UNIVERSITY OF COPENHAGEN
FACULTY OR DEPARTMENT

Undergraduate Thesis

Jens Pedersen

Statistical Arbitrage in the U.S. Equities Market

Pairs Trading in Standard & Poor's 500

Supervisor: Heino Bohn Nielsen

Submitted on: 29 April 2021

Number of characters: 48.667 (20 np)

Statistical Arbitrage in the U.S. Equities Market

Pairs Trading in Standard & Poor's 500

Jens Pedersen
University of Copenhagen
Department of Economics

May 2021

Abstract

This paper applies a statistical arbitrage algorithm on the U.S. equities index, Standard and Poor's 500, with daily historical prices, from January 2010 to December 2019. The algorithm constructs a portfolio that trades stocks matched into pairs using a wild bootstrap co-integration approach. The empirical results from the algorithm show that the strategy yields an annualised return of 1.17 per cent and a Sharpe Ratio of 1.10. Our findings corroborate previous literature on the topic of modest returns for the strategy compared to a benchmark index. However, the pairs trading portfolio exhibits less volatility, as the risk-adjusted returns are greater.

Keywords: Statistical arbitrage, pairs trading, vector autoregressive model, co-integration, wild bootstrap, algorithmic trading

JEL classification: C32, C58 and G17

Table of Contents

1 Introduction **3**

2 Theoretical Framework **5**
2.1 The Trading Strategy 5
2.2 The Vector Autoregressive Model 6
2.3 Model Specification 7
2.4 Co-integration 7
2.5 Rank Determination 10

3 Methodological Framework **15**
3.1 Data 15
3.2 The Statistical Algorithm 16
3.3 The Trading Algorithm 19

4 Empirical Results **20**

5 Discussion **21**
5.1 The Normality Assumption & Z-score 21
5.2 The Trading Information & Liquidity Problems 21
5.3 Pairs Trading Dual-listed Firms 21

6 Conclusion **22**

7 References **23**

8 Appendices **25**

1 Introduction

Pairs trading is a statistical arbitrage strategy, pioneered by Nunzo Tartaglia and a group of physicists, mathematicians and computer scientists at Morgan Stanley in the mid-1980s (Gatev et al., 2005). The strategy is built on monitoring pairs of shares whose prices are assumed to be driven by the same economic forces, and trading on the temporary deviations from a supposed long-run equilibrium. The risk-free nature of the strategy arises from the opening of opposing positions for each trade, i.e., by going long in the relatively undervalued stock and short in the relatively overvalued stock, and thus, a profit is made by closing the positions when the spread mean-reverts (Lin et al., 2006).

Traditional methods of pairs trading have sought to identify pairs based on correlation and non-parametric decision rules. In this paper, we select pairs based on the presence of a co-integrating relationship as proposed by Lin et al. (2006), Caldeira and Moura (2013), and Nielsen and Rahbek (2020) to derive a precise and dynamic definition of the long-run equilibrium price spread.

In the presence of co-integration between pairs of stocks that have prices that are non-stationary, i.e., representing a stochastic process, there exists a vector $\beta = (1,b)'$ such that a linear combination

$$s_t = \beta' X_t = p_t + bq_t \quad (1.1)$$

[1]

is a stationary process, i.e., have mean and variance that is constant over time. From this, it is possible to exploit the long-run equilibrium relationship between the stocks from the relative short-term misalignment in prices by the aforementioned strategy (Caldeira and Moura, 2013; Nielsen and Rahbek, 2020).

This paper extends the work on co-integration in pairs trading by Gatev et al. (2005), Lin et al. (2006), Avellaneda and Lee (2010), and Caldeira and Moura (2013) by easing the requirement of shares from different sectors not being traded, as long as they satisfy the co-integration criterion. This is implemented, as this paper assumes that these requirements have been applied as a measure against deficiencies in the underlying statistical techniques used. Thus, this paper aggravates the conditions for the selection process of the pairs, as models are now tested for model misspecification, i.e., autocorrelation, and bootstrapped version of the rank test is instated, strengthening the robustness of the pairs in the traded portfolio.

In this paper, the sample covers adjusted daily stock prices on the Standard & Poor's 500 index (S&P 500) from January 2010 to December 2019, a total of 721,131 observations.

[1] Here p_t and q_t denotes the prices or log-prices of the respective stocks.

Findings

In our analysis, we find that our statistical arbitrage strategy generates an annualised return of 1.17 per cent with a Sharpe Ratio of 1.10. A long position in the S&P 500 in the same period, would have generated 30.43 per cent in annualised return with a Sharpe Ratio of 0.76. This result show that co-integrated is profitable, albeit much less so than a simple long position in the index. However, the pairs trading strategy exhibit lower risk, as the risk-adjusted return (as evidenced by the Sharpe Ratio) is greater. Our findings consolidate prior literature on the topic of co-integrated pairs trading.

Programs and data

All material, including data, programs and code can be found at https://github.com/jeopedersen/Bachelor-s-Thesis.git.

The entire decomposition and data cleansing for the algorithm, i.e., the pairs trading code, has been written in the statistical computing language R or Microsoft Excel. The daily adjusted prices have been obtained from Bloomberg L.P. software.

Acknowledgements

I would like to thank Prof. Heino Bohn Nielsen for great supervision. Furthermore, I would like to thank Ph.D. Anton Rask Lundborg Hansen for helping me when stuck with coding.

2 Theoretical Framework

In the following section, this paper presents the theoretical framework, providing the foundation for the pairs trading strategy. This paper gives a thorough description of the framework including derivations when necessary. Additional derivations are enclosed in the appendix.

First, a brief introduction to the trading strategy. Second, an overview of the vector autoregression model and how it is specified. Third, a theoretical introduction to the concept of co-integration, and lastly, an introduction to the rank determination of the co-integrated vector autoregressive model.

2.1 The Trading Strategy

To execute the pairs trading strategy, this paper computes the deviation from the long-run equilibrium. This deviation is quantified as a z-score, defined as

$$z_t = \frac{s_t - \bar{s_T}}{\sqrt{T^{-1}\sum_{t=1}^{T}(s_t - \bar{s}_T^2)}}, \quad \text{where} \quad \bar{s_T} = T^{-1}\sum_{t=1}^{T} s_t \tag{2.1}$$

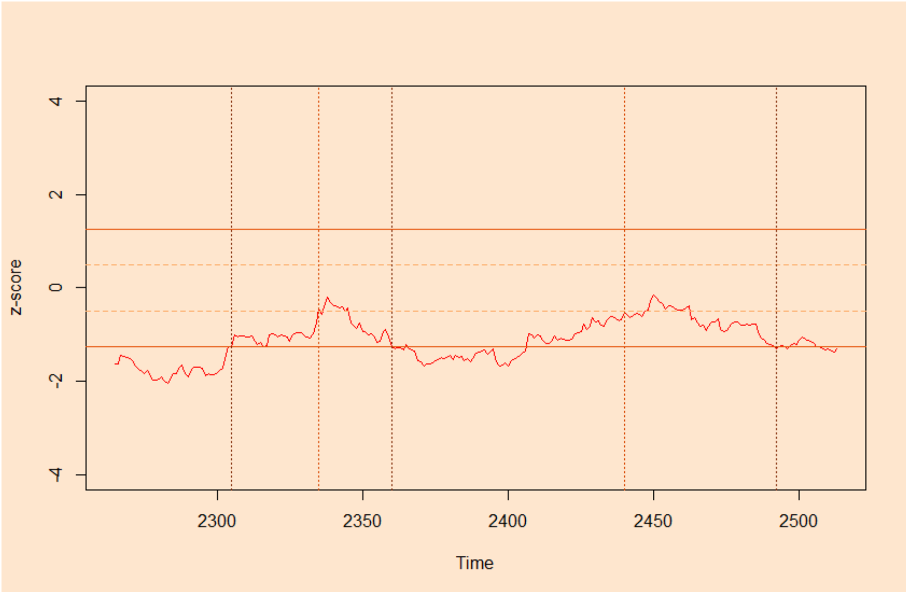
and $s_t = p_t + \hat{b}q_t$, where $\hat{\beta} = (1,\hat{b})'$ is the estimated co-integrated vector. Thus, the above equation, namely the z-score, measures the distance to the long-run equilibrium of the co-integrated residual in units of standard deviation, i.e., how far the pair is from the theoretical equilibrium (Avellaneda and Lee, 2010; Nielsen and Rahbek, 2020). From here on, if the deviation at time T is larger than our pre-determined threshold, $|z_t| > k$, this paper opens the position at prices p_T and q_T. If z_T is positive and above k, p_T is sold and q_T is bought. The opposite is true if z_T is negative. During the holding period, i.e., $T+1, T+2, \ldots, T+h$ the spread is monitored. If the deviation from the long-run equilibrium mean-reverts at some $T+h$, i.e., $z_{T+h} = 0$, the position is closed with a net return given by

$$(p_{t+h} - p_t) + b(q_{t+h} - q_t) = (p_{t+h} + bq_{t+h}) - (p_t + bq_t) = s_{t+h} - s_t \tag{2.2}$$

Equation (2.2) is the total (log) return of the pairs trading portfolio[2]. In addition to this, this paper implements a stop-loss preventing the deviation to increase excessively, as the co-integration property may have broken down. If z_{T+h} is larger than our threshold, $|z_T| > k + \delta$ [3] the stop-loss is instated (Caldeira and Moura, 2013; Nielsen and Rahbek, 2020). In Figure 1, a pair found in the data generating process (DGP) is presented, as an illustrative example of the trading strategy.

[2]The same market value of stocks are bought and sold.
[3]$\delta > 0$.

Figure 1: Schematic evolution of the z-score and the associated trading rules for Google Inc. and UDR Inc.



Vertical dark brown dotted lines illustrates a point of "Buy to open". Vertical orange dotted lines illustrates a point of closing the "Buy to open" position.

2.2 The Vector Autoregressive Model

The vector autoregressive (VAR) model is a generalisation of the univariate autoregressive model that models current observations of a variable based on prior observations. The general case, VAR(k), where k denotes the amount of lag augmentations and X_t is a p-dimensional vector, is given by

$$X_t = \mu + A_1 X_{t-1} +, \dots, A_k X_{t-k} + \epsilon_t, \quad t = 1, 2, \dots, T, \tag{2.3}$$

where μ is a constant term, $A_i \in \mathbf{R}^{p \times p}$, $X_0, \dots, X_{k+1}$ is fixed and e_t i.i.d. $N(0, \Omega)$ (Nielsen and Rahbek, 2020).

2.2.1 The Process of the Vector Autoregressive Model

In order to denote the process of the VAR model, we can look at the characteristic polynomial, i.e., if the process is stationary or non-stationary. For the VAR(1), we have that characteristic polynomial is given by

$$A(z) = I_p - A_z \tag{2.4}$$

and has a unit root, i.e., $z = 1$ if

$$|A(1)| = |I_p - A| = 0$$

In the academic literature, a stationary process is denoted as $I(0)$ and a non-stationary, i.e., unit root process, as $I(1)$ (Nielsen and Rahbek, 2020).

2.3 Model Specification

To ensure that a VAR(k) model is well-specified, an issue that needs to be addressed is the lag augmentation, k. This paper follows the process of choosing the Akaike information criterion (AIC), based on the sequential increasing of the lag order up to a VAR(k)-process proposed by Akaike (1969). Furthermore, to ensure that our model is well-specified, this paper examines residual autocorrelation using the Breusch-Godfrey Lagrange Multiplier (BG) test.

2.3.1 Akaike Information Criterion

To obtain an adequately parsimonious model, it is appropriate to balance model fit against the complexity of the model. To select the appropriate set of lags for an autoregressive model, this paper utilises the AIC, defined as

$$AIC = \log(\hat{\sigma}^2) + \frac{2 * k}{T} \tag{2.5}$$

Where k is the number of estimated parameters and T is the number observations, i.e., sample size. The objective is to select the autoregressive model, which has the smallest value of AIC, i.e., the smallest prediction error and best combination of fit and parsimony (Akaike (1969) and Nielsen (2020a)).

2.3.2 Breusch-Godfrey Lagrange Multiplier Test

To ensure that the residuals of the VAR(k) do not suffer from autocorrelation the BG test is instated. The BG test is based upon the following auxiliary regression

$$\hat{\mathbf{u}}_t = A_1 \mathbf{y}_{t-1} + \ldots + A_p \mathbf{y}_{t-p} + CD_t + B_1 \hat{\mathbf{u}}_{t-1} + \ldots + B_h \hat{\mathbf{u}}_{t-h} + \varepsilon_t \tag{2.6}$$

The null hypothesis is defined as

$$H_0 : B_1 = \ldots = B_h = 0$$

and correspondingly the alternative hypothesis is defined as

$$H_1 : \exists \; B_i \neq 0 \quad \text{for} \quad i = 1, 2, \ldots, h$$

Furthermore, the test statistic is defined as

$$LM_h = T * R^2 \approx \chi^2(j) \quad ,$$

where T is the number observations in the basic series and R^2 comes from the estimated regression in (2.9). From the above, it is seen that the LM test statistic follows a χ^2-distribution with j degrees of freedom (DF) (Breusch, 1978).

2.4 Co-integration

To examine the existence of a long-run equilibrium between a pair of stocks, this paper uses co-integration, which states that non-stationary processes can have linear combinations that are stationary (Engle and Granger, 1987). Furthermore, Engle and Granger (1987) showed that the error-correction formulation was equivalent to the phenomenon of co-integration and, thus, very useful in forecasting and probabilistic analysis.

In this section, this paper formulates the mathematical structure of a co-integrated vector autogressive model (CVAR), the assumption for the Granger Representation Theorem (GRT) and the role of the deterministic term, i.e., linear trend, in the CVAR.

2.4.1 The Mathematical Structure of Co-integration

We consider an autoregressive time series and recall that if the autoregressive time series has a unit root and is a random walk, i.e., $I(1)$, then the moving average solution can be written as a random walk component, a stationary process and a contribution from the initial values.

For example, we have a two dimensional VAR(1) process X_t given by

$$\Delta X_t = \alpha\beta' X_{t-1} + \epsilon_t \tag{2.7}$$

Where $\alpha = (-1{,}0)', \beta = (1, -1)'$ and $\epsilon_t = (\epsilon_{1_t}, \epsilon_{2t})'$ i.i.d. $N_2(0, \Omega)$ or in vector form

$$\begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}' \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

Or

$$\begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{pmatrix} = \begin{pmatrix} -(X_{1t-1} - X_{2t-1}) \\ 0 \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \tag{2.8}$$

We note that $\alpha_\perp = (0{,}1)'$, and thus, we have,

$$\alpha_\perp^{'} \Delta X_t = \Delta X_{2t} = \alpha_\perp^{'}(\alpha\beta' X_{t_1} + \epsilon_t) = \alpha_\perp^{'} = \epsilon_{2t} \tag{2.9}$$

From this, we have $X_{2t} = \sum_{i=1}^{t} \epsilon_{2i} + X_{2,0}$[4] is equivalent to the sum of a random walk and the initial value, and thus, an $I(1)$ process.

We can now isolate $\beta' X_t$ from (2.7)

$$\beta' \Delta X_t = \beta' \alpha \beta' X_{t-1} + \beta' \epsilon_t \leftrightarrow$$
$$\Delta X_{1t} - \Delta X_{2t} = -(X1t - 1 - X_{2t-1}) + \epsilon_{1t} - \epsilon_{2t},$$

such that

$$\beta' X_t = X_{1t} - X_{2t} \leftrightarrow$$

$$\beta' \epsilon_t = \epsilon_{1t} - \epsilon_{2t} \tag{2.10}$$

We have equation (2.10) is the spread between X_{1t} and X_{2t}, and is i.i.d. Gaussian, and thus, stationary and asymptotically stable. From here, we can collect the terms, i.e.,

$$\begin{pmatrix} X_{1t} - X_{2t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} X_t = \begin{pmatrix} \epsilon_{1t} - \epsilon_{2t} \\ \sum_{i=1}^{t} \epsilon_{2i} + X_{2,0} \end{pmatrix}$$

[4] $X_{2,0}$ is equal variable X_2 in period 0.

We can use that

$$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

With some algebraic manipulation ad modum Nielsen and Rahbek (2020), we find that

$$X_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \sum_{i=1}^{t} \epsilon_{2t} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} (\epsilon_{1t} - \epsilon_{2t}) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} X_{2,0}$$

Following the notation introduced, we can write

$$X_t = \beta_\perp \alpha'_\perp \sum_{i=1}^{t} \epsilon_t + \alpha\beta'\epsilon_t + \beta_\perp \alpha'_\perp X_0 \tag{2.11}$$

Thus, we have now shown that X_t has a representation as the sum of a random walk, $\alpha'_\perp \sum_{i=1}^{t} \epsilon_t$, a stationary process, $\beta'\epsilon_t$ and the initial value, and is a non-stationary $I(1)$ process. Moreover, we see that the linear combination, $\beta' X_t$ is asymptotically stable due to the fact that $\beta'\beta_\perp = 0$ (Nielsen and Rahbek, 2020).

In accordance with the subject of matter, namely, pairs trading, where we have that $X_t = (p_t, q_t)'$, X_t is error-correcting to $\beta' X_{t-1}$ with an adjustment vector, α, which can be written as

$$\Delta X_t = \alpha\beta' X_t t - 1 + \epsilon_t \tag{2.12}$$

It is important to note that this example holds for co-integrating relations that are i.i.d., however, this may not be the case and, thus, Assumption 1 is necessary when conducting analysis for co-integrating relations that are not i.i.d.

Assumption 1 *Let X_t be a vector autoregressive process with characteristic polynomial $A(z)$, $z \in \mathbf{C}$. Assume that $A(z)$ has exactly $(p - r)$ roots at $z = 1$ and the remaining roots are larger than one in absolute value, $|z| > 1$*

Apart from being an important assumption for the analysis of data that is not i.i.d., Assumption 1 is also a key point for GRT, which states exactly what we have just shown, i.e., that a co-integrated vector autoregressive process can be decomposed into four components, namely, a random walk, a stationary process, a deterministic part, and a term that depends on the initial values. For more on the GRT see Johansen (2004) (Nielsen and Rahbek, 2020).

2.4.2 Deterministic Terms

In this paper, we extend our equation (2.3) such that a linear trend is allowed in our VAR(k) model. Allowing for deterministic terms, e.g., a linear trend enables us to model variables that we believe are stationary around a deterministic linear trend. The interpretation is that the linear combination $\beta' X_t$ cancels our stochastic trends, but not deterministic trends (Johansen, 2004; Nielsen, 2020c).

For example, we often see that economic variables are trending, i.e., they have a tendency to systematically increase or decrease over time. Thus, this trending behaviour will change the unconditional expectation over time, which is not in accordance with our assumption of stationarity (Nielsen, 2020b).

We consider the initial VAR(1) with a linear trend $\tau \in \mathbf{R}^p$ where $\Pi = \alpha\beta'$ and $\alpha, \beta \in \mathbf{R}^{p\times r}$[5].

$$\Delta X_t = \Pi X_{t-1} + \tau t + \mu + \epsilon_t, \quad t = 1{,}2, \ldots, T \tag{2.13}$$

This leads to consider the model $H_{r,l}$, which is given by

$$\Delta X_t = \alpha\beta' X_{t-1} + \alpha\tau_l' t + \mu + \epsilon_t \tag{2.14}$$

Where $\tau_l' \in \mathbf{R}^r$. We denote $H^0_{r,l}$, $H^0_{r,l} \subseteq H_{r,l}$, and Assumption 3.1 must hold. We can then say that under $H^0_{r,l}$, X_t has the representation

$$X_t = C \sum_{i=1}^{t} \epsilon_t + C^* t + C_0 + C^* S_{t,l} \tag{2.15}$$

Where the co-integrating relations, i.e., $S_{t,l} = \beta' X_t$ are asymptotically stable around a linear trend, namely,

$$S_{t,l} = (I_r + \beta'\alpha) S_{t-1,l} + \beta'\tau t + \beta'\mu + \epsilon_t \tag{2.16}$$

Thus, we see that X_t is trending $I(1)$, and $\beta' X_t$ also has a linear trend (Nielsen and Rahbek, 2020).

2.5 Rank Determination

In this section, we introduce the term rank and reduced rank regression (RRR) model, which is a multivariate model with a coefficient matrix with reduced rank. Furthermore, rank determination by rank test and a bootstrapped version of the rank test is presented.

2.5.1 Reduced Rank Regression

The reduced rank regression is an explicit estimation method in multivariate regressions, which accounts for reduced rank restrictions of the coefficient matrix (Johansen, 2004). It is necessary to introduce RRR, as for asymptotic inference and hypothesis testing, we have that one of two hypotheses is whether the multivariate model, i.e., in this case the VAR, has reduced rank, r, and given that it has reduced rank, r, the linear restrictions on the co-integrating vectors β.

We consider the unrestricted VAR(1) model given by

$$\Delta X_t = \Pi X_{t-1} + \epsilon_t$$

With X_0 given and ϵ_t being i.i.d. $N_p(0, \Omega)$. Using that $Y_t = \Delta X_t$ and $Z_t = X_{t-1}$, we have can write the model as

$$Y_t = \Pi Z_t + \epsilon_t$$

[5] See Appendix A for a definition of Rank and Subspaces.

In order to describe our above model, we have to introduce the notation for product moments

$$S_{yy} = \frac{1}{T}\sum_{t=1}^{T} Y_t Y_t^{'} = \frac{1}{T}\sum_{t=1}^{T} \Delta X_t \Delta X^{'}{}_t$$
$$S_{yz} = \frac{1}{T}\sum_{t=1}^{T} Y_t Z_t^{'} = \frac{1}{T}\sum_{t=1}^{T} \Delta X_t X^{'}{}_{t-1}$$
$$S_{zz} = \frac{1}{T}\sum_{t=1}^{T} Z_t Z_t^{'} = \frac{1}{T}\sum_{t=1}^{T} X_t X^{'}{}_{t-1}$$

From this, it follows that the unrestricted maximum likelihood (ML) estimator of Π is given by the ordinary least squares (OLS) estimator.

$$\hat{\Pi} = S_{yz}S_{zz}^{-1} = \left(\frac{1}{T}\sum_{t=1}^{T}\Delta X_t X_{t-1}^{'}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1}X_{t-1}^{'}\right)^{-1}, \tag{2.17}$$

From this regression (2.17), we can form the residuals, $\hat{\epsilon}_t = \hat{\Pi} Z_t$, and we derive the variance of the residuals

$$\hat{\Omega} = S_{yy} - S_{yz}S_{zz}^{-1}S_{zy} \tag{2.18}$$

A more detailed derivation of the variance can be found in Appendix B.

Under the reduced rank restriction, H_r, we have the model is given by

$$H_r : \Delta X_t = \Pi X_{t-1} + \epsilon \quad \text{with} \quad \Pi = \alpha\beta'$$

We have the that H_r is non-linear, and thus, we can not use the usual linear regression results, and therefore, we have to deduce the ML estimator by solving the eigenvalue problem. The derivations for solving the eigenvalue problem and can be found in Appendix C. From Appendix C, i.e., the solved eigenvalue problem, we have a RRR, and we can now determine $\hat{\beta}$, and from $\hat{\beta}$, we can now estimate $\hat{\alpha}$ and $\hat{\Omega}$ (Johansen, 2004; Nielsen and Rahbek, 2020).

2.5.2 Rank Test

The first hypothesis that the coefficient matrix, Π, has reduced rank is by definition given by

$$LR_r := LR(H_r|H_p) = -T \cdot \log \sum_{i=r+1}^{p} (1 - \hat{\lambda}_i) \tag{2.19}$$

The LR test statistic for hypothesis H_r against the unrestricted model H_p is given by $LR_r = 2 - \log Q$, where Q denotes the ratio of the maximised likelihood functions. The computation of the test statistics LR_i for $i = 0,1,\ldots,p-1$ is based on solving the eigenvalue problem. The rank test[6] is performed in steps, which estimate the rank of Π. First, it is tested whether Π has $rank = 0$, i.e.,

[6]This test is also known as Trace Test.

no co-integration, against the alternative hypothesis Π has $rank \geq 2$. Provided that the null hypothesis cannot be rejected, the sequence stops, as Π is a matrix of zeros. However, the test will continue if the null hypothesis is rejected and it tests for $rank(\Pi) = 1$ against $1 < rank(\Pi) \geq 2$. If the null hypothesis cannot be rejected, it is concluded that a co-integrating relationship exists. However, if the null hypothesis is rejected, the sequence continues (Nielsen and Rahbek, 2020).

The distribution of test statistic, LR_r, is delimited by the process of our VAR(k) model. This is due to the fact that when evaluating the LR_r, all previous cases have been rejected, and hence the H_{r-1}, or rank Π less than or equal to $r-1$, does not hold. Thus, it is sufficient to only consider the LR_r under the H_r^0, where the limiting distribution is given by a dimensional standard Brownian motion, $DF_{p-r}(\mathbf{W})$[7].

An example for two stock prices, i.e., $X_t = (p_t, q_t)'$, the bivariate VAR model gives $LR_0 > c_0$, and hence rank $r = 0$ is rejected. Next, we have that $LR_1 < c_1$ and H_1 is not rejected, as H_0, i.e., rank equal to zero, and H_1, i.e., rank less than or equal to 1, was accepted, implies $\hat{r} = 1$.

2.5.3 A Bootstrap Version of the Rank Test

The asymptotic rank test reviewed above compares LR statistics, LR_r, with critical values from the asymptotic distribution, $DF_{p-r}(\mathbf{W})$. Not only is this distribution ineffective with small samples[8], but it can also deviate from the model formulation, e.g., conditional and unconditional hetereoskedasticity may shift the actual distribution (Nielsen and Rahbek, 2020). Thus, this paper introduce Cavaliere et al. (2010)'s suggestion to simulate the finite sample distribution under the null hypothesis H_r^0 using a bootstrap approach. The bootstrap algorithm is given by

1. For the observed data, $\{X_t\}_{t=0}^T$, estimate the model and calculate the rank test statistics for LR_r, for $r = 0, 1, \ldots, p-1$. For each of the hypotheses, H_r, $r = 0,1,\ldots,p-1$, simulate the distribution under H_r^0 following:

2. Impose the hypothesis $H_r^0 : rank(\Pi) = r$, and then estimate the restricted model

$$\Delta X_t = \hat{\alpha}^{(r)} \hat{\beta}^{(r)\prime} X_{t-1} + \hat{\epsilon}_t^{(r)}$$

   Recenter the residuals as

$$\hat{\epsilon}_t^{(r,c)} = \hat{\epsilon}_t^{(r)} - T^{-1} \sum_{t=1}^{T} \hat{\epsilon}_t^{(r)}$$

3. Use the estimates under the null hypothesis, $\hat{\alpha}^{(r)}$ and $\hat{\beta}^{(r)}$, as well as a sequence of bootstrap shocks, $\{\epsilon_t^*\}_{t=1}^T$, to recursively generate a bootstrap

[7]The mathematics for the distribution of the univariate autoregressive model is beyond the scope of this paper. See Johansen (2004) and Nielsen and Rahbek (2020) for a detailed analysis the limiting distribution.

[8]The actual finite-sample distribution may be different from the $DF_{p-r}(\mathbf{W})$, and thus ineffective in smaller samples.

sample

$$\Delta X_t^* = \hat{\alpha}^{(r)} \hat{\beta^{(r)}}' X_{t-1}^* + \epsilon_t^*, \tag{2.20}$$

with $X_0^* = X_0$. To define the bootstrap shock, ϵ_t^*, we use the so-called wild bootstrap, which will mimic any patterns of heteroskedasticity in the original data, and thus, this approach is robust towards autoregressive conditional heteroskedasticity (ARCH).

$$\epsilon_t^* = \hat{\epsilon}_t^{(r,c)} \eta_t \tag{2.21}$$

The choice for the distribution of the bootstrap errors, η_t, should be such that the errors have the same properties as the actual errors. For financial data, e.g., stock returns, it is well-known that these errors are heavy-tailed (Bradley and Taqqu, 2001). Generally, the wild bootstraps validity would require that the bootstrap shocks, ϵ_T^* have zero mean and unit variance. Thus, distributions such as the Gaussian or the Mammen's two point distribution would undermine the validity, as Lemma 1 from Cavaliere et al. (2020) states that the bootstrap must replicate the particular conditional distributions of the original statistics. Therefore, the two-point Rademacher distribution, with skewness zero and kurtosis one is chosen for η_t. The two-point Rademacher distribution outcome is

$$\eta_t = \begin{cases} 1 & \text{with probability} \quad \frac{1}{2} \\ -1 & \text{with probability} \quad \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

Furthermore, the distribution assumes symmetry and matches all even moments, and thus, ensuring that kurtosis is not exaggerated (Davidson and Flachaire, 2008).

For the bootstrap data, $\{X_t^*\}_{t=0}^T$, estimate the model to check that Assumption 1 is satisfied, and then calculate the bootstrap rank test statistic, i.e.,

$$LR_r^* = -T \log \sum_{i=r+1}^{p} (1 - \hat{\lambda}_i^*),$$

where $\hat{\lambda}_1^* > \hat{\lambda}_2^* > \cdots > \hat{\lambda}_p^*$ are the eigenvalues obtained for the bootstrap sample. We now have that the obtained LR_r^*, is a realisation from the relevant distribution, i.e., two-point Rademacher, under the H_r^0. However, if Assumption 1 is not satisfied and there is an explosive root, i.e., an $I(2)$ process[9], in the estimated model for the bootstrap sample, the sample is skipped and the process is initialised again from step three.

4. The above process is repeated a large number of times, B. Nielsen and Rahbek (2021) show that the wild bootstrap test is consistent and the

[9] An explosive root or an explosive process can be interpreted as the rate of change of the random walk. However, this is beyond the scope of this paper. For more on this topic see Johansen (2004).

bootstrap inference is robust towards heteroskedasticity at replications, $B > 250$, while the asymptotic test is not. Thus, we obtain

$$LR_r^{*(1)}, LR_r^{*(2)}, \dots, LR_r^{*(b)}. \tag{2.22}$$

The bootstrap p-value of the original statistic, LR_r is calculated as

$$p_r^* = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}(LR_r^{*(b)} > LR_r), \tag{2.23}$$

where $\mathbf{1}(\cdot)$ is the indicator function.

5. If the p-value is larger than a specified significance level, e.g., $p_r^* \geq 0.05$, we accept the hypothesis $H_r : rank(\Pi)$ and the bootstrap simulation sets $\hat{r} = r$ and stops.

   If the p-value is smaller than the specified significance level, H_r is rejected, and the rank is increased by one under the hypothesis, and steps two to four are repeated.

   If all reduced ranks, $r = 0, 1, \dots, p-1$, are rejected, we accept the unrestricted VAR, H_p.

3 Methodological Framework

In this section, this paper presents the applied methodology, which lays the foundation for our pairs trading algorithm. First, we present our data and its decomposition. Second, we present the statistical measures, i.e., the selection process for pairs. Last, we present the criteria for executing the trades.

3.1 Data

The empirical analysis for a pairs trading strategy should utilise a high frequency data set. Furthermore, an important characteristic for pairs trading is liquid assets, as illiquidity increases the slippage effect, i.e., the situation where a seller or buyer receives a different execution price than intended. In addition to this, illiquid assets involve higher operational costs and decreases the options to rent the assets, e.g., for short-selling (Caldeira and Moura, 2013)[10].

The data set consists of 10 years of historical daily adjusted closing prices for 287 firms on the S&P 500, from 1 January 2010 to 26 December 2019[11]. The S&P is structured such that a firms must adhere to certain criteria, e.g., a minimum monthly trading volume of 250,000 shares for six months up until an evaluation date. This and other criteria entails that some firms may enter and exit the index. Therefore, only firms that have been on the index the entire period are considered for the analysis (S&P Global, 2021). Furthermore, to ensure that our statistical and trading algorithm is consistent, firms with missing observations are discarded. Thus, this results in a total of 721,231 observations, 287 firms and 41,041 different combinations of pairs.

The period from 1 January 2010 to 31 December 2018, will serve as an in-sample, i.e., training period, whereas the period from 1 January 2019 to 26 December 2019 will serve as an out-of-sample, i.e., testing period. Regularly, in pairs trading, an intuitive relationship between the shares or the driving economic forces needs to exist, not only a co-integrating relationship, i.e., to combat unreliable (spurious) pairs. (Gatev et al., 2005). Thus, the motivation for choosing the S&P 500 is that the possible pairs that may arise from the co-integration tests also have the same fundamental basis, i.e., the same underlying economic forces that drive the growth or decline in line with Caldeira and Moura (2013). In Figure 2, the evolution of the S&P 500 in the data period is presented. It is clear that an upward trend exists, which also supports the choice of adding a linear deterministic trend to the co-integrated model.

[10] This paper do not account for operational costs when trading, however, transaction costs would occur should this strategy be executed genuinely.

[11] All data have been gathered from Bloomberg unless specifically stated.

Figure 2: Evolution of daily adjusted closing prices for the S&P 500

From 1 January 2010 to December 24 2019. Data from Yahoo Finance.

3.2 The Statistical Algorithm

In the following section, a thorough review of the selection criteria for the pairs is presented. First, we present how the statistical models are built and tested. Second, we test for the existence of a co-integrating relationship. Last, we make inference on the equilibrium relationship within the pairs, and thus, creating our portfolio of pairs to trade. All hypotheses are tested at a significance level of 5 per cent.

3.2.1 The Statistical Models

To initially model the variation in our data (training period), we construct 41,041 VAR(k) models with a linear deterministic trend, but no constant term, i.e.,

$$X_t = \tau t + A_1 X_{t-1} + \cdots + A_k X_{t-k} + \epsilon_t, \quad t = 1, \ldots, T \tag{3.1}$$

For each VAR model, the algorithm starts with $k = 20$ and for each VAR model chooses the minimum AIC value for k for the specific model. Next, the algorithm examines whether the respective VAR(k) model has autocorrelation. If the algorithm finds that a VAR(k) model suffers from autocorrelation at a significance level greater than our specified p-value threshold, the model, i.e., pair, is discarded.

3.2.2 The Co-integrating Relationship

To determine the co-integration rank, the algorithm runs a co-integration bootstrap test, i.e., the wild bootstrap. An example of a simulation the algorithm runs is shown in Table 1.

```
-------------------Test of cointegration rank-----------------

B = 399, N = 2265, k = 2, p = 2
model: restricted linear trend (3)

Testing rank r against k:
------------------------wild bootstrap-----------------------
 r      Q  wb.pv
 0 34.999 0.0175 *
 1 11.169 0.1125
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
-------------------------------------------------------------

Sequential test conclusion (signif = 0.05):
Wild bootstrap selected rank: 1

All of the simulations were successful
```

Table 1: Wild Bootstrap Test Results: Selected Rank = 1

From Table 1, we first notice that the bootstrap results are calculated with $B = 399$, which is clearly larger than $B > 250$, which according to Nielsen and Rahbek (2021) is enough to ensure a consistent and robust model. Furthermore, the algorithm draws its resampling based on the Rademacher distribution. We observe that with a sample size, $N = 2265$, that the bootstrap critical value, Q is 34.99 for LR_0 and 11.17 for LR_1. Thus, the hypothesis of no co-integration is rejected, and the wild bootstrap selected rank is 1.

The algorithm discards 36.800 models with autocorrelation or where no co-integrating relationship could be found. Thus, 4241 models are selected for the pairs trading portfolio.

3.2.3 Inference on the Equilibrium Relation

From the estimation results from our reduced rank models, which we found during our wild bootstrap test, we built an equilibrium relation of the co-integrated vector[12]. Even though a pair has a co-integrating relationship, our algorithm adds some restrictions on the parameters of the remaining 4241 models to determine whether they can enter our trading portfolio. These restrictions are based on the beta, β, values of the normalised co-integrating vector and the speed of adjustment of our co-integrated system, i.e., the half-life[13] of our pairs following Caldeira and Moura (2013). The beta restriction is set to

$$\beta = [-0.5, 2]$$

Thus, models with beta values smaller than -0.5 and larger than 2 are discarded.

The half-life restriction is determined by the second largest eigenvalue in our companion matrix (see Edelman and Murakami (1995) for a detailed analysis of the companion matrix). Usually, half-life is found through the differential equation known as the Ornstein-Uhlenbeck formula, however, we can exploit that we know the value of the second largest eigenvalue in the companion matrix for the respective VAR(k) model, and thus, we can approximate the half-life of mean reversion through

$$HL = \frac{\log(2)}{(1-\lambda)},$$

where λ denotes the second largest eigenvalue (Steffen et al., 2014). In Appendix D, a thorough mathematical example of the speed of adjustment in a co-integrated system and eigenvalues is presented. In this paper, the algorithm discards any pairs where the second largest eigenvalue, λ, is larger than 0.97. Thus, the maximum half-life for any given pair in the algorithms portfolio is

$$\begin{aligned} HL &= \frac{\log(2)}{(1-0.97)} \\ &= 23.10 \end{aligned}$$

This approximate measure shows that at least 50 per cent of the initial deviation should revert after 23 trading days. The average half-life in Caldeira and Moura (2013)'s paper is 10 trading days. However, they do not account for the error-correction, α, of both stocks, and thus, their half-life estimates need to be roughly doubled as stated by Ercolao and Skogman (2019). Therefore, our half-life estimates are roughly equivalent. From introduction of these restrictions, our algorithm discards 4230 VAR(k) models, and thus, leaves 11. The 11 pairs and their respective long-run equilibrium can be found in Table 4 in Appendix E.

[12]The co-integrating vector is normalised such that the left-hand side of the equation is equal to 1.

[13]Half-Life is the expected time to revert half of the pairs deviation from the mean measured in trading days.

3.3 The Trading Algorithm

In the following section, a review of the trading algorithm is presented. First, a brief introduction is given into the structuring of the trading algorithm. Second, the profitability computation methods are outlined.

3.3.1 Executing the Trading Strategy

The trading algorithm is structured as mentioned in the theoretical framework, i.e., the algorithm computes the deviation from the long-run equilibrium based on a z-score from the training period, and applies it on the test period. The specific threshold values can be found in Table 2 and figures of trades for each pair can be found in Appendix F.

Threshold Value	Trading Decision	Explanation
$z_t < -1.25$	Buy to open	Buy p_t and sell q_t
$z_t > 1.25$	Sell to open	Buy q_t and sell p_t
$z_t = -0.5$	Close buy to open	Sell p_t and buy q_t
$z_t = 0.5$	Close sell to open	Sell q_t and buy p_t
$\|z_t\| > 3.25$	Stop-loss	Close both positions

Table 2: Trading Signals

The ratio of shares bought to shares sold may vary depending on investor preference. However, this paper follows the approach of Gatev et al. (2005), Caldeira and Moura (2013), and Nielsen and Rahbek (2020), i.e., to construct capital neutral portfolio, by using the proceeds from the short sell to finance the long leg of the spread. However, the problem is how to construct such a capital neutral portfolio. Lin et al. (2006) suggest constructing the capital neutral portfolio using the co-integrating coefficient as the hedge ratio. Thus, the threshold values for entry and exit determines the profit per trade. The algorithm assumes the following relationship

$$(p_{t+h} - p_t) + b(q_{t+h} - q_t) = (p_{t+h} + bq_{t+h}) - (p_t + bq_t) = s_{t+h} - s_t \tag{3.2}$$

Thus, the total (log) return of our portfolio is determined by the spread at period t to $t + h$. The return on the overall portfolio is computed as a simple mean of the return to each individual pair, and thus, assuming that all pairs are equally weighted.

4 Empirical Results

In the following section, this paper presents the empirical results. Summary statistics for the trading algorithm and an overview of the stock sectors are also presented. This is compared to a long position in the S&P 500 in the same period.

In Table 3, summary statistics for the algorithm is presented.

No. of observations	721,131
No. of observations used for training	649,955
No. of possible trading days (test period)	248
No. of possible co-integrated pairs	41,041
No. of pairs in portfolio	11
Annualised return	1.17 %
Annualised volatility	1.06 %
Annualised Sharpe Ratio	1.10
Max. return	2.08 %
Min. return	-1.37 %

Table 3: Summary Statistics for Trading Algorithm

From Table 3, we see that from a possible 41,041 pairs, our algorithm builds a portfolio on the criteria mentioned in the theoretical and methodological framework, leading to 11 pairs. We see that the annualised return of the portfolio is 1.17 per cent and the annualised volatility (standard deviation) is 1.06 per cent, which leads to a Sharpe Ratio[14], i.e., the risk-adjusted return, of 1.10 per cent. Moreover, we see that maximum return for a pair is 2.08 per cent and the minimum return -1.37 per cent. Furthermore, in Figure 3, in Appendix G, the individual returns for the pairs are presented.

In Table 5, Appendix E, a summary of the 16 unique stocks in the trading portfolio is presented. From the table, we see that eight sectors and 13 industries are represented with a clear overweight of 37.5 per cent in the utilities sector. The utilities sector refers to a category of companies that provides basic amenities, e.g., electricity, gas, and sewage services. Moreover, the utilities sector often provides public service and is thus less likely to experience financial distress. The typical investor buys stocks in the utilities sector as a long-term investment strategy focusing on dividends and income stability. Furthermore, the sector tends to do well during macroeconomic downturns (Gatev et al., 2005).

Comparing a long position in the S&P 500 during the same period as the test period for our pairs trading algorithm, a long position would have generated 30.43 per cent in annualised return with a three-year Sharpe Ratio of 0.76[15].

[14] The Sharpe Ratio is computed as the annualised return divided by the annualised volatility.

[15] https://www.morningstar.com/indexes/spi/spx/risk

5 Discussion

In this section, this paper discuss issues regarding the theoretical framework, methodological framework and topics for further research. First, a discussion on the normality assumptions in the theoretical framework. Second, a discussion of the trading information and liquidity problems with the trading strategy. Last, a topic for further research is introduced, namely pairs trading dual-listed firms.

5.1 The Normality Assumption & Z-score

In traditional portfolio theory, e.g., Markowitz (1952), a key assumption is the normal distribution. Divergence from normality is usually tested in normality tests, e.g., Jarque-Bera. However, it is widely documented that financial variables do not follow a normal distribution (Costa and Iezzi, 2005). Thus, including a normality test in the misspecification tests for our VAR models could have the immediate consequence that they would be discarded. Furthermore, a caveat regarding the z-score is that the z-score is less informative with asymmetric distributions because in skewed distributions observations that lie on opposite side of the mean can have the same absolute z-score despite one being more/less probable than the other, and thus, a normality test could have increased the validity of the z-score.

5.2 The Trading Information & Liquidity Problems

In this paper, the executed trades are made on information about the previous trading day's adjusted close price. If the share price were to open at another level than the previous trading day, then we would have that the deviation from which the z-score is based upon is incorrect. This could mean that the adjustment to the equilibrium has taken place already. Consequently, it could be insightful to investigate a pairs trading strategy tick-by-tick, i.e., with data outlining the stock development after each trade. Furthermore, this paper does not account for transaction or operational costs for this strategy, which - ceteris paribus - would decrease the return. Moreover, this paper does not consider that the trading strategy requires an unlimited amount of capital, as every trade is executed. Even though a stop-loss is instated if the deviation increases or decreases excessively over time, one could argue that liquidity problems could occur much sooner.

5.3 Pairs Trading Dual-listed Firms

For further research, it could be interesting to test the notion of co-integrated pairs trading on dual-listed firms (DLF), e.g., Royal Dutch Shell. In the event of co-integration in DLFs, the intuitive economic association is present, which is prevalent in the academic literature, e.g., Gatev et al. (2005). De Jong et al. (2009) shows that a simple trading strategy, i.e., computing the theoretical equilibrium as the number of shares outstanding for both pairs of the DLF and fixing the current and future equity flow at a specified ratio, produce abnormal returns up to 10 per cent in annualised returns. As a result, it could be fascinating to investigate the risk-adjusted return for such co-integrated portfolios.

6 Conclusion

In this paper, we employ a statistical arbitrage trading algorithm in line with Gatev et al. (2005), Avellaneda and Lee (2010), Caldeira and Moura (2013), and Nielsen and Rahbek (2020), which trades pairs of stocks listed on the largest index in world, Standard and Poor's 500. Mean-reversion patterns are identified under the co-integration framework, which includes a wild bootstrap analysis of the co-integration rank. Thus, the co-integration relation for identified pairs are robust towards autoregressive conditional heteroskedasticity and the linear deterministic trend. From this framework, the algorithm builds a pairs trading portfolio around 11 pairs out of 41,041 possible.

This paper's findings corroborate Do and Faff (2012), who find that pairs trading in the U.S. equities market from 2009 and onward, is profitable, albeit at very modest levels. Our trading algorithm generates a modest return of 1.17 per cent in our test period, without accounting for transactions costs, which would significantly decrease the return. In addition to this, this paper's findings coincide well with Do and Faff (2012), who also find that a pairs trading strategy exhibits a lower risk, e.g., our portfolio's Sharpe Ratio is 1.10 compared to a three year Sharpe Ratio for the S&P 500 at 0.76. Furthermore, this paper's findings also corroborate Gatev et al. (2005), who find that there is a predominance of assets in the utilities sector among top pairs.

This paper acknowledges some of the drawbacks that are predominant in the trading strategy. For example, not accounting for transactions costs, as transaction costs for a pairs trading strategy can be considered an essential component, which the performance would heavily rely on. Moreover, on the theoretical part, accounting for structural shifts in the residual spreads of the pairs is a strenuous task, and thus, pairs that at some point could be co-integrated may break down. Consequently, it is suggested that further research is put into the residual behaviour such that trading signals are more accurate. However, the theoretical backing of this strategy is strong and the empirical results are quite promising.

7 References

Markowitz, H. (1952). Portfolio Selection. *American Finance Association*, *7*(1), 77–91. http://%20//links.jstor.org%20/sici?%20Sici%20=%200022-1082%25%2028195203%25297%253A1%25%203C77%253APS%25%203E2.0.CO%25%203B2-1

Akaike, H. (1969). Fitting Autoregressive Models for Prediction, 243–247.

Breusch, L. T. (1978). Testing for Autocorrelation in Dynamic Linear Models. *Australian Economic Papers*, *17*(31), 334–355. https://doi.org/10.1111/j.1467-8454.1978.tb00635.x

Engle, R. F., & Granger, C. W. J. (1987). Co-Integration and Error Correction : Representation , Estimation , and Testing. *Econometrica*, *55*(2), 251–276. https://www.jstor.org/stable/1913236?seq=1

Edelman, A., & Murakami, H. (1995). Polynomial Roots from Companion Matrix Eigenvalues. *Mathematics of Computation*, *64*(210), 763. https://doi.org/10.2307/2153450

Bradley, B. O., & Taqqu, M. S. (2001). *Financial Risk and Heavy Tails* (tech. rep.). https://doi.org/10.1016/b978-044450896-6.50004-2

Johansen, S. (2004). *Cointegration: an overview* (tech. rep. November).

Costa, M., & Iezzi, S. (2005). *Normality Test for Financial Variables* (tech. rep.).

Gatev, E. G., Goetzmann, W. N., & Rouwenhorst, K. G. (2005). Pairs Trading: Performance of a Relative Value Arbitrage Rule. *SSRN Electronic Journal*, (08). https://doi.org/10.2139/ssrn.141615

Lin, Y. X., McCrae, M., & Gulati, C. (2006). Loss protection in pairs trading through minimum profit bounds: A cointegration approach. *Journal of Applied Mathematics and Decision Sciences*, *2006*(May), 1–14. https://doi.org/10.1155/JAMDS/2006/73803

Davidson, R., & Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, *146*(1), 162–169. https://doi.org/10.1016/j.jeconom.2008.08.003

De Jong, A., Rosenthal, L., & Van Dijk, M. A. (2009). The risk and return of arbitrage in dual-listed companies. *Review of Finance*, *13*(3), 495–520. https://doi.org/10.1093/rof/rfn031

Avellaneda, M., & Lee, J. H. (2010). Statistical Arbitrage in the U.S. Equities Market. *Quantitative Finance*, *10*(7), 761–782. https://doi.org/10.1080/14697680903124632

Cavaliere, G., Rahbek, A., & Taylor, A. M. R. (2010). *Bootstrap Sequential Determination of the Co-integration Rank in VAR Models* (No. *10-07*).

Do, B., & Faff, R. (2012). Are pairs trading profits robust to trading costs? *Journal of Financial Research*, *35*(2), 261–287. https://doi.org/10.1111/j.1475-6803.2012.01317.x

Caldeira, J., & Moura, G. V. (2013). Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy. *SSRN Electronic Journal*, (January 2005), 1–28. https://doi.org/10.2139/ssrn.2196391

Steffen, J., Held, P., & Kruse, R. (2014). Cointegration Analysis of Financial Time Series Data.

Ercolao, M., & Skogman, L. (2019). *Trading Index Volatility by the Pair: A pairs trading strategy built on fractionally cointegrated realized volatilities* (Doctoral dissertation).

Cavaliere, G., Nielsen, H. B., & Rahbek, A. (2020). Bootstrapping Noncausal Autoregressions: With Applications to Explosive Bubble Modeling. *Journal of Business and Economic Statistics*, *38*(1), 55–67. https://doi.org/10.1080/07350015.2018.1448830

Nielsen, H. B. (2020a). Chapter 4: Univariate Models for Stationary Economic Time Series.

Nielsen, H. B. (2020b). Chapter 7: Non-Stationary Time Series and Unit Root Testing.

Nielsen, H. B. (2020c). Chapter 8: Analysis of Non-Stationary and Co-integrated Time Series.

Nielsen, H. B., & Rahbek, A. (2020). *Introduction to the Theory for Co-integration Analysis in Vector Autoregressions* (tech. rep.).

Nielsen, H. B. (2021). Eigenvalues and the Speed of Adjustment in a Co-integrated System.

Nielsen, H. B., & Rahbek, A. (2021). *Bootstrap Testing: Illustration and Implementation for the AR(1)* (tech. rep. No. 1).

S&P Global. (2021). S&P U.S. Indices Methodology.

8 Appendices

Appendix A - Rank and Subspaces

Let M be a $m \times n$ matrix of rank r. then with rank$(M) = r$, it holds that

1. $\text{rank}(M) = r \leq \min(m,n)$
2. $\text{rank}(M) = \text{rank}(MM') = \text{rank}(M'M) = \text{rank}(M') = r$
3. $\text{rank}(MB) \leq \min(\text{rank}(M), \text{rank}(B))$
4. $\text{rank}(MB) = \text{rank}(M) = r \quad \text{if} \quad B \quad \text{is a} \quad n \times n \quad \text{matrix of full rank}$

For the column space of M

$$span(M) = Mx|x \in \mathbf{R}^n$$

it holds that

$$\begin{aligned} dim(span(M)) &= rank(M) = r \\ span(M) &= span(MM') \end{aligned}$$

Furthermore, the direct sum applies, that is

$$\begin{aligned} m &= dim(span(MM)) + dim(span(M)_\perp) \\ &= r + (m - r) \end{aligned}$$

This is used explicitly in co-integration. Let α be a $m \times r$ matrix with rank so that $span(\alpha) = span(M)$. Then define $\alpha_\perp$ as a $m \times (m-r)$ matrix of full rank $(m - r)$, such that $span(\alpha_\perp = span(M)_\perp$, i.e.,

$$\alpha' \alpha_\perp = 0 \quad \text{and} \quad span(\alpha, \alpha_\perp) = \mathbf{R}^m$$

In particular, the orthogonal projection is often applied

$$I_m = \alpha(\alpha'\alpha)^{-1}\alpha' + \alpha_\perp(\alpha_\perp^{'}\alpha_\perp)^{-}1\alpha^{'}{}_\perp$$

as stated by Nielsen and Rahbek (2020).

Appendix B - Variance of the Residuals in the VAR(1)

$$\begin{aligned}
\hat{\Omega} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t^{'} \\
&= \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{\Pi}Z_t)(Y_t^{'} - Z_t^{'}\hat{\Pi}') \\
&= \frac{1}{T}\sum_{t=1}^{T}(Y_tY_t^{'} - Y_tZ_t^{'}\hat{\Pi}^{'} - \hat{\Pi}Z_tY_t^{'} + \hat{\Pi}Z_tZ_t^{'}\hat{\Pi}^{'}) \\
&= S_{yy} - S_{yz}\hat{\Pi}^{'} - \hat{\Pi}S_yz + \hat{\Pi}S_{zz}\hat{\Pi}^{'} \\
&= S_{yy} - S_{yz}S_{zz}^{-1}S_{zy} - S_{yz}S_{zz}^{-1}S_{zy} + S_{yz}S_{zz}^{-1}S_{zz}S_{zz}^{-1}S_{zy} \\
&= S_{yy} - S_{yz}S_{zz}^{-1}Szy
\end{aligned}$$

Appendix C - Solving the Eigenvalue Problem

Theorem 1 *Under H_r the ML estimator of β, $\hat{\beta}$, is found by solving the eigenvalue problem*

$$|\lambda S_{zz} - S_{yz}^{-1} S_{yz}| = 0 \tag{8.1}$$

with eigenvalues

$$1 > \hat{\lambda}_1 > \cdots > \hat{\lambda}_r > \cdots > \hat{\lambda}_p > 0$$

and with corresponding eigenvectors $\hat{V} = (\hat{v}_1, \ldots, \hat{v}_p$, for which

$$\hat{V}' S_{zz} \hat{V} = I_p \quad \text{and} \quad \hat{V}' S_{zy} S_{yy}^{-1} S_{yz} \hat{V} = \hat{A},$$

where $\hat{A} = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$. Then we have

$$\hat{\beta} = (\hat{v}_1, \ldots, \hat{v}_r). \tag{8.2}$$

Furthermore, we have the estimators for α and Ω are

$$\hat{\alpha} = S_{yz} \hat{\beta} (\hat{\beta}' S_{zz} \hat{\beta})^{-1}, \tag{8.3}$$

$$\hat{\Omega} = S_{yy} - S_{yz} \hat{\beta} (\hat{\beta}' S_{zz} \hat{\beta})^{-1} \hat{\beta}' S_{zy} = S_{yy \cdot \hat{\beta}} \tag{8.4}$$

Thus, we now have the maximised value of the log-likelihood function given by

$$\log L_{max}(\hat{\alpha}, \hat{\beta}, \hat{\Omega}) = -\frac{T}{2} \log(|S_{yy}|) - \frac{T}{2} \sum_{i=1}^{r} \log(1 - \hat{\lambda}_i) \tag{8.5}$$

We note that (8.3) and (8.4) are found by OLS regression with β known, and the non-linear restriction implies that in order to find $\hat{\beta}$ an eigenvalue problem has to be solved (Nielsen and Rahbek, 2020). See Johansen (2004) and Nielsen and Rahbek (2020) for proof of Theorem 1.

Appendix D - Eigenvalues and the Speed of Adjustment in a Co-integrated System

The model is given by

$$\begin{aligned}
\Delta X_t &= \alpha\beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \epsilon_t && (8.6)\\
X_t - X_{t-1} &= \alpha\beta' X_{t-1} + \Gamma_1 (X_{t-1} - X_{t-2})\epsilon_t && (8.7)\\
X_t &= (\alpha\beta' + I + \Gamma_1) X_{t-1} - \Gamma_1 X_{t-2} + \epsilon_t, && (8.8)
\end{aligned}$$

and thus, we have the companion form

$$A = \begin{pmatrix} \alpha\beta' + I + \Gamma_1 & -\Gamma_1 \\ I & 0 \end{pmatrix}. \tag{8.9}$$

An example

$$\alpha = \begin{pmatrix} -0.5 \\ 0.25 \end{pmatrix} \quad \beta' = \begin{pmatrix} 1 & -1 \end{pmatrix} \quad \Gamma_1 = 0 \tag{8.10}$$

Thus, we have that

$$\begin{aligned}
\alpha\beta' + I + \Gamma_1 &= \begin{pmatrix} -0.5 \\ 0.25 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} && (8.11)\\
&= \begin{pmatrix} 0.5 & 0.5 \\ 0.25 & 0.75 \end{pmatrix} && (8.12)
\end{aligned}$$

And the companion form

$$A = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{8.13}$$

Thus, we have the eigenvalues: $1, 0.25, 0, 0$, which is in line with the speed of adjustment of the two α-values, which are $-0.5, 0.25$, which is a combined adjustment of 0.75 per time period. Therefore, we can conclude that it is possible to use the second largest eigenvalue in the companion matrix as an indicator of the speed of adjustment for the co-integrated system. However, there is some notion about the amount of lags in the co-integrated system, which can increase the speed of adjustment. Nevertheless, this would only diminish the half-life of the co-integrated pair, and thus, the mean-reverting process is shortened[16] (Nielsen, 2021).

[16] The mathematics of this is beyond the scope of this paper, and would not serve much purpose for the understanding of the speed of adjustment for the co-integrated pair.

Appendix E - Pairs in the Trading Portfolio

```
AIZ.UN.Equity = 1.185 * FE.UN.Equity + 0
CMS.UN.Equity = 0.416 * MSFT.UW.Equity - 0.001
DTE.UN.Equity = 0.307 * EXPD.UW.Equity - 0.001
DTE.UN.Equity = 0.465 * ROL.UN.Equity - 0.001
ES.UN.Equity = 0.406 * MSFT.UW.Equity - 0.001
FE.UN.Equity = 1.309 * GOOGL.UW.Equity- 0.001
FE.UN.Equity = 1.961 * ICE.UN.Equity - 0.001
FRT.UN.Equity = 1.542 * ROL.UN.Equity - 0.001
NEE.UN.Equity = 0.508 * UDR.UN.Equity - 0.001
NI.UN.Equity = 1.035 * ROL.UN.Equity - 0.002
RSG.UN.Equity = 0.751 * TDG.UN.Equity- 0.001
```

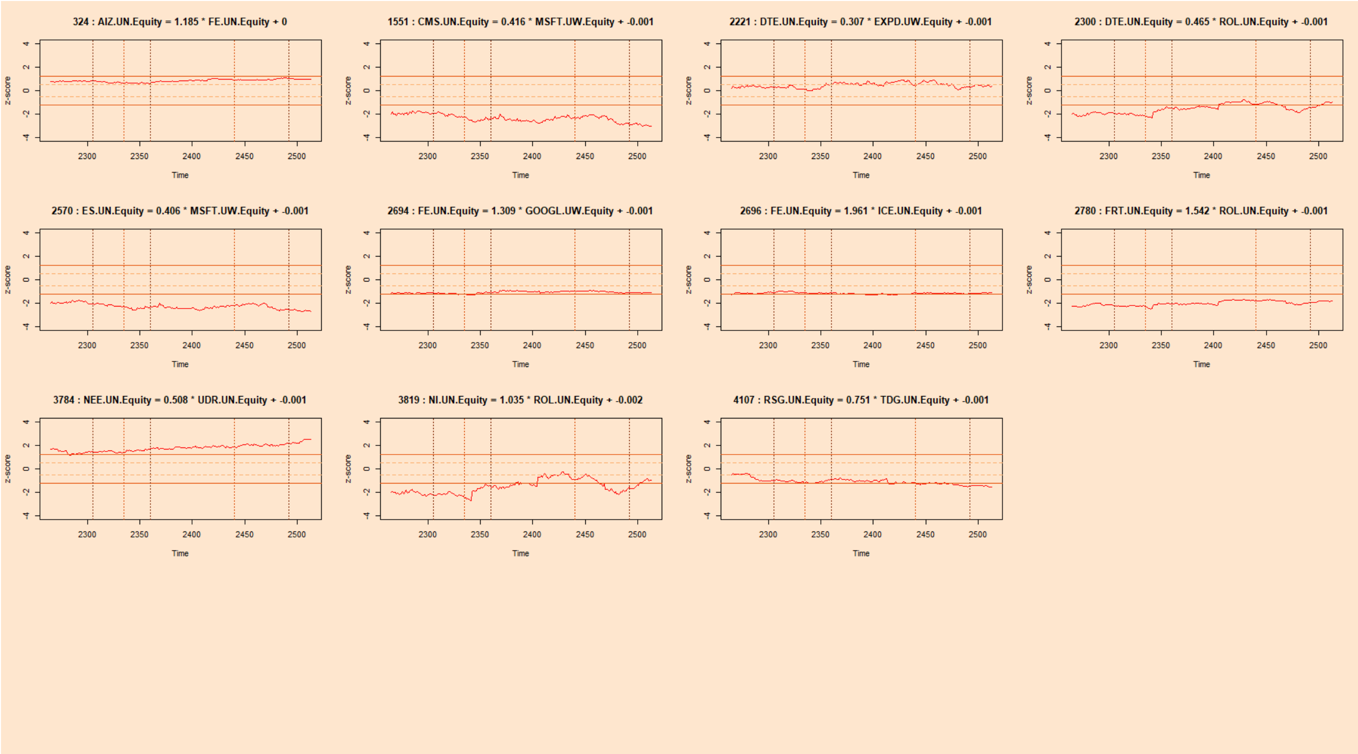
Table 4: Long-run Equilibrium of Pairs in the Trading Portfolio

Ticker	**Sector**	**Industry**
AIZ	Financial Services	Insurance—Specialty
CMS	Utilities	Utilities—Regulated Electric
DTE	Utilities	Utilities—Regulated Electric
ES	Utilities	Utilities—Regulated Electric
EXPD	Industrials	Integrated Freight & Logistics
FE	Utilities	Utilities—Diversified
GOOGL	Communication Services	Internet Content & Information
ICE	Financial Services	Financial Data & Stock Exchanges
MFST	Technology	Software—Infrastructure
NEE	Utilities	Utilities—Regulated Electric
NI	Utilities	Utilities—Regulated Gas
ROL	Consumer Cyclical	Personal Services
RSG	Industrials	Waste Management
TDG	Industrials	Aerospace & Defense
UDR	Real Estate	REIT—Residential
VTR	Real Estate	REIT—Healthcare Facilities

Table 5: Summary of Stock Sectors

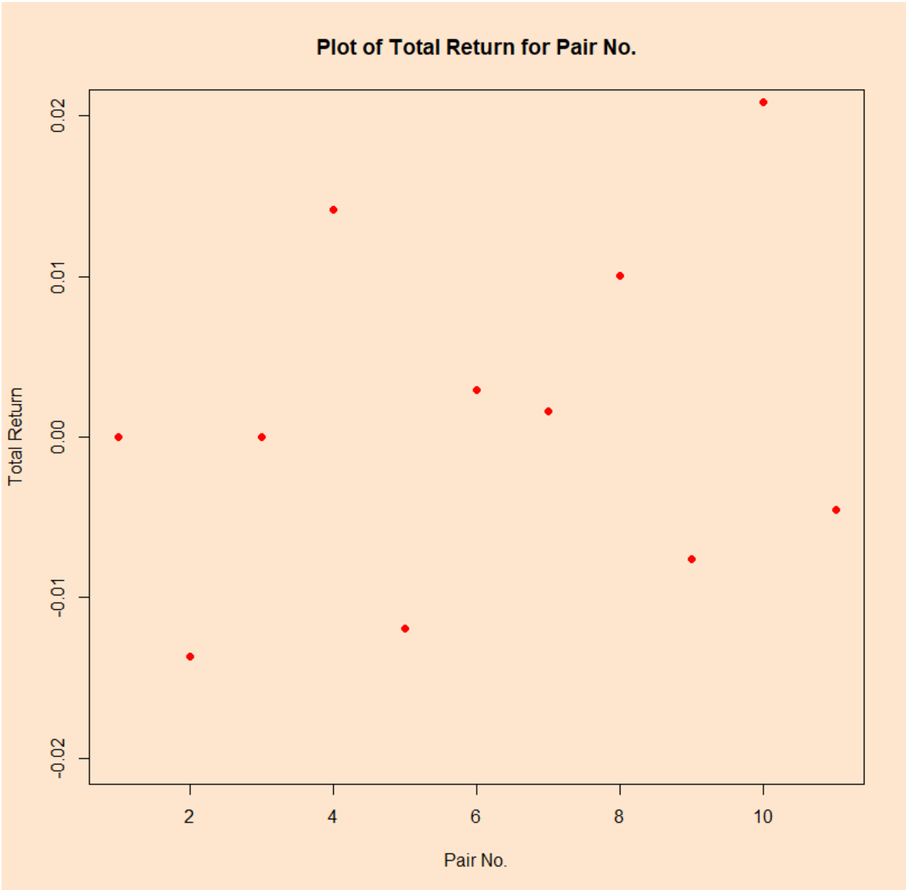
Appendix F - Executed Trades for the Pairs in the Portfolio

Figure 3: Executed Trades for the Pairs in the Portfolio



Appendix G - Plot of Total Return for Pairs

Figure 4: Plot of Total Return for Pairs



Number on x-axis indicates the number of pair.