abs(as.numeric(Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]]) +
Minneslista[[i]][1,4]*(1.5/2*Spreadlist[[i]][1,2] - 1*Spreadlist[[i]][1,2])

  Minneslista[[i]][1,4] <- 0

  Minneslista[[i]][1,1] <- 0

  Minneslista[[i]][1,3] <- 0

  Minneslista[[i]][1,2] <- 0

  TickerTradeLoss <- TickerTradeLoss +1

  Vafan <- Vafan +1

  M <- matrix(nrow = 1, ncol = 1)

  M[1,1] <- i

  Vafanlist[[length(Vafanlist) + 1]] <- M

}

```

#new stoploss, trade open, stoploss changing to tighter.

```

if ((j - Minneslista[[i]][1,3]) == trunc(0.25*Durationlist[[i]][1,8]) & Minneslista[[i]][1,1] == 1 &
1.5/2*Spreadlist[[i]][1,2] < abs(SpreadTradinglist[[i]][j])){

  Return <- Return + (1.5/2*Spreadlist[[i]][1,2] -
abs(SpreadTradinglist[[i]][j]))/(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]]+abs(as.numeric(
Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]])

  Bank <- Bank + Minneslista[[i]][1,4]*(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]] +
abs(as.numeric(Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]]) +
Minneslista[[i]][1,4]*(1.5/2*Spreadlist[[i]][1,2] - abs(SpreadTradinglist[[i]][j]))

  Minneslista[[i]][1,4] <- 0

  Minneslista[[i]][1,1] <- 0

  Minneslista[[i]][1,3] <- 0

  Minneslista[[i]][1,2] <- 0

  TickerTradeLoss <- TickerTradeLoss +1

  Vafan <- Vafan +1

  M <- matrix(nrow = 1, ncol = 1)

```

```

M[1,1] <- i

Vafanlist[[length(Vafanlist) + 1]] <- M
}

#stop loss level two, trade open - making money

if ((j - Minneslista[[i]][1,3]) > trunc(0.25*Durationlist[[i]][1,8]) & (j - Minneslista[[i]][1,3]) <=
trunc(1.2*Durationlist[[i]][1,8]) & Minneslista[[i]][1,1] == 1 & 1.5/2*Spreadlist[[i]][1,2] <
abs(SpreadTradinglist[[i]][j])){

  Return <- Return + (1.5/2*Spreadlist[[i]][1,2] -
1.5/2*Spreadlist[[i]][1,2])/(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]]+abs(as.numeric(Tra
dinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]])

  Bank <- Bank + Minneslista[[i]][1,4]*(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]] +
abs(as.numeric(Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]]) +
Minneslista[[i]][1,4]*(1.5/2*Spreadlist[[i]][1,2] - 1.5/2*Spreadlist[[i]][1,2])

  Minneslista[[i]][1,4] <- 0

  Minneslista[[i]][1,1] <- 0

  Minneslista[[i]][1,3] <- 0

  Minneslista[[i]][1,2] <- 0

  TickerTradeLoss <- TickerTradeLoss +1

  Vafan <- Vafan +1

  M <- matrix(nrow = 1, ncol = 1)

  M[1,1] <- i

  Vafanlist[[length(Vafanlist) + 1]] <- M
}

#stopp loss, dependent on time. We have been in the trade too long and close trade

if (Minneslista[[i]][1,1] == 1 & (j - Minneslista[[i]][1,3]) == trunc(1.2*Durationlist[[i]][1,8]) & j <
length(Constrading2[,1])){

  Return <- Return + (1.5/2*Spreadlist[[i]][1,2] -
abs(SpreadTradinglist[[i]][j]))/(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]]+abs(as.numeric(
Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]])

  Bank <- Bank + Minneslista[[i]][1,4]*(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]] +
abs(as.numeric(Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]]) +
Minneslista[[i]][1,4]*(1.5/2*Spreadlist[[i]][1,2] - abs(SpreadTradinglist[[i]][j]))

```



```

Minneslista[[i]][1,4] <- 0
Minneslista[[i]][1,1] <- 0
Minneslista[[i]][1,3] <- 0
Minneslista[[i]][1,2] <- 0
TickerTradeProfit <- TickerTradeProfit +1
Vafan <- Vafan +1
M <- matrix(nrow = 1, ncol = 1)
M[1,1] <- i
Vafanlist[[length(Vafanlist) + 1]] <- M
}

#if trade open last day
if (Minneslista[[i]][1,1] == 1 & j == length(Constrading2[,1]))
{Minneslista[[i]][1,1] <- 0

#if 1.5 SD < lower than absolut value of price
if (1.5/2*Spreadlist[[i]][1,2] < abs(SpreadTradinglist[[i]][j])){TickerTradeLoss <- TickerTradeLoss +1
c <- c +1

Return <- Return + (1.5/2*Spreadlist[[i]][1,2] -
abs(SpreadTradinglist[[i]][j]))/(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]]+abs(as.numeric(
Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]])

Bank <- Bank + Minneslista[[i]][1,4]*(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]] +
abs(as.numeric(Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]]) +
Minneslista[[i]][1,4]*(1.5/2*Spreadlist[[i]][1,2] - abs(SpreadTradinglist[[i]][j]))

Minneslista[[i]][1,4] <- 0
M <- matrix(nrow = 1, ncol = 4)
M[1,1] <- Spreadlist[[i]][1,2] - abs(SpreadTradinglist[[i]][j])
M[1,2] <- Tradinglist[[i]][1,1]
M[1,3] <- Tradinglist[[i]][1,2]
M[1,4] <- Tradinglist[[i]][1,3]
Loserlist[[length(Loserlist)+1]] <- M
Minneslista[[i]][1,3] <- 0

```

```

if (c < 10){
  M <- matrix(nrow = 1, ncol = 1)
  M[1,1] <- i
  Loserlist2[[length(Loserlist2) + 1]] <- M}

} #Trade closed, counting losses.

#if trade open last day and we are "in the money".
else {TickerTradeProfit <- TickerTradeProfit +1

  Return <- Return + (1.5/2*Spreadlist[[i]][1,2] -
abs(SpreadTradinglist[[i]][j]))/(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]]+abs(as.numeric(
Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]])

  Bank <- Bank + Minneslista[[i]][1,4]*(Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,1]] +
abs(as.numeric(Tradinglist[[i]][1,3]))*Constrading2[Minneslista[[i]][1,3],Tradinglist[[i]][1,2]]) +
Minneslista[[i]][1,4]*(1.5/2*Spreadlist[[i]][1,2] - abs(SpreadTradinglist[[i]][j]))

  Minneslista[[i]][1,4] <- 0
}

Minneslista[[i]][1,3] <- 0}

}}#close day close pair

TidTillAvslut <- DagariVunnaTrades/TickerYay

#creating net return per trade measure
Net <- Return/TickerTrade
Net <- Net*100

### Code for creating plots, not active
#plot(Constrading2[,Tradinglist[[13]][1,1]] -
as.numeric(Tradinglist[[13]][1,3])*Constrading2[,Tradinglist[[13]][1,2]] - Spreadlist[[13]][1,1])

#lines(Constrading2[,Tradinglist[[13]][1,1]] -
as.numeric(Tradinglist[[13]][1,3])*Constrading2[,Tradinglist[[13]][1,2]] - Spreadlist[[3]][1,1])

```

```

#abline(h = Spreadlist[[13]][1,2], col = "blue", lwd = 2)

#abline(h = -Spreadlist[[13]][1,2], col = "blue", lwd = 2)


#plot(1:nrow(Constrading2), Constrading2[,Tradinglist[[3]][1,1]] - as.numeric(Tradinglist[[3]][1,3]) *
Constrading2[,Tradinglist[[3]][1,2]])

#lines(Constrading2[,Tradinglist[[3]][1,1]] - as.numeric(Tradinglist[[3]][1,3]) *
Constrading2[,Tradinglist[[3]][1,2]])

#plot(1:nrow(Cons2), Cons3[,Tradinglist[[23]][1,2]], col = "blue")

#lines(Cons3[,Tradinglist[[23]][1,2]]*as.numeric(Tradinglist[[23]][1,3]), col = "blue")

#lines(Cons3[,Tradinglist[[6]][1,1]] - as.numeric(Tradinglist[[6]][1,3])*Cons3[,Tradinglist[[6]][1,2]] -
Spreadlist[[6]][1,1])

#abline(h = mean(Cons2[,Tradinglist[[3]][1,1]] - as.numeric(Tradinglist[[3]][1,3]) *
Cons2[,Tradinglist[[3]][1,2]])+Spreadlist[[3]][1,2], col = "blue", lwd = 2)

#abline(h = mean(Cons2[,Tradinglist[[3]][1,1]] - as.numeric(Tradinglist[[3]][1,3]) *
Cons2[,Tradinglist[[3]][1,2]])-Spreadlist[[3]][1,2], col = "blue", lwd = 2)

#abline(h = mean(Cons2[,Tradinglist[[3]][1,1]] - as.numeric(Tradinglist[[3]][1,3]) *
Cons2[,Tradinglist[[3]][1,2]])+1.5/2*Spreadlist[[3]][1,2], col = "red", lwd = 2)

#abline(h = mean(Cons2[,Tradinglist[[3]][1,1]] - as.numeric(Tradinglist[[3]][1,3]) *
Cons2[,Tradinglist[[3]][1,2]])-1.5/2*Spreadlist[[3]][1,2], col = "red", lwd = 2)

#abline(h = mean(Cons2[,Tradinglist[[3]][1,1]] - as.numeric(Tradinglist[[3]][1,3]) *
Cons2[,Tradinglist[[3]][1,2]])+0.25*Spreadlist[[3]][1,2], col = "green", lwd = 2)

#abline(h = mean(Cons2[,Tradinglist[[3]][1,1]] - as.numeric(Tradinglist[[3]][1,3]) *
Cons2[,Tradinglist[[3]][1,2]])-0.25*Spreadlist[[3]][1,2], col = "green", lwd = 2)

#abline(h = mean(Cons2[,Tradinglist[[3]][1,1]] - as.numeric(Tradinglist[[3]][1,3]) *
Cons2[,Tradinglist[[3]][1,2]])-0*Spreadlist[[3]][1,2], col = "purple", lwd = 2)

#plot(x,y1,type="l",col="red")

#lines(x,y2,col="green")

```