UNIVERSITY OF THE AEGEAN

The Aspect of Pairs Trading for Statistical Arbitrage

Markaki Angeliki

Supervisor: Asc. Prof. Xanthopoylos Stylianos
Co-Supervisor: Asc. Prof. Zimeras Stylianos
Co-Supervisor: Asc. Prof. Xatzopoylos Petros

Department of Statistics and Actuarial - Financial Mathematics
School of Sciences
University of the Aegean

2022

Contents

Abstract

The main thing that distinguishes financial time series analysis from other time series analysis, is that the former behave in a very unique way. This is due to the existence of unit roots. In this thesis we will present the fundamental theory behind unit root tests, spurious regression and cointegration. In the application we implement different strategies and measure their performance. This is in order to showcase the concept of statistical arbitrage and compare different strategies, such as the long only strategy, the ADL strategy and spread strategy (pairs trading).

Introduction

Financial time series analysis is concerned with theory and practice of asset valuation over time. A key feature that distinguishes financial time series analysis from other time series analysis, is that the former behave in a very unique way. This, intrigued a majority of researchers to focus their study on financial time series. This behaviour is a result of the stochastic trend that emerges from financial time series, along with the returns' stylized facts. That is why financial time series demand to be treated with special care in order to achieve proper estimations. The existence of unit roots is a thorn in the flesh for the majority of statistical approaches. This creates the need for specific techniques and approaches in order to address these issues.

In financial time series, pairs trading is a strategy that involves matching a long position with a short position in two assets with high correlation. The advantage of pairs trading, when it performs as expected, apart from the profit investors make, is the reduction of potential losses which result in an improved risk profile. In addition, in order for the pairs trading outcome to be valid, it requires a high statistical correlation between the assets. The pair can also involve assets from different industries as long as they are correlated.

When a strategy involves opening both a long and a short position simultaneously, one can take advantage of the, potential, price difference between these correlated assets and achieve statistical arbitrage. Statistical arbitrage is not strictly limited to two assets, since investors can perform it to a group of correlated assets. The risk behind statistical arbitrage strategy is that no one can know in advance how long the mispricing will last and how wide the spread will be.

Both concepts have been studied and analysed extensively in literature, both in financial and in statistical perspective. The most significant ones are the works by Bent E. Sorensen, Ruey S. Tsay, Soren Johansen, Peter J. Brockwell, Richard A. Davis.

This work consists of the theoretical presentation of the above terms as well as their practical application. In the theoretic part we present the fundamental definitions and properties one should know before the application. In the practical application part we used the statistical package R and performed different strategies to achieve statistical arbitrage.

The outline of this thesis is as follows. Chapter 1 begins with an introduction to time series. We present the definitions of the statistical measures of time series. The chapter continues with the definition of stationarity and when a time series is called integrated. At the end we present the four types of convergence, which are the convergence surely, convergence in probability, convergence in distribution and convergence in mean and also the term of ergodicity.

In chapter 2 we present fundamental measures of financial time series. The chapter begins with the definition of returns and the different types of them. The main focus of this chapter is the presentation of the stylized facts of financial returns. These stylized facts are volatility clusters, fat tails and nonlinear dependence.

Chapter 3 focuses on the violations of assumptions that lead to the phenomenon of spurious regression. At the first half of the chapter, we present the behaviour that one should expect from regression models applied on time series, from autoregressive models and the Autoregressive Distributed Lag (ADL) model. In addition, we discuss about how one can pick the right lag length and the Durbin- Watson test used for checking for correlation. At the second half of the chapter, we present the two cases of the very important spurious regression giving an example for each of them. The chapter ends with a reference of the spurious regression from the multivariate perspective and some ideas on how spurious regression can be cured.

Chapter 4 contains the concept of cointegration, and it mostly focuses on the unit root tests. Unit root test are very important for financial time series and here we present the Dickey- Fuller test, the Phillips- Perron test, the Zivot- Andrews test and the KPSS test.

Finally in chapter 5 we have a pair trading application. We used the statistical package R to demonstrate it.

Chapter 1

Time Series

Definition 1. *Time Series [7]*
A time series is a set of observations x_t, each one being recorded at a specified time t.

A discrete time series is one in which the set T_0 of times at which observations are made is a discrete set, as is the case for example when observations are made at fixed time intervals.

Continuous time series are obtained when observations are recorded continuously over some time interval, e.g. when $T_0 = [0, 1]$. We shall use the notation $x(t)$ rather than x_t if we wish to indicate specifically that observations are recorded continuously.

1.0.1 Differences

Definition 2. *Differences [7]*
We define the first difference operator ∇ by

$$\nabla X_t = X_t - X_{t-1} = (1 - \mathbf{B})X_t$$

where $\mathbf{B}$ is the backward shift operator,

$$\mathbf{B}X_t = X_{t-1}, \qquad t \in \mathbb{N} - \{1\}$$

Powers of the operators $\mathbf{B}$ and ∇ are defined in the obvious way, i.e. $\mathbf{B}^j(X_t) = X_{t-j}$ and $\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t))$, $\;j \in \mathbb{N}$ with $\nabla^0(X_t) = X_t$. Polynomials in $\mathbf{B}$ and ∇ are manipulated in precisely the same way as polynomial functions of real variables. For example,

$$\nabla^2 X_t = \nabla(\nabla X_t) = (1 - \mathbf{B})(1 - \mathbf{B})X_t = (1 - 2\mathbf{B} + \mathbf{B}^2)X_t = X_t - 2X_{t-1} + X_{t-2}$$

Equivalent to the backward shift operator is the lag operator

$$\mathbf{L}(X_t) = X_{t-1}, \qquad t \in \mathbb{N} - \{1\}$$

Remark 1. *It may be convenient to work with the first difference in logarithms of a series. We denote this by $\nabla log(X_t) = log(X_t) - log(X_{t-1})$.*
There are some reasons why we prefer to use the natural logarithm over the actual value.

It is common to use log differences in finance and one of the main reasons is the independence of the direction of change, an example of which will be given in log returns section. There are some economic measures, like GDP and Industrial production, that appear to have an exponential growth over time and log transformation takes that under consideration.

1.0.2 Statistical Measures

Definition 3. *Mean [84]*
The ℓth moment of a continuous random variable X is defined as

$$E(X^\ell) = \int_{-\infty}^{+\infty} x^\ell f(x)dx$$

where E stands for expectation and $f(x)$ is the probability density function of X. The first moment is called the mean or expectation of X. It measures the central location of the distribution. For time series we denote the mean function $\mu_t = E(X_t)$.

Definition 4. *Central Moment [84]*
The ℓth central moment of X is defined as

$$E((X-\mu_X)^\ell) = \int_{-\infty}^{+\infty} (x-\mu_X)^\ell f(x)dx$$

where $\mu_X = E(X)$

Definition 5. *Second Central Moment [84]*
The second central moment, denoted by σ_x^2 , measures the variability of X and is called the variance of X. The positive square root, σ_x, of variance is the standard deviation of X. For asset returns, variance (or standard deviation) is a measure of uncertainty and, hence, is often used as a risk measure.

$$Var_(X) = \sigma_X^2 = E((X-\mu_X)^2) = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f(x)dx$$

Definition 6. *Third Moment [84]*
The third moment measures the symmetry of X with respect to its mean. The skewness is defined as

$$S(X) = E\left(\frac{(X-\mu_X)^3}{\sigma_X^3}\right)$$

Definition 7. *Forth Moment [84]*
The fourth moment measures the degree of peakedness of the distribution of X and it is called the kurtosis. The kurtosis of X is defined as

$$K(X) = E\left(\frac{(X-\mu_X)^4}{\sigma_X^4}\right)$$

Remark 2. *The quantity $K(X) - 3$ is called the excess kurtosis because $K(X) = 3$ for a normal distribution. Thus, the excess kurtosis of a normal random variable is zero.*

Remark 3. *In statistics, skewness and kurtosis, are often used to summarize the extent of asymmetry and tail thickness. Usually in finance, the first fourth central moments of a random variable are used to describe the behaviour of asset returns, not that higher order moments are not important but they are much harder to study.*

Definition 8. *Covariance function [84]*

The covariance function of X and Y is

$$\gamma(X, Y) = Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Definition 9. *Correlation Function [7]*

The correlation function of X, Y is

$$\rho \equiv \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = Cor(X, Y)$$

1.0.3 Time Series Measures

Definition 10. *Mean [15]*

One way of describing a time series is to specify the joint probability distribution of $X_{t_1}, \ldots, X_{t_n}$ for any set of times $t_1, \ldots, t_n$ and any value of n. We will denote $\mu_t = \mu(t) = E(X(t))$ as the mean function of the X_t time series at a given time t.

We denote $\sigma_X^2(t) = Var(X_t)$ as the variance of the time series X_t at a given time t.

Definition 11. *AutoCovariance Function [7]*

One can assume that autocovariance depends on the distance between the two time instances and not their position. This means that we are interested in the lag h between t and $t + h$

$$\begin{aligned}\gamma_X(h) = \gamma_X(t, t+h) &= Cov(X_t, X_{t+h}) \\ &= E\left((X_t - \mu_X(t))(X_{t+h} - \mu_X(t+h))\right)\end{aligned}$$

For $h = 0$ we have

$$\gamma_X(0) = E\left((X_t - \mu_X(t))(X_{t+0} - \mu_X(t+0))\right) = Cov(X_t, X_{t+0}) = Var(X_t)$$

Definition 12. *Autocorrelation Function [7]*

The autocorrelation function (ACF) of X_t at lag h is

$$\rho_X(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)} = \frac{\gamma_X(t, t+h)}{\gamma_X(0)} = \frac{Cov(X_t, X_{t+h})}{Var(X_t)} = Cor(X_{t+h}, X_t)$$

1.0.4 Strictly stationarity

Definition 13. *Strictly Stationarity [8]*

The time series $\{X_t\}_{t\in\mathbb{Z}}$ is said to be strictly stationary if the joint distributions of $(X_{t_1}, \ldots, X_{t_k})'$ and $(X_{t_1+h}, \ldots, X'_{t_k+h})$ are the same for all positive integers k and for all $t_1, \ldots, t_k, \quad h \in \mathbb{Z}$.

Definition 14. *Weakly Stationarity [8]*

The time series $\{X_t\}_{t\in\mathbb{Z}}$, is said to be stationary if

1. $E(X_t)^2 < \infty \quad$ *for all $t \in \mathbb{Z}$*
2. $E(X_t) = m \quad$ *independent of t for all $t \in \mathbb{Z}$*
3. $\gamma_X(h) = \gamma_X(t, t+h) \quad$ *independent of t for all $h, t \in \mathbb{Z}$*

1.0.5 Integrated

As we have already discussed, the time series can be non- stationary in terms of mean and variance. Crucial are the non- stationary processes that are integrated, which have the basic property that by differentiating them we obtain stationary processes.

Economic time series are not usually stationary but their relative differences, or the differences when we measure the variable in logarithms, are stationary.

The stationary processes are differentiated from the integrated, because the integrated processes have the form in which dependency disappears over time.

When a time series Y_t has a unit autoregressive root, Y_t is integrated of order one. The integrated order determines the number of differences needed to obtain a stationary process. This is often denoted by $Y_t \sim I(1)$. We simply say that Y_t is $I(1)$. If Y_t is $I(1)$, its first difference Y_t is stationary.

It is sometimes necessary to differentiate more than once to obtain a stationary process. Y_t is $I(2)$ when Y_t needs to be differentiated twice in order to obtain a stationary series. Using the notation introduced here, Y_t is $I(2)$, its first difference ∇Y_t is $I(1)$ and its second difference $\nabla^2 Y_t$ is stationary. Integrated processes of order two can be seen as a generalization of integrated processes of order one but where the slope of the growth line, instead of being fixed, varies over time. Generalizing, we say that a process is integrated of order $d \geq 0$, and we denote it by $I(d)$, when upon differentiating it d times a stationary process is obtained. When Y_t is stationary, it is integrated of order 0 so Y_t is $I(0)$.

The long-memory processes are stationary processes where the autocorrelations decay much more slowly over time than in the case of the ARMA processes or in the integrated processes.

1.0.6 Convergences

Theorem 1. *Weak Law of Large Numbers [36]*
If $X_t \in \mathbb{R}$ are i.i.d and $E(X) < \infty$, then as $n \to \infty$,

$$\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t \underset{p}{\to} E[Y]$$

Theorem 2. *Central Limit Theory [36]*
If $X_t \in \mathbb{R}$ are i.i.d and $E(X)^2 < \infty$, then as $n \to \infty$,

$$\sqrt{n}(\bar{X} - \mu) \underset{d}{\to} N(0, \sigma^2)$$

where $\mu = E(X)$ and $\sigma^2 = E((X - \mu)^2)$.

Remark 4. *A rate of convergence in $\frac{1}{\sqrt{n}}$ in the sense of the Distribution Function (DF) metric is established for sequences $\{X_t\}$ of independent identically distributed random variables such that $E(X_1^3) < \infty$.*

Definition 15. *Convergence surely [36]*
Let (Ω, F, P) be a probability space. Let $X_1, X_2, \ldots$ be a sequence of random variables on (Ω, F, P). Let X be another random variable on (Ω, F, P). We say that X_n converges almost surely (or, with probability 1) to X if

$$\lim_{n\to\infty} P(\{\omega : X_n(\omega) = X(\omega)\}) = 1$$

In this case, we denoted $X_n \overset{a.s.}{\to} X$.

Definition 16. *Convergence in probability [36]*

A sequence of random variables $X_n \in \mathbb{R}$ converges in probability to X as $n \to \infty$, denoted $X_n \underset{p}{\to} X$ or alternatively $p\lim_{n\to\infty} X_n = X$, if for all $\delta > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \leq \delta) = 1$$

Equivalent for all $\epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$$

We call X the probability limit of X_n.

Definition 17. *Convergence in distribution [36]*

Let X_n be a sequence of random variables with distributions $F_n(u) = P(X_n \leq u)$. We say that X_n converges in distribution to X as $n \to \infty$, denoted $X_n \underset{d}{\to} X$, if for all u at which $F(u) = P(X \leq u)$ is continuous, $F_n(u) \to F(u)$ as $n \to \infty$. We refer to X and its distribution $F(u)$ as the asymptotic distribution, large sample distribution, or limit distribution of X_n.

Definition 18. *Convergence in Mean [36]*

Let $p \geq 1$ be a fixed number. A sequence of random variables $X_1, X_2, X_3, \ldots$ converges in the p^{th} mean or in the L^p norm to a random variable X, shown by $X_n \underset{L^p}{\to} X$, if

$$\lim_{n\to\infty} E(|X_n - X|^p) = 0$$

If $p = 2$, it is called the mean-square convergence, and it is shown by $X_n \underset{m.s.}{\to} X$.

This is a really strong type of convergence for random variables in the sense that

$$\begin{array}{c} X_n \overset{a.s.}{\to} X \Rightarrow X_n \underset{p}{\to} X \Rightarrow X_n \underset{d}{\to} X \\ \Uparrow \\ X_n \underset{L^p}{\to} X \end{array}$$

$X_n \underset{p}{\to} X \Rightarrow X_n \overset{a.s.}{\to} X$ if exist a subsequence $(X_n)_n$ that converge.

$X_n \underset{d}{\to} X \Rightarrow X_n \underset{p}{\to} X$ only if $x = c$ where c is constant.

$X_n \underset{p}{\to} X \Rightarrow X_n \underset{L^p}{\to} X$ only if X is bounded, $X \leq Y$ and $E(Y) < \infty$.

1.0.7 Ergodicity

Theorem 3. *Ergodic [36]*

A stationary series $X_t \in \mathbb{R}$ is ergodic if and only if for all events A and B

$$\lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} P(A_l \cap B_l) = P(A)P(B)$$

Theorem 4. *Ergodic Theorem [36]*
If $X_t \in \mathbb{R}$ is strictly stationary, ergodic, and $E(X) < \infty$, then as $n \to \infty$,

$$E(\bar{X} - \mu) \to 0$$

and

$$\bar{X} \underset{p}{\to} \mu$$

where $\mu = E(X)$.

Chapter 2

Time Series in Financial Theory

2.0.1 Returns

In financial time series we study the returns of an asset instead of the prices. This happens because returns are complete scale-free summary of opportunity. In addition, the returns of an asset have more attractive statistical properties than prices. These reasons were firstly presented by Campbell et al. (1997) [10].

We denote P_t as the asset price at time t. It is common for t to be referred to a day, but that is not necessary because it can be referred to any frequency (week, month etc.)

Definition 19. *One-Period Simple Return [84]*

Holding the asset for one period from date $t-1$ to date t would result in a simple gross return

$$1 + R_t = \frac{P_t}{P_{t-1}}$$

or

$$P_t = P_{t-1}(1 + R_t)$$

The corresponding one- period simple net return or simple return is

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Definition 20. *Multiperiod Simple Return [84]*

Holding the asset for k periods between dates $t-k$ and t gives a k-period simple gross return

$$\begin{aligned}
1 + R_t[k] = \frac{P_t}{P_{t-k}} &= \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \cdots \times \frac{P_{t-k+1}}{P_{t-k}} \\
&= (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}) \\
&= \prod_{j=0}^{k-1} (1 + R_{t-j})
\end{aligned}$$

The corresponding k-period simple return is $R_t[k] = \frac{P_t - P_{t-k}}{P_{t-k}}$.

Definition 21. *Natural logarithm [84]*

The natural logarithm of the simple gross return of an asset is called the continuously compounded return or log return:

$$r_t = \ln(1 + R_t) = \ln \frac{P_t}{P_{t-1}} = \ln P_t - \ln P_{t-1}$$

Next we will introduce the term of portfolio return.

Definition 22. *Portfolio [84]*

The simple net return of a portfolio consisting of N assets is a weighted average of the simple net returns of the assets involved, where the weight on each asset is the percentage of the portfolio's value invested in that asset. Let p be a portfolio that places weight w_i on asset i. Then, the simple return of p at time t is

$$R_{p,t} = \sum_{i=1}^{N} w_i R_{it},$$

where R_{it} is the simple return of asset i.

The relationships between simple return R_t and log return r_t are

$$r_t = \ln(1 + R_t), \quad R_t = e^{r_t} - 1$$

If the returns R_t and r_t are in percentages, then

$$r_t = 100 \ln\left(1 + \frac{R_t}{100}\right), \quad R_t = 100\left(e^{\frac{r_t}{100}} - 1\right)$$

Why we prefer log returns instead of simple ones

The continuously compounded multi-period return is simply the sum of continuously compounded one-period returns involved. Second, statistical properties of log returns are more tractable. Third, they are symmetric, while simple returns are not. For example, an investment of 100$ that yields a simple return of 50% followed by a simple return of -50% will result in 75$

$$\begin{aligned}
R_t &= \frac{P_t - P_{t-1}}{P_{t-1}} \\
&\Rightarrow 0.5 = \frac{P_t - 100}{100} \\
&\Rightarrow 50 = P_t - 100 \\
&\Rightarrow P_t = 150
\end{aligned}$$

$$\begin{aligned}
R_{t+1} &= \frac{P_{t+1} - P_t}{P_t} \\
&\Rightarrow -0.5 = \frac{P_{t+1} - 150}{150} \\
&\Rightarrow -75 = P_{t+1} - 150 \\
&\Rightarrow P_{t+1} = 75
\end{aligned}$$

An investment of 100$ that yields a continuously compounded return of 50% followed by a continuously compounded return of -50% will remain at 100$

$$
\begin{aligned}
r_t &= \ln \frac{P_t}{P_{t-1}} \\
&\Rightarrow 0.5 = \ln \frac{P_t}{100} \\
&\Rightarrow 0.5 = \ln P_t - \ln 100 \\
&\Rightarrow \ln P_t = 0.5 + \ln 100 \\
&\Rightarrow P_t = e^{0.5+\ln 100}
\end{aligned}
$$

$$
\begin{aligned}
r_{t+1} &= \ln \frac{P_{t+1}}{P_t} \\
&\Rightarrow -0.5 = \ln P_{t+1} - \ln(e^{0.5+\ln 100}) \\
&\Rightarrow -0.5 + 0.5 + \ln 100 = \ln P_{t+1} \\
&\Rightarrow \ln 100 = \ln P_{t+1} \\
&\Rightarrow P_{t+1} = 100
\end{aligned}
$$

2.0.2 Stylized Facts

Stylized facts are, generally speaking, statistical properties that appear to be present in many empirical asset returns. It is important to be aware of them because when building models that are supposed to represent asset price dynamics, the models must be able to capture/replicate these properties.

These stylized facts are

- Volatility clusters
- Fat tails
- Nonlinear dependence

and they will be further analysed.

Volatility Clusters

A way to measure the uncertainty of the market is volatility, which is defined as the standard deviation of returns.

Volatility can be divided into two cases, the unconditional and the conditional volatility. In general unconditional volatility is defined as volatility over an entire time period and denoted by σ. Conditional volatility is defined as volatility in a given time period, based on the past and defined by σ_t.

The phenomenon where data appear to have different variance in different time periods is called volatility clusters.

One can consider the squared logarithmic returns as good proxies for volatility.

Fat tails

In finance it is very common to encounter with time series that exhibit fat tails.

Definition 23. *Fat tails A random variable is said to have fat tails if it exhibits more extreme outcomes than a normally distributed random variable with the same mean and variance.*

The concept of fat tails was first presented by Mandelbrot (1963) [52] and Fama (1963, 1965) [26] [27]. Fat tails indicate that there is larger probability for extreme values to appear than the normal distribution. In other words, an asset is more likely to have very high and very low returns. This event would have a huge impact in the predictive capability of the model.

In the field of risk management the assumption of normal distribution may end up in a catastrophic underestimation of the risk. On the other hand, techniques that do not assume normal distribution, are complicated and one should be very cautious because if they are used wrong may, we may end up to incorrect outcomes.

Nonlinear dependence

Most statistical models assume that the relationship between different returns is linear. We say that the returns on two assets X and Y are linearly dependent if the conditional expectation $E(Y|X)$ is a linear function of X. This linear dependency can be easily measured and described by using Pearson's correlation coefficient ρ.

Note that if $E(Y|X)$ is not an expression of the linear function of X then ρ does not completely capture the dependence between the two variables.

2.0.3 Example

For better understanding the above theoretical approach we will use the Nasdaq 100 stock market index to analyse. The Nasdaq 100 is an index consisting of the 100 largest non-financial companies, consisting with industrial, technology, telecommunication etc. The data are from 2007-04-26 till 2021-04-09.

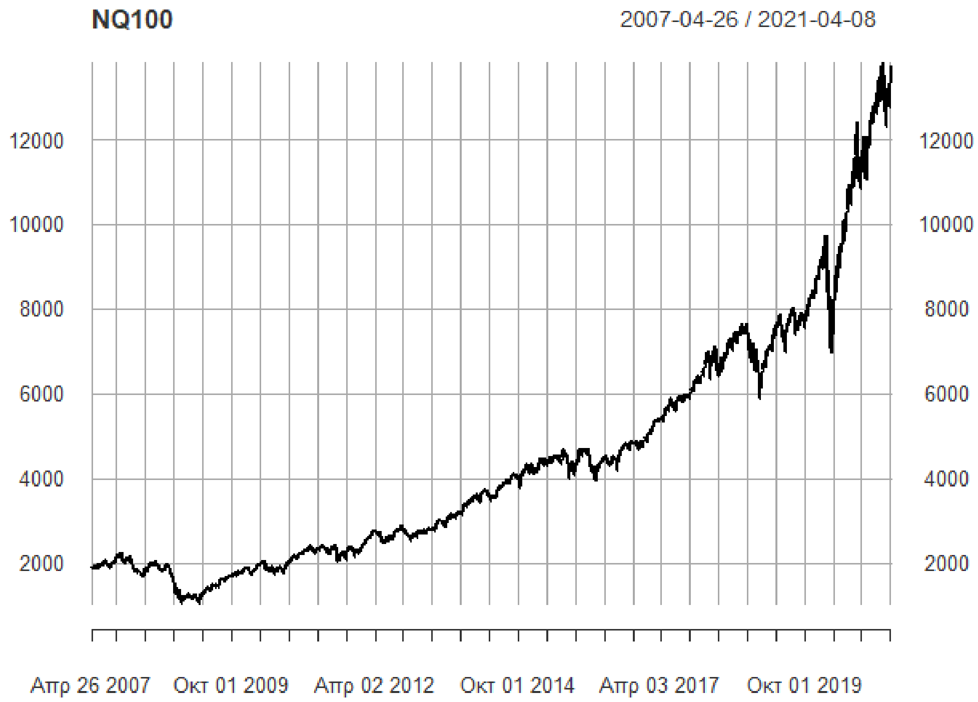

Figure 2.1: Nasdaq

As we can see from the graph of Nasdaq 100 shows an upturn through this specific time period. This upturn indicates the existence of time trend in the data. We can clearly observe a sudden drop around the time when Corona-virus outbreaks.

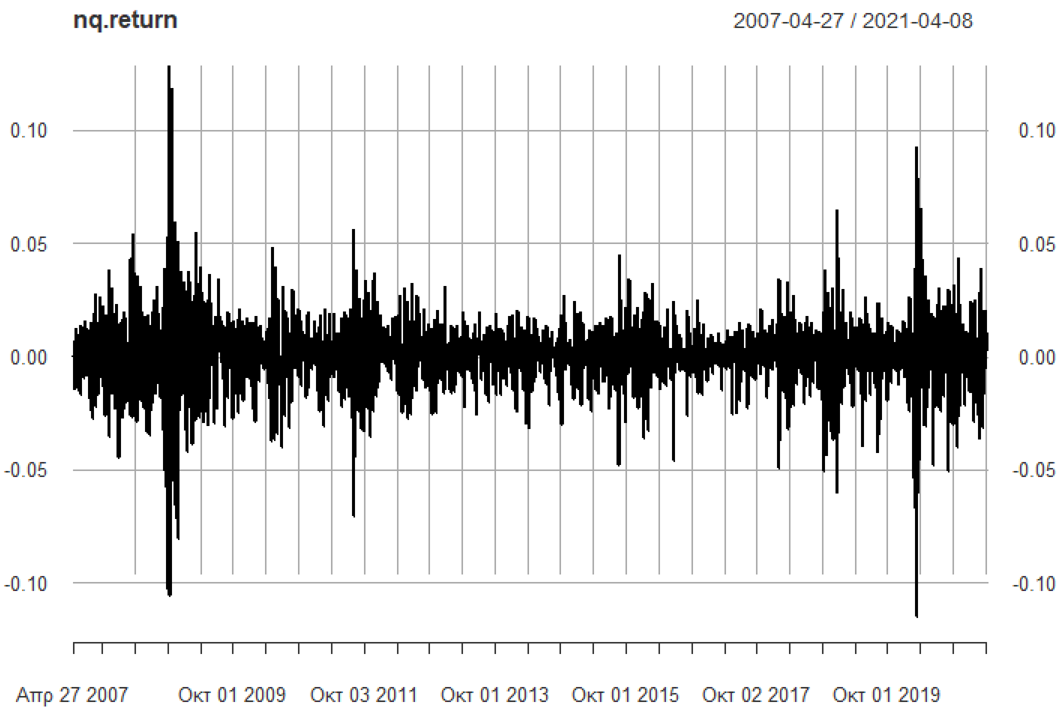

Figure 2.2: Returns

We calculate and plot the returns of Nasdaq 100 using the statistical package R. The first thing one can notice is the existence of multiple volatility clusters. The most notable one can be spotted at the worldwide financial crisis on 2008. Another incident that affected Nasdaq 100 is the American president elections on 2016. The most recent burst of volatility occurred due to the Corona-virus in 2019.

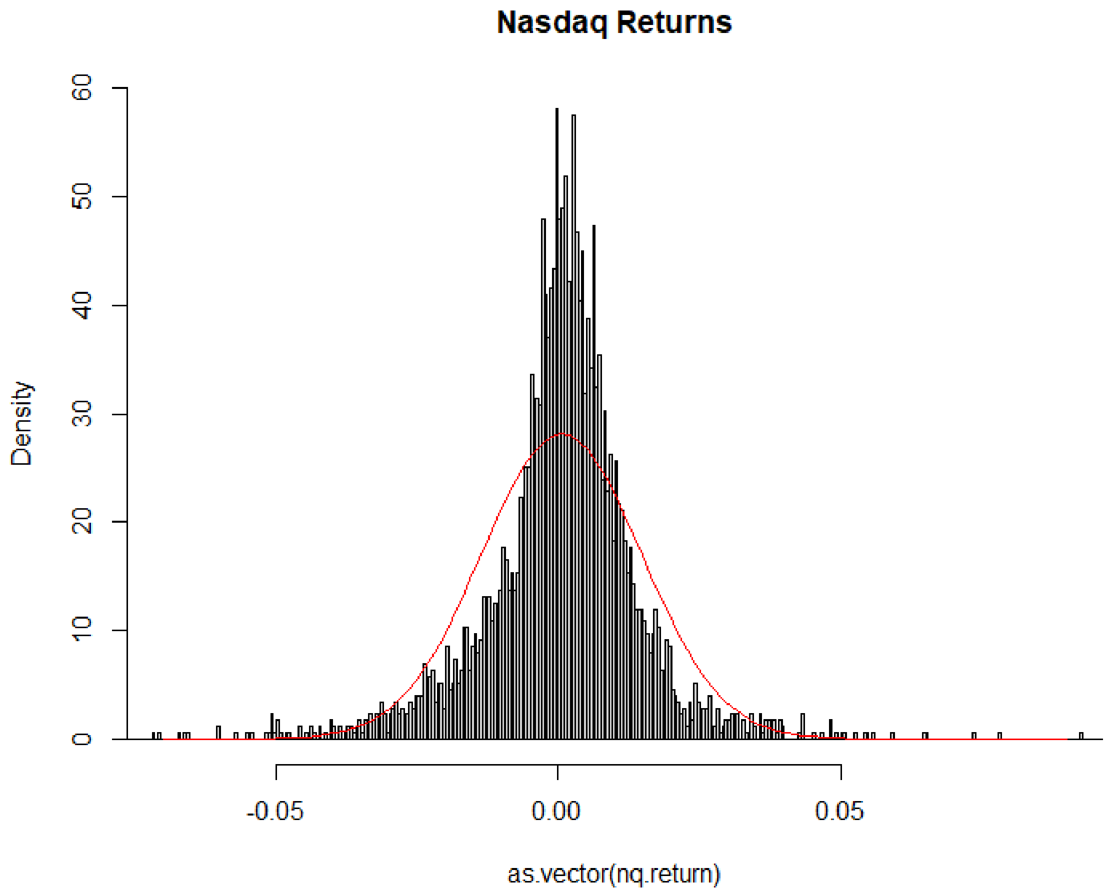

Figure 2.3: Histogram

The histogram appears slightly skewed to the left. This is supported by the skewness value, which acts as a measure of the asymmetry of the probability distribution.The skewness of the returns computed by R is -0.35 when the actual value for the data to be normal distributed should be 0. Also, the peak of the returns' distribution exceeds by far the theoretical bell shaped normal. The kurtosis value computed by R and is 10.11.

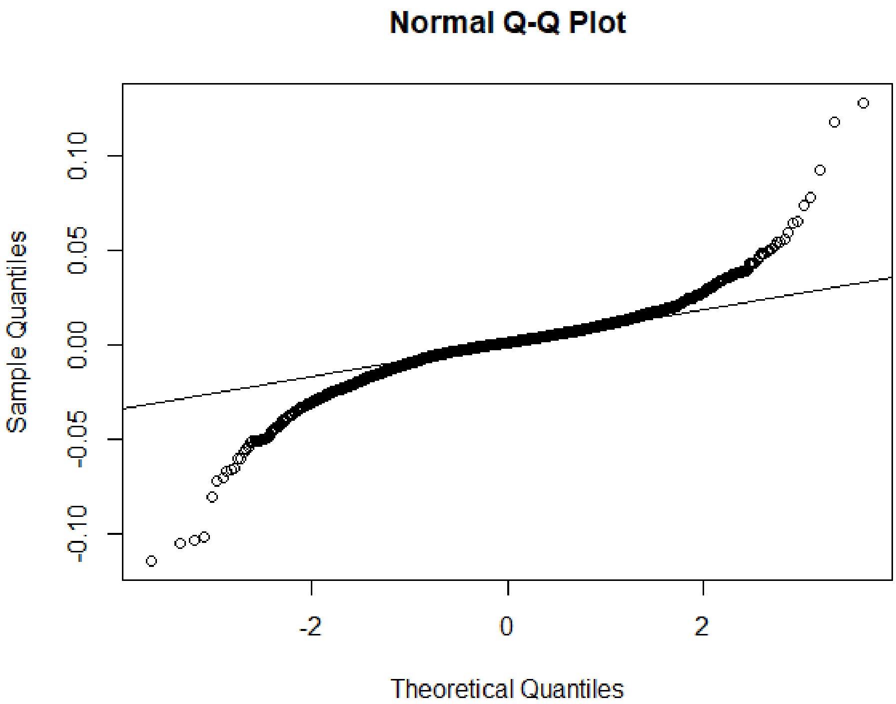

Figure 2.4: Normal QQ-plot

From the QQ-plot the data appear to fit well in the center but as we move outwards, the tails drift further from the normal distribution. The QQ-plot comes to support the evidence of peakedness and fat tails from the histogram. On the contrary the QQ-plot provides no clear clue for skewness as both tails seem to drift apart from the normal distribution line. Also we cannot see any change in variance as we did in the returns plot.

Maybe these fat tails indicate that the returns follow Student-t distribution.

Ruey S. Tsay (2014) [84] presented the distributional properties of returns indicating that Student-t distribution should be used for testing.

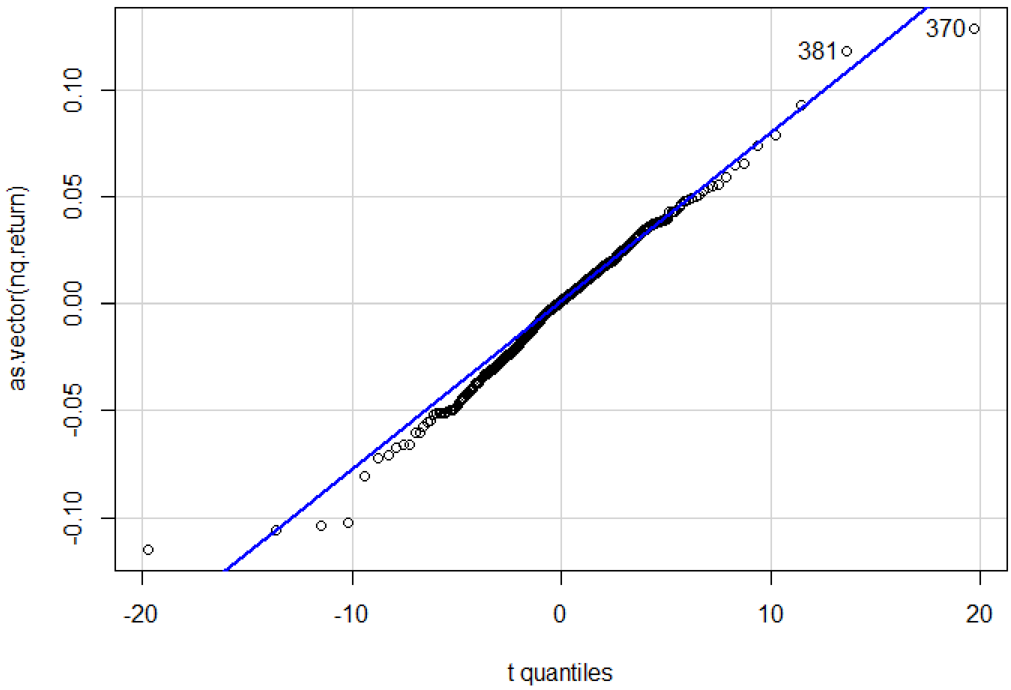

Figure 2.5: Student t(3)

After executing QQ-plot for Student-t for different degrees of freedom we conclude that the returns fit better at 3 degrees of freedom.

From the Student-t with 3 degrees of freedom figure, the returns seem to be closer to the line while the downside is still quite fat. There are also two leverage values.

Another thing one can notice is the two enumerated observations, which are identified as outliers. These are observed at the financial crisis period of 2008 and it is expected to have outliers in that period.

Next we need a statistical test to ensure that the data do not follow normal distribution. Since we observe fat tails the appropriate test would be Jarque and Bera's test, as it focuses on the tails of the distribution.

Jarque and Bera's test (1987) [39] is a goodness-of-fit test and compares the combination of skewness and kurtosis of the data with the corresponding values of the normal distribution. The JB-statistic is asymptotically distributed as a chi-squared random variable with 2 degrees of freedom.

Jarque and Bera Normality test

X-squared	df	p-value
15032	2	$2.2 \cdot 10^{-16}$

The p-value of the test is $2.2 \cdots 10^{-16}$, thus we reject the null hypothesis of normality, as it was expected from the normal QQ-plot.

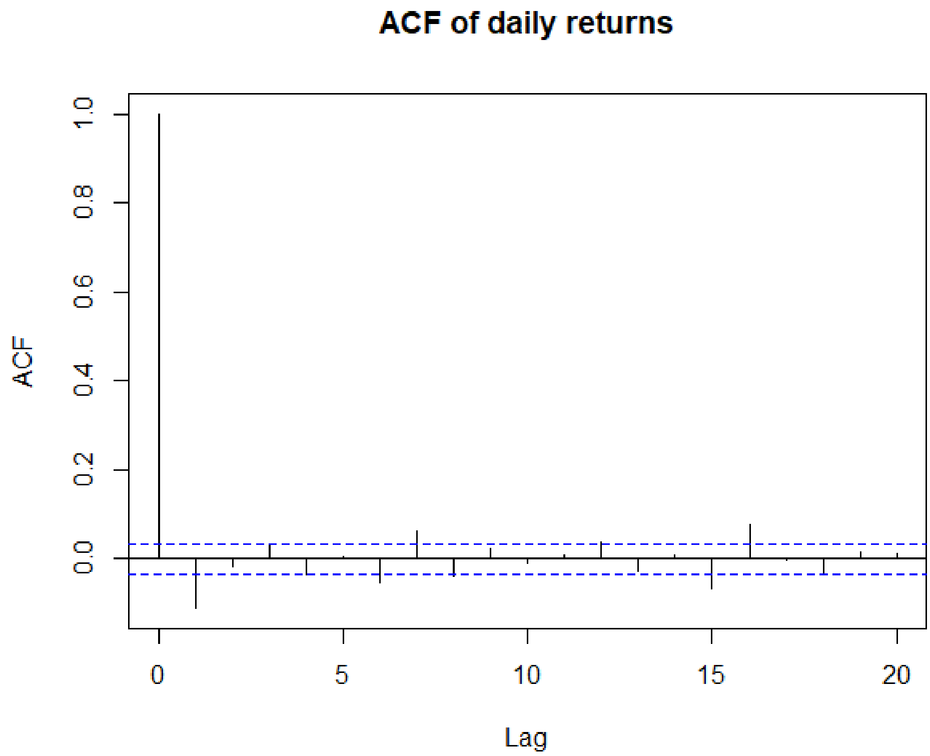

Figure 2.6: ACF plot

From the above graph of the autocorrelation function we can see the autocorrelations for different lags. The blue dot lines around zero implicate the confidence interval. Thus, every vertical line, which represents a lag, reaches outside the interval is statistically significant. In this case the preferable lag would be 16. The first value is equal to 1, as it is expected, because it is the autocorrelation at the present time. From the ACF we can see that the autocorrelation is negative and also there is no evidence for seasonality.

```
library(quantmod) # Load the package
library(tseries)

getSymbols("NQ=F", from="2007-04-26", to="2021-04-09", src = "yahoo")
NQ100 <- `NQ=F`[,6] #NASDAQ100
rm(`NQ=F`)


plot(NQ100)


NQ100 <- na.omit(NQ100)
summary(NQ100)
nq.return <- diff(log(NQ100))
summary(nq.return)
nq.return <- na.omit(nq.return)

plot(nq.return)

jarque.bera.test(nq.return)


####fat tails
nq.vec <- as.vector(nq.return)
qqnorm(as.vector(nq.return))
qqline(as.vector(nq.return))
```

```
## t-distribution
qqPlot(as.vector(nq.return),distribution="t", df=3,envelope=F)


hist(as.vector(nq.return), freq = F, breaks =600,xlim=c(-0.07,0.09), main="Nasdaq Returns")
curve(dnorm(x,mean(nq.return),sd(nq.return)),add=TRUE, col="red")



library(MASS)
library(stats)
q = acf(nq.return,20)
plot(q1,main="ACF of squared daily returns")
```

Chapter 3

Linear Regression

3.1 Linear Regression Model

Consider the simple linear regression model for cross-sectional data $y = \beta_0 + \beta_1 x + \epsilon$
Recall the assumptions that need to hold

1. The model is in the following format
$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$
2. $E(\epsilon_i|X) = 0$
3. $\sigma^2(x) = \sigma^2$
4. $Cov(\epsilon_i \ , \ \epsilon_s) = 0, \qquad \forall i \neq s$
5. $\epsilon_i \sim N(0, \sigma^2)$

Equivalent to the cross-sectional data we can use the linear model in time series and regress the Y time series on the X time series. Note that in this case the observations are in chronological order and cannot be reordered.

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

For the simple linear model we have

$$\begin{aligned}
y_1 &= \beta_1 + \beta_2 x_1 + \epsilon_1 \\
y_2 &= \beta_1 + \beta_2 x_2 + \epsilon_2 \\
y_3 &= \beta_1 + \beta_2 x_3 + \epsilon_3 \\
&\vdots \\
y_{T-1} &= \beta_1 + \beta_2 x_{T-1} + \epsilon_{T-1} \\
y_T &= \beta_1 + \beta_2 x_T + \epsilon_T
\end{aligned}$$

$$
\begin{aligned}
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}_{T\times 1} &= \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{T\times 1} \beta_1 + \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix}_{T\times 1} \beta_2 + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}_{T\times 1} \\
&= \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_T \end{pmatrix}_{T\times 2} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}_{2\times 1} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}_{T\times 1} \\
&= X_{T\times 2}\boldsymbol{\beta}_{2\times 1} + \boldsymbol{\epsilon}_{T\times 1}
\end{aligned}
$$

The last one is the linear model written in matrices.

3.1.1 Multi Linear regression

Consider the following independent variables

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + \epsilon_t$$

Equivalent to the simple regression we have

$$
\begin{aligned}
y_1 &= \beta_1 + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_k x_{1k} + \epsilon_1 \\
y_2 &= \beta_1 + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_k x_{2k} + \epsilon_2 \\
y_3 &= \beta_1 + \beta_2 x_{32} + \beta_3 x_{33} + \cdots + \beta_k x_{3k} + \epsilon_3 \\
&\vdots \\
y_T &= \beta_1 + \beta_2 x_{T2} + \beta_3 x_{T3} + \cdots + \beta_k x_{Tk} + \epsilon_T
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}_{T\times 1} &= \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{T\times 1} \beta_1 + \begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{T2} \end{pmatrix}_{T\times 1} \beta_2 + \begin{pmatrix} x_{13} \\ x_{23} \\ \vdots \\ x_{T3} \end{pmatrix}_{T\times 1} \beta_3 + \cdots + \begin{pmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{Tk} \end{pmatrix}_{T\times 1} \beta_k + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}_{T\times 1} \\
&= \begin{pmatrix} 1 & x_{12} & x_{13} & \dots & x_{1k} \\ 1 & x_{22} & x_{23} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{T2} & x_{T3} & \dots & x_{Tk} \end{pmatrix}_{T\times k} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}_{k\times 1} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}_{T\times 1} \\
&= X_{T\times k}\boldsymbol{\beta}_{k\times 1} + \boldsymbol{\epsilon}_{T\times 1}
\end{aligned}
$$

The last one is the vectorized form of the multiple linear regression.
Assumption

- The variables (Y, X) satisfy the linear regression equation

$$Y = \beta X' + \epsilon \tag{3.1}$$

$$E[\epsilon|X] = 0 \tag{3.2}$$

- The variables have finite second moments

$$E[Y^2] < \infty$$
$$E\|X\|^2 < \infty$$

- An invertible design matrix

$$E[XX'] > 0$$

We will consider both the general case of heteroskedastic regression where the conditional variance $E[\epsilon^2|X] = \sigma^2(X)$ is unrestricted, and the specialized case of homoskedastic regression where the conditional variance is constant. In the latter case we add the following assumption.

Homoskedastic Linear Regression Model In addition to Assumption 4.2

$$E[\epsilon^2|X] = \sigma^2(X) = \sigma^2 \tag{3.3}$$

is independent of X.

3.1.2 AR(1)

In finance and in time series in general, the past values of a variable may have an impact on the present and on the future values. That is why it is very useful to regress time series Y_t, not on exogenous variables but on its lagged values.

Autoregressive model of order 1 or AR(1) is a simple linear regression model, where Y_t is the dependent variable and Y_{t-1} is the explanatory variable and has a statistically significant lag-1 autocorrelation which indicates that the lagged Y_{t-1} might be useful in predicting Y_t.

Formula:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

$$E(Y_t) = \phi_0 + \phi_1 E(Y_{t-1}) \qquad (1)$$
$$Var(Y_t) = \gamma_0$$

Under stationarity

$$E(Y_t) = E(Y_{t-1}) \overset{(1)}{\Rightarrow} E(Y_t) = \phi_0 + \phi_1 E(Y_t) \Rightarrow (1-\phi_1)E(Y_t) = \phi_0 \Rightarrow E(X_t) = \frac{\phi_0}{1-\phi_1} \quad \phi_1 \neq 1$$

$$Var(Y_t) = \phi_1^2 Var(Y_{t-1}) + \sigma_\epsilon^2$$

As we already know we can generalize an AR(1) model with an AR(p) and study the influence of the past values on the present one.

3.1.3 ADL

A combination of the linear regression model and an autoregression model is the Autogressive Distributed Lag (ADL) model. According to Hank et al. (2019) [34], an ADL(p,q) model assumes that a time series Y_t can be represented by a linear function of p of its lagged values and q lags of another time series X_t:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \cdots + \delta_q X_{t-q} + u_t$$

is an autoregressive distributed lag model with p lags of Y_t and q lags of X_t where

$$E(u_t | Y_{t-1}, Y_{t-2}, \ldots, X_{t-1}, X_{t-2}, \ldots) = 0$$

3.1.4 Lag Length

It may seem, and it's not completely wrong, that with many predictors comes great predictive capability. The first thing that one should be cautious about is the overfitting. The second one is that many predictors imply many estimations which consequently add more errors to the model. These make no exception for the ADL and AR models. That is why one should carefully choose the lag length. There are statistical methods that are helpful to determine how many lags should be included as regressors.

1. **The F-test**
   Estimate an AR(p) model and test the significance of the largest lag(s). If the test rejects, drop the respective lag(s) from the model. This approach has the tendency to produce models where the order is too large: in a significance test we always face the risk of rejecting a true null hypothesis.

2. **Use of AIC and BIC**
   To counter the issue of producing too large models, one could use an information criterion to choose the lag length.

   - The Bayes information criterion (BIC):

   $$BIC(p) = \log\left(\frac{SSR(p)}{T}\right) + +(p+1)\frac{\log(T)}{T}$$

   - The Akaike information criterion (AIC):

   $$AIC(p) = \log\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{2}{T}$$

Both criteria are estimators of the optimal lag length p. The lag order $\hat{p}$ that minimizes the respective criterion is called the BIC estimate or the AIC estimate of the optimal model order. The basic idea of both criteria is that the SSR decreases as additional lags are added to the model such that the first term decreases whereas the second increases as the lag order grows. One can show that the BIC is a consistent estimator of the true lag order while the AIC is not which is due to the differing factors in the second addend. Nevertheless, both estimators are used in practice where the AIC is sometimes used as an alternative when the BIC yields a model with "too few" lags.

3.1.5 Durbin-Watson test

Consider the simple linear model

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t$$

where the residuals are in an AR(1) form

$$\epsilon_t = \rho\epsilon_{t-1} + e_t$$

Let $\hat{\epsilon}_t$ be the residuals resulting from the LS estimation, then the DW-test is based on the following statistical criterion

$$
\begin{aligned}
DW &= \frac{\sum_{t=2}^{T} (\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^{T} \hat{\epsilon}_t^2} \\
&= \frac{\sum_{t=2}^{T} \hat{\epsilon_t}^2 + \sum_{t=2}^{T} \hat{\epsilon_{t-1}}^2 - 2\sum_{t=2}^{T} \hat{\epsilon_t}\hat{\epsilon_{t-1}}}{\sum_{t=1}^{T} \hat{\epsilon_t}^2} \\
&\approx 1 + \frac{\sum_{t=2}^{T} \hat{\epsilon_{t-1}}^2}{\sum_{t=1}^{T} \hat{\epsilon_t}^2} - 2\frac{\sum_{t=2}^{T} \hat{\epsilon_t}\hat{\epsilon_{t-1}}}{\sum_{t=1}^{T} \hat{\epsilon_t}^2} \\
&\approx 1 + 1 - 2\hat{\rho} = 2(1 - \hat{\rho}) \qquad (3.4)
\end{aligned}
$$

for $\frac{\sum_{t=2}^{T} \hat{\epsilon_t}^2}{\sum_{t=1}^{T} \hat{\epsilon_t}^2} \approx 1$, $\frac{\sum_{t=2}^{T} \hat{\epsilon_{t-1}}^2}{\sum_{t=1}^{T} \hat{\epsilon_t}^2} \approx 1$ and $\frac{\sum_{t=2}^{T} \hat{\epsilon_t}\hat{\epsilon_{t-1}}}{\sum_{t=1}^{T} \hat{\epsilon_t}^2}$ which is an approach of the LS estimator $\hat{\rho}$ of the coefficient ρ.

Based on the 3.4 equation we have the following facts for the relationship between the value of DW and ρ

1. If $\hat{\rho} = 0$ there is no correlation and $DW \approx 2$

2. If $\hat{\rho} = 1$ we have perfect positive correlation and $DW \approx 0$. This value is called lower-L and is denoted as DW_L.

3. If $\hat{\rho} = -1$ there is negative correlation and $DW \approx 4$. This value is called upper-U and is denoted as DW_U.

The null hypothesis is then

$$H_0 : DW = 2$$

and there are 2 possible alternative hypothesis for this test with the first one being

$$H_a : DW < 2$$

which implies that there is positive correlation between the variables and the second one being

$$H_a : DW > 2$$

which implies that there is negative correlation between the variables.

In specific

- If $DW < DW_L$ we reject the null hypothesis of no autocorrelation and accept that there is positive autocorrelation.

- If $DW > (4 - DW_L)$ we reject the null hypothesis of no autocorrelation and accept that there is negative autocorrelation.

- If $DW_U < DW < (4 - DW_U)$ we accept the null hypothesis of no autocorrelation.

- If $DW_L < DW < (4 - DW_U)$ or if $(4 - DW_U) < DW < (4 - DW_L)$ the test is inconclusive.

3.2 Spurious Regression

3.2.1 Case 1 of Spurious regression

In time series analysis, sometimes, we would like to know whether changes in a variable will have an impact on changes of other variables. Basically we would like to know if changes on X causes changes on Y. This is the concept of causality. Testing causality among variables is one of the most important and, yet, one of the most difficult issues in economics.

The two most difficult challenges are when the correlation does not imply causality and that there always exists the possibility of ignored common factors. The causal relationship among variables might disappear when the previously ignored common causes are considered.

Definition 24. *The Granger causality test [34]*

The Granger causality test (1969) [28] is an F-test of the null hypothesis that all lags of a variable X included in a time series regression model do not have predictive power for Y_t. The Granger causality test does not test whether X actually causes Y but whether the included lags are informative in terms of predicting Y.

This misconception of correlation and causality leads us to spurious regression.

Another reason for the arise of spurious regression is the nonstationarity of time series. This means that two statistically independent series, if both unit root processes, are likely to fool traditional statistical analysis by appearing to be statistically related by both graphically and traditional statistical tests.

The phenomenon was observed and named by Granger and Newbold (1974) [30] and explained using the theory of non-stationary time series by Phillips (1986) [62].

Definition 25. *Spurious [34]*

When two stochastically trending time series are regressed onto each other, the estimated relationship may appear highly significant using conventional normal critical values although the series are unrelated. This is what econometricians call a spurious relationship.

These stochastically trending time series could be two independent random walks with zero mean and variance equal to 1. Random walks have a weird ability to fool casual analysis.

Consider the following random walk processes

$$Y_t = Y_{t-1} + \epsilon_{1t} \tag{3.5}$$
$$X_t = X_{t-1} + \epsilon_{2t} \tag{3.6}$$

where $(\epsilon_{1t}, \epsilon_{2t})$ are i.i.d., mean zero, mutually uncorrelated, and normalized to have unit variance. Let Y_t^* and X_t^* denote demeaned versions of $Y_t = Y_{t-1}$ and $X_t = X_{t-1}$. From the FCLT they satisfy

$$\left(\frac{1}{\sqrt{T}} Y^*_{\lfloor Tr \rfloor} \, , \, \frac{1}{\sqrt{T}} X^*_{\lfloor Tr \rfloor}\right) \xrightarrow{d} (W_1^*(r) \, , \, W_2^*(r))$$

where $W_1^*(r)$ and $W_2^*(r)$ are demeaned Brownian motions.

Applying the CMT the sample correlation has the asymptotic distribution

$$\hat{\rho} = \frac{\frac{1}{n^2}\sum_{i=1}^{n} Y_i^* X_i^*}{\sqrt{\frac{1}{n^2}\sum_{i=1}^{n} Y_i^{*2}}\sqrt{\frac{1}{n^2}\sum_{i=1}^{n} X_i^{*2}}} \underset{d}{\to} \frac{\int_0^1 W_1^* W_2^*}{\sqrt{\int_0^1 W_1^{*2}}\sqrt{\int_0^1 W_2^{*2}}}$$

The right-hand-side is a random variable. Furthermore it is also non-degenerate (indeed, it is non-zero with probability one). Thus the sample correlation $\hat{\rho}$ remains random in large samples.

Example

```
##### first form spurious----------

rm(list=ls())
cat("\014")


# by construction y and x are two independent random walks
set.seed(2492)
n <- 5000
y <- c()
x <- c()
y[1] <- x[1] <- 0

for (i in 2:n) {
  y[i] <- y[i-1] + rnorm(1)
  x[i] <- x[i-1] + rnorm(1)
}


fit1 <- lm( y ~ x )
summary(fit1)

library(car)
durbinWatsonTest(fit1)
acf(fit1$residuals)


fit2 <- lm(diff(y) ~ diff(x))
summary(fit2)
durbinWatsonTest(fit2)
acf(fit2$residuals)
```

First we consider a model

$$y_t = \beta x_t + \epsilon_t,$$

where y_t and x_t are two independent random walks.

Table 3.1:

	Dependent variable:
	y
x	-0.787^{***}
	(0.003)
Constant	25.282^{***}
	(0.403)
Observations	5,000
R^2	0.923
Adjusted R^2	0.923
Residual Std. Error	14.211 (df = 4998)
F Statistic	$6.0162 \cdot 10^5$ *** (df = 1; 4998)
Note:	*p<0.1; **p<0.05; ***p<0.01

We expect the coefficient of X converge to zero, because the two random walks are independent. In this case, it is quite significant since we have a low p-value. One could falsely assume, due to high R^2, that this model has interpretive capability. On the contrary Durbin-Watson test indicates that the residuals are serially correlated.

Autocorrelation test Durbin-Watson

H_0 : residuals are not autocorrelated		H_a : residuals are autocorrelated of order 1	
lag	Autocorrelation	D-W Statistic	p- value
1	0.9957	0.0079	0

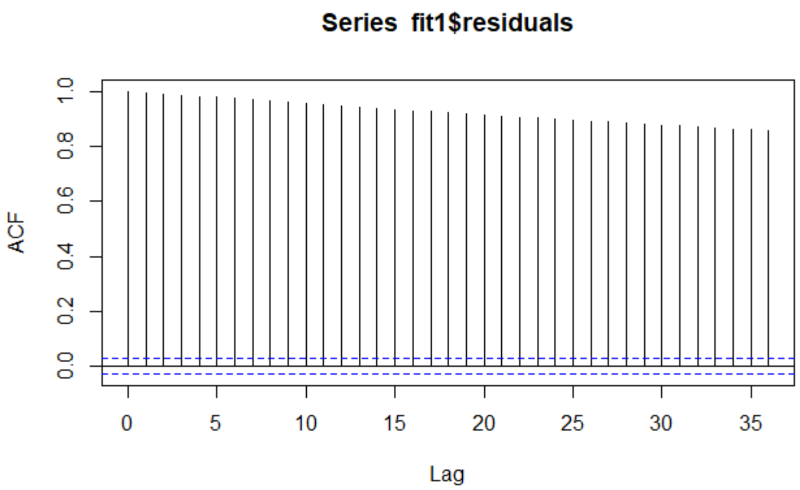

Figure 3.1: ACF of the residuals

These are strong evidences that we have spurious regression. In order to be sure, we consider a model of the first differences

$$\Delta y_t = \beta \Delta x_t + \epsilon_t$$

Table 3.2:

	Dependent variable:
	diff(y)
diff(x)	−0.015
	(0.014)
Constant	0.035**
	(0.014)
Observations	4,999
R^2	0.0002
Adjusted R^2	0.00003
Residual Std. Error	1.003 (df = 4997)
F Statistic	1.148 (df = 1; 4997)
Note:	*p<0.1; **p<0.05; ***p<0.01

The first thing one can easily notice is the drastically fall of the R^2. This means that the model can predict squat. Now the Durbin-Watson test implies that the residuals are serially uncorrelated.

Autocorrelation test Durbin-Watson

H_0 : residuals are not autocorrelated H_a : residuals are autocorrelated of order 1

lag	Autocorrelation	D-W Statistic	p- value
1	0.0186	1.9624	0.182

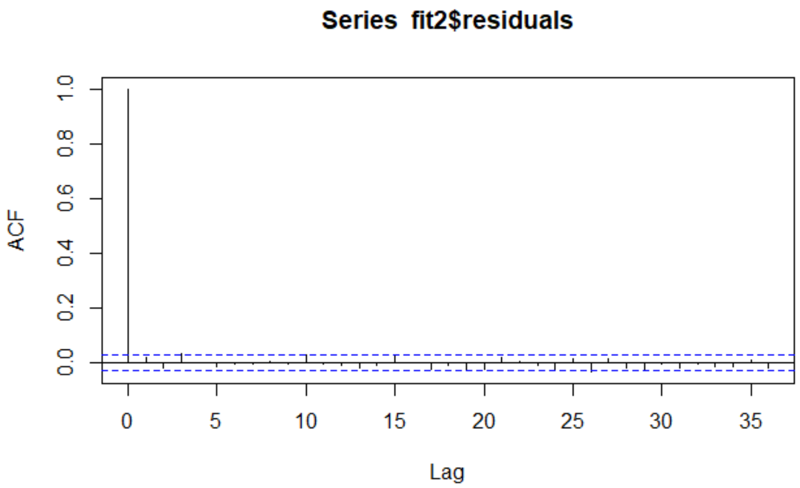

Figure 3.2: ACF of the residuals

The correlogram supports the evidence from the DW-test.

3.2.2 Case 2 of Spurious Regression

Sorensen (2019) [78] presented another case of spurious regression. Again, consider two random walks x_t and y_t but in this case y_t is simulated using x_t. Even though there actually is a relation between x_t and y_t, this is also a spurious regression.

Example

```
##########second form spurious--------

rm(list=ls())
cat("\014")


set.seed(601)
n <- 5000
y <- c()
x <- c()
y[1] <- x[1] <- 0

for (i in 2:n) {
  y[i] <- y[i-1] + rnorm(1)
  x[i] <- x[i-1] + rnorm(1)
}

y <- y + 0.5 * x
fit3 <- lm( y ~ x )
summary(fit3)

library(car)
durbinWatsonTest(fit3)
acf(fit3$residuals)


fit4 <- lm(diff(y) ~ diff(x))
summary(fit4)
durbinWatsonTest(fit4)
acf(fit4$residuals)
```

First we consider a model

$$y_t = \beta x_t + \epsilon_t,$$

where y_t and x_t are two random walks. The relationship between the two random walks emerge from the structure of y_t, where $y_t = 0.5x_t$.

Table 3.3:

	Dependent variable:
	y
x	1.054^{***}
	(0.003)
Constant	6.619^{***}
	(0.256)
Observations	5,000
R^2	0.966
Adjusted R^2	0.966
Residual Std. Error	9.935 (df = 4998)
F Statistic	$1.4209 \cdot 10^5$ *** (df = 1; 4998)
Note:	*p<0.1; **p<0.05; ***p<0.01

At first we observe that R^2 is very big, this implies that the model has high interpretive capability. On the contrary Durbin-Watson test indicates that the residuals are serially correlated.

Autocorrelation test Durbin-Watson

H_0 : residuals are not autocorrelated H_a : residuals are autocorrelated of order 1

lag	Autocorrelation	D-W Statistic	p- value
1	0.9933	0.0132	0

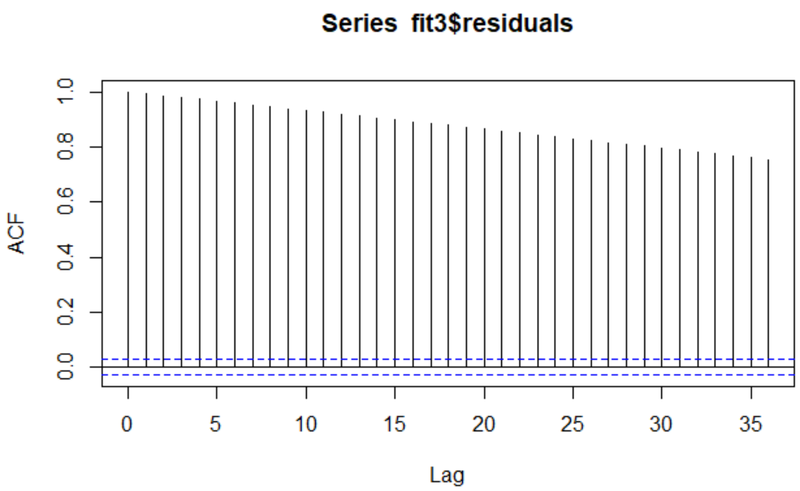

Figure 3.3: ACF of the residuals

These are strong evidences that we have spurious regression. In order to be sure, we consider a model of the first differences

$$\Delta y_t = \beta \Delta x_t + \epsilon_t$$

Table 3.4:

	Dependent variable:
	diff(y)
diff(x)	0.514***
	(0.014)
Constant	−0.016
	(0.014)
Observations	4,999
R^2	0.203
Adjusted R^2	0.202
Residual Std. Error	1.010 (df = 4997)
F Statistic	$1.269 \cdot 10^5$ *** (df = 1; 4997)
Note:	*p<0.1; **p<0.05; ***p<0.01

The first thing one can easily notice is the drastically fall of the R^2. This means that the model can predict squat. Now the Durbin-Watson test implies that the residuals are serially uncorrelated.

Autocorrelation test Durbin-Watson

H_0 : residuals are not autocorrelated		H_a : residuals are autocorrelated of order 1	
lag	Autocorrelation	D-W Statistic	p- value
1	−0.0120	2.0240	0.41

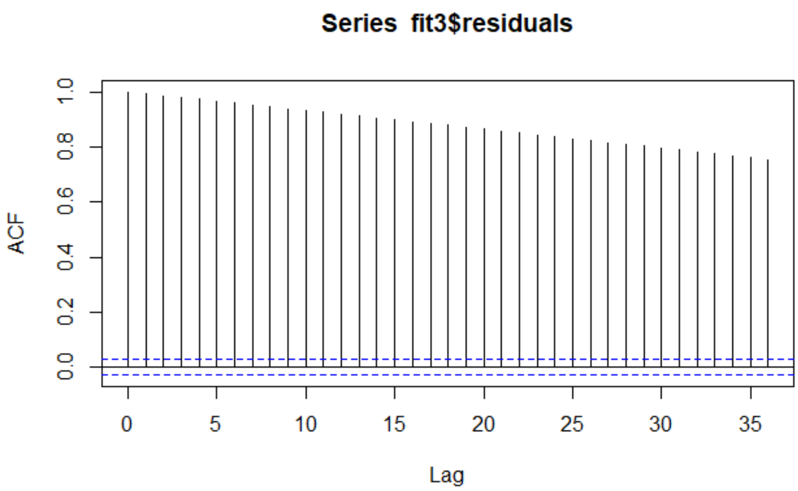

Figure 3.4: ACF of the residuals

The correlogram supports the evidence from the DW-test.

3.2.3 Multivariate Spurious Regression

Spurious Regression

Of course the previous concept can be applied when there are more than one variables. If some or all of the variables in the regression are I(1) then the usual statistical results may or may not hold, like in the spurious regression case.

Although there is a sweet point where nonstationarity does not imply spurious regression, that is the case of cointegration.

Let $\boldsymbol{Y_t} = (y_{1t}, \cdots, y_{nt})'$ denote an $(n \times 1)$ vector of I(1) time series that are not cointegrated. Using the partition $\boldsymbol{Y_t} = (y_{1t}, \boldsymbol{Y_{2t}'})'$, consider the least squares regression of y_{1t} on $\boldsymbol{Y_{2t}}$ giving the fitted model

$$y_{1t} = \hat{\boldsymbol{\beta}}_2' \boldsymbol{Y_{2t}} + \hat{\epsilon}_t \tag{3.7}$$

Since y_{1t} is not cointegrated with $\boldsymbol{Y_{2t}}$ 3.7 is a spurious regression and the true value of $\boldsymbol{\beta}_2$ is zero. The following results about the behaviour of $\hat{\boldsymbol{\beta}}_2$ in the spurious regression 3.7 are due to Phillips (1986) [62]

- $\hat{\boldsymbol{\beta}}_2$ does not converge in probability to zero but instead converges in distribution to a non-normal random variable not necessarily centered at zero. This is the spurious regression phenomenon.
- The usual OLS t-statistics for testing that the elements of $\boldsymbol{\beta}_2$ are zero diverge to $\pm\infty$ as $T \to \infty$. Hence, with a large enough sample it will appear that $\boldsymbol{Y_t}$ is cointegrated when it is not if the usual asymptotic normal inference is used.
- The usual R^2 from the regression converges to unity as $T \to \infty$ so that the model will appear to fit well even though it is misspecified.
- Regression with I(1) data only makes sense when the data are cointegrated.

3.3 Cures for Spurious Regressions

There are three ways in which the problems associated with spurious regressions can be avoided. The first approach is to include lagged values of both the dependent and independent variable in the regression. For example,consider an OLS regression of y_t on x_t being the following model

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t \tag{3.8}$$

Equivalent,

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + \beta_3 x_{t-1} + \epsilon_t \tag{3.9}$$

It can be shown that OLS estimation of 3.9 yields consistent estimates of all of the parameters. The coefficients $\hat{\beta}_{2T}$ and $\hat{\beta}_{3t}$ each individually converge at rate $\sqrt{T}$ to a Gaussian distribution, and the t test of the hypothesis that $\beta_2 = 0$ is asymptotically $N(0,1)$, as is the t test of the hypothesis that $\beta_3 = 0$. However, an F test of the joint null hypothesis where β_2 and β_3 are both zero has a non-standard limiting distribution. Hence, including lagged values in the regression is sufficient to solve many of the problems associated with spurious regressions, although tests of some hypotheses will still involve non-standard distributions.

A second approach is to differentiate the data before estimating the relation, as in

$$\Delta y_t = \beta_0 + \beta_2 \Delta x_t + \epsilon_t \tag{3.10}$$

Clearly, since the regressors and error term ϵ_t are all I(0) for this regression under the null hypothesis, $\hat{\beta}_{0T}$ and $\hat{\beta}_{2T}$ both converge at rate $\sqrt{T}$ to Gaussian variables. Any t or F test based on 3.10 has the usual limiting Gaussian or $\mathcal{X}^2$ distribution.

Because the specification 3.10 avoids the spurious regression problem as well as the nonstandard distributions for certain hypotheses associated with the levels regression 3.8, many researchers recommend routinely differentiating apparently nonstationary variables before estimating regressions. While this could be the perfect solution for the spurious problem, there are two different situations in which it might be inappropriate. First, if the data are really stationary, then the first differences method can result in a misspecified regression. Second, even if both y_t and x_t are truly I(1) processes, there is an interesting class of models for which the bivariate dynamic relation between y and x will be misspecified if the researcher simply differentiates both y and x. This class of models, known as cointegrated processes.

Chapter 4

Statistical Arbitrage in Pairs Trading

4.1 Stochastic Processes

The vast majority of financial time series like, interest rates, foreign exchange rates, and asset price series exhibit non-stationarity. The non-stationarity of a price series is mostly attributable to the absence of a set price level, as well as the existence of inflation. In literature these series are encountered by the name of unit root non-stationary time series and trend non-stationary. We prefer the second term as it is more representative. The random-walk model is the most well-known example of trend non-stationary time series.

As we have already examined, random walk is a stochastic process and can be produced by adding a binary variable values ($S_t = S_{t-1} + X_t$) for discrete time. The binary random variable is not the only way to produce the random walk. We can instead add the values of a Normally distributed random variable.

The recursive form is defined as

$$y_t = y_{t-1} + u_t$$

where $u_t \sim N(0, \sigma^2)$. This is nothing but a special case of an AR(1) model, with $\phi_1 = 1$. As we already know AR(1) in order for it to be stationary it needs to have $|\phi_1| < 1$.

Consider an AR(1) model with polynomial equation

$$\begin{aligned}
&y_t = \phi_1 y_{t-1} \Rightarrow \\
&y_t - \phi_1 y_{t-1} = 0 \Rightarrow \\
&(1 - \phi_1 \boldsymbol{B}) y_t = 0
\end{aligned}$$

so, in order for y_t to be stationary the absolute value of the characteristic root $1 - \phi_1 z = 0$, ϕ_1 must be less than 1.

Now, let's consider an AR(2) model with polynomial equation

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t, \qquad u_t \sim (0, \sigma^2)$$

$$\begin{aligned}
E(y_t) &= \phi_0 + \phi_1 E(y_{t-1}) + \phi_2 E(y_{t-2}) + E(u_t) \Rightarrow \\
\mu &= \phi_0 + \phi_1 \mu + \phi_2 \mu + 0 \Rightarrow \\
\mu &= \frac{\phi_0}{1 - \phi_1 - \phi_2}
\end{aligned}$$

where the restriction $\phi_1 + \phi_2 \neq 1$ must hold.

An AR(2) series in order to be stationary it needs to satisfy the second-order difference polynomial equation

$$\begin{aligned} &y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} \Rightarrow \\ &y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = 0 \Rightarrow \\ &(1 - \phi_1 \boldsymbol{B} - \phi_2 \boldsymbol{B}^2) y_t = 0 \end{aligned}$$

$$1 - \phi_1 z - \phi_2 z^2 = 0$$

The solutions of the polynomial equation are referred to as the characteristic roots of the AR(2) model.

$$z_{1,2} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4 \cdot 1 \cdot \phi_2}}{-2\phi_2}$$

If the characteristic roots z_1, z_2 are real valued, then the second-order difference equation of the model can be factored such as $(1 - z_1\boldsymbol{B})(1 - z_2\boldsymbol{B})$ the AR(2) model can be regarded as an AR(1) model that operates on top of another AR(1) model. If the discriminant, $\phi_1^2 + 4\phi_2$, is negative, the characteristic roots are complex numbers (called a complex conjugate pair).

The stationarity condition of an AR(2) time series is that the norm of the characteristic roots are less than 1.

Since ARMA is a combination of AR and MA models we can generalize the previous concept into an ARMA(p,q) model

$$\begin{aligned} &y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = \phi_0 + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \\ &(1 - \phi_1 \boldsymbol{B} - \cdots - \phi_p \boldsymbol{B}^p) y_t = \phi_0 + (1 - \theta_1 \boldsymbol{B} - \cdots - \theta_q \boldsymbol{B}^q) a_t, \quad a_t \text{ is a white noise series.} \end{aligned}$$

we obtain a characteristic equation similar to the one of an AR model. If the absolute value of all the characteristic roots are less than 1, then the ARMA model is stationary.

When the autoregressive or moving average polynomial of an ARMA model has a root on or around the unit circle[1], the unit root problem in time series occurs. A unit root in either of these polynomials has significant modeling implications. A root near 1 in the autoregressive polynomial, for example, indicates that the data should be differentiated before fitting an ARMA model, but a root near 1 in the moving-average polynomial shows that the data have been over-differentiated.

According to Lalley's notes [48] Brownian motion can be viewed as a limit of rescaled simple random walks.

Consider a continuous-time stochastic process $\{W_n(t)\}_{t \geq 0}$ for $n \geq 1$

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{1 \leq j \leq \lfloor nt \rfloor} u_j$$

[1]unit circle is the circle consisting all the complex numbers with norm equal to 1.

where $u_1, u_2, \cdots$ is a sequence of i.i.d. random variables with mean zero 0 and variance 1.

This is a random walk with jumps of size $u_k/\sqrt{n}$ at times k/n, where $k \in \mathbb{Z}_+$. Since the random variables u_j are independent, the increments of $W_n(t)$ are independent. From the Central Limit Theorem, for large n the distribution of the increments converges to a $N(0,t)$. Hence, for $n \to \infty$, one can embrace the idea that the distribution of $W_n(t)$ converges to the distribution of a standard Brownian motion.

4.1.1 Functional Central Limit Theorem

In previous chapter we presented the simple version of Central Limit Theorem (CLT): if u_t i.i.d. with mean zero and variance σ^2, then the sample mean $\mu_{u_T} = \frac{1}{T}\sum_{t=1}^{T} u_t$ satisfies

$$\sqrt{T}\mu_{u_T} \xrightarrow{L} N(0,\sigma^2)$$

Consider now an estimator based on the following principle: When given a sample of size T, we calculate the mean of the first half of the sample and throw out the rest of the observations:

$$\mu_{u_{T/2}} = \frac{1}{\lfloor T/2 \rfloor}\sum_{t=1}^{\lfloor T/2 \rfloor} u_t$$

where $\lfloor T/2 \rfloor$ denotes the largest integer that is less than or equal to $T/2$.

This estimator would also satisfy the CLT and moreover would be independent of an estimator that uses only the second half of the sample.

Consider a variable $X_T(r)$ from the sample mean of the first rth fraction of observations, defined by

$$X_T(r) = \frac{1}{T}\sum_{t=1}^{\lfloor Tr \rfloor} u_t$$

For any given realization, $X_T(r)$ is a step function in r, with

$$X_T(r) = \begin{cases} 0 & \text{for } 0 \le r < 1/T \\ \frac{u_1}{T} & \text{for } 1/T \le r < 2/T \\ \frac{u_1+u_2}{T} & \text{for } 2/T \le r < 3/T \\ \vdots & \\ \frac{u_1+\cdots+u_T}{T} & \text{for } r = 1 \end{cases}$$

Then

$$\sqrt{T}X_T(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr \rfloor} u_t = \frac{\sqrt{\lfloor Tr \rfloor}}{\sqrt{T}}\frac{1}{\sqrt{\lfloor Tr \rfloor}}\sum_{t=1}^{\lfloor Tr \rfloor} u_t \tag{4.1}$$

But

$$\frac{1}{\sqrt{\lfloor Tr \rfloor}}\sum_{t=1}^{\lfloor Tr \rfloor} u_t \xrightarrow{L} N(0,r)$$

by the CLT while $\frac{\sqrt{\lfloor Tr \rfloor}}{\sqrt{T}} \to \sqrt{r}$. Hence, the asymptotic distribution is

$$\sqrt{T} X_T(r) \xrightarrow{L} N(0, r\sigma^2)$$

and

$$\sqrt{T}\frac{X_T(r)}{\sigma} \xrightarrow{L} N(0, r). \tag{4.2}$$

If we were similarly to consider the behaviour of a sample mean based on observations $\lfloor Tr_1 \rfloor$ through $\lfloor Tr_2 \rfloor$ for $r_2 > r_1$, we would conclude that this too is asymptotically Normal,

$$\sqrt{T}\frac{X_T(r_2 - X_T(r_1)}{\sigma} \xrightarrow{L} N(0, r_2 - r_1)$$

and it is independent of the estimator in 4.2, provided that $r < r_1$. It thus should not be surprising that the sequence of stochastic functions $\left\{\sqrt{T}\frac{X_T(\cdot)}{\sigma}\right\}_{T=1}^{\infty}$ has an asymptotic probability law that is described by standard Brownian motion $W(\cdot)$:

$$\sqrt{T}\frac{X_T(\cdot)}{\sigma} \xrightarrow{L} W(\cdot) \tag{4.3}$$

The expression $X_T(\cdot)$ denotes a random function while $X_T(r)$ denotes the value that function assumes at date r. This means that $X_T(\cdot)$ is a function, while $X_T(r)$ is a random variable.

Result 4.3 is known as the Functional Central Limit Theorem (FCLT). The derivation here assumed that u_t was i.i.d.

Note that at $r = 1$, the function $X_T(r)$ in 4.1 is the sample mean. Thus, for $r = 1$ the function in 4.3 is the simple CLT.

$$\sqrt{T}\frac{X_T(1)}{\sigma} = \frac{1}{\sigma\sqrt{T}}\sum_{t=1}^{T} u_t \xrightarrow{L} W(1) \sim N(0, 1)$$

Continuous Mapping Theorem

It is known that if $\{x_T\}_{T=1}^{\infty}$ is a sequence of random variables with $x_T \xrightarrow{L} x$ and if $g : \mathbb{R} \to \mathbb{R}$ is a continuous function, then $g(x_T) \xrightarrow{L} g(x)$. A similar result holds for sequences of random functions. Here, the analog to the function $g(\cdot)$ is a continuous functional, which could associate a real random variable y with the stochastic function $S(\cdot)$. For example, $y = \int_0^1 S(r)dr$ and $y = \int_0^1 [S(r)]^2 dr$ represent continuous functional. The continuous mapping theorem (A.3 p.276) of Hall and Heyde (1980) [32] the states that if $S_T(\cdot) \xrightarrow{L} S(\cdot)$ and $g(\cdot)$ is a continuous functional, then $g(S_T(\cdot)) \xrightarrow{L} g(S(\cdot))$.

The continuous mapping theorem also applies to a continuous functional $g(\cdot)$ that maps a continuous bounded function on $[0, 1]$ into another continuous bounded function on $[0, 1]$. For example, the function whose value at r is a positive constant σ times $h(r)$ represents the result of applying the continuous functional $g[h(\cdot)] = \sigma h(\cdot)$ to $h(\cdot)$. Thus, it follows from 4.3 that

$$\sqrt{T}X_T(\cdot) \xrightarrow{L} \sigma W(\cdot) \tag{4.4}$$

Recalling that $W(r) \sim N(0, r)$, result 4.4 implies that $\sqrt{T}X_T(r) \approx N(0, \sigma^2 r)$. As an example, consider the function $S_T(\cdot)$ whose values at r given by

$$S_T(r) = [\sqrt{T}X_T(r)]^2 \tag{4.5}$$

Since $\sqrt{T}X_T(\cdot) \xrightarrow{L} \sigma W(\cdot)$, it follows that

$$S_T(\cdot) \xrightarrow{L} \sigma^2[W(\cdot)]^2$$

In other words, if the value $W(r)$ from a realization of standard Brownian motion at every date r is squared and then multiplied by σ^2, the resulting continuous-time process would follow essentially the same probability law as does the continuous-time process defined by $S_T(r)$ in 4.5 for T sufficiently large.

4.2 Asymptotic Properties

4.2.1 Asymptotic Properties of a First-order Autoregression when the True Coefficient is Unity

Hamilton (1994) [33] provides us, relying on the seminal work of Phillips (1987) [63], with the calculations of the asymptotic distribution of some basic stochastic processes with unit roots.

Proposition 1. *Suppose that ξ_t, follows a random walk without drift,*

$$\xi_t = \xi_{t-1} + u_t,$$

where $\xi_0 = 0$ and $\{u_t\}$ is an i.i.d. sequence with mean zero and variance σ^2. Then

1.
$$T^{-1/2}\sum_{t=1}^{T} u_t \xrightarrow{L} \sigma W(1)$$

2.
$$T^{-1}\sum_{t=1}^{T} \xi_{t-1}u_t \xrightarrow{L} \frac{1}{2}\sigma[W(1)^2 - 1]$$

3.
$$T^{-3/2}\sum_{t=1}^{T} tu_t \xrightarrow{L} \sigma W(1) - \sigma\int_0^1 W(r)dr$$

4.
$$T^{-3/2}\sum_{t=1}^{T} \xi_{t-1} \xrightarrow{L} \sigma\int_0^1 W(r)dr$$

5.
$$T^{-2}\sum_{t=1}^{T}\xi_{t-1}^{2}\xrightarrow{L}\sigma\int_0^1[W(r)]^2dr$$

6.
$$T^{-5/2}\sum_{t=1}^{T}t\xi_{t-1}\xrightarrow{L}\sigma\int_0^1 rW(r)dr$$

7.
$$T^{-3}\sum_{t=1}^{T}t\xi_{t-1}^{2}\xrightarrow{L}\sigma\int_0^1 r[W(r)]^2dr$$

8.
$$T^{-(\nu+1)}\sum_{t=1}^{T}t^{\nu}\xrightarrow{L}\frac{1}{\nu+1}\qquad \nu=0,1,\ldots$$

As Phillips (1987) [63] mentions we attain the same results when the initial value ξ_0, is any fixed value or drawn from a specified distribution and also when the ξ_0 was equal to zero.

All the results from the Proposition 1 are written in terms of the same functional standard Brownian motion, denoted by $W(r)$. In this way all the results are correlated. The Proposition 1 can also be useful in calculating the asymptotic distributions of some basic stochastic processes with unit roots.

Case 1. No Constant Term or Time Trend in the Regression

Consider the following AR(1) model,

$$y_t=\phi_1 y_{t-1}+u_t$$

where u_t, is i.i.d. with mean zero and variance σ^2. We are interested in the properties of the OLS estimate

$$\hat{\phi}_{1T}=\frac{\sum_{t=1}^{T}y_{t-1}y_t}{\sum_{t=1}^{T}y_{t-1}^2}$$

when the true value of ϕ_1 is unity. From previous equation, the deviation of the OLS estimate from the true value is characterized by

$$T(\hat{\phi}_{1T}-1)=\frac{T^{-1}\sum_{t=1}^{T}y_{t-1}u_t}{T^{-2}\sum_{t=1}^{T}y_{t-1}^2} \tag{4.6}$$

if the true value of $\phi_1=1$, then

$$y_t=y_0+u_1+u_2+\cdots+u_t$$

Since the initial value y_0 does not affect the asymptotic distribution, we can rewrite the result (2) of the Proposition 1 as

$$T^{-1}\sum_{t=1}^{T} y_{t-1}u_t \xrightarrow{L} \frac{1}{2}\sigma[[W(1)]^2 - 1], \tag{4.7}$$

while from result (5) we have

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{L} \sigma \int_0^1 [W(r)]^2 dr \tag{4.8}$$

Since 4.6 is a continuous function of 4.7 and 4.8, subsequently from result (3) of Proposition 1 under the null hypothesis that $\phi_1 = 1$, the OLS estimator, $\hat{\phi}_{1T}$, is characterized by

$$T(\hat{\phi}_{1T} - 1) \xrightarrow{L} \frac{\frac{1}{2}([W(1)]^2 - 1)}{\int_0^1 [W(r)]^2 dr} \tag{4.9}$$

Note that $[W(1)]^2$ is a $\mathcal{X}_1^2$ variable. The probability that a $\mathcal{X}_1^2$ variable is less than unity is 0.68, and since the denominator of 4.9 must be positive, the probability that $\hat{\phi}_{1T} - 1$ is negative approaches 0.68 as T becomes large. In other words, in two-thirds of the samples generated by a random walk, the estimate $\hat{\phi}_{1T}$ will be less than the true value of unity. Moreover, in those samples for which $[W(1)]^2$ is large, the denominator of 4.9 will be large as well. The result is that the limiting distribution of $T(\hat{\phi}_{1T} - 1)$ is skewed to the left.
Recall that in the stationary case when $|\phi_1| < 1$, the estimate $\hat{\phi}_{1T}$ is downward biased in small samples. Even so, in the stationary case the limiting distribution of $\sqrt{T}(\hat{\phi}_{1T} - \phi_1)$ is symmetric around zero. By contrast, when the true value of ϕ_1 is unity, even the limiting distribution of $T(\hat{\phi}_{1T} - 1)$ is asymmetric, with negative values twice as likely as positive values.

Case 2. Constant Term but No Time Trend Included in the Regression. Hence, The True Process Is a Random Walk

Considering an AR(1) model involving a constant term of the following form

$$y_t = \phi_0 + \phi_1 y_{t-1} + u_t \tag{4.10}$$

with u_t, i.i.d. with mean zero and variance σ^2.
We will investigate the asymptotic properties of the OLS estimators

$$\begin{bmatrix} \hat{\phi}_{0T} \\ \hat{\phi}_{1T} \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} y_{t-1}y_t \end{bmatrix}, \tag{4.11}$$

under the null hypothesis that $\phi_0 = 0$ and $\phi_1 = 1$. Note that the deviation of an estimated OLS coefficient vector ($\boldsymbol{b_T}$) from the true value ($\boldsymbol{\beta}$) given by

$$\mathbf{b}_T - \boldsymbol{\beta} = \left[\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right]^{-1} \left[\sum_{t=1}^{T} \mathbf{x}_t \mathbf{u}_t\right] \tag{4.12}$$

Equivalently,

$$\begin{bmatrix} \hat{\phi}_{0T} \\ \hat{\phi}_{1T} - 1 \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} u_t \\ \sum_{t=1}^{T} y_{t-1} u_t \end{bmatrix}, \tag{4.13}$$

Using the result (4) of the Proposition 1 we obtain the following:

$$T^{-3/2} \sum_{t=1}^{T} y_{t-1} \xrightarrow{L} \sigma \int_0^1 W(r)dr \tag{4.14}$$

Case 3. Constant Term but No Time Trend Included in the Regression.

Considering an AR(1) model with the true process being a random walk with drift:

$$y_t = \phi_0 + y_{t-1} + u_t \tag{4.15}$$

where the true value of ϕ_0 is not zero.

This change in the assumption of the true process form, will have a huge impact on the asymptotic distribution, note that

$$y_t = \phi_0 + y_{t-1} + u_t \Rightarrow \tag{4.16}$$

$$y_t = y_0 + \phi_0 t + \sum_{t=1}^{t} u_t \tag{4.17}$$

Denote that $\xi_t = \sum_{t=1}^{t} u_t$ with $\xi_0 = 0$.

Consider the sum

$$\sum_{t=1}^{T} y_{t-1} = \sum_{t=1}^{T} [y_0 + \phi_0(t-1) + \xi_{t-1}] \tag{4.18}$$

where the term $\sum_{t=1}^{T} y_{t-1}$ in 4.18 can be rewritten as Ty_0, and if this is divided by T, the result will be a fixed value. The second term, $\sum \phi_0(t-1)$, using the result (8) of the Proposition 1 converges to

$$T^{-2} \sum_{t=1}^{T} \phi_0(t-1) \rightarrow \frac{\phi_0}{2}$$

The third term converges:

$$T^{-3/2} \sum_{t=1}^{T} \xi_{t-1} \xrightarrow{L} \sigma \int_0^1 W(r)dr,$$

as a result of Proposition 1 (4).

Hamilton (1994) [33] in his book (p.495-497) calculates in detail the following result:

$$\begin{pmatrix} T^{1/2}(\hat{\phi}_{0T} - \phi_0) \\ T^{3/2}(\hat{\phi}_{1T} - \phi_1) \end{pmatrix} \xrightarrow{L} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \frac{a}{2} \\ \frac{a}{2} & \frac{a^2}{3} \end{pmatrix}^{-1} \sigma^2 \begin{pmatrix} 1 & \frac{a}{2} \\ \frac{a}{2} & \frac{a^2}{3} \end{pmatrix} \begin{pmatrix} 1 & \frac{a}{2} \\ \frac{a}{2} & \frac{a^2}{3} \end{pmatrix}^{-1} \right) = N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \frac{a}{2} \\ \frac{a}{2} & \frac{a^2}{3} \end{pmatrix}^{-1} \right)$$

Case 4. Constant Term and Time Trend Included in the Regression.

Considering an AR(1) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + \delta t + u_t \tag{4.19}$$

with ϕ_0 being the constant term and δt being the time trend. Also, u_t is i.i.d. with mean zero and variance σ^2.

If ϕ_0 is nonzero, y_{t-1} would be asymptotically equivalent to a time trend. Since a time trend is already included as a separate variable in the regression, this would make the explanatory variables collinear in large samples.

Note that the regression model of 4.15 can equivalently be written as

$$\begin{aligned} y_t &= (1-\phi_1)\phi_0 + \phi_1(y_{t-1} - \phi_0(t-1)) + (\delta + \phi_1\phi_0)t + u_t \\ &= \phi_0' + \phi_1'\xi_{t-1} + \delta t + u_t \end{aligned} \tag{4.20}$$

where $\phi_0' = (1-\phi_1)\phi_0$, $\phi_1' = \phi_1$, $\delta' = (\delta + \phi_1\phi_0)$, and $\xi_t = y_t - \phi_0 t$. Moreover, under the null hypothesis that $\phi_1 = 1$ and $\delta = 0$,

$$\xi_t = y_0 + u_1 + u_2 + \cdots + u_t$$

that is, ξ_t, is the random walk described in Proposition 1. Consider, a hypothetical regression of y_t, on a constant, ξ_{t-1}, and a time trend, producing the OLS estimates

$$\begin{bmatrix} \hat{\phi}_{0T}' \\ \hat{\phi}_{1T}' \\ \hat{\delta}_T' \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^T \xi_{t-1} & \sum_{t=1}^T t \\ \sum_{t=1}^T \xi_{t-1} & \sum_{t=1}^T \xi_{t-1}^2 & \sum_{t=1}^T \xi_{t-1} t \\ \sum_{t=1}^T t & \sum_{t=1}^T t\xi_{t-1} & \sum_{t=1}^T t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^T y_t \\ \sum_{t=1}^T \xi_{t-1} y_t \\ \sum_{t=1}^T t y_t \end{bmatrix} \tag{4.21}$$

The maintained hypothesis is that $\phi_1 = 1$, and $\delta = 0$, which in the transformed system would mean $\phi_0' = 0$, $\phi_1' = 1$, and $\delta' = \phi_0$. The deviations of the OLS estimates from these true values are given by

$$\begin{bmatrix} \hat{\phi}_{0T}' \\ \hat{\phi}_{1T}' - 1 \\ \hat{\delta}_T' - \phi_0 \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^T \xi_{t-1} & \sum_{t=1}^T t \\ \sum_{t=1}^T \xi_{t-1} & \sum_{t=1}^T \xi_{t-1}^2 & \sum_{t=1}^T \xi_{t-1} t \\ \sum_{t=1}^T t & \sum_{t=1}^T t\xi_{t-1} & \sum_{t=1}^T t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^T u_t \\ \sum_{t=1}^T \xi_{t-1} u_t \\ \sum_{t=1}^T t u_t \end{bmatrix} \tag{4.22}$$

4.2.2 Asymptotic Results for Unit Root Processes with General Serial Correlation

This section generalizes Proposition 1 to allow for serial correlation.

Proposition 2. *Let*

$$u_t = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_t - j \tag{4.23}$$

where

$$E(\epsilon_t) = 0$$

$$E(\epsilon_t \epsilon_\tau) = \begin{cases} \sigma^2 & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases}$$

$$\sum_{j=0}^{\infty} j|\psi_j| < \infty \tag{4.24}$$

Then

$$u_1 + u_2 + \cdots + u_t = \psi(1)(\epsilon_1 + \epsilon_2 + \cdots + \epsilon_t) + \eta_t - \eta_0, \tag{4.25}$$

where $\psi(1) = \sum_{j=0}^{\infty} \psi$, $\eta_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j}$, $a_j = -(\psi_{j+1} + \psi_{j+2} + \dots)$ *and* $\sum_{j=0}^{\infty} |a_j| < \infty$

Notice that if y_0 is an $I(1)$ process y_t whose first difference is given by u_t, or

$$\Delta y_t = u_t$$

then

$$y_t = u_1 + u_2 + \cdots + u_t + y_0 = \psi(1)(\epsilon_1 + \epsilon_2 + \cdots + \epsilon_t) + \eta_t + \eta_0 + y_0$$

Proposition 2 thus states that any $I(1)$ process whose first difference satisfies 4.23 and 4.24 can be written as the sum of a random walk, $\psi(1)(\epsilon_1 + \epsilon_2 + \cdots + \epsilon_t)$, initial conditions $(y_0 - \eta_0)$, and a stationary process (η_t). This observation was first made by Beveridge and Nelson (1981) [5], and 4.25 is sometimes referred to as the Beveridge-Nelson decomposition.

Notice that η_t, is a stationary process. An important implication of this is that if 4.25 is divided by $\sqrt{t}$, only the first term $\frac{1}{\sqrt{t}}\psi(1)(\epsilon_1 + \epsilon_2 + \cdots + \epsilon_t)$ should matter for the distribution of $\frac{1}{\sqrt{t}}\psi(1)(u_1 + u_2 + \cdots + u_t)$ as $t \to \infty$.

Proposition 3. *Let* $u_t = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ *where* $\sum_{j=0}^{\infty} j|\psi| < \infty$ *and* $\{\epsilon_t\}$ *is an i.i.d. sequence with mean zero, variance* σ^2 *and finite fourth moment. Define*

$$\gamma_j = E(u_t u_{t-j}) = \sigma \sum_{s=0}^{\infty} \psi_s \psi_{s+j} \quad \textit{for } j = 0, 1, 2, \cdots \tag{4.26}$$

$$\lambda = \sigma \sum_{j=0}^{\infty} \psi_j = \sigma\psi(1)$$

$$\xi_t = u_1 + u_2 + \cdots + u_t \textit{for } t = 1, 2, \cdots, T \tag{4.27}$$

with $\xi_0 = 0$ *Then*

1. $T^{-1/2} \sum_{t=1}^{T} u_t \xrightarrow{L} \lambda W(1)$
2. $T^{-1/2} \sum_{t=1}^{T} u_{t-j}\epsilon_t \xrightarrow{L} N(0, \sigma^2\gamma_0) \quad \textit{for } j = 1, 2, \cdots$
3. $T^{-1} \sum_{t=1}^{T} u_t u_{t-j} \xrightarrow{p} \gamma_j \quad \textit{for } j = 0, 1, 2, \cdots$
4. $T^{-1} \sum_{t=1}^{T} \xi_{t-1}\epsilon_t \xrightarrow{L} \frac{1}{2}\sigma\lambda\{[W(1)^2 - 1]\}$
5. $$T^{-1} \sum_{t=1}^{T} \xi_{t-1} u_{t-j} \xrightarrow{L} \begin{cases} \frac{1}{2}\{\lambda^2[W(1)]^2 - \gamma_0\} & \textit{for } j = 0 \\ \frac{1}{2}\{\lambda^2[W(1)]^2 - \gamma_0\} + \gamma_0 + \gamma_1 + \gamma_2 + \cdots + \gamma_{j-1} & \textit{for } j = 1, 2, \cdots \end{cases}$$
6. $T^{-3/2} \sum_{t=1}^{T} \xi_{t-1} \xrightarrow{L} \lambda \int_0^1 W(r)dr$

7. $T^{-3/2}\sum_{t=1}^{T} tu_{t-j} \xrightarrow{L} \lambda\{W(1) - \int_0^1 W(r)dr\}$ *for* $j = 0, 1, 2, \cdots$

8. $T^{-2}\sum_{t=1}^{T} \xi_{t-1}^2 \xrightarrow{L} \lambda^2 \int_0^1 [W(r)]^2 dr$

9. $T^{-5/2}\sum_{t=1}^{T} t\xi_{t-1} \xrightarrow{L} \lambda \int_0^1 rW(r)dr$

10. $T^{-3}\sum_{t=1}^{T} t\xi_{t-1}^2 \xrightarrow{L} \lambda^2 \int_0^1 r[W(r)]^2 dr$

11. $T^{-(\nu+1)}\sum_{t=1}^{T} t^{\nu} \rightarrow \frac{1}{\nu+1} \qquad \nu = 0, 1, \cdots$

4.3 Cointegration

Assume a pair of time series, x_t and y_t that they tend to "move together". There may be an economic theory involved that forces these pairs to move in a similar way. Of course, there may be some reasons that drive them apart but the economic theory that is involved is stronger. Thus, they will return to their original equilibrium.

Davidson et al. (1978) [20] presented models that in the long-run equilibrium should not drift arbitarily far from each other.

The equilibrium will hold if $y_t \sim I(0)$, and $x_t \sim I(0)$ then $y_t - \beta x_t = u_t \sim I(0)$. But there are cases when the equilibrium relationship holds even if the time series are not stationary.

These cases, where the time series are non-stationary, but their linear combination provides stationary residuals are called cointegrated time series.

Definition 26. *Cointegration [78]*

Let $x_t \sim I(1)$ and $y_t \sim I(1)$ be two time series. They are said to be cointegrated if there exists a parameter β such that

$$u_t = y_t - \beta x_t$$

is a stationary process.

The u_t is referred as the equilibrium error and represents the deviation from the long-run equilibrium. The co-movement of the time series causes their residuals to have a mean-reverting behaviour and get closer to the equilibrium line.

The formal mathematical expression of the equilibrium concept, ends up to cointegration model. In essence a pair of two I(1) times series with common trend, in a way that they cancel each other's trend and their linear combination comes out as stationary process.

First we will need a model. Consider the regression

$$y_t = \beta_0 + \beta_1 x_t + u_t \tag{4.28}$$

In the case where the x_t and y_t are cointegrated the coefficient of x_t, $\hat{\beta}_1$, converges to the true value of rate T. Then, the OLS method provides super-consistent estimator.

If x_t is a random walk and the residuals are serially uncorrelated, we note that $T \cdot (\hat{\beta}_1 - \beta_1)$ is asymptotically distributed to a function of two independent Brownian motions, $\frac{\int_0^1 B_2 dB_1}{\int_0^1 B_2^2 dt}$ with zero mean. It is worth mentioning that the standard t-test is asymptotically normally distributed.

If there is serial correlation on residuals or there is a correlation between the x_t and u_s for some s, $T \cdot (\hat{\beta}_1 - \beta_1)$ asymptotically does not follow a symmetric distribution. Therefore the t-statistic is no longer asymptotically normally distributed.

Up until this point, we dealt with the simple case where the cointegrated time series are integrated of order 1. The attention now turns to a more generic case, where the two cointegrated time series are integrated of order $d > 1$. Let, $x_t \sim I(d)$ and $y_t \sim I(d)$. This implies that d-differences are needed in order for x_t and y_t to be stationary. The x_t and y_t are said to be co-integrated of order $CI(d,p)$ if both of them are integrated of order d and the residuals are integrate of order $d-p$. This means that $x_t \sim I(d)$, $y_t \sim I(d)$ and $y_t - \beta x_t = u_t \sim I(d-p)$. Of course this requires $p \leq d$. It is easy to see that the simple case of $I(1)$ can be presented as a $CI(1,1)$. The integration order of x_t and y_t is $d = 1$, the integration order of the residuals is $p = 1 \Rightarrow d - p = 0$.

4.3.1 Normalization

The cointegration vector $\boldsymbol{\beta}$ in 3.3 is not sui generis since for any nonzero scalar c the linear combination $c\boldsymbol{\beta}'\boldsymbol{Y_t} = \boldsymbol{\beta}^{*\prime}\boldsymbol{Y_t} \sim I(0)$. Hence, some normalization assumption is required to uniquely identify $\boldsymbol{\beta}$. A typical normalization is

$$\boldsymbol{\beta} = (1, -\beta_2, \cdots, -\beta_n)'$$

so that the cointegration relationship may be expressed as

$$\boldsymbol{\beta}'\boldsymbol{Y_t} = y_{1t} - \beta_2 y_{2t} - \cdots - \beta_n y_{nt} \sim I(0)$$

or

$$y_{1t} = \beta_2 y_{2t} + \cdots + \beta_n y_{nt} + u_t \tag{4.29}$$

where $u_t \sim I(0)$. In 4.29, the error term u_t is often referred to as the disequilibrium error or the cointegrating residual. In long-run equilibrium, the disequilibrium error u_t is zero and the long-run equilibrium relationship is

$$y_{1t} = \beta_2 y_{2t} + \cdots + \beta_n y_{nt}$$

Note that one should be very careful when choosing the variable to label as y_{1t} because β_1 must not be zero. One approach that avoids this normalization problem is the full-information maximum likelihood estimate proposed by Johansen (1988, 1991) [40] [41].

4.4 Unit Root test

4.4.1 An Alternative Representation of an AR(p) Process

Consider the following AR(p) model,

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) y_t = \epsilon_t, \tag{4.30}$$

where $\{\epsilon_t\}$ is an i.i.d. sequence with mean zero, variance σ^2, we can rewrite the 4.30

$$[(1-\phi^* L) - (\zeta_1 L + \zeta_2 L^2 + \cdots + \zeta_{p-1} L^{p-1})(1-L)]y_t = \epsilon_t \tag{4.31}$$

or

$$y_t = \phi^* y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + \epsilon_t \tag{4.32}$$

where $\phi^* = \phi_1 + \phi_2 + \cdots + \phi_p$ and $\zeta_j = -[\phi_{j+1} + \phi_{j+2} + \ldots \phi_p]$ for $j = 1, 2, \ldots, p-1$.

From 4.32 one can see that y_{t-1} is $I(1)$, while all of the other regressors $(\Delta y_{t-1}, \ldots, \Delta y_{t-p+1})$ are $I(0)$.

This canonical form proposed by Fuller (1976) [21] requires no specific sample size and it is easier to estimate the parameters by direct OLS estimation.

4.4.2 Summary of Asymptotic Results for an Estimated Autoregression that Includes a Constant Term

The preceding analysis applies to OLS estimation of

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + a + \phi^* y_{t-1} + \epsilon_t$$

under the assumption that the true value of a is zero and the true value of p is 1.

Let ϵ_t be i.i.d. with mean zero, variance σ^2, and the roots of

$$(1 - \zeta_1 u - \zeta_2 u^2 - \cdots - \zeta_{p-1} u^{p-1}) = 0$$

are outside the unit circle. It was seen that the estimates $\hat{\zeta}_1, \hat{\zeta}_2, \ldots, \hat{\zeta}_{p-1}$ converge at rate $\sqrt{T}$ to Gaussian variates, and standard t or F tests for hypotheses about these coefficients have the usual limiting Gaussian or $\mathcal{X}^2$ distributions. The estimates $\hat{a}$ and $\hat{p}$ converge at rates $\sqrt{T}$ and T, respectively, to nonstandard distributions.

When the autoregression includes lagged changes as here, tests for a unit root based on the value of p, t tests, or F tests are described as augmented Dickey-Fuller tests.

Case 1: Estimated regression: $y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + a + \phi^* y_{t-1} + \epsilon_t$
True process: same specification as estimated regression with $\phi^* = 1$
Any t or F test involving $\zeta_1, \zeta_2, \ldots, \zeta_{p-1}$, can be compared with the usual t or F tables for an asymptotically valid test.
Z_{DF} has the same asymptotic distribution as the variable described under the heading Case 1 in Table B.
OLS t test of $\phi^* = 1$ has the same asymptotic distribution as the variable described under Case 1 in Table B.

Case 2: Estimated regression: $y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + a + \phi^* y_{t-1} + \epsilon_t$
True process: same specification as estimated regression with $a = 0$ and $\phi^* = 1$
Any t or F test involving $\zeta_1, \zeta_2, \ldots, \zeta_{p-1}$, can be compared with the usual t or F tables for an asymptotically valid test.
Z_{DF} has the same asymptotic distribution as the variable described under Case 2 in Table B.

OLS t test of $\phi^* = 1$ has the same asymptotic distribution as the variable described under Case 2 in Table B.
OLS F test of joint hypothesis that $a = 0$ and $\phi^* = 1$ has the same asymptotic distribution as the variable described under Case 2 in Table B.

Case 3: Estimated regression: $y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + a + \phi^* y_{t-1} + \epsilon_t$
True process: same specification as estimated regression with $a \neq 0$ and $\rho = 1$
$\hat{\phi}^*_T$ converges at rate $T^{3/2}$ to a Gaussian variable; all other estimated coefficients converge at rate $T^{1/2}$ to Gaussian variables.
Any t or F test involving any coefficients from the regression can be compared with the usual t or F tables for an asymptotically valid test.

Case 4: Estimated regression: $y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + a + \phi^* y_{t-1} + \epsilon_t$
True process: same specification as estimated regression with a any value, $\phi^* = 1$,and $\delta = 0$
Any t or F test involving $\zeta_1, \zeta_2, \ldots, \zeta_{p-1}$, can be compared with the usual t or F tables for an asymptotically valid test.
Z_{DF} has the same asymptotic distribution as the variable described under Case 4 in Table B.
OLS t test of $\phi^* = 1$ has the same asymptotic distribution as the variable described under Case 4 in Table B.
OLS F test of joint hypothesis that $\phi^* = 1$ and $\delta = 0$ has the same asymptotic distribution as the variable described under Case 4 in Table B.

4.4.3 Unit Root AR(p) Processes with p Unknown

Many proposals have been made for how to handle the cases where the p is unknown but finite, for example in an ARIMA$(p, 1, 0)$ model.

Instinctively the first approach is to estimate $y_t = \boldsymbol{\beta} \boldsymbol{x}'_t$ with p taken to be some predetermine upper bound $\hat{p}$. In order to determine the order of AR model we will use a backward stepwise procedure starting with the hypothesis that only the $zeta_{\hat{p}-1} = 0$. Obviously, we can use the OLS t-test. If the null hypothesis is not rejected we will continue with F-test to examine if $\zeta_{\hat{p}-1} = 0$ and $\zeta_{\hat{p}-2} = 0$ simultaneously. We continue this procedure until the null hypothesis $\zeta_{\hat{p}-1} = 0$, $\zeta_{\hat{p}-2} = 0$, $\ldots$, $\zeta_{\hat{p}-\ell} = 0$ is rejected. The recommended regression is then

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-\ell} \Delta y_{t-(p+\ell)} + a + \phi^* y_{t-1} + \delta t$$

Of course across the literature there are other approaches to estimate p, like Hall (1991) [31] and Said and Dickey (1984) [70]

4.4.4 Dickey - Fuller

A more systematic approach for testing the presence of a unit root of the autoregressive polynomial in order to decide whether or not a time series should be differentiated is the approach that was pioneered by Dickey and Fuller (1979) [21].

To illustrate the important statistical issues associated with autoregressive unit root tests, consider the simple AR(1) model

$$y_t = \phi y_{t-1} + \epsilon_t \quad \text{where } \epsilon_t \sim WN(0, \sigma^2)$$

The hypotheses of interest are

$$H_0 : \phi = 1(\text{unit root in } \phi(z) = 0) \to y_t \sim I(1)$$
$$H_1 : |\phi| < 1 \to y_t \sim I(0)$$

The test statistic is

$$t_{\phi=1} = \frac{\hat{\phi} - 1}{SE(\hat{\phi})}$$

where $\hat{\phi}$ is the least square estimate and $SE(\hat{\phi})$ is the standard error estimate.

The limiting distribution of $t_{\phi=1}$ is called the Dickey-Fuller (DF) distribution and does not have a closed form representation. Consequently, quantiles of the distribution must be computed by numerical approximation or by simulation.

The unit root tests described above are valid if the time series y_t is well characterized by an $AR(1)$ with white noise errors. Many financial time series, however, have a more complicated dynamic structure than when captured by a simple $AR(1)$ model. Said and Dickey (1984) [70] augment the basic autoregressive unit root test to accommodate general $ARMA(p, q)$ models with unknown orders and their test is referred to as the augmented Dickey Fuller (ADF) test. The ADF test tests the null hypothesis that a time series y_t is $I(1)$ against the alternative that it is $I(0)$, assuming that the dynamics in the data have an ARMA structure. The ADF test is based on estimating the test regression

$$\nabla y_t = \gamma y_{t-1} + \sum_{j=1}^{p-1} \psi_j \nabla y_{t-j} + w_t$$

where $\gamma = \sum_{j=1}^{p} \phi_j - 1$ and $\psi_j = -\sum_{j=1}^{p} \phi_i$ for $j = 2, \ldots p$.

Remark 5. *One will have to deal with the problem of choosing the optimal lag length p, in order to implement the ADF test. If p is too small then the remaining serial correlation in the errors will bias the test. If p is too large the degrees of freedom will be too small, and so the test would not be powerful enough.*

A useful rule of thumb for determining $p_{\max}$, suggested by Schwert (1989) [71], is

$$p_{\max} = \left[12 \cdot \left(\frac{T}{100} \right)^{1/4} \right] \tag{4.33}$$

4.4.5 Summary of Dickey-Fuller Tests in the absence of Serial Correlation

We have already examined the asymptotic properties of the OLS in various cases of the presence or not of constant trend and time trend in Cases 1-4.

In order to construct the null hypothesis for unit root testing, one should choose the appropriate Case. In case where the choice is not clear one should find a more generic way to construct the null hypothesis. For example it is suggested to use Case 4 for a series with trend, while Case 2 is more appropriate when there is no significant trend.

Case 1: Estimated regression: $y_t = \phi_1 y_{t-1} + u_t$
True process: $y_t = y_{t-1} + u_t$ $u_t \sim$ iid $N(0, \sigma^2)$
$T(\hat{\phi}_{1T} - 1)$ has the distribution described under the heading Case 1 in Table B.
$(\hat{\phi}_{1T} - 1)/\hat{\sigma}_{\hat{\phi}_{1T}}$ has the distribution described under Case 1 in Table B.

Case 2: Estimated regression: $y_t = \phi_0 + \phi_1 y_{t-1} + u_t$
True process: $y_t = y_{t-1} + u_t$ $u_t \sim$ iid $N(0, \sigma^2)$
$T(\hat{\phi}_{1T} - 1)$ has the distribution described under the Case 2 in Table B.
$(\hat{\phi}_{1T} - 1)/\hat{\sigma}_{\hat{\phi}_{1T}}$ has the distribution described under Case 2 in Table B.
OLS F test of joint hypothesis that $\phi_0 = 0$ and $\phi_1 = 1$ has the distribution described under Case 2 in Table B.

Case 3: Estimated regression: $y_t = \phi_0 + \phi_1 y_{t-1} + u_t$
True process: $y_t = \phi_0 + y_{t-1} + u_t$, $\phi_0 \neq 0$ and $u_t \sim$ iid $N(0, \sigma^2)$
$(\hat{\phi}_{1T} - 1)/\hat{\sigma}_{\hat{\phi}_{1T}} \xrightarrow{L} N(0, 1)$

Case 4: Estimated regression: $y_t = \phi_0 + \phi_1 y_{t-1} + \delta t + u_t$
True process: $y_t = \phi_0 + y_{t-1} + u_t$ $u_t \sim$ iid $N(0, \sigma^2)$
$T(\hat{\phi}_{1T} - 1)$ has the distribution described under the Case 4 in Table B.
$(\hat{\phi}_{1T} - 1)/\hat{\sigma}_{\hat{\phi}_{1T}}$ has the distribution described under Case 4 in Table B.
OLS F test of joint hypothesis that $\phi_1 = 1$ and $\delta = 0$ has the distribution described under Case 4 in Table B.

4.4.6 Phillips-Perron

The augmented DF is motivated by the need to generate iid errors. An alternative strategy for allowing errors that are not iid is that of Phillips and Perron (1988) [67], known as the Phillips–Perron (PP) unit root test.

A great advantage of Philips-Perron test is that it is non-parametric. The Phillips-Perron (PP) unit root tests differ from the ADF tests mainly in how they deal with serial correlation (does not require to select the level) and heteroskedasticity in the errors (HAC type corrections).

In particular, where the ADF tests use a parametric autoregression to approximate the ARMA structure of the errors in the test regression, the PP tests ignore any serial correlation in the test regression. The test regression for the PP tests is

$$\Delta y_t = \boldsymbol{z' D_t} + \beta y_{t-1} + u_t$$

where $\boldsymbol{D_t}$ is a vector of deterministic terms (constant, trend etc.), u_t is $I(0)$ and may be heteroskedastic. The PP tests correct for any serial correlation and heteroskedasticity in the errors u_t of the test regression by directly modifying the test statistics $t_{\beta=0}$ and $T\hat{\beta}$. These modified statistics, denoted Z_t and Z_β, are given by

$$Z_t = \left(\frac{\hat{\sigma}^2}{\hat{\lambda}^2}\right)^{1/2} \cdot t_{\beta=0} - \frac{1}{2}\left(\frac{\hat{\lambda}^2 - \hat{\sigma}^2}{\hat{\lambda}^2}\right) \cdot \left(\frac{T \cdot SE(\hat{\beta})}{\hat{\sigma}^2}\right)$$

$$Z_\pi = T\hat{\beta} - \frac{1}{2}\frac{T^2 \cdot SE(\hat{\beta})}{\hat{\sigma}^2}\left(\hat{\lambda}^2 - \hat{\sigma}^2\right)$$

Phillips-Perron test does not add lags, but uses Newey-West HAC robust standard errors to calculate test statistic.

The terms $\hat{\sigma}^2$ and $\hat{\lambda}^2$ are consistent estimates of the variance parameters

$$\sigma^2 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E[u_t^2]$$

$$\lambda^2 = \lim_{T\to\infty} \sum_{t=1}^{T} E[T^{-1} S_T^2]$$

where $S_T = \sum_{t=1}^{T} u_t$.

The main disadvantage of the PP test is that it is based on asymptotic theory. And it also shares disadvantages of ADF tests: sensitivity to structural breaks, poor small sample power too often resulting in unit root conclusions.

Remark 6. *The consensus ordains that ADF test is more preferable in most cases because PP-test needs large sample of data. However, in our times large data is not a problem. It is, also, non- parametric i.e. it does not assume any distribution. In addition, it uses the HAC estimators for the variance and so it is robust. Lastly, it does not assume any model form such as the ARMA model.*

4.4.7 Summary of Phillips-Perron Test for Unit Roots

Case 1: Estimated regression: $y_t = \beta y_{t-1} + u_t$
True process: $y_t = y_{t-1} + u_t$
Z_β has the same asymptotic distribution as the variable described under the heading Case 1 in Table B.
Z_t has the same asymptotic distribution as the variable described under Case 1 in Table B.

Case 2: Estimated regression: $y_t = a + \beta y_{t-1} + u_t$
True process: $y_t = y_{t-1} + u_t$
Z_β has the same asymptotic distribution as the variable described under the heading Case 2 in Table B.
Z_t has the same asymptotic distribution as the variable described under Case 2 in Table B.

Case 4: Estimated regression: $y_t = a + \beta y_{t-1} + \delta t + u_t$
True process: $y_t = a + y_{t-1} + u_t$ for any a
Z_β has the same asymptotic distribution as the variable described under the heading Case 4 in Table B.
Z_t has the same asymptotic distribution as the variable described under Case 4 in Table B.

4.4.8 Zivot

Structural Break

The assumption of stationarity implies that the parameters of regression models i.e. the mean and the variance are constant over time.

However, in econometrics, we define the structural break as an unexpected change over time in these parameters.

This, can lead to forecasting errors and unreliability of the model in general.

There are several types of structural breaks with some of them being when there is a known number of breaks in mean with unknown break points. Another type is when there is an unknown number of breaks in mean with unknown break points and also when there are breaks in variance.

The Chow (1960) [16] test, was one of the first tests which set the foundation for structural break testing. Its theory is hinged on that if parameters are constant then out-of-sample forecasts should be unbiased.

$$H_0 : \beta_1 = \beta_2 \quad \text{meaning there is no structural break}$$
$$H_a : \beta_1 \neq \beta_2 \quad \text{meaning there is a known structural break at time} T_b$$

The advantages of the Chow test are that it is easy to implement and that the F-statistic has a standard distribution.

A disadvantage of the Chow test is that the break point must be predetermined earlier in order to implement the test.

We introduce a simple example for better understanding the structural breaks. We simulate data from Normal distribution and divide them into two subsets.

The first subset was generated from $N(0,1)$ and the second was generated from $N(2,2)$.

That way, we have a clear structural break for presentational purposes.

We calculate the F-test, which essentially is a Chow test, for all possible break points.

$$F(T_1) = \frac{(SSE - SSE(T_1))/k}{SSE(T_1)(n-m)}$$

where,
SSE is the Sum of Squares Error,
$\underset{\sim}{\sigma}^2 = \frac{1}{n-k}(\underset{\sim}{e}'\underset{\sim}{e})$,
$\underset{\sim}{e} = Y - X\underset{\sim}{\beta}$
k = number of regressors before SB ($t \leq T_1$)
and m = all parameters.

We practice the supF-test and extract the maximum F-score from all the possible break points.

```
set.seed(300)
x1 <- rnorm(1000)
x2 <- rnorm(1000, mean = 2, sd = 2^0.5)
X <- c(x1, x2)
ourts1 <- ts(X)
library(strucchange)
```

```
test2 <- Fstats(ourts1~1)
 #Gets a sequence of fstatistics for all possible break points
ourts1.fs <- test2$Fstats
 #These are the fstats
bp.ourts1 <- breakpoints(ourts1~1)
#Gets the breakpoint based on the F-stats
plot(ourts1)
 #plots the series myts1
lines(bp.ourts1)
#plots the break date implied by the sup F test
bd.ourts1 <- breakdates(bp.ourts1)
sctest(test2)
 #Obtains a p-value for the implied breakpoint
```

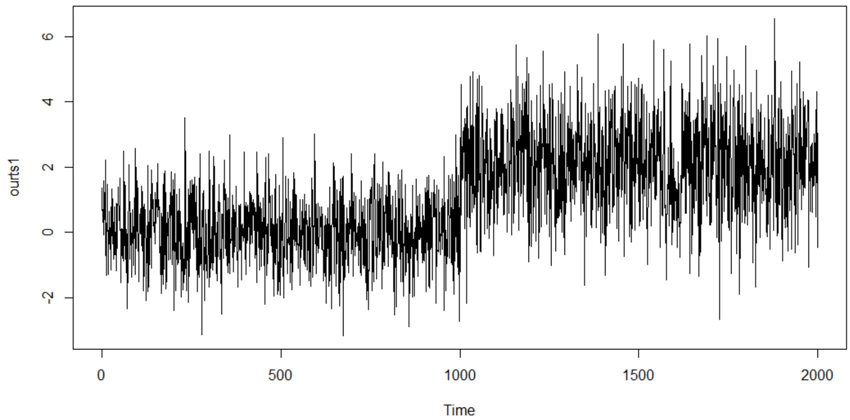

Figure 4.1: Original data

The graph above presents the data from both subsets and we clearly see the relocation of the mean and the variance that there is due to the way that the model was built.

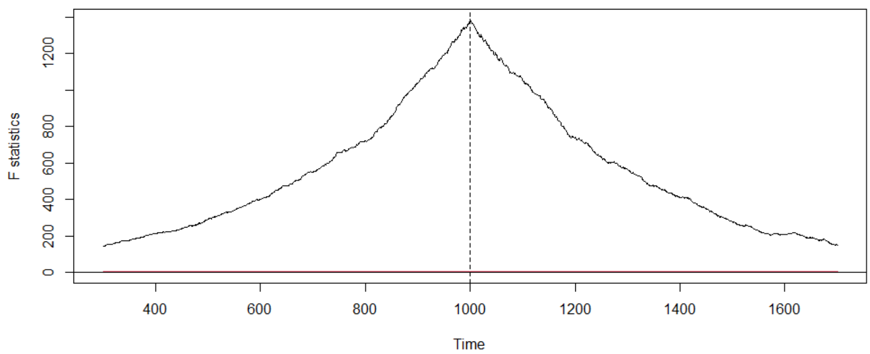

Figure 4.2: F- score plot

In the above figure, we have illustrated the F-test values and as it was expected the maximum F- score appears exactly where the structural break is and it is also the point that the supF test indicates as the change point.

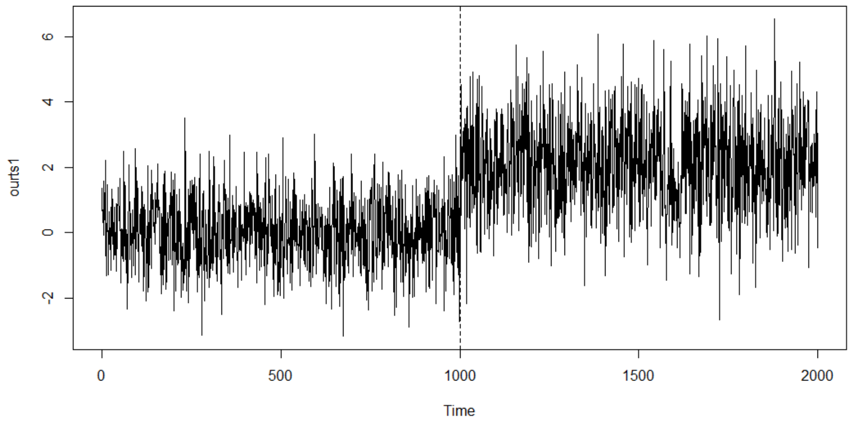

Figure 4.3: Structural break simulation

Zivot and Andrews (1992) [50] endogenous structural break test is a sequential ADF test which utilizes the full sample and uses a different dummy variable for each possible break date. The break date is selected where the t-statistic from the ADF test of unit root is at a minimum (most negative). Consequently a break date will be chosen where the evidence is the least favourable for the unit root null. The critical values in Zivot and Andrews (1992) [50] are different to the critical values in Perron (1989) [59]. The difference is due to the fact that the selecting of the time of the break is treated as the outcome of an estimation procedure, rather than predetermined exogenously. Zivot and Andrews' test provides more evidence for unit roots than Perron's test.

An extension of the Zivot and Andrews' test is the Lumsdaine and Papell's test, which allows two structural breaks under the alternative hypothesis of the unit root test and additionally allows for breaks in level and trend.

The derivation of critical values on ZA and Lumsdaine and Papell (1998) [50] assumes no breaks under the null hypothesis. This assumption may lead to conclude incorrectly (spuriously) reject H_0 (unit root) when, in fact, the series is difference-stationary with breaks.

4.4.9 Stationarity test KPSS

All the unit root tests we have encountered by now have a null hypothesis of non-stationarity. Kwiatkowski et al. (1992) [47] proposed a unit root test which has a null hypothesis of stationarity, this test is widely known as KPSS test. They consider the following model

$$y_t = z' D_t + y_{2t} + u_t \tag{4.34}$$

$$y_{2t} = y_{2,t-1} + \epsilon_t, \quad \epsilon_t \sim WN(0, \sigma_\epsilon^2) \tag{4.35}$$

where D_t contains deterministic components (constant or constant plus time trend), u_t is $I(0)$ and may be heteroskedastic, y_{2t} is a random walk. The null hypothesis of stationarity is constructed as $H_0 : \sigma_\epsilon^2 = 0$, that indicates that y_{2t} is a constant. The KPSS test statistic is given by

$$KPSS = \frac{T^{-2} \sum_{t=1}^{T} \hat{S}_t^2}{\hat{\lambda}^2} \tag{4.36}$$

where $\hat{S}_t^2 = \sum_{j=1}^{t} \hat{u}_j$, $\hat{u}_j$ is the residual of a regression of y_t on D_t and $\hat{\lambda}^2$ is a consistent estimate of the long-run variance of u_t using $\hat{u}_t$.

The stationary test is a one-sided right-tailed test so that one rejects the null of stationarity at the $100 \cdot a\%$ level if the KPSS test statistic 4.36 is greater than the $100 \cdot (1-a)\%$ quantile from the appropriate asymptotic distribution, which can be found in Zivot (2007) [89].

Chapter 5

Application

We tried to find asset in order to make our study. Our first approach was the stocks where we check their correlation, unfortunately we could not find any strong correlation relationship. Our second approach was to use futures, where we found some, but we arbitrarily choose Nasdaq and Dow Jones Index to make our study.

5.1 Nasdaq and Dow Jones Index

5.1.1 Prices

For the application we are using the Nasdaq 100 and the Dow Jones Index from 01/01/2005 to 01/01/2022. We collected our data from yahoo finance. The Nasdaq 100 is an index which involves the 100 largest non-financial companies, consisting of industrial, technology, telecommunication etc. Dow Jones Index is an index containing 30 prominent companies listed on stock exchanges in the United States.

In order to check whether our approach is consistent, and not just the result of over-fitting, we will separate our data into two samples. The first sample, called in-sample period, is used for constructing the model and is from 01/01/2005 to 31/12/2019. The other one, called out-of-sample period, is used for testing the predictive capability of the model and is from 01/01/2020 to 01/01/2022.

Figure 5.1: NQ time series

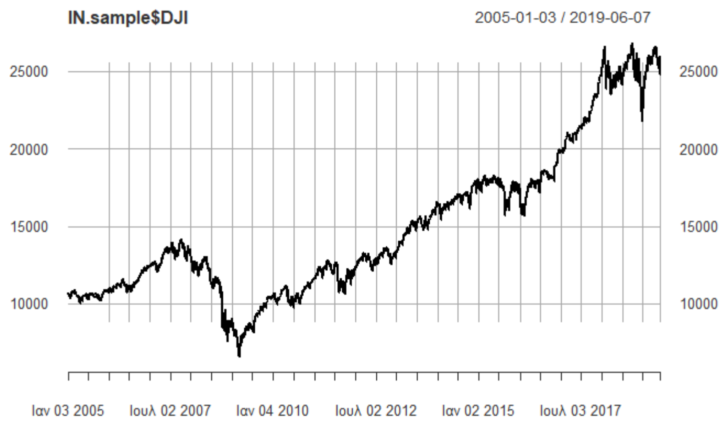

Figure 5.2: DJI time series

The graphs for the NQ prices and the Dow Jones Index (DJI) prices move quite similarly while both show an upward trend, especially after 2009. Also, there is a significance drop around 2008, which is expected since the worldwide financial crisis happened in that period.

	1pct	5pct	10pct
τ_3	-3.96	-3.41	-3.12
ϕ_2	6.09	4.68	4.03
ϕ_3	8.27	6.25	5.34
	τ_3	ϕ_2	ϕ_3
statistic	-1.6185	3.8915	2.9163

Table 5.1: ADF test for NQ prices

In order to check if the time series of the prices are stationary we chose the Augmented Dickey-Fuller (ADF). The NQ τ_3-statistic is $-1.61 > -3.41$ for the 5% significance level.

This indicates that there is unit root and so the time series of the prices is not stationary. The $\phi_2 = 3.89 < 4.68$ and the $\phi_3 = 2.91 < 6.25$ indicate that there is no drift and no trend.

Phillips-Perron test

H_0 : Existence of Unit Root		H_a : Stationary
Dickey Fuller Z_a	Truncation lag parameter	p- value
-4.8799	9	0.8376

Table 5.2: PP test for NQ prices

Another test that we use is the Phillips-Perron test which has a null hypothesis of the existence of unit root. The p-value of the test, for the prices of NQ, is 0.837, which is greater than the 5% significance level, and indicates the existence of unit root.

	1pct	5pct	10pct
τ_3	-3.96	-3.41	-3.12
ϕ_2	6.09	4.68	4.03
ϕ_3	8.27	6.25	5.34
	τ_3	ϕ_2	ϕ_3
statistic	-1.6565	2.6612	2.3283

Table 5.3: ADF test for DJI prices

The DJI τ_3-statistic is $-1.65 > -3.41$ for the 5% significance level. This indicates that there is unit root and so the time series of the prices is not stationary. The $\phi_2 = 2.66 < 4.68$ and the $\phi_3 = 2.32 < 6.25$ indicate that there is no drift and no trend.

Phillips-Perron test

H_0 : Existence of Unit Root		H_a : Stationary
Dickey Fuller Z_a	Truncation lag parameter	p- value
-5.3214	9	0.8129

Table 5.4: PP test for DJI prices

From the Phillips-Perron test which has a null hypothesis of the existence of unit root, we have the p-value of the test, for the prices of DJI, is 0.813, which is greater than the 5% significance level, and indicates the existence of unit root.

Regression

We construct a simple linear regression model with the NQ being the dependent variable and the DJI being the independent variable. For the regression we use the prices of the two assets.

The regression has the following form:

$$NQ_t = \beta_0 + \beta_1 DJI_t + \epsilon_t$$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	$-2.208 \cdot 10^3$	16.70	-132.2	$2 \cdot 10^{-16}$
DJI	0.3653	$1.063 \cdot 10^{-3}$	343.8	$2 \cdot 10^{-16}$
R^2	0.9703	Adj R^2	0.9702	
F-statistic:	$1.182 \cdot 10^5$	p-value	$2.2 \cdot 10^{-16}$	

Table 5.5: Summary of the regression

The R^2 indicates that the 97.03% can be explained from the regression. F-test's p-value suggests that the coefficients are not simultaneously zero, so they are all statistically significant.

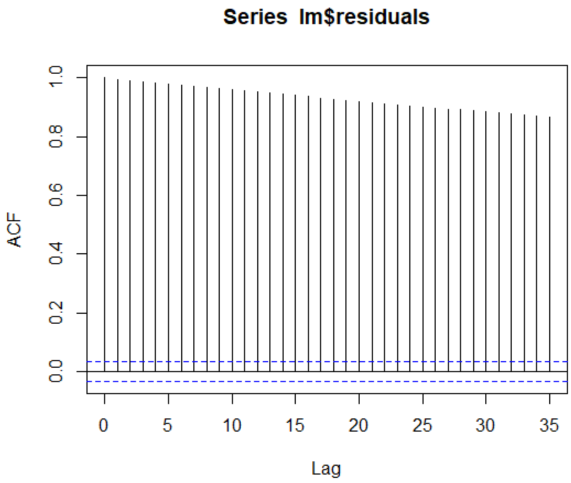

Figure 5.3: ACF of the residuals

The above graph of ACF shows strong autocorrelation, in the residuals, between the present time and the lagged values. Autocorrelation decays slowly to zero as the lag increases. This is expected as there is a unit root in the time series of the prices.

Autocorrelation test Durbin-Watson

H_0 : residuals are not autocorrelated		H_a : residuals are autocorrelated of order 1
lag	D-W Statistic	p- value
1	0.010177	$2.2 \cdot 10^{-16}$

The Durbin Watson test suggests a strong positive autocorrelation between the residuals. This supports the evidences from the ACF.

Phillips-Perron test

H_0 : Existence of Unit Root		H_a : Stationary
Dickey Fuller Z_a	Truncation lag parameter	p- value
-17.267	9	0.1468

Table 5.6: PP test for model Residuals

The Phillips-Perron test's p-value is higher than the usual significance level and so we fail to reject null hypothesis. In other words there is unit root in the residuals of the regression.

Logarithmic Prices

We continue the analysis constructing a simple linear regression model with the NQ being the dependent variable and the DJI being the independent variable. For the regression we use the logarithmic prices of the two assets.

The regression has the following form:

$$log(NQ_t) = \beta_0 + \beta_1 log(DJI_t) + \epsilon_t$$

	Estimate	Std. Error	t value	Pr(>$\|t\|$)
(Intercept)	-7.876622	0.061257	-128.6	$2 \cdot 10^{-16}$
log(DJI)	1.654434	0.006402	258.4	$2 \cdot 10^{-16}$

R^2	0.9486	Adj R^2	0.9485
F-statistic:	$6.678 \cdot 10^4$	p-value	$2.2 \cdot 10^{-16}$

Table 5.7: Summary of the regression

From the summary of the regression on the logarithmic prices, we can see that the $R^2 = 94.86\%$. Also, the DJI is statistically significant, since the p-value of the F-test is less than 5%. In addition, the correlation between these assets is 0.9739.

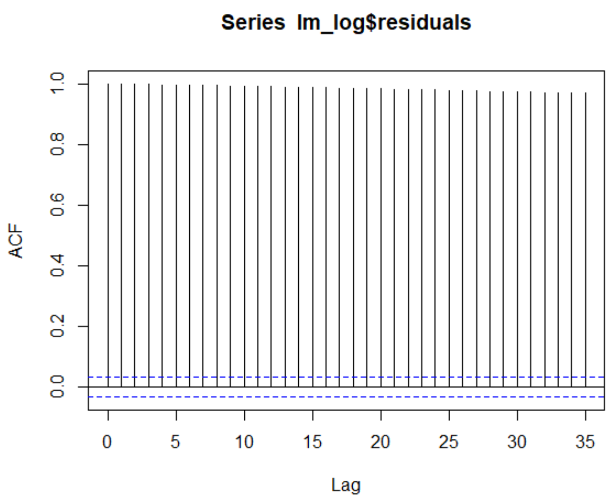

Figure 5.4: ACF of the residuals

The above graph of ACF, on the residuals of the logarithmic prices, shows strong autocorrelation between the present time and the lagged values. Autocorrelation decays slowly to zero as the lag increases.

Autocorrelation test Durbin-Watson

H_0 : residuals are not autocorrelated		H_a : residuals are autocorrelated of order 1
lag	D-W Statistic	p- value
1	0.00057369	$2.2 \cdot 10^{-16}$

Durbin-Watson test, for the prices' residuals, supports the above ACF results because it shows autocorrelation.

Phillips-Perron test

H_0 : Existence of Unit Root		H_a : Stationary
Dickey Fuller Z_a	Truncation lag parameter	p- value
-24.183	9	0.02975

Table 5.8: PP test for model Residuals

Lastly, using the Phillips-Perron test on the residuals we conclude that they are stationary since $p = 0.02975 < 0.05$ significance level.

Spread

In pairs trading we can use the spread of the time series of the logarithmic prices see if it is mean reverting, i.e. if the two time series of the pair are oscillating around a long run equilibrium. From the linear regression we obtained the 1.654434 coefficient for the DJI.

$$spread = NQ - 0.3653 * DJI$$

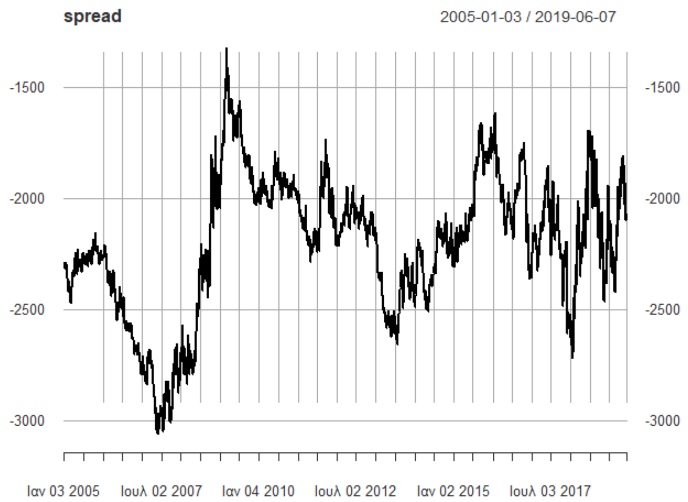

Figure 5.5: Spreads of the prices between NQ and DJI

	1pct	5pct	10pct
τ_3	-3.96	-3.41	-3.12
ϕ_2	6.09	4.68	4.03
ϕ_3	8.27	6.25	5.34

	τ_3	ϕ_2	ϕ_3
statistic	-3.059	3.1252	4.6796

Table 5.9: ADF test for the spread

The ADF test for the spread shows that τ_3-statistic is $-3.059 > -3.41$ for the 5% significance level. This indicates that there is unit root. The $\phi_2 = 3.1252 < 4.68$ indicates that there is drift.

Phillips-Perron test

H_0 : Existence of Unit Root		H_a : Stationary
Dickey Fuller Z_a	Truncation lag parameter	p- value
-17.266	9	0.1469

Table 5.10: PP test for the spread

The p-value of the Phillips-Perron test, on spread, is 0.1469 which is less than the 5% significance level, and so, we reject the null hypothesis. Therefore, the spread of the time series is not stationary.

Spread for the Logarithmic Prices

$$spread = log(NQ) - 1.654434 * log(DJI)$$

Figure 5.6: Spreads of the logarithmic prices between NQ and DJI

The plot of the spread shows a clear trend and a spike going downwards around 2008.

	1pct	5pct	10pct
τ_3	-3.96	-3.41	-3.12
ϕ_2	6.09	4.68	4.03
ϕ_3	8.27	6.25	5.34

	τ_3	ϕ_2	ϕ_3
statistic	-3.7405	6.1484	7.1598

Table 5.11: ADF test for the spread

The ADF test for the spread shows that τ_3-statistic is $-3.7405 < -3.41$ for the 5% significance level. This indicates that there is no unit root and so the trend is stationary. The $\phi_2 = 6.1484 > 4.68$ indicates that there is drift.

Phillips-Perron test

H_0 : Existence of Unit Root		H_a : Stationary
Dickey Fuller Z_a	Truncation lag parameter	p- value
-24.183	9	0.02975

Table 5.12: PP test for the spread

The p-value of the Phillips-Perron test, on spread, is 0.02975 which is less than the 5% significance level, and so, we reject the null hypothesis. Therefore, the spread of the time series is stationary.

5.1.2 Daily Logarithmic Returns

As it was mentioned in a previous chapter, in financial time series it is preferable to work with the daily logarithmic returns over the simple ones. For this reason, we calculated the first difference of the daily logarithmic prices.

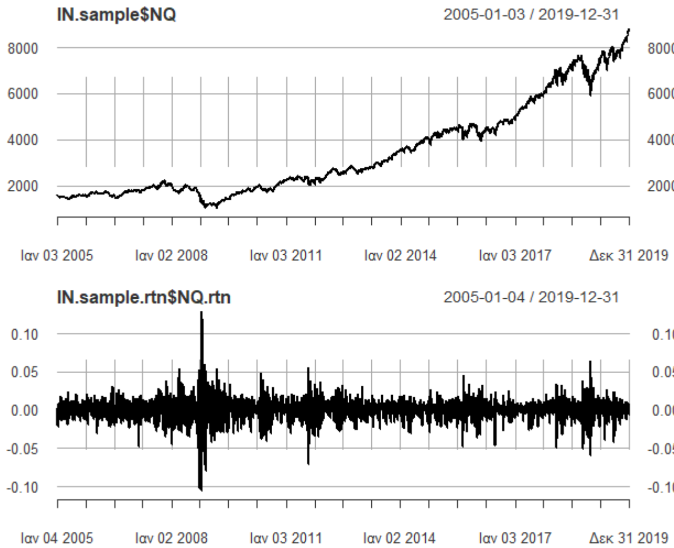

Figure 5.7: Prices vs Daily Logarithmic Returns of NQ

From the above graph we can see the difference between the prices and the daily logarithmic returns of NQ. In the daily logarithmic returns plot it is easy to detect the spikes that occur in various dates with the biggest one being around 2008. The formation of volatility clusters is also quite observable. Apart from 2008, which is when the worldwide financial crisis happened, there are also smaller clusters around 2011 and 2016, when the Japan earthquake and tsunami, and the American president elections, happened respectively.

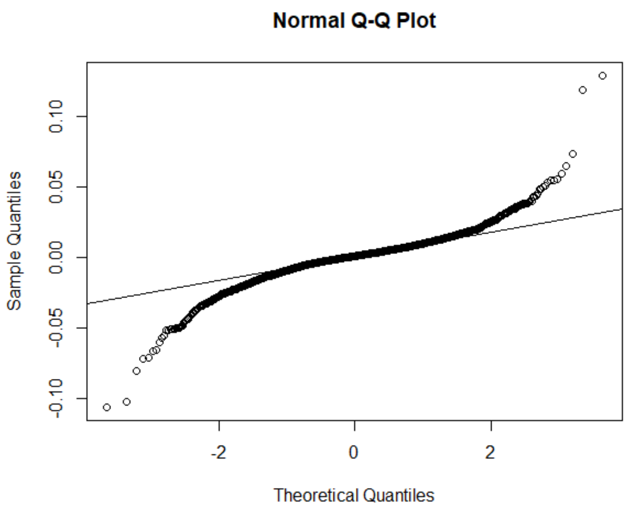

Figure 5.8: Daily Logarithmic Returns of NQ QQ plot for Normal Distribution

From the QQ-plot, of the daily logarithmic returns of NQ, the data appear to fit well in the center but as we move outwards, we have heavy tails that drift further from the normal distribution.

It provides no clear clue for skewness as both tails seem to drift apart from the normal distribution line. Also we cannot see any change in variance as we did in the daily logarithmic returns plot.

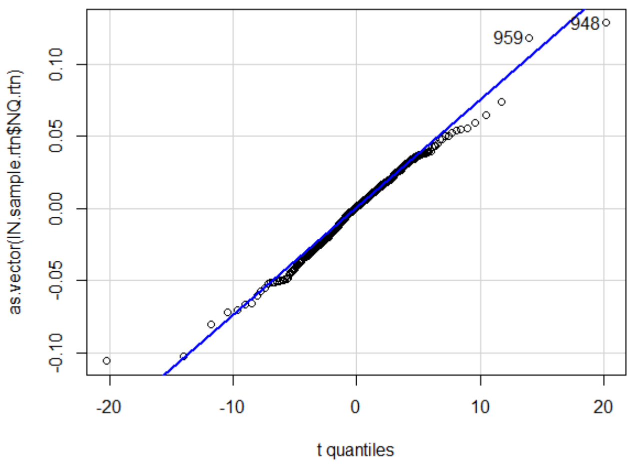

Figure 5.9: Daily Logarithmic Returns of NQ QQ plot for Student's t-Distribution 3df

From the Student-t with 3 degrees of freedom figure, the daily logarithmic returns of NQ seem to be closer to the line while the downside tail is still quite fat. This indicates that we over-estimate the right side and under-estimate the left side.

Another thing one can notice is the two enumerated observations, which are identified as outliers. Even though, R does not point out any outliers on the left side it does not mean they are insignificant.

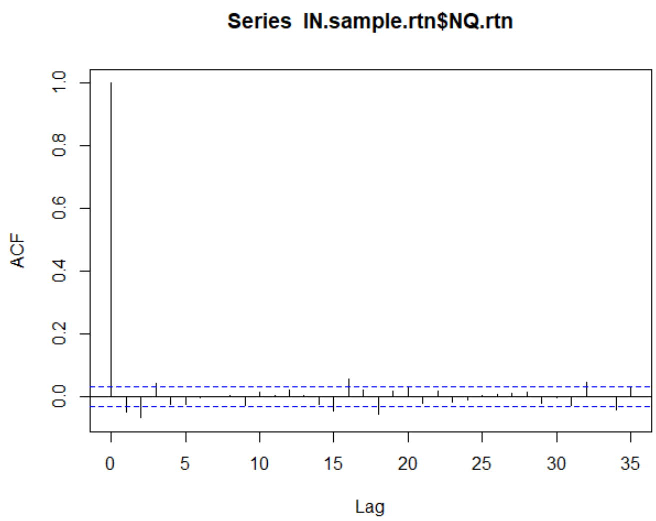

Figure 5.10: ACF of the Daily Logarithmic Returns of NQ

The ACF figure for the daily logarithmic returns of NQ shows that there is negative autocorrelation between the present and the lagged values of NQ.

	1pct	5pct	10pct
τ_1	-2.58	-1.95	-1.62

	τ_1
statistic	-46.5854

Table 5.13: ADF test on Daily Logarithmic Returns of NQ

Testing the daily logarithmic returns of NQ for stationarity we choose to use the ADF test for 5% significance level. The conclusion that derives from the table is that τ_1-statistic, which is equal to -46.5854, is less than the critical value and so the daily logarithmic returns of NQ are stationary.

Phillips-Perron test

H_0 : Existence of Unit Root		H_a : Stationary
Dickey Fuller Z_a	Truncation lag parameter	p- value
-3580.2	9	0.01

Table 5.14: PP test for Daily Logarithmic Returns of NQ

The Phillips-Perron test for NQ daily logarithmic returns comes to support that evidence of stationarity as the p-value is less than the 5% significance level. Thus, we reject the null hypothesis, of existence of unit root, so the daily logarithmic returns are stationary.

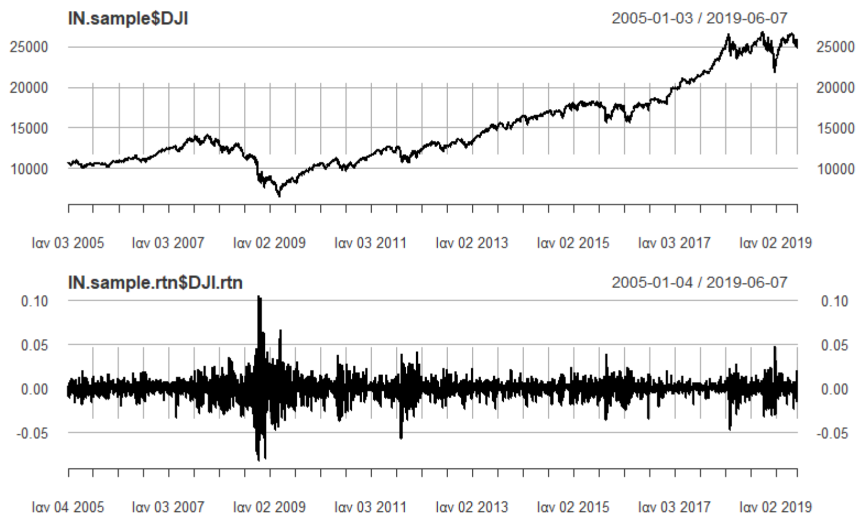

Figure 5.11: Prices vs Daily Logarithmic Returns of DJI

We continue the same analysis for the daily logarithmic returns of DJI.

From the above graph we can see the difference between the prices and the daily logarithmic returns of DJI. In the daily logarithmic returns plot it is easy to detect the spikes that occur in various dates with the biggest being around 2008. It is also quite observable the formation of volatility clusters. Apart from the 2008, there are also clusters around 2011 and 2016.

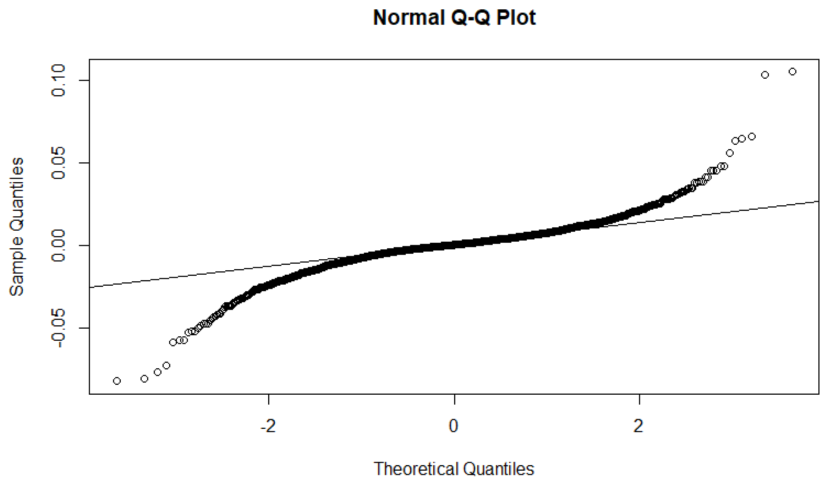

Figure 5.12: Daily Logarithmic Returns of DJI QQ plot for Normal Distribution

From the QQ-plot, of the daily logarithmic returns of DJI, the data appear to fit well in the center but as we move outwards, we have heavy tails that drift further from the normal distribution.

It provides no clear clue for skewness as both tails seem to drift apart from the normal distribution line. Also we cannot see any change in variance as we did in the daily logarithmic returns plot.

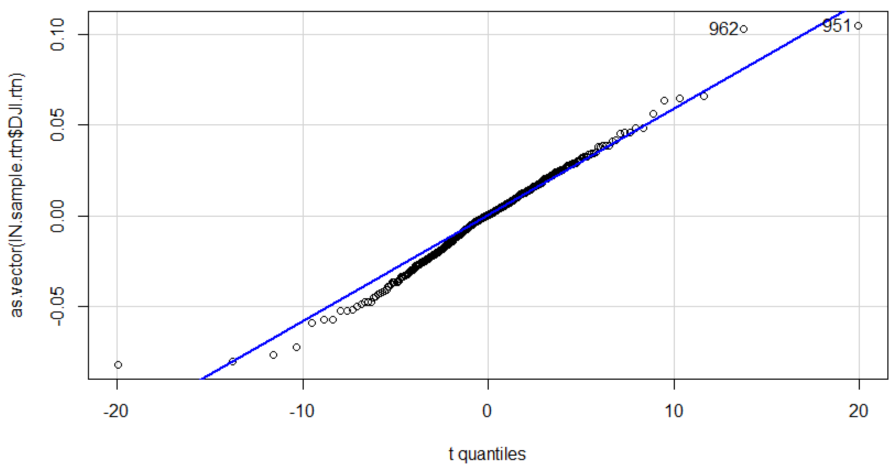

Figure 5.13: Daily Logarithmic Returns of DJI QQ plot for Student's t-Distribution with 3 df

The Student's t-Distribution with 3 degrees of freedom fits better than the Normal Distribution for the daily logarithmic returns of DJI and also we have heavy tails.

Another thing one can notice is the enumerated observations, which are identified as outliers.

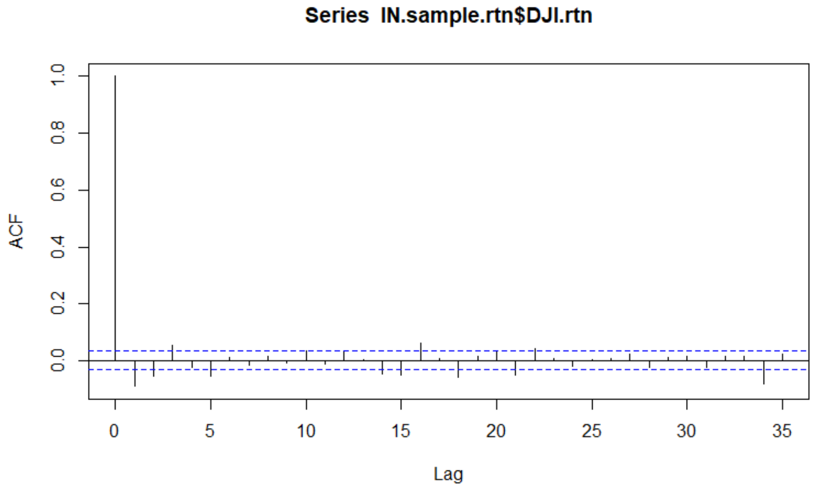

Figure 5.14: ACF of the Daily Logarithmic Returns of DJI

The ACF figure for the daily logarithmic returns of DJI shows that there is negative autocorrelation between the present and the lagged values of DJI.

	1pct	5pct	10pct
τ_1	-2.58	-1.95	-1.62

	τ_1
statistic	-47.2411

Table 5.15: ADF test on Daily Logarithmic Returns of DJI

Testing the daily logarithmic returns of DJI for stationarity we choose to use the ADF test for 5% significance level. The conclusion that derives from the table is that τ_1-statistic, which is equal to -47.2411, is less than the critical value and so the daily logarithmic returns of DJI are stationary.

Phillips-Perron test

H_0 : Existence of Unit Root		H_a : Stationary
Dickey Fuller Z_a	Truncation lag parameter	p- value
-3720.3	9	0.01

Table 5.16: PP test for Daily Logarithmic Returns of DJI

The Phillips-Perron test for DJI daily logarithmic returns comes to support the evidence of stationarity as the p-value is less than the 5% significance level. Thus, we reject the null hypothesis, of existence of unit root, so the daily logarithmic returns are stationary.

Regression

Once again we construct a simple linear regression model with the daily logarithmic returns of NQ being the dependent variable and the daily logarithmic returns DJI being the independent variable. The regression has the following form:

$$NQ_t = \beta_1 DJI_t + \epsilon_t$$

	Estimate	Std. Error	t value	Pr(>|t|)
DJI.rtn	1.021474	0.009715	105.1	$2 \cdot 10^{-16}$

R^2	0.7532	Adj R^2	0.7531
F-statistic:	$1.105 \cdot 10^4$	p-value	$2.2 \cdot 10^{-16}$

Table 5.17: Summary of the regression

From the summary of the regression on the daily logarithmic returns, we can see that the $R^2 = 75\%$. Also, the coefficient of the DJI is statistically significant, since the p-value of the F-test is less than 5%. In addition, the correlation between these assets is positive, since it is 0.8677972.

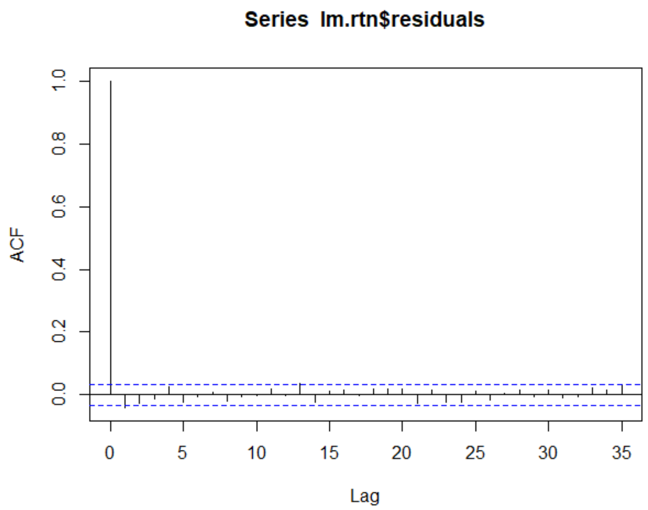

Figure 5.15: ACF

The above graph of the residuals' ACF shows negative autocorrelation between the present time and its lagged values.

Autocorrelation test Durbin-Watson

H_0 : residuals are not autocorrelated		H_a : residuals are autocorrelated of order 1
lag	D-W Statistic	p- value
1	2.0798	0.9919

The Durbin-Watson test of the residuals shows slightly negative autocorrelation and support of the evidence from ACF.

Phillips-Perron test

H_0 : Existence of Unit Root		H_a : Stationary
Dickey Fuller Z_a	Truncation lag parameter	p- value
-3632.9	9	0.01

Table 5.18: PP test for the residuals

The Phillips-Perron test indicates that there is no unit root.

5.1.3 ADL

Continuing the analysis, we construct the Autogressive Distributed Lag (ADL) model with $p = 0$ and $q = 1$. In essence, we want to estimate the present value of the daily logarithmic returns of NQ using the previous values of the daily logarithmic returns of DJI. The model is in the following form:

$$NQ_t = \beta_1 DJI_{t-1} + \epsilon_t$$

	Estimate	Std. Error	t value	Pr(>|t|)
lag(DJI.rtn)	-0.08530	0.01951	-4.373	$1.26 \cdot 10^{-5}$
R^2	0.005253	Adj R^2	0.004978	
F-statistic:	19.13	p-value	$1.258 \cdot 10^{-5}$	

Table 5.19: Summary of the regression for the lagged Daily Logarithmic Returns

From the summary of the regression, we can see that the $R^2 = 0.5253\%$. Also, the intercept and the coefficient of the DJI are both statistically significant, since the p-value of the F-test is less than 5%.

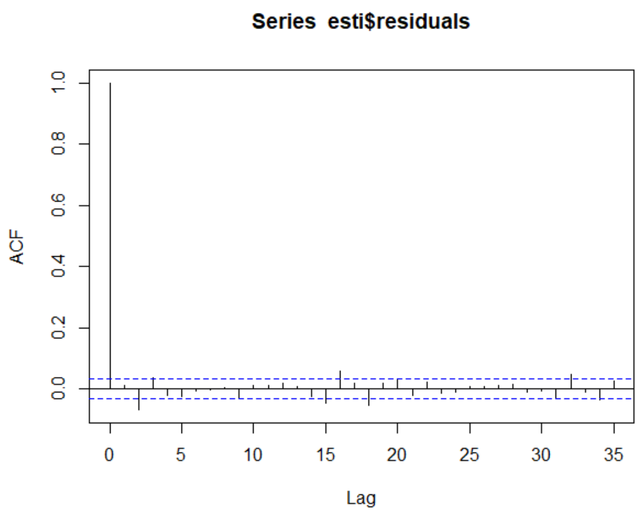

Figure 5.16: ACF lagged model

Autocorrelation test Durbin-Watson

H_0 : residuals are not autocorrelated		H_a : residuals are autocorrelated of order 1
lag	D-W Statistic	p- value
1	1.9783	0.2569

Both the ACf and the Durbin-Watson test indicate that the residuals of the daily logarithmic returns are slightly positively correlated, which is expected due to the mean reverting.

Phillips-Perron test		
H_0 : Existence of Unit Root		H_a : Stationary
Dickey Fuller Z_a	Truncation lag parameter	p- value
-3375.6	9	0.01

Table 5.20: PP test for the residuals of the ADL model

The Phillips-Perron test indicates that there is no unit root.

5.1.4 Backtesting

A basic strategy for checking if the model has forecasting capability is the back-testing process.

Back-testing is very useful in detecting the weaknesses of forecasting models but it does not provide any information for what causes these weaknesses.

Because the data, we are working with, are daily the look-back period for the back-testing procedure is 4 years while the burn in period will be the first 2 years. By throwing away 2 years,through the burn-in period we help the procedure to run normally afterwards.

While the look back period is 4 years because we start with a burn in period of 2 years, this translates to an expanding window that starts with a length of 2 years and expands to its length to 4 years.

Those 2 and 4 years are basically 500 and 1000 long length observation, respectively, because the "working days" of the market in a year are approximately 250.

We will use 3 different strategies and check their performance in the in sample period and in the out of sample period.

Prices

Strategy 1: Statistical Arbitrage

For this strategy we used the linear regression of the prices of NQ and DJI, we produce the spreads and we chose to use the ± 2 standard deviation as a measure for when we will be in long/short position. This time we let the model to choose when we open or close the position.

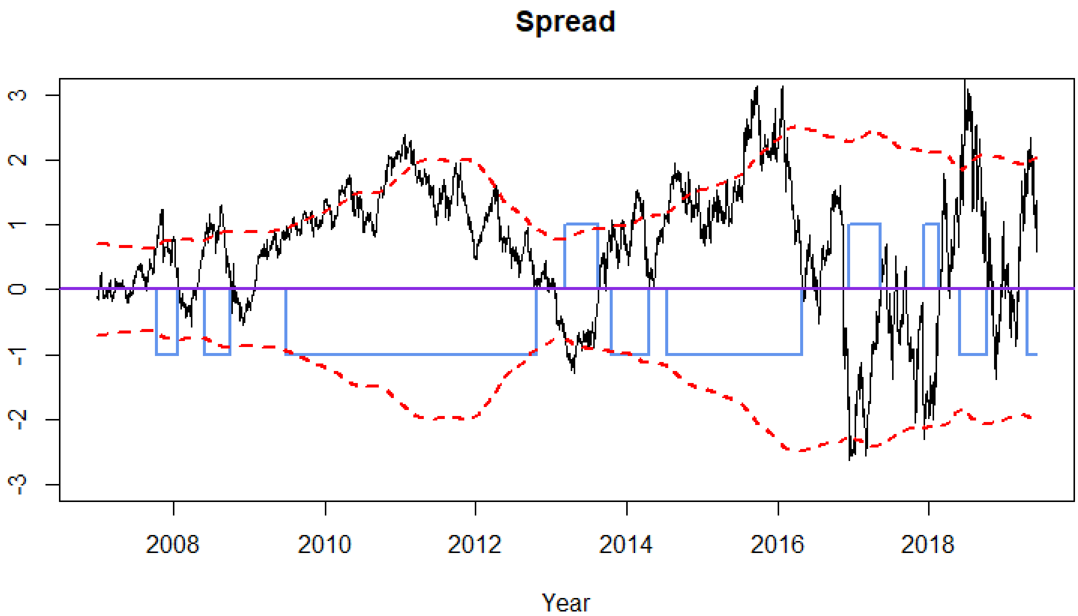

Figure 5.17: Spread and position for Strategy 1

In this graph we can see the spread of NQ and DJI. The blue line represents the position we took in this strategy. The dashed red lines stand for the two standard deviations, where we took positions and the middle line is where we close our positions.

Strategy 2

Using the linear regression of the prices of NQ and DJI, we produce the spreads and we chose to use the ±2 standard deviation as a measure for when we will be in long/short position, i.e. values in absolute less than 2 standard deviation will be ignored.

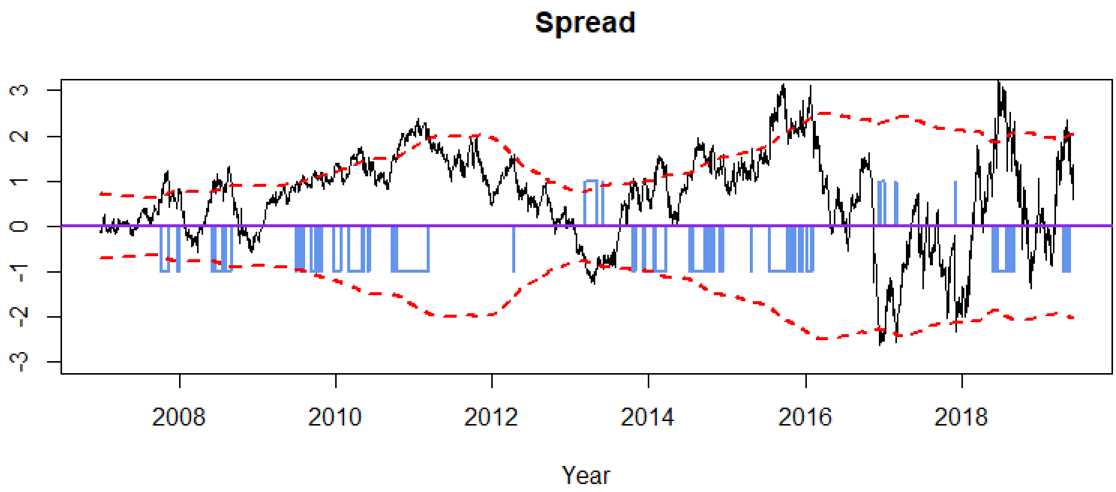

Figure 5.18: Spread and position for Strategy 2

Strategy 3

In this strategy we use the same method as before but this time we hold our position until we swap from long to short or from short to long. For example, we take our position if spread is greater than 2 standard deviation and keep that position until spread is less than -2 standard deviation.

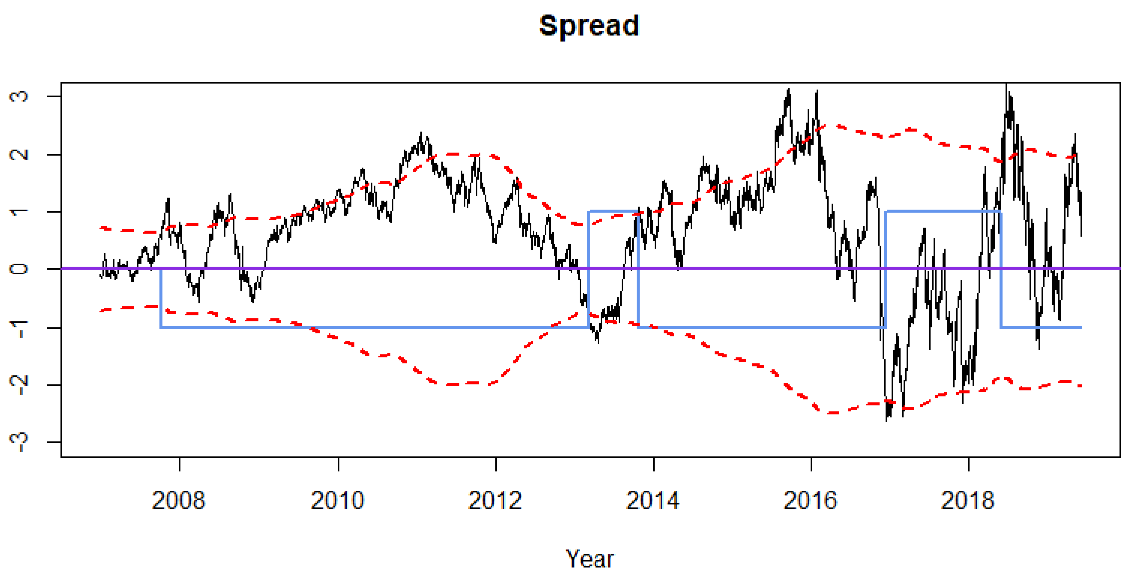

Figure 5.19: Spread and position for Strategy 3

Log Prices

Strategy 1: Statistical Arbitrage

For this strategy we used the linear regression of the logarithmic prices of NQ and DJI, we produce the spreads and we chose to use the ±2 standard deviation as a measure for when we will be in long/short position. This time we let the model to choose when we open or close the position.

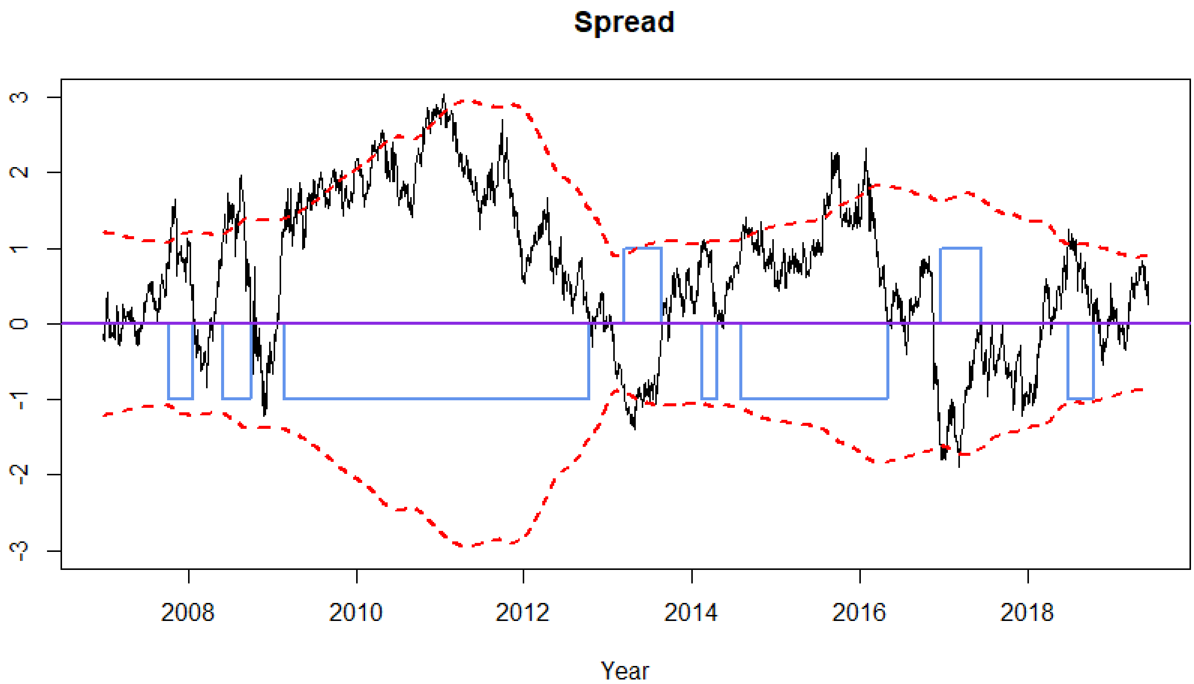

Figure 5.20: Spread and position for Strategy 1

In this graph we can see the spread of logarithmic NQ and DJI. The blue line represents the position we took in this strategy. The dotted red lines stand for the two standard deviations, where we took positions and the middle line is where we close our positions.

Strategy 2

Using the linear regression of the logarithmic prices of NQ and DJI, we produce the spreads and we chose to use the ± 2 standard deviation as a measure for when we will be in long/short position, i.e. values in absolute less than 2 standard deviation will be ignored.

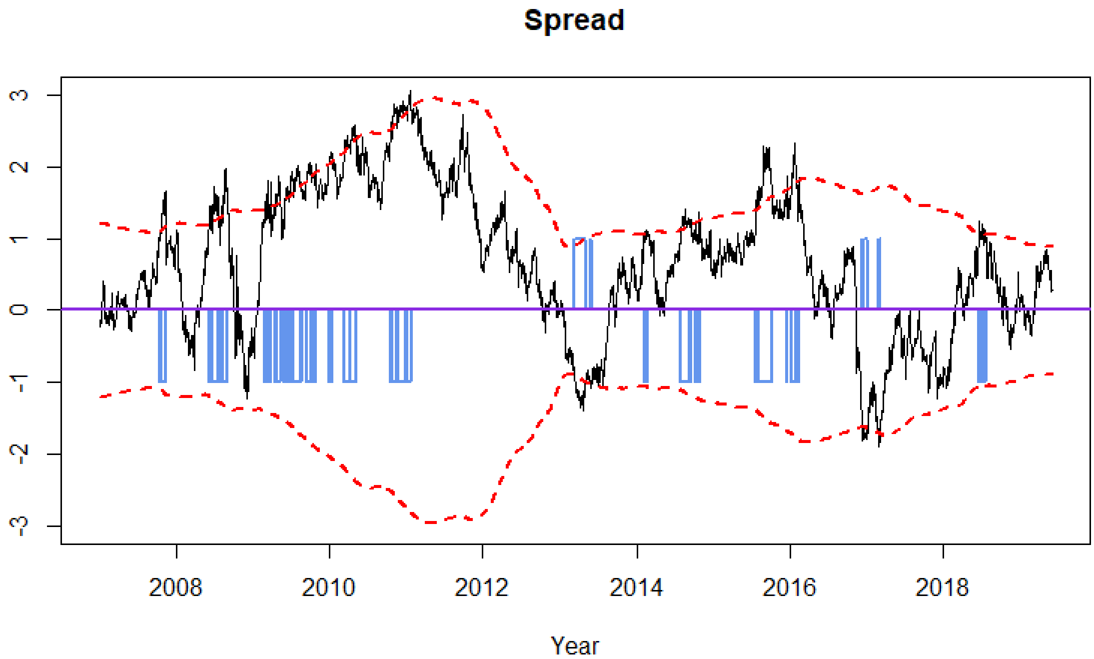

Figure 5.21: Spread and position for Strategy 2

Strategy 3

In this strategy we use the same method as before but this time we hold our position until we swap from long to short or from short to long. For example, we take our position if spread is greater than 2 standard deviation and keep that position until spread is less than -2 standard deviation.

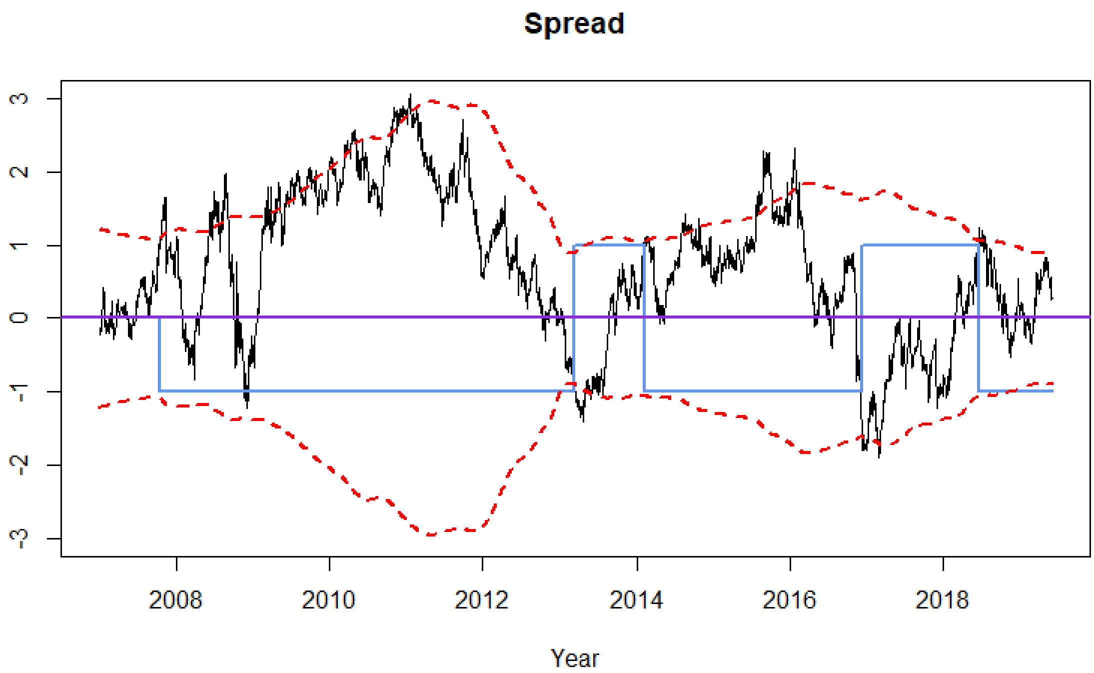

Figure 5.22: Spread and position for Strategy 3

5.1.5 Out of sample

Prices

Statistical Arbitrage for the Out of sample Period

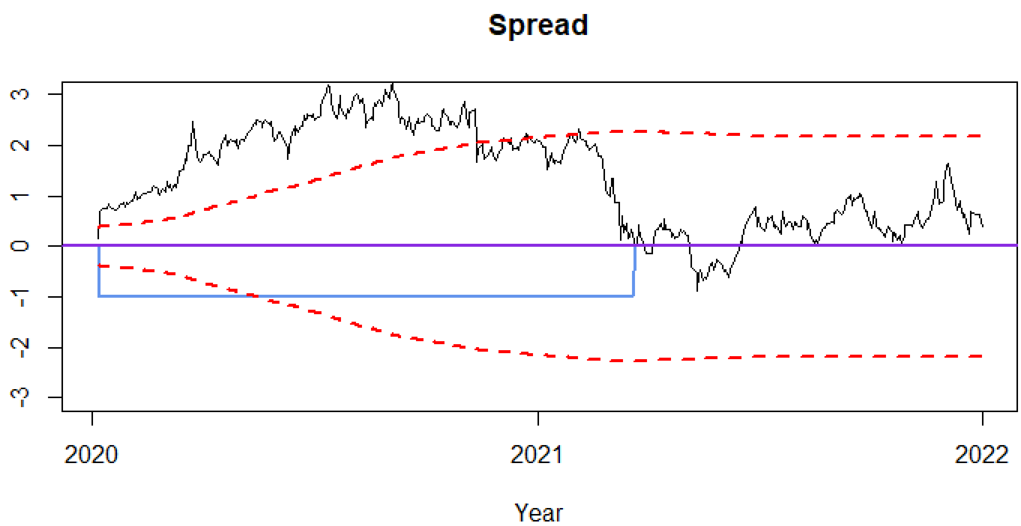

Figure 5.23: Spread and position for Strategy 1

In this graph we can see the spread of NQ and DJI. The blue line represents the position we took in this strategy. The dashed red lines stand for the two standard deviations, where we took positions and the middle line is where we close our positions.

Strategy 2 for the Out of sample Period

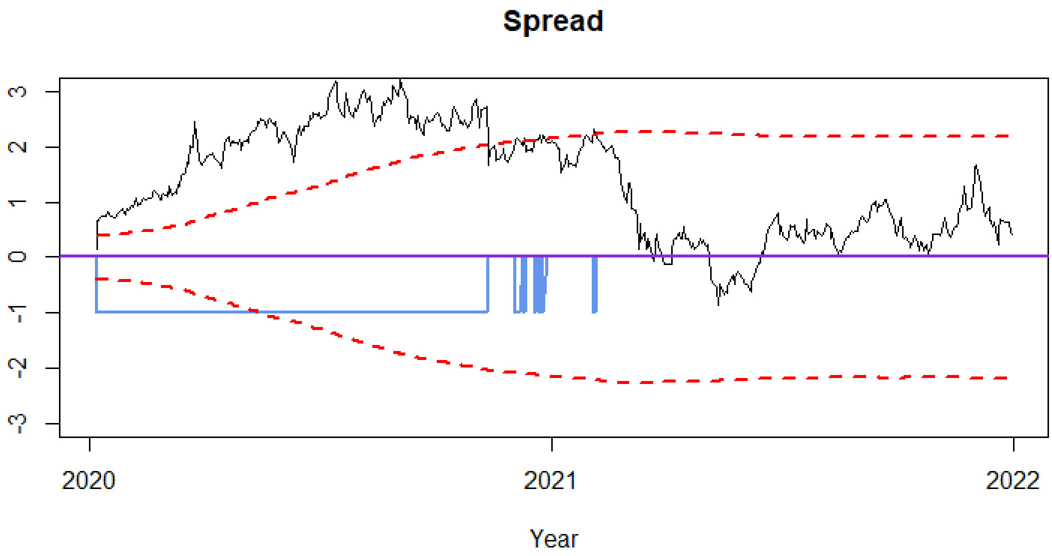

Figure 5.24: Spread and position for Strategy 2

Strategy 3 for the Out of sample Period

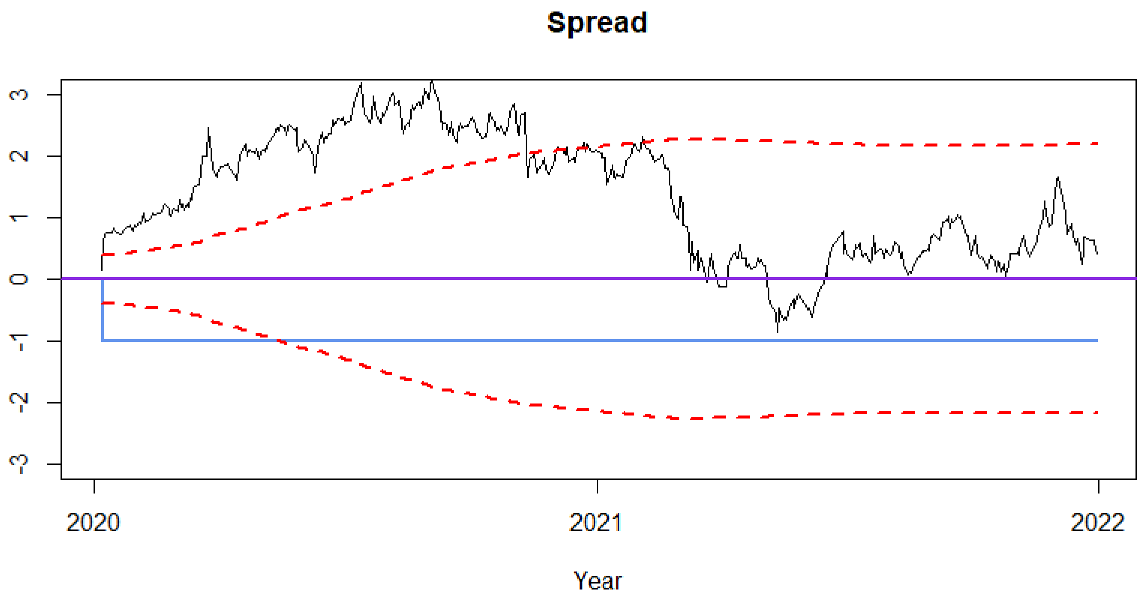

Figure 5.25: Spread and position for Strategy 1

Log Prices

Statistical Arbitrage for the Out of sample Period

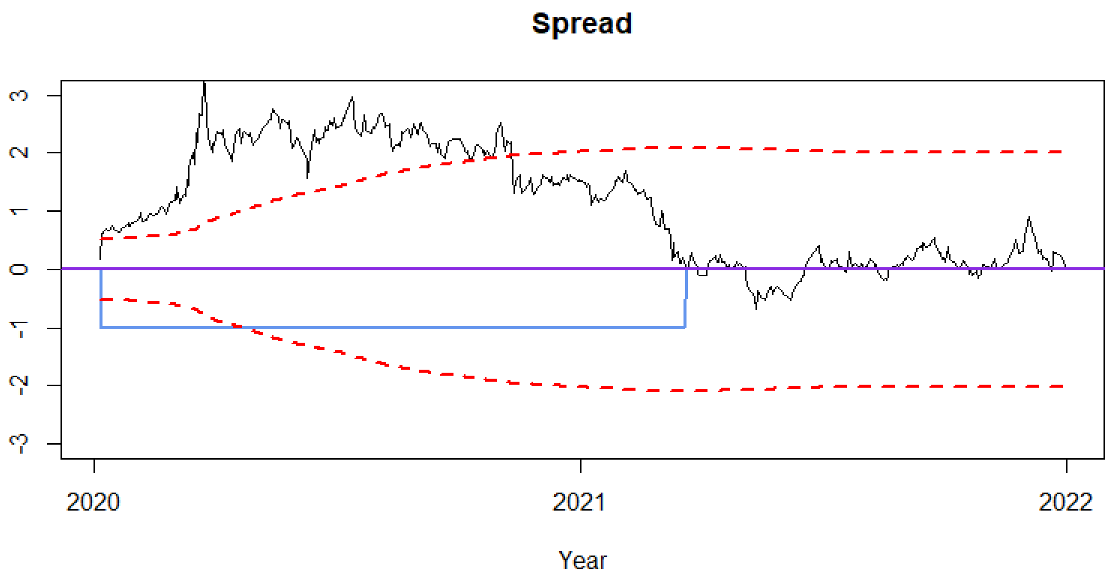

Figure 5.26: Spread and position for Strategy 1

In this graph we can see the spread of NQ and DJI. The blue line represents the position we took in this strategy. The dashed red lines stand for the two standard deviations, where we took positions and the middle line is where we close our positions.

Strategy 2 for the Out of sample Period

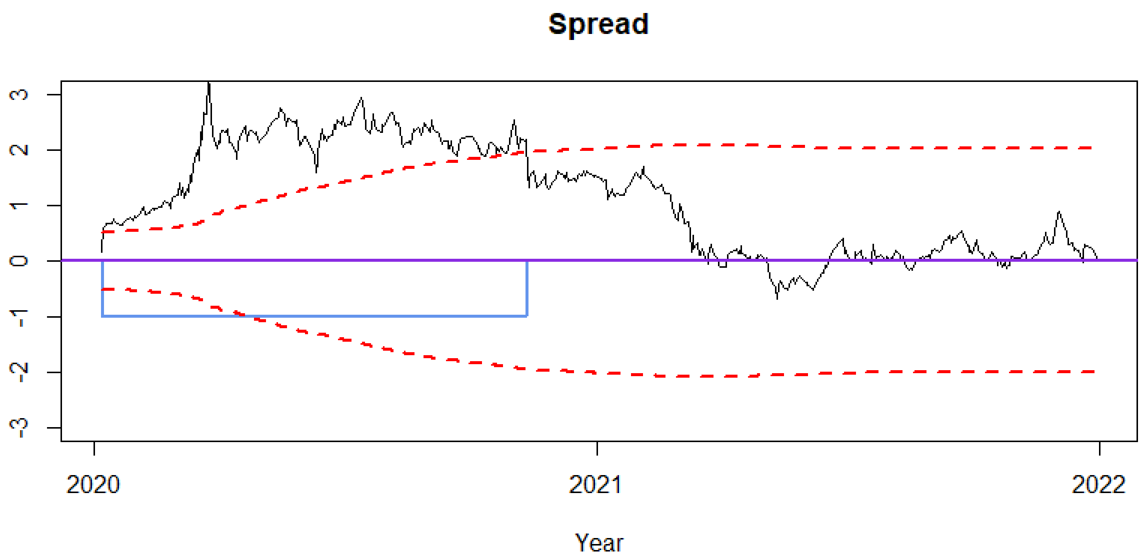

Figure 5.27: Spread and position for Strategy 2

Strategy 3 for the Out of sample Period

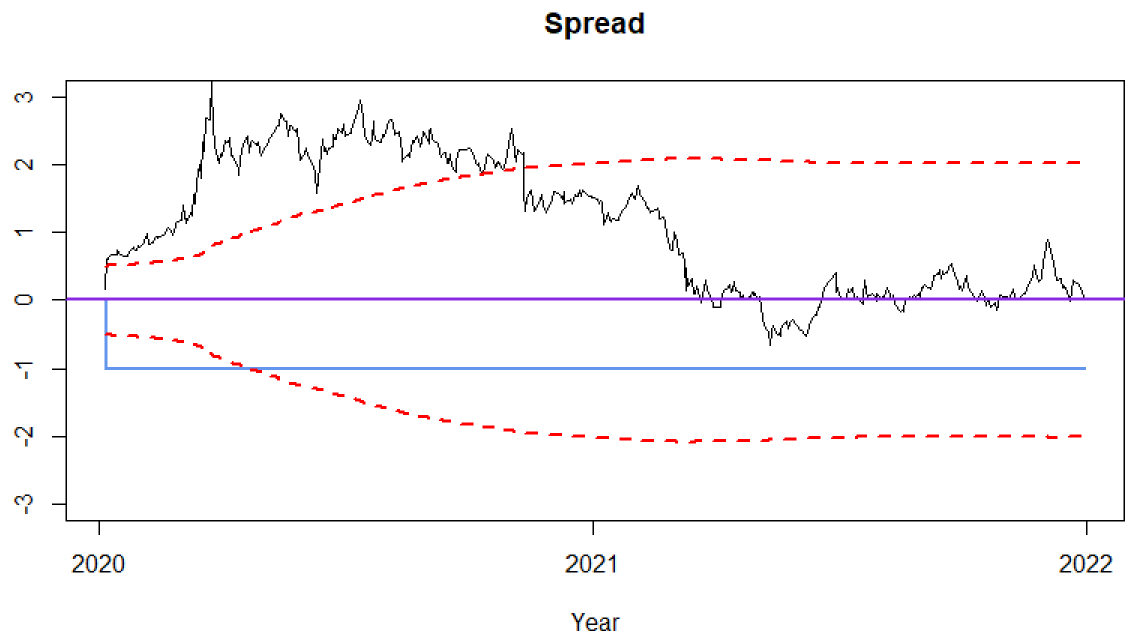

Figure 5.28: Spread and position for Strategy 3

5.2 Investigation for possible pairs

Now lets consider the stocks of Dow Jones Index for the period 1/1/2014 until 1/1/2022. We will test for possible pairs using the Augmented Dickey Fuller unit root test on the residuals, which we consider as the spread, of a simple linear regression between these pairs. The data were imported from yahoo finance and have been divided into four groups, where each group contains a two year period. Our goal is to see if there are any pairs appropriate for statistical arbitrage and also if the relationship between them holds for the next period. The code that have been used here is a modification of the original R-code.

As it turned out we could not found any appropriate pairs for the period 2014-2016. In the period 2016-2018 we were able to find 15 pairs, which we can use in our strategies.

2016-2018

	Stock 1	Stock 2	correlation	beta	p-value
1	MMM	HON	0.98	0.10	0.01
2	UNH	CAT	0.98	1.38	0.04
3	JPM	AXP	0.97	-0.60	0.03
4	V	BA	0.97	2.00	0.02
5	UNH	AXP	0.97	0.30	0.01
6	MSFT	CAT	0.97	1.11	0.01
7	UNH	HON	0.97	-2.50	0.04
8	MMM	JNJ	0.97	-1.05	0.03
9	MSFT	BA	0.96	1.34	0.05
10	V	HON	0.96	-0.99	0.03
11	UNH	BA	0.96	1.70	0.04
12	UNH	MSFT	0.96	0.27	0.03
13	HON	HD	0.96	-0.19	0.02
14	MSFT	AXP	0.95	0.26	0.03
15	CAT	AXP	0.95	-1.08	0.04

For the next period, 2018-2020, we found two pairs completely different from those in the previous one.

2018-2020

	Stock 1	Stock 2	correlation	beta	p-value
1	V	MSFT	0.99	0.87	0.01
2	V	AXP	0.95	-1.79	0.01

In the last period 6 pairs seem to fulfil the requirements to be used in our strategies. Note that the two of them JP Morgan with American Express and United Health Group with Microsoft appeared again in the period 2016-2018.

2020-2022

	Stock 1	Stock 2	correlation	beta	p-value
1	JPM	AXP	0.97	0.55	0.04
2	GS	AXP	0.97	-0.33	0.02
3	TRV	JPM	0.96	1.51	0.01
4	UNH	MSFT	0.95	1.65	0.01
5	TRV	HON	0.95	0.51	0.01
6	UNH	HD	0.95	0.71	0.01

Conclusion

In the previous chapters we presented the theory relevant to the pairs trading and the statistical arbitrage. More specifically the concepts of cointegration, spurious regression and the unit root tests. We then continued with more practical applications. These applications are based on spread between two assets and how this spread can be utilized for the pair trading. This strategy is based on the placement of opposite positions (one long and one short), in order to achieve a hedged position, which is on averaged profitable due to the exploitation of a mispricing between two highly correlated assets. We managed to present different strategies used in pair trading, such as the long only strategy, the spread strategy and the ADL strategy.

For the application we used the technological indexes, videlicet the Nasdaq (NQ), and industrial indexes, videlicet the Dow Jones Index (DJI). We can see that there are time periods when the technological companies are preferable or time periods when the industrial indexes are preferable. The decision that investors are enquired to make, about which one they will prefer, has to do with market conditions and is also the reason why on specific time periods one will outperform the other.

Our aim, through this application, was to highlight what statistical arbitrage is, i.e. the on average achievement of profit using statistical models and how one can construct it.

Finally, we should mention that our results may not seem very appealing, because we wanted to focus on the statistical aspect and outcome rather than create more complex and complete trading algorithms. If one wishes, to do so, there are a number of trading techniques that can be used for better results, like risk-adjustment, optimal execution etc.

Appendix A

R code

```
############## NASDAQ AND DJI ------------
rm(list=ls())
cat("\014")

library(quantmod) # Load the package
library(tseries)
library(readxl)
library(car)
library(dplyr)
library(urca)
library(lmtest)


getSymbols("NQ=F", from="2005-01-01", to="2022-01-01", src = "yahoo")
NQ <- `NQ=F`[,6] #NASDAQ100
rm(`NQ=F`)
NQ <- na.omit(NQ)
NQ.rtn <- diff(log(NQ))
NQ.rtn <- na.omit(NQ.rtn)


getSymbols("DJI", from="2005-01-01", to="2022-01-01", src = "yahoo")
DJI <- DJI$DJI.Adjusted #Dow Jones Industrial Average
DJI <- na.omit(DJI)
DJI.rtn <- diff(log(DJI))
DJI.rtn <- na.omit(DJI.rtn)


data.full <- merge(NQ, DJI, join = "inner")
colnames(data.full) <- c('NQ', 'DJI')
IN.sample <- window(data.full, start = "2005-01-03", end = "2019-12-31")
OUT.sample <- window(data.full, start = "2020-01-01", end = "2022-01-01")
summary(IN.sample)

data.full.rtn <- merge(NQ.rtn, DJI.rtn, join = "inner")
summary(data.full.rtn)
colnames(data.full.rtn) <- c('NQ.rtn', 'DJI.rtn')
IN.sample.rtn <- window(data.full.rtn, start = "2005-01-04", end = "2019-12-31")
OUT.sample.rtn <- window(data.full.rtn,  start = "2020-01-01", end = "2022-01-01")


###### prices----------

plot(IN.sample$NQ)
plot(IN.sample$DJI)


summary(ur.df(IN.sample$NQ, type="trend", selectlags="BIC"))
summary(ur.df(IN.sample$DJI, type="trend", selectlags="BIC"))

pp.test(IN.sample$NQ)
pp.test(IN.sample$DJI)
```

```
lm_log <- lm(log(IN.sample$NQ) ~ log(IN.sample$DJI) )
summary(lm_log)

acf(lm_log$residuals)

lm_test2 <- lm(as.vector(log(IN.sample$NQ)) ~ as.vector(log(IN.sample$DJI)) )
lmtest::dwtest(lm_test2)
pp.test(as.vector(lm_log$residuals))

cor(log(IN.sample$NQ), log(IN.sample$DJI))


##### spread for the prices
lm <- lm(IN.sample$NQ ~ IN.sample$DJI)
summary(lm)

spread <- IN.sample$NQ - 0.3653 * IN.sample$DJI
plot(spread)


###spread for the log prices
spread <- log(IN.sample$NQ) -1.654434 * log(IN.sample$DJI)
plot(spread)
summary(ur.df(spread, type="trend", selectlags="BIC"))
pp.test(as.vector(spread))

#####returns--------

par(mfrow=c(2,1))
plot(IN.sample$NQ)
plot(IN.sample.rtn$NQ.rtn)
par(mfrow=c(1,1))

par(mfrow=c(2,1))
plot(IN.sample$DJI)
plot(IN.sample.rtn$DJI.rtn)
par(mfrow=c(1,1))


qqnorm(as.vector(IN.sample.rtn$NQ.rtn))
qqline(as.vector(IN.sample.rtn$NQ.rtn))
qqPlot(as.vector(IN.sample.rtn$NQ.rtn),distribution="t", df=3,envelope=F)

qqnorm(as.vector(IN.sample.rtn$DJI.rtn))
qqline(as.vector(IN.sample.rtn$DJI.rtn))
qqPlot(as.vector(IN.sample.rtn$DJI.rtn),distribution="t", df=3,envelope=F)

acf(IN.sample.rtn$NQ.rtn)
acf(IN.sample.rtn$DJI.rtn)

summary(ur.df(IN.sample.rtn$NQ.rtn, type="none", selectlags="BIC"))
summary(ur.df(IN.sample.rtn$DJI.rtn, type="none", selectlags="BIC"))

pp.test(IN.sample.rtn$NQ.rtn)
pp.test(IN.sample.rtn$DJI.rtn)

cor(IN.sample.rtn$NQ.rtn, IN.sample.rtn$DJI.rtn)

lm.rtn <- lm(IN.sample.rtn$NQ.rtn ~ -1 + IN.sample.rtn$DJI.rtn)
summary(lm.rtn)

acf(lm.rtn$residuals)

lm.rtnd <- lm(as.vector(IN.sample.rtn$NQ.rtn) ~ as.vector(IN.sample.rtn$DJI.rtn))
dwtest(lm.rtnd)

pp.test(lm.rtn$residuals)


#####ADL----------

l_dji_rtn <- stats::lag(IN.sample.rtn$DJI.rtn)
```

```
esti <- lm(IN.sample.rtn$NQ.rtn[2:length(l_dji_rtn),] ~ -1 + l_dji_rtn[2:length(l_dji_rtn),] )
summary(esti)
acf(esti$residuals)

estid <- lm(as.vector(IN.sample.rtn$NQ.rtn[2:length(l_dji_rtn),]) ~ l_dji_rtn[2:length(l_dji_rtn),] )
dwtest(estid)

pp.test(as.vector(esti$residuals))


####### Backtesting --------------


######### PRICES -------------


####### STRATEGY 1 ----------


T <- length(IN.sample$NQ) # number of observations for return y
pred <- matrix(nrow=T, ncol=1)
spread <- matrix(nrow=T, ncol=1)
volat <- matrix(nrow=T, ncol=1)
coef<- matrix(nrow=T, ncol=1)
pred[1:500, ] <- 0
wm <- 500 #burnin
we <- 1000 #look-back

for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
  t2 = t - 1; # end of the data window
  window <- IN.sample$NQ[t1:t2] # data for estimation
  lm_p <- lm(IN.sample$NQ[t1:t2] ~ IN.sample$DJI[t1:t2] )
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <- lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1]  <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
    pred[t,1]  <- 1
  }
  else{
    if ((pred[t-1,1] == 1) & (lm_p$residuals[length(lm_p$residuals)] > 0 )){
      pred[t,1] <- 0
    }
    else if ( (pred[t-1,1] == -1) & (lm_p$residuals[length(lm_p$residuals)] < 0 ) ) {
      pred[t,1] <- 0
    }
    else if ( (pred[t-1,1] == -1) & (lm_p$residuals[length(lm_p$residuals)] > 0 ) ) {
      pred[t,1] <- -1
    }
    else if ( (pred[t-1,1] == 1) & (lm_p$residuals[length(lm_p$residuals)] < 0 ) ) {
      pred[t,1] <- 1
    }
    else if ( (pred[t-1,1] == 0) & (lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)) ) {
      pred[t,1] <-  - 1
    }
    else if ( (pred[t-1,1] == 0) & (lm_p$residuals[length(lm_p$residuals)] < -2 * sd(temp_res)) ) {
      pred[t,1] <- 1
    }
    else {
      pred[t,1] <- 0
    }
  }
}
```

```r
sprd <- spread[-c(3625),1]
sprd2 <- sprd[501:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)

position <- pred[-c(3625),1]
position2 <- position[501:length(position)]


vol <- volat[-c(3625),1]
vol2 <- vol[501:length(vol)]


ts_spread <- xts(sprd3, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_pred <- xts(position2, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
     xlab= "Year", ylab = " ", main = "Spread")
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], (-2) * ts_volat, type="l",
       col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)



#### Strategy 2 ------


T <- length(IN.sample$NQ) # number of observations for return y
pred <- matrix(nrow=T, ncol=1)
spread <- matrix(nrow=T, ncol=1)
volat <- matrix(nrow=T, ncol=1)
coef <- matrix(nrow=T, ncol=1)
wm <- 500 #burnin
we <- 1000 #look-back

for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
  t2 = t - 1; # end of the data window
  window <- IN.sample$NQ[t1:t2] # data for estimation
  lm_p <- lm(IN.sample$NQ[t1:t2] ~  IN.sample$DJI[t1:t2] )
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <- lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1]  <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
    pred[t,1]  <- 1
  }
  else{
    pred[t,1] <- 0
  }
}



sprd <- spread[-c(3625),1]
sprd2 <- sprd[501:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)
```

```
position <- pred[-c(3625),1]
position2 <- position[501:length(position)]


vol <- volat[-c(3625),1]
vol2 <- vol[501:length(vol)]


ts_spread <- xts(sprd3, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_pred <- xts(position2, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
     xlab= "Year", ylab = " ", main = "Spread")
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], (-2) * ts_volat, type="l",
      col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)



######### STRATEGY 3 -----------


T <- length(IN.sample$NQ) # number of observations for return y
pred <- matrix(nrow=T,ncol=1)
spread <- matrix(nrow=T,ncol=1)
volat <- matrix(nrow=T,ncol=1)
coef <- matrix(nrow = T, ncol = 1)
wm <- 500 #burnin
we <- 1000 #look-back

for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
  t2 = t - 1; # end of the data window
  window <- IN.sample$NQ[t1:t2] # data for estimation
  lm_p <- lm(IN.sample$NQ[t1:t2] ~ IN.sample$DJI[t1:t2] )
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <- lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1]  <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
    pred[t,1]  <- 1
  }
  else{
    pred[t,1] <- NA
  }
}



sprd <- spread[-c(3625),1]
sprd2 <- sprd[501:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)

position <- pred2[-c(3625),1]
position2 <- position[501:length(position)]
```

```
vol <- volat[-c(3625),1]
vol2 <- vol[501:length(vol)]


ts_spread <- xts(sprd3, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_pred <- xts(position2, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
     xlab= "Year", ylab = " ", main = "Spread")
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], (-2) * ts_volat, type="l",
      col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)



########### LOG PRICES ---------------

####### STRATEGY 1 --------



T <- length(IN.sample$NQ) # number of observations for return y
pred <- matrix(nrow=T, ncol=1)
spread <- matrix(nrow=T, ncol=1)
volat <- matrix(nrow=T, ncol=1)
coef <- matrix(nrow = T, ncol = 1)
pred[1:500, ] <- 0
wm <- 500 #burnin
we <- 1000 #look-back

for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
  t2 = t - 1; # end of the data window
  window <- IN.sample$NQ[t1:t2] # data for estimation
  lm_p <- lm(log(IN.sample$NQ[t1:t2]) ~ log(IN.sample$DJI[t1:t2]) )
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <- lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1]  <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
    pred[t,1]  <- 1
  }
  else{
    if ((pred[t-1,1] == 1) & (lm_p$residuals[length(lm_p$residuals)] > 0 )){
      pred[t,1] <- 0
    }
    else if ( (pred[t-1,1] == -1) & (lm_p$residuals[length(lm_p$residuals)] < 0 ) ) {
      pred[t,1] <- 0
    }
    else if ( (pred[t-1,1] == -1) & (lm_p$residuals[length(lm_p$residuals)] > 0 ) ) {
      pred[t,1] <- -1
    }
    else if ( (pred[t-1,1] == 1) & (lm_p$residuals[length(lm_p$residuals)] < 0 ) ) {
      pred[t,1] <- 1
    }
    else if ( (pred[t-1,1] == 0) & (lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)) ) {
      pred[t,1] <-  - 1
```

```
    }
    else if ( (pred[t-1,1] == 0) & (lm_p$residuals[length(lm_p$residuals)] < -2 * sd(temp_res)) ) {
      pred[t,1] <- 1
    }
    else {
      pred[t,1] <- 0
    }
  }
}



sprd <- spread[-c(3625),1]
sprd2 <- sprd[501:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)

position <- pred[-c(3625),1]
position2 <- position[501:length(position)]


vol <- volat[-c(3625),1]
vol2 <- vol[501:length(vol)]


ts_spread <- xts(sprd3, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_pred <- xts(position2, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
     xlab= "Year", ylab = " ", main = "Spread")
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], (-2) * ts_volat, type="l",
      col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)



#### Strategy 2 ------


T <- length(IN.sample$NQ) # number of observations for return y
pred <- matrix(nrow=T,ncol=1)
spread <- matrix(nrow=T,ncol=1)
volat <- matrix(nrow=T,ncol=1)
coef <- matrix(nrow = T, ncol=1)
wm <- 500 #burnin
we <- 1000 #look-back

for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
  t2 = t - 1; # end of the data window
  window <- IN.sample$NQ[t1:t2] # data for estimation
  lm_p <- lm(log(IN.sample$NQ[t1:t2]) ~ log(IN.sample$DJI[t1:t2]) )
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <- lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1] <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
```

```
    pred[t,1]  <- 1
  }
  else{
    pred[t,1] <- 0
  }
}



sprd <- spread[-c(3625),1]
sprd2 <- sprd[501:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)

position <- pred[-c(3625),1]
position2 <- position[501:length(position)]


vol <- volat[-c(3625),1]
vol2 <- vol[501:length(vol)]


ts_spread <- xts(sprd3, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_pred <- xts(position2, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
           xlab= "Year", ylab = " ", main = "Spread")
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], (-2) * ts_volat, type="l",
      col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)


######### STRATEGY 3 ------------


T <- length(IN.sample$NQ) # number of observations for return y
pred <- matrix(nrow=T, ncol=1)
spread <- matrix(nrow=T, ncol=1)
volat <- matrix(nrow=T, ncol=1)
coef <- matrix(nrow = T, ncol = 1)
wm <- 500 #burnin
we <- 1000 #look-back

for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
  t2 = t - 1; # end of the data window
  window <- IN.sample$NQ[t1:t2] # data for estimation
  lm_p <- lm(log(IN.sample$NQ[t1:t2]) ~ log(IN.sample$DJI[t1:t2]) )
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <-lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1]  <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
    pred[t,1]  <- 1
  }
  else{
    pred[t,1] <- NA
  }
```

```
}



sprd <- spread[-c(3625),1]
sprd2 <- sprd[501:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)

position <- pred2[-c(3625),1]
position2 <- position[501:length(position)]


vol <- volat[-c(3625),1]
vol2 <- vol[501:length(vol)]


ts_spread <- xts(sprd3, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_pred <- xts(position2, order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]] )

plot(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
     xlab= "Year", ylab = " ", main = "Spread")
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(IN.sample.rtn)[501:dim(IN.sample.rtn)[1]], (-2) * ts_volat, type="l",
      col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)



########### OUT OF SAMPLE -------

###### Prices


###### STRATEGY 1 ----------


T <- length(data.full$NQ) # number of observations for return y
pred <- matrix(nrow=T,ncol=1)
spread <- matrix(nrow=T,ncol=1)
volat <- matrix(nrow=T,ncol=1)
coef <- matrix(nrow=T, ncol=1)
pred[1:3625, ] <- 0
wm <- 3625 #burnin
we <- 1000 #look-back

for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
  t2 = t - 1; # end of the data window
  window <- data.full$NQ[t1:t2] # data for estimation
  lm_p <- lm(data.full$NQ[t1:t2] ~ data.full$DJI[t1:t2])
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <- lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1]  <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
    pred[t,1]  <- 1
  }
  else{
```

```
    if ((pred[t-1,1] == 1) & (lm_p$residuals[length(lm_p$residuals)] > 0 )){
      pred[t,1] <- 0
    }
    else if ( (pred[t-1,1] == -1) & (lm_p$residuals[length(lm_p$residuals)] < 0 ) ) {
      pred[t,1] <- 0
    }
    else if ( (pred[t-1,1] == -1) & (lm_p$residuals[length(lm_p$residuals)] > 0 ) ) {
      pred[t,1] <- -1
    }
    else if ( (pred[t-1,1] == 1) & (lm_p$residuals[length(lm_p$residuals)] < 0 ) ) {
      pred[t,1] <- 1
    }
    else if ( (pred[t-1,1] == 0) & (lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)) ) {
      pred[t,1] <-  - 1
    }
    else if ( (pred[t-1,1] == 0) & (lm_p$residuals[length(lm_p$residuals)] < -2 * sd(temp_res)) ) {
      pred[t,1] <- 1
    }
    else {
      pred[t,1] <- 0
    }
  }
}




sprd <- spread[-c(4106),1]
sprd2 <- sprd[3626:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)

position <- pred[-c(4106),1]
position2 <- position[3626:length(position)]


vol <- volat[-c(4106),1]
vol2 <- vol[3626:length(vol)]


ts_spread <- xts(sprd3, order.by = index(data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_pred <- xts(position2, order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]] )

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index( data.full.rtn)[3626:dim( data.full.rtn)[1]]

plot(index( data.full.rtn)[3626:dim( data.full.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
     xlab= "Year", ylab = " ", main = "Spread")
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], (-2) * ts_volat, type="l",
      col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)




####Strategy 2 ------

T <- length(data.full$NQ) # number of observations for return y
pred <- matrix(nrow=T, ncol=1)
spread <- matrix(nrow=T, ncol=1)
volat <- matrix(nrow=T, ncol=1)
coef <- matrix(nrow=T, ncol=1)
wm <- 3625 #burnin
we <- 1000 #look-back
```

```
for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
  t2 = t - 1; # end of the data window
  window <- data.full$NQ[t1:t2] # data for estimation
  lm_p <- lm(data.full$NQ[t1:t2] ~ data.full$DJI[t1:t2] )
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <- lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1]  <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
    pred[t,1]  <- 1
  }
  else{
    pred[t,1] <- 0
  }
}



sprd <- spread[-c(4106),1]
sprd2 <- sprd[3626:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)

position <- pred[-c(4106),1]
position2 <- position[3626:length(position)]


vol <- volat[-c(4106),1]
vol2 <- vol[3626:length(vol)]


ts_spread <- xts(sprd3, order.by = index(data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_pred <- xts(position2, order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]] )

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index( data.full.rtn)[3626:dim( data.full.rtn)[1]]

plot(index( data.full.rtn)[3626:dim( data.full.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
     xlab= "Year", ylab = " ", main = "Spread")
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], (-2) * ts_volat, type="l",
      col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)





####### Strategy 3 -----------

T <- length(data.full$NQ) # number of observations for return y
pred <- matrix(nrow=T,ncol=1)
spread <- matrix(nrow=T,ncol=1)
volat <- matrix(nrow=T,ncol=1)
coef <- matrix(nrow=T, ncol=1)
wm <- 3625 #burnin
we <- 1000 #look-back

for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
```

```
  t2 = t - 1; # end of the data window
  window <- data.full$NQ[t1:t2] # data for estimation
  lm_p <- lm(data.full$NQ[t1:t2] ~ data.full$DJI[t1:t2] )
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <- lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1]  <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
    pred[t,1]  <- 1
  }
  else{
    pred[t,1] <- NA
  }
}



sprd <- spread[-c(4106),1]
sprd2 <- sprd[3626:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)

position <- pred2[-c(4106),1]
position2 <- position[3626:length(position)]


vol <- volat[-c(4106),1]
vol2 <- vol[3626:length(vol)]


ts_spread <- xts(sprd3, order.by = index(data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_pred <- xts(position2, order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]])

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index( data.full.rtn)[3626:dim( data.full.rtn)[1]]

plot(index( data.full.rtn)[3626:dim( data.full.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
     xlab= "Year", ylab = " ", main = "Spread")
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], (-2) * ts_volat, type="l",
      col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)


##### LOG PRICES --------


###### STRATEGY 1 -------------



T <- length(data.full$NQ) # number of observations for return y
pred <- matrix(nrow=T, ncol=1)
spread <- matrix(nrow=T, ncol=1)
volat <- matrix(nrow=T, ncol=1)
coef <- matrix(nrow =T, ncol=1)
pred[1:3625, ] <- 0
wm <- 3625 #burnin
we <- 1000 #look-back

for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
```

```
  t2 = t - 1; # end of the data window
  window <- data.full$NQ[t1:t2] # data for estimation
  lm_p <- lm(log(data.full$NQ[t1:t2]) ~  log(data.full$DJI[t1:t2]) )
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <- lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1]  <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
    pred[t,1]  <- 1
  }
  else{
    if ((pred[t-1,1] == 1) & (lm_p$residuals[length(lm_p$residuals)] > 0 )){
      pred[t,1] <- 0
    }
    else if ( (pred[t-1,1] == -1) & (lm_p$residuals[length(lm_p$residuals)] < 0 ) ) {
      pred[t,1] <- 0
    }
    else if ( (pred[t-1,1] == -1) & (lm_p$residuals[length(lm_p$residuals)] > 0 ) ) {
      pred[t,1] <- -1
    }
    else if ( (pred[t-1,1] == 1) & (lm_p$residuals[length(lm_p$residuals)] < 0 ) ) {
      pred[t,1] <- 1
    }
    else if ( (pred[t-1,1] == 0) & (lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)) ) {
      pred[t,1] <-  - 1
    }
    else if ( (pred[t-1,1] == 0) & (lm_p$residuals[length(lm_p$residuals)] < -2 * sd(temp_res)) ) {
      pred[t,1] <- 1
    }
    else {
      pred[t,1] <- 0
    }
  }
}



sprd <- spread[-c(4106),1]
sprd2 <- sprd[3626:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)

position <- pred[-c(4106),1]
position2 <- position[3626:length(position)]


vol <- volat[-c(4106),1]
vol2 <- vol[3626:length(vol)]


ts_spread <- xts(sprd3, order.by = index(data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_pred <- xts(position2, order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]] )

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index( data.full.rtn)[3626:dim( data.full.rtn)[1]]

plot(index( data.full.rtn)[3626:dim( data.full.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
     xlab= "Year", ylab = " ", main = "Spread")
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], (-2) * ts_volat, type="l",
      col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)
```

```
####Strategy 2 ------

T <- length(data.full$NQ) # number of observations for return y
pred <- matrix(nrow=T,ncol=1)
spread <- matrix(nrow=T,ncol=1)
volat <- matrix(nrow=T,ncol=1)
coef <-matrix(nrow=T, ncol=1)
wm <- 3625 #burnin
we <- 1000 #look-back

for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
  t2 = t - 1; # end of the data window
  window <- data.full$NQ[t1:t2] # data for estimation
  lm_p <- lm(log(data.full$NQ[t1:t2]) ~ log(data.full$DJI[t1:t2]) )
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <- lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1]  <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
    pred[t,1]  <- 1
  }
  else{
    pred[t,1] <- 0
  }
}



sprd <- spread[-c(4106),1]
sprd2 <- sprd[3626:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)

position <- pred[-c(4106),1]
position2 <- position[3626:length(position)]


vol <- volat[-c(4106),1]
vol2 <- vol[3626:length(vol)]


ts_spread <- xts(sprd3, order.by = index(data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_pred <- xts(position2, order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]] )

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index( data.full.rtn)[3626:dim( data.full.rtn)[1]]

plot(index( data.full.rtn)[3626:dim( data.full.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
     xlab= "Year", ylab = " ", main = "Spread")
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], (-2) * ts_volat, type="l",
      col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)
```

```
####### Strategy 3 -----------

T <- length(data.full$NQ) # number of observations for return y
pred <- matrix(nrow=T,ncol=1)
spread <- matrix(nrow=T,ncol=1)
volat <- matrix(nrow=T,ncol=1)
coef <- matrix(nrow=T, ncol=1)
wm <- 3625 #burnin
we <- 1000 #look-back

for (t in (wm + 1) : T){
  t1 = t - we; # start of the data window
  if(t1 < 1){ t1 = 1}
  t2 = t - 1; # end of the data window
  window <- data.full$NQ[t1:t2] # data for estimation
  lm_p <- lm(log(data.full$NQ[t1:t2]) ~ log(data.full$DJI[t1:t2]) )
  temp_res <- lm_p$residuals
  volat[t,1] <- sd(temp_res)
  spread[t,1] <- lm_p$residuals[length(lm_p$residuals)]
  coef[t,1] <- lm_p$coefficients[2]
  if(lm_p$residuals[length(lm_p$residuals)] > 2 * sd(temp_res)){
    pred[t,1]  <- (- 1)
  }
  else if ( lm_p$residuals[length(lm_p$residuals)] < (- 2) * sd(temp_res) ) {
    pred[t,1]  <- 1
  }
  else{
    pred[t,1] <- NA
  }
}


sprd <- spread[-c(4106),1]
sprd2 <- sprd[3626:length(sprd)]
sprd3 <- sprd2 / sd(sprd2)

position <- pred2[-c(4106),1]
position2 <- position[3626:length(position)]


vol <- volat[-c(4106),1]
vol2 <- vol[3626:length(vol)]


ts_spread <- xts(sprd3, order.by = index(data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_pred <- xts(position2, order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]] )
ts_volat <- xts( (vol2 / sd(sprd2)), order.by = index( data.full.rtn)[3626:dim( data.full.rtn)[1]] )

plot(ts_spread)
plot(ts_pred, main= "Positions")



frame <- data.frame(ts_spread, ts_pred, row.names = index( data.full.rtn)[3626:dim( data.full.rtn)[1]]

plot(index( data.full.rtn)[3626:dim( data.full.rtn)[1]], frame[, 1], type="l", ylim= c(-3,3),
     xlab= "Year", ylab = " ", main = "Spread")
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], frame[, 2], type="l",
      col = "cornflowerblue",lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], 2 * ts_volat, type="l",
      col = "red", lty = 2,lwd = 2)
lines(index(data.full.rtn)[3626:dim(data.full.rtn)[1]], (-2) * ts_volat, type="l",
      col = "red",lwd = 2, lty = 2 )
abline(h=0, col="blueviolet",lwd = 2)
```

```
#####Investigation for possible pairs-------
rm(list=ls())
cat("\014")

library(quantmod) # Load the package
library(tseries)
library(readxl)
library(car)
library(dplyr)
library(urca)
library(lmtest)
library(tidyverse)

mySymbols <- c('AAPL', 'AMGN', 'AXP', 'BA', 'CAT', 'CRM', 'CSCO', 'CVX', 'DIS',
               'GS', 'HD', 'HON', 'IBM', 'INTC', 'JNJ', 'JPM','KO', 'MCD', 'MMM',
               'MRK', 'MSFT', 'NKE', 'PG', 'TRV', 'UNH', 'V', 'VZ', 'WBA', 'WMT')
myFut <-lapply(mySymbols, function(x) {getSymbols(x, from="2014-01-01", to="2022-01-01",
               periodicity = "daily", auto.assign=FALSE)} )


names(myFut) <- mySymbols
closePrices <- lapply(myFut, Cl)
closePrices <- do.call(merge, closePrices)
names(closePrices)<-sub("\\.Close", "", names(closePrices))
head(closePrices)
closePrices <- na.omit(closePrices)
logPrices <- log(closePrices)
logPrices <- na.omit(logPrices)



### 2014-2016 --


train<-logPrices[1:504]


# get the correlation of each pair
left_side<-NULL
right_side<-NULL
correlation<-NULL
beta<-NULL
pvalue<-NULL
for (i in 1:length(mySymbols)) {
  for (j in 1:length(mySymbols)) {
    if (i>j) {
      left_side<-c(left_side, mySymbols[i])
      right_side<-c(right_side, mySymbols[j])
      correlation<-c(correlation, cor(train[,i], train[,j]))
      m<-lm(train[,i]~train[,j])
      beta<-c(beta, as.numeric(coef(m)[1]))
      # get the mispricings of the spread
      sprd<-residuals(m)
      # adf test
      pvalue<-c(pvalue, adf.test(sprd, alternative="stationary", k=0)$p.value)
    }
  }
}

df<-data.frame(left_side, right_side, correlation, beta, pvalue)
mypairs1<-df%>%filter(pvalue<= 0.05, correlation >0.95)%>%arrange(-correlation)
mypairs1




### 2016-2018 --
```

```
train<-logPrices[505:1007]



# get the correlation of each pair
left_side<-NULL
right_side<-NULL
correlation<-NULL
beta<-NULL
pvalue<-NULL
for (i in 1:length(mySymbols)) {
  for (j in 1:length(mySymbols)) {
    if (i>j) {
      left_side<-c(left_side, mySymbols[i])
      right_side<-c(right_side, mySymbols[j])
      correlation<-c(correlation, cor(train[,i], train[,j]))
      m<-lm(train[,i]~train[,j])
      beta<-c(beta, as.numeric(coef(m)[1]))
      # get the mispricings of the spread
      sprd<-residuals(m)
      # adf test
      pvalue<-c(pvalue, adf.test(sprd, alternative="stationary", k=0)$p.value)
    }
  }
}

df<-data.frame(left_side, right_side, correlation, beta, pvalue)
mypairs2<-df%>%filter(pvalue<= 0.05, correlation>0.95)%>%arrange(-correlation)
mypairs2



### 2018-2020 --

train<-logPrices[1008:1510]



# get the correlation of each pair
left_side<-NULL
right_side<-NULL
correlation<-NULL
beta<-NULL
pvalue<-NULL
for (i in 1:length(mySymbols)) {
  for (j in 1:length(mySymbols)) {
    if (i>j) {
      left_side<-c(left_side, mySymbols[i])
      right_side<-c(right_side, mySymbols[j])
      correlation<-c(correlation, cor(train[,i], train[,j]))
      m<-lm(train[,i]~train[,j])
      beta<-c(beta, as.numeric(coef(m)[1]))
      # get the mispricings of the spread
      sprd<-residuals(m)
      # adf test
      pvalue<-c(pvalue, adf.test(sprd, alternative="stationary", k=0)$p.value)
    }
  }
}

df<-data.frame(left_side, right_side, correlation, beta, pvalue)
mypairs3<-df%>%filter(pvalue<= 0.05, correlation>0.95)%>%arrange(-correlation)
mypairs3


### 2020-2022 --
```

```
train<-logPrices[1510:2015]



# get the correlation of each pair
left_side<-NULL
right_side<-NULL
correlation<-NULL
beta<-NULL
pvalue<-NULL
for (i in 1:length(mySymbols)) {
  for (j in 1:length(mySymbols)) {
    if (i>j) {
      left_side<-c(left_side, mySymbols[i])
      right_side<-c(right_side, mySymbols[j])
      correlation<-c(correlation, cor(train[,i], train[,j]))
      m<-lm(train[,i]~train[,j])
      beta<-c(beta, as.numeric(coef(m)[1]))
      # get the mispricings of the spread
      sprd<-residuals(m)
      # adf test
      pvalue<-c(pvalue, adf.test(sprd, alternative="stationary", k=0)$p.value)
    }
  }
}

df<-data.frame(left_side, right_side, correlation, beta, pvalue)
mypairs4<-df%>%filter(pvalue<= 0.05, correlation>0.95)%>%arrange(-correlation)
mypairs4
```

Appendix B

Tables

CHI-SQUARED PERCENTAGE POINTS (continued)											
ν	0.1%	0.5%	1.0%	2.5%	5.0%	10.0%	12.5%	20.0%	25.0%	33.3%	50.0%
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337
26	9.222	11.160	12.198	13.844	15.379	17.292	18.021	19.820	20.843	22.399	25.336
27	9.803	11.808	12.879	14.573	16.151	18.114	18.861	20.703	21.749	23.339	26.336
28	10.391	12.461	13.565	15.308	16.928	18.939	19.704	21.588	22.657	24.280	27.336
29	10.986	13.121	14.256	16.047	17.708	19.768	20.550	22.475	23.567	25.222	28.336
30	11.588	13.787	14.953	16.791	18.493	20.599	21.399	23.364	24.478	26.165	29.336
35	14.688	17.192	18.509	20.569	22.465	24.797	25.678	27.836	29.054	30.894	34.336
40	17.916	20.707	22.164	24.433	26.509	29.051	30.008	32.345	33.660	35.643	39.335
45	21.251	24.311	25.901	28.366	30.612	33.350	34.379	36.884	38.291	40.407	44.335
50	24.674	27.991	29.707	32.357	34.764	37.689	38.785	41.449	42.942	45.184	49.335
55	28.173	31.735	33.570	36.398	38.958	42.060	43.220	46.036	47.610	49.972	54.335
60	31.738	35.534	37.485	40.482	43.188	46.459	47.680	50.641	52.294	54.770	59.335

Table B.1: $\mathcal{X}^2$ distribution

CHI-SQUARED PERCENTAGE POINTS (continued)											
ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	60.275	66.619
40	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661
55	57.016	58.945	61.665	63.577	67.211	68.796	73.311	77.380	82.292	85.749	93.168
60	62.135	64.147	66.981	68.972	72.751	74.397	79.082	83.298	88.379	91.952	99.607

STUDENT'S t PERCENTAGE POINTS											
ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.255	0.434	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.434	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307
45	0.255	0.434	0.680	0.850	1.165	1.301	1.679	2.014	2.412	2.690	3.281
50	0.255	0.433	0.679	0.849	1.164	1.299	1.676	2.009	2.403	2.678	3.261
55	0.255	0.433	0.679	0.848	1.163	1.297	1.673	2.004	2.396	2.668	3.245
60	0.254	0.433	0.679	0.848	1.162	1.296	1.671	2.000	2.390	2.660	3.232
∞	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090

Table B.2: Student's t percentage points

		PERCENTAGE POINTS OF THE F DISTRIBUTION													
$\nu_2 \backslash \nu_l$		2	3	4	5	6	7	8	10	12	15	20	30	50	∞
1	0.500	1.50	1.71	1.82	1.89	1.94	1.98	2.00	2.04	2.07	2.09	2.12	2.15	2.17	2.20
	0.600	2.63	2.93	3.09	3.20	3.27	3.32	3.36	3.41	3.45	3.48	3.52	3.56	3.59	3.64
	0.667	4.00	4.42	4.64	4.78	4.88	4.95	5.00	5.08	5.13	5.18	5.24	5.29	5.33	5.39
	0.750	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.32	9.41	9.50	9.58	9.67	9.74	9.85
	0.800	12.0	13.1	13.6	14.0	14.3	14.4	14.6	14.8	14.9	15.0	15.2	15.3	15.4	15.6
2	0.500	1.00	1.13	1.21	1.25	1.28	1.30	1.32	1.35	1.36	1.38	1.39	1.41	1.42	1.44
	0.600	1.50	1.64	1.72	1.76	1.80	1.82	1.84	1.86	1.88	1.89	1.91	1.92	1.94	1.96
	0.667	2.00	2.15	2.22	2.27	2.30	2.33	2.34	2.37	2.38	2.40	2.42	2.43	2.45	2.47
	0.750	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.38	3.39	3.41	3.43	3.44	3.46	3.48
	0.800	4.00	4.16	4.24	4.28	4.32	4.34	4.36	4.38	4.40	4.42	4.43	4.45	4.47	4.48
3	0.500	0.88	1.00	1.06	1.10	1.13	1.15	1.16	1.18	1.20	1.21	1.23	1.24	1.25	1.27
	0.600	1.26	1.37	1.43	1.47	1.49	1.51	1.52	1.54	1.55	1.56	1.57	1.58	1.59	1.60
	0.667	1.62	1.72	1.77	1.80	1.82	1.83	1.84	1.86	1.87	1.88	1.89	1.90	1.90	1.91
	0.750	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.45	2.46	2.46	2.47	2.47	2.47
	0.800	2.89	2.94	2.96	2.97	2.97	2.97	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98
4	0.500	0.83	0.94	1.00	1.04	1.06	1.08	1.09	1.11	1.13	1.14	1.15	1.16	1.18	1.19
	0.600	1.16	1.26	1.31	1.34	1.36	1.37	1.38	1.40	1.41	1.42	1.43	1.43	1.44	1.45
	0.667	1.46	1.55	1.58	1.61	1.62	1.63	1.64	1.65	1.65	1.66	1.67	1.67	1.68	1.68
	0.750	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
	0.800	2.47	2.48	2.48	2.48	2.47	2.47	2.47	2.46	2.46	2.45	2.44	2.44	2.43	2.43
5	0.500	0.80	0.91	0.96	1.00	1.02	1.04	1.05	1.07	1.09	1.10	1.11	1.12	1.13	1.15
	0.600	1.11	1.20	1.24	1.27	1.29	1.30	1.31	1.32	1.33	1.34	1.34	1.35	1.36	1.37
	0.667	1.38	1.45	1.48	1.50	1.51	1.52	1.53	1.53	1.54	1.54	1.54	1.55	1.55	1.55
	0.750	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.87
	0.800	2.26	2.25	2.24	2.23	2.22	2.21	2.20	2.19	2.18	2.18	2.17	2.16	2.15	2.13
6	0.500	0.78	0.89	0.94	0.98	1.00	1.02	1.03	1.05	1.06	1.07	1.08	1.10	1.11	1.12
	0.600	1.07	1.16	1.20	1.22	1.24	1.25	1.26	1.27	1.28	1.29	1.29	1.30	1.31	1.31
	0.667	1.33	1.39	1.42	1.44	1.44	1.45	1.45	1.46	1.46	1.47	1.47	1.47	1.47	1.47
	0.750	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.76	1.76	1.75	1.75	1.74
	0.800	2.13	2.11	2.09	2.08	2.06	2.05	2.04	2.03	2.02	2.01	2.00	1.98	1.97	1.95
7	0.500	0.77	0.87	0.93	0.96	0.98	1.00	1.01	1.03	1.04	1.05	1.07	1.08	1.09	1.10
	0.600	1.05	1.13	1.17	1.19	1.21	1.22	1.23	1.24	1.24	1.25	1.26	1.26	1.27	1.27
	0.667	1.29	1.35	1.38	1.39	1.40	1.40	1.41	1.41	1.41	1.41	1.41	1.42	1.42	1.42
	0.750	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.68	1.68	1.67	1.66	1.66	1.65
	0.800	2.04	2.02	1.99	1.97	1.96	1.94	1.93	1.92	1.91	1.89	1.88	1.86	1.85	1.83
8	0.500	0.76	0.86	0.91	0.95	0.97	0.99	1.00	1.02	1.03	1.04	1.05	1.07	1.07	1.09
	0.600	1.03	1.11	1.15	1.17	1.19	1.20	1.20	1.21	1.22	1.22	1.23	1.24	1.24	1.25
	0.667	1.26	1.32	1.35	1.36	1.36	1.37	1.37	1.37	1.37	1.38	1.38	1.38	1.37	1.37
	0.750	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.62	1.62	1.61	1.60	1.59	1.58
	0.800	1.98	1.95	1.92	1.90	1.88	1.87	1.86	1.84	1.83	1.81	1.80	1.78	1.76	1.74

Table B.3: F Distribution percentage points

		PERCENTAGE POINTS OF THE F DISTRIBUTION (continued)													
$\nu_2 \backslash \nu_l$		2	3	4	5	6	7	8	10	12	15	20	30	50	∞
9	0.500	0.75	0.85	0.91	0.94	0.96	0.98	0.99	1.01	1.02	1.03	1.04	1.05	1.06	1.08
	0.600	1.02	1.10	1.13	1.15	1.17	1.18	1.18	1.19	1.20	1.21	1.21	1.22	1.22	1.22
	0.667	1.24	1.30	1.32	1.33	1.34	1.34	1.34	1.34	1.35	1.35	1.35	1.34	1.34	1.34
	0.750	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.58	1.57	1.56	1.55	1.54	1.53
	0.800	1.93	1.90	1.87	1.85	1.83	1.81	1.80	1.78	1.76	1.75	1.73	1.71	1.70	1.67
10	0.500	0.74	0.85	0.90	0.93	0.95	0.97	0.98	1.00	1.01	1.02	1.03	1.05	1.06	1.07
	0.600	1.01	1.08	1.12	1.14	1.15	1.16	1.17	1.18	1.18	1.19	1.19	1.20	1.20	1.21
	0.667	1.23	1.28	1.30	1.31	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.31
	0.750	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.55	1.54	1.53	1.52	1.51	1.50	1.48
	0.800	1.90	1.86	1.83	1.80	1.78	1.77	1.75	1.73	1.72	1.70	1.68	1.66	1.65	1.62
11	0.500	0.74	0.84	0.89	0.93	0.95	0.96	0.98	0.99	1.01	1.02	1.03	1.04	1.05	1.06
	0.600	1.00	1.07	1.11	1.13	1.14	1.15	1.16	1.17	1.17	1.18	1.18	1.18	1.19	1.19
	0.667	1.22	1.27	1.29	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.29
	0.750	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.45
	0.800	1.87	1.83	1.80	1.77	1.75	1.73	1.72	1.69	1.68	1.66	1.64	1.62	1.60	1.57
12	0.500	0.73	0.84	0.89	0.92	0.94	0.96	0.97	0.99	1.00	1.01	1.02	1.03	1.04	1.06
	0.600	0.99	1.07	1.10	1.12	1.13	1.14	1.15	1.16	1.16	1.17	1.17	1.17	1.18	1.18
	0.667	1.21	1.26	1.27	1.28	1.29	1.29	1.29	1.29	1.29	1.29	1.29	1.28	1.28	1.27
	0.750	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.45	1.44	1.42
	0.800	1.85	1.80	1.77	1.74	1.72	1.70	1.69	1.66	1.65	1.63	1.61	1.59	1.57	1.54
13	0.500	0.73	0.83	0.88	0.92	0.94	0.96	0.97	0.98	1.00	1.01	1.02	1.03	1.04	1.05
	0.600	0.98	1.06	1.09	1.11	1.13	1.13	1.14	1.15	1.15	1.16	1.16	1.16	1.17	1.17
	0.667	1.20	1.25	1.26	1.27	1.28	1.28	1.28	1.28	1.28	1.28	1.27	1.27	1.27	1.26
	0.750	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.43	1.42	1.40
	0.800	1.83	1.78	1.75	1.72	1.69	1.68	1.66	1.64	1.62	1.60	1.58	1.56	1.54	1.51
14	0.500	0.73	0.83	0.88	0.91	0.94	0.95	0.96	0.98	0.99	1.00	1.01	1.03	1.04	1.05
	0.600	0.98	1.05	1.09	1.11	1.12	1.13	1.13	1.14	1.14	1.15	1.15	1.16	1.16	1.16
	0.667	1.19	1.24	1.26	1.26	1.27	1.27	1.27	1.27	1.27	1.26	1.26	1.26	1.25	1.24
	0.750	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.41	1.40	1.38
	0.800	1.81	1.76	1.73	1.70	1.67	1.65	1.64	1.62	1.60	1.58	1.56	1.53	1.51	1.48
15	0.500	0.73	0.83	0.88	0.91	0.93	0.95	0.96	0.98	0.99	1.00	1.01	1.02	1.03	1.05
	0.600	0.97	1.05	1.08	1.10	1.11	1.12	1.13	1.13	1.14	1.14	1.15	1.15	1.15	1.15
	0.667	1.18	1.23	1.25	1.25	1.26	1.26	1.26	1.26	1.26	1.25	1.25	1.25	1.24	1.23
	0.750	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.41	1.40	1.38	1.36
	0.800	1.80	1.75	1.71	1.68	1.66	1.64	1.62	1.60	1.58	1.56	1.54	1.51	1.49	1.46
16	0.500	0.72	0.82	0.88	0.91	0.93	0.95	0.96	0.97	0.99	1.00	1.01	1.02	1.03	1.04
	0.600	0.97	1.04	1.08	1.10	1.11	1.12	1.12	1.13	1.13	1.14	1.14	1.14	1.14	1.14
	0.667	1.18	1.22	1.24	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.24	1.24	1.23	1.22
	0.750	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.43	1.41	1.40	1.38	1.37	1.34
	0.800	1.78	1.74	1.70	1.67	1.64	1.62	1.61	1.58	1.56	1.54	1.52	1.49	1.47	1.43
17	0.500	0.72	0.82	0.87	0.91	0.93	0.94	0.96	0.97	0.98	0.99	1.01	1.02	1.03	1.04
	0.600	0.97	1.04	1.07	1.09	1.10	1.11	1.12	1.12	1.13	1.13	1.13	1.14	1.14	1.14
	0.667	1.17	1.22	1.23	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.23	1.23	1.22	1.21
	0.750	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.41	1.40	1.39	1.37	1.36	1.33
	0.800	1.77	1.72	1.68	1.65	1.63	1.61	1.59	1.57	1.55	1.53	1.50	1.48	1.46	1.42

		PERCENTAGE POINTS OF THE F DISTRIBUTION (continued)													
$\nu_2 \backslash \nu_l$		2	3	4	5	6	7	8	10	12	15	20	30	50	∞
18	0.500	0.72	0.82	0.87	0.90	0.93	0.94	0.95	0.97	0.98	0.99	1.00	1.02	1.02	1.04
	0.600	0.96	1.04	1.07	1.09	1.10	1.11	1.11	1.12	1.12	1.13	1.13	1.13	1.13	1.13
	0.667	1.17	1.21	1.23	1.24	1.24	1.24	1.24	1.24	1.23	1.23	1.23	1.22	1.22	1.21
	0.750	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.40	1.39	1.38	1.36	1.34	1.32
	0.800	1.76	1.71	1.67	1.64	1.62	1.60	1.58	1.55	1.53	1.51	1.49	1.46	1.44	1.40
19	0.500	0.72	0.82	0.87	0.90	0.92	0.94	0.95	0.97	0.98	0.99	1.00	1.01	1.02	1.04
	0.600	0.96	1.03	1.07	1.09	1.10	1.10	1.11	1.12	1.12	1.12	1.13	1.13	1.13	1.13
	0.667	1.16	1.21	1.22	1.23	1.23	1.23	1.23	1.23	1.23	1.23	1.22	1.22	1.21	1.20
	0.750	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.40	1.38	1.37	1.35	1.33	1.30
	0.800	1.75	1.70	1.66	1.63	1.61	1.58	1.57	1.54	1.52	1.50	1.48	1.45	1.43	1.39
20	0.500	0.72	0.82	0.87	0.90	0.92	0.94	0.95	0.97	0.98	0.99	1.00	1.01	1.02	1.03
	0.600	0.96	1.03	1.06	1.08	1.09	1.10	1.11	1.11	1.12	1.12	1.12	1.12	1.12	1.12
	0.667	1.16	1.21	1.22	1.23	1.23	1.23	1.23	1.23	1.22	1.22	1.22	1.21	1.20	1.19
	0.750	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.40	1.39	1.37	1.36	1.34	1.32	1.29
	0.800	1.75	1.70	1.65	1.62	1.60	1.58	1.56	1.53	1.51	1.49	1.47	1.44	1.41	1.37
21	0.500	0.72	0.81	0.87	0.90	0.92	0.94	0.95	0.96	0.98	0.99	1.00	1.01	1.02	1.03
	0.600	0.96	1.03	1.06	1.08	1.09	1.10	1.10	1.11	1.11	1.12	1.12	1.12	1.12	1.12
	0.667	1.16	1.20	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.21	1.20	1.20	1.19
	0.750	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.39	1.38	1.37	1.35	1.33	1.32	1.28
	0.800	1.74	1.69	1.65	1.61	1.59	1.57	1.55	1.52	1.50	1.48	1.46	1.43	1.40	1.36
22	0.500	0.72	0.81	0.87	0.90	0.92	0.93	0.95	0.96	0.97	0.99	1.00	1.01	1.02	1.03
	0.600	0.96	1.03	1.06	1.08	1.09	1.10	1.10	1.11	1.11	1.11	1.12	1.12	1.12	1.12
	0.667	1.16	1.20	1.21	1.22	1.22	1.22	1.22	1.22	1.21	1.21	1.21	1.20	1.19	1.18
	0.750	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.37	1.36	1.34	1.32	1.31	1.28
	0.800	1.73	1.68	1.64	1.61	1.58	1.56	1.54	1.51	1.49	1.47	1.45	1.42	1.39	1.35
23	0.500	0.71	0.81	0.86	0.90	0.92	0.93	0.95	0.96	0.97	0.98	1.00	1.01	1.02	1.03
	0.600	0.95	1.02	1.06	1.07	1.09	1.09	1.10	1.10	1.11	1.11	1.11	1.11	1.11	1.11
	0.667	1.15	1.20	1.21	1.22	1.22	1.22	1.22	1.21	1.21	1.21	1.20	1.19	1.19	1.17
	0.750	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.38	1.37	1.35	1.34	1.32	1.30	1.27
	0.800	1.73	1.68	1.63	1.60	1.57	1.55	1.53	1.51	1.49	1.46	1.44	1.41	1.38	1.34
24	0.500	0.71	0.81	0.86	0.90	0.92	0.93	0.94	0.96	0.97	0.98	0.99	1.01	1.01	1.03
	0.600	0.95	1.02	1.06	1.07	1.08	1.09	1.10	1.10	1.10	1.11	1.11	1.11	1.11	1.11
	0.667	1.15	1.19	1.21	1.21	1.21	1.21	1.21	1.21	1.21	1.20	1.20	1.19	1.18	1.17
	0.750	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.36	1.35	1.33	1.31	1.29	1.26
	0.800	1.72	1.67	1.63	1.59	1.57	1.55	1.53	1.50	1.48	1.46	1.43	1.40	1.38	1.33
25	0.500	0.71	0.81	0.86	0.89	0.92	0.93	0.94	0.96	0.97	0.98	0.99	1.00	1.01	1.03
	0.600	0.95	1.02	1.05	1.07	1.08	1.09	1.09	1.10	1.10	1.11	1.11	1.11	1.11	1.11
	0.667	1.15	1.19	1.21	1.21	1.21	1.21	1.21	1.21	1.20	1.20	1.19	1.19	1.18	1.16
	0.750	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.37	1.36	1.34	1.33	1.31	1.29	1.25
	0.800	1.72	1.66	1.62	1.59	1.56	1.54	1.52	1.49	1.47	1.45	1.42	1.39	1.37	1.32
26	0.500	0.71	0.81	0.86	0.89	0.91	0.93	0.94	0.96	0.97	0.98	0.99	1.00	1.01	1.03
	0.600	0.95	1.02	1.05	1.07	1.08	1.09	1.09	1.10	1.10	1.10	1.10	1.11	1.11	1.10
	0.667	1.15	1.19	1.20	1.21	1.21	1.21	1.21	1.20	1.20	1.20	1.19	1.18	1.18	1.16
	0.750	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.35	1.34	1.32	1.30	1.28	1.25
	0.800	1.71	1.66	1.62	1.58	1.56	1.53	1.52	1.49	1.47	1.44	1.42	1.39	1.36	1.31

Critical Values for the Phillips-Perron Z_ρ, Test and for the Dickey-Fuller Test Based on Estimated OLS Autoregrassive Coefficient

Sample Size T	Probability that $(\rho-1)$, is less than entry							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
Case 1								
25	-11.9	-9.3	-7.3	-5,3	1.01	1.40	1.79	2.28
50	-12.9	-9.9	-7.7	-5.5	0.97	1.35	1.70	2.16
100	-13.3	-10.2	-7.9	-5.6	0.95	1.31	1.65	2.09
250	-13.6	-10.3	-8.0	-5.7	0.93	1.28	1.62	2.04
500	-13.7	-10.4	-8.0	-5.7	0.93	1.28	1.61	2.04
∞	-13.8	-10.5	-8.1	-5.7	0.93	1.28	1.60	2.03
Case 2								
25	-17.2	-14.6	-12.5	-10.2	-0.76	0.01	0.65	1.40
50	-18.9	-15.7	-13.3	-10.7	-0.81	-0.07	0.53	1.22
100	-19.8	-163	-13.7	-11.0	-0.83	-0.10	0.47	1.14
250	-20.3	-16.6	-14.0	-11.2	-0.84	-0.12	0.43	1.09
500	-20.5	-16.8	-140	-112	-0.84	-0.13	0.42	1.06
∞	-20.7	-16.9	-14.1	-11.3	-0.85	-0.13	0.41	1.04
Case 4								
25	-22.5	-19.9	-17.9	-15.6	-3.66	-2.51	-1.53	-0.43
50	-25.7	-22.4	-19.8	-16.8	-3.71	-2.60	-1.66	-0.65
100	-27.4	-23.6	-20.7	-17.5	-3.74	-2.62	-1.73	-0.75
250	-28.4	-24.4	-21.3	-18.0	-3.75	-2.64	-1.78	-0.82
500	-28.9	-24.8	-21.5	-18.1	-3.76	-2.65	-1.78	-0.84
∞	-29.5	-25.1	-21.8	-18.3	-3.77	-2.66	-1.79	-0.87

Table B.4: Source: Wayne A. Fuller, Introduction to Statistical Time Series, Wiley, New York, 1976, p. 371.

Critical Values for the Phillips-Perron Z_t, Test and for the Dickey-Fuller Test Based on Estimated OLS t Statistic

Sample Size T	Probability that $(\rho-1)/\hat{\sigma}_\beta$, is less than entry							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
Case 1								
25	- 2.66	-2.26	- 1.95	- 1.60	0.92	1.33	1.70	2.16
50	- 2.62	- 2.25	-1.95	- 1.61	0.91	1.31	1.66	2.08
100	-2.60-	-2.24	-1.95	-1.61	0.90	1.29	1.64	2.03
250	- 2.58	- 2.23	-1.95	- 1.62	0.89	1.29	1.63	2.01
500	- 2.58	-2.23	-1,95	- 1.62	0.89	1.28	1.62	2.00
∞	-2.58	- 2.23	-1.95	- 1.62	0.89	1.28	1.62	2.00
Case 2								
25	-3.75	-3.33	- 3.00	- 2.63	- 0.37	0.00	0.34	0.72
50	- 3.58	-3,22	-2.93	- 2.60	- 0.40	- 0.03	0.29	0.66
100	-3.51	-3.17	-2.89	-2.58	- 0.42	-0.05	0.26	0.63
250	- 3.46	- 3.14	- 2.88	- 2.57	-0.42	- 0.06	0.24	0.62
500	-3.44	- 3.13	- 2.87	-2.57	- 0.43	- 0.07	0.24	0.61
∞	- 3.43	- 3.12	-2.86	- 2.57	- 0.44	- 0.07	0.23	0.60
Case 4								
25	- 4.38	-3.95	- 3.60	-3.24	-1.14	- 0.80	-0.50	-0.15
50	- 4.15	-3.80	-3.50	-3.18	-1.19	-0.87	-0.58	-0.24
100	- 4.04	3.73	-3.45	-3.15	- 1.22	-0.90	- 0.62	- 0.28
250	-3.99	-3.69	- 3.43	-3.13	- 1.23	-0.92	- 0.64	-0.31
500	-3.98	- 3.68	3.42	-3.13	- 1.24	-0.93	- 0.65	-0.32
∞	- 3.96	- 3.66	-3.41	-3.12	-1.25	- 0.94	- 0.66	- 0.33

Table B.5: Source: Wayne A. Fuller, Introduction to Statistical Time Series, Wiley, New York, 1976, p. 373.

Critical Values for the Dickey-Fuller Test
Based on Estimated OLS F Statistic

Sample Size T	Probability that F is less than entry							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.999
Case 2								
25	0.29	0.38	0.49	0.65	4.12	5.18	6.30	7.88
50	0.29	0.39	0.50	0.66	3.94	4.86	5.80	7.06
100	0.29	0.39	0.50	0.67	3.86	4.71	5.57	6.70
250	0.30	0.39	0.51	0.67	3.81	4.63	5.45	6.52
500	0.30	0.39	0.51	0.67	3.79	4.61	5.41	6.47
∞	0.30	0.40	0.51	0.67	3.78	4.59	5.38	6.43
Case 4								
25	0.74	0.90	1.08	1.33	5.91	7.24	8.65	10.61
50	0.76	0.93	1.11	1.37	5.61	6.73	7.81	9.31
100	0.76	0.94	1.12	1.38	5.47	6.49	7.44	8.73
250	0.76	0.94	1.13	1.39	5.39	6.34	7.25	8.43
500	0.76	0.94	1.13	1.39	5.36	6.30	7.20	8.34
∞	0.77	0.94	1.13	1.39	5.34	6.25	7.16	8.27

Table B.6: Source: David A. Dickey and Wayne A. Fuller, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," Econometrica 49 (1981), p. 1063.

Critical Values for the Phillips Z_ρ Statistic
When Applied to Residuals from Spurious Cointegrating Regression

Number of right-hand variables in regression, excluding trend or constant	Sample Size T	Probability that $\rho - 1)/\hat{\sigma}_\beta$, is less than entry						
		0.01	0.025	0.05	0.075	0.1	0.125	0.150
Case 1								
1	500	-22.8	-18.9	-15.6	-13.8	12.5	-11.6	-10.7
2	500	-29.3	-25.2	-21.5	-19.6	-18.2	-17.0	-16.0
3	500	-36.2	-31.5	-27.9	-25.5	-23.9	-22.6	-21.5
4	500	-42.9	-37.5	-33.5	-30.9	-28.9	-27.4	-26.2
5	500	-48.5	-42.5	-38.1	-35.5	-33.8	-32.3	-30.9
Case 2								
1	500	-28.3	-23.8	-20.5	-18.5	-17.0	-15.9	-14.9
2	500	-34.2	-29.7	-26.1	-23.9	-22.2	-21.0	-19.9
3	500	-41.1	-35.7	-32.1	-29.5	-27.6	-26.2	-25.1
4	500	-47.5	-41.6	-37.2	-34.7	-32.7	-31.2	-29.9
5	500	-52.2	-46.5	-41.9	-39.1	-37.0	-35.5	-34.2
Case 4								
1	500	-28.9	-24.8	-21.5	-	-18.1	-	-
2	500	-35.4	-30.8	-27.1	-24.8	-23.2	-21.8	-20.8
3	500	-40.3	-36.1	-32.2	-29.7	-27.8	-26.5	-25.3
4	500	-47.4	-42.6	-37.7	-35.0	-33.2	-31.7	-30.3
5	500	-53.6	-47.1	-42.5	-39.7	-37.7	-36.0	-34.6

Table B.7: Source: P. C. B. Phillips and S. Ouliaris, "Asymptotic Properties of Residual Based Tests for Cointegration", Econometrica 58 (1990), pp. 189-90. Also Wayne A. Fuller, Introduction to Statistical Time Series, Wiley, New York, 1976, p. 371.

Critical Values for the Phillips Z, Statistic or the Dickey-Fuller t Statistic When Applied to Residuals from Spurious Cointegrating Regression

Number of right-hand variables in regression, excluding trend or constant	**Sample Size** T	**Probability that** $(\hat{\rho}-1)/\hat{\sigma}_{\beta}$**, is less than entry**						
		0.01	0.025	0.05	0.075	0.1	0.125	0.150
					Case 1			
1	500	-3.39	-3.05	-2.76	-2.58	-2.45	-2.35	-2.26
2	500	-3.84	3.55	-3.27	-3.11	-2.99	-2.88	-2.79
3	500	-4.30	-3.99	-3.74	-3.57	-3.44	-3.35	-3.26
4	500	-4.67	-4.38	-413	-3.95	-3.81	-3.71	-3.61
5	500	-4.99	-4.67	-440	-4.25	-4.14	-4.04	-3.94
					Case 2			
1	500	-3.96	-3.64	-3.37	-3.20	-3.07	-2.96	-2.86
2	500	-4.31	-4.02	-3.77	-3.58	-3.45	-3.35	-3.26
3	500	-4.73	-4.37.	-4.11	-3.96	-3.83	-3.73	-3.65
4	500	-5.07	-4.71	-4.45	-429	-4.16	-4.05	-3.96
5	500	-5.28	-4.98	-4.71	-4.56	-4.43	-4.33	-4.24
					Case 4			
1	500	-3.98	-3.68	-3.42	-	-3.13	-	-
2	500	-4.36	-4.07	-3.80	-3.65	-3.52	3.42	3.33
3	500	-4.65	-4.39	-4.16	-3.98	-3.84	-3.74	-3.66
4	500	-5.04	-4.77	-4.49	-4.32	-4.20	-4.08	-4.00
5	500	-5.36	-5.02	-4.74	-4.58	-4.46	-4.36	-4.28

Table B.8: Source: C. B. Phillips and S. Ouliaris, "Asymptotic Properties of Residual Based Tests for Cointegration", Econometrica 58 (1990), p. 190. Also Wayne A. Fuller, Introduction to Statistical Time Series, Wiley, New York, 1976, p. 373.

Bibliography

[1] Sung K Ahn and Gregory C Reinsel. Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American statistical association*, 85(411):813–823, 1990.

[2] Richard T Baillie and David D Selover. Cointegration and models of exchange rate determination. *International Journal of Forecasting*, 3(1):43–51, 1987.

[3] Anindya Banerjee, Juan J Dolado, John W Galbraith, and David Hendry. *Co-integration, error correction, and the econometric analysis of non-stationary data.* Oxford university press, 1993.

[4] Josef Barunik. Lecture notes: Introduction to cointegration. 2010.

[5] Stephen Beveridge and Charles R Nelson. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary economics*, 7(2):151–174, 1981.

[6] P Bickel, P Diggle, S Fienberg, U Gather, I Olkin, and S Zeger. *Springer series in statistics.* Springer, 2009.

[7] Peter J Brockwell, Peter J Brockwell, Richard A Davis, and Richard A Davis. *Introduction to time series and forecasting.* Springer, 2016.

[8] Peter J Brockwell and Richard A Davis. *Time series: theory and methods.* Springer Science & Business Media, 2009.

[9] Andrius Buteikis. Lecture notes: Multivariate models: Granger causality, var and vecm models. 2017.

[10] John Y Campbell, Andrew W Lo, and A Craig MacKinlay. *The econometrics of financial markets.* princeton University press, 2012.

[11] John Y Campbell and Robert J Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228, 1988.

[12] John Y Campbell and Robert J Shiller. Stock prices, earnings, and expected dividends. *the Journal of Finance*, 43(3):661–676, 1988.

[13] Mehmet Caner and Lutz Kilian. Size distortions of tests of the null hypothesis of stationarity: evidence and implications for the ppp debate. *Journal of International Money and Finance*, 20(5):639–657, 2001.

[14] John C. Chao. Lecture notes: Additional topics in time series. 2017.

[15] Chris Chatfield. *The analysis of time series: an introduction.* Chapman and Hall/CRC, 2003.

[16] Gregory C Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, pages 591–605, 1960.

[17] Dean Corbae and Sam Ouliaris. Cointegration and tests of purchasing power parity. *The Review of Economics and Statistics*, pages 508–511, 1988.

[18] Paul SP Cowpertwait and Andrew V Metcalfe. *Introductory time series with R.* Springer Science & Business Media, 2009.

[19] Jon Danielsson. *Financial risk forecasting: The theory and practice of forecasting market risk with implementation in R and Matlab*, volume 588. John Wiley & Sons, 2011.

[20] James EH Davidson, David F Hendry, Frank Srba, and Stephen Yeo. Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the united kingdom. *The Economic Journal*, pages 661–692, 1978.

[21] David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

[22] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[23] BE Elliot, TJ Rothenberg, and JH Stock. Efficient tests of the unit root hypothesis. *Econometrica*, 64(8):13–36, 1996.

[24] Robert F Engle and Clive WJ Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

[25] Robert F Engle and Byung Sam Yoo. Forecasting and testing in co-integrated systems. *Journal of econometrics*, 35(1):143–159, 1987.

[26] Eugene F Fama. Mandelbrot and the stable paretian hypothesis. *The journal of business*, 36(4):420–429, 1963.

[27] Eugene F Fama. The behavior of stock-market prices. *The journal of Business*, 38(1):34–105, 1965.

[28] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

[29] Clive WJ Granger. Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and control*, 2:329–352, 1980.

[30] Clive WJ Granger and Paul Newbold. Spurious regressions in econometrics. *Journal of econometrics*, 2(2):111–120, 1974.

[31] Geoffrey Hall. Perceptual and associative learning. *Oxford psychology series*, 18:29–107, 1991.

[32] Peter Hall and CC Heyde. Rates of convergence in the martingale central limit theorem. *The Annals of Probability*, pages 395–404, 1981.

[33] James Douglas Hamilton. *Time series analysis*. Princeton university press, 1994.

[34] Christoph Hanck, Martin Arnold, Alexander Gerber, and Martin Schmelzer. Introduction to econometrics with r. *University of Duisburg-Essen*, 2019.

[35] Bruce E Hansen. Testing for parameter instability in linear models. *Journal of policy Modeling*, 14(4):517–533, 1992.

[36] Bruce E. Hansen. *Econometrics*. 2021.

[37] Niles Hansen. Do producer services induce regional economic development? *Journal of regional Science*, 30(4):465–476, 1990.

[38] John K. Hunter. Lecture notes: Metric spaces. 2020.

[39] Carlos M Jarque and Anil K Bera. A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, pages 163–172, 1987.

[40] Søren Johansen. Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2-3):231–254, 1988.

[41] Søren Johansen. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: journal of the Econometric Society*, pages 1551–1580, 1991.

[42] Søren Johansen et al. *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press on Demand, 1995.

[43] Soren Johansen and Katarina Juselius. Testing structural hypotheses in a multivariate cointegration analysis of the ppp and the uip for uk. *Journal of econometrics*, 53(1-3):211–244, 1992.

[44] Robert King, Charles I Plosser, James H Stock, and Mark W Watson. Stochastic trends and economic fluctuations, 1987.

[45] Gebhard Kirchgässner, Jürgen Wolters, and Uwe Hassler. *Introduction to modern time series analysis*. Springer Science & Business Media, 2012.

[46] Jeroen JM Kremers. Us federal indebtedness and the conduct of fiscal policy. *Journal of Monetary economics*, 23(2):219–238, 1989.

[47] Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178, 1992.

[48] Steve Lalley. Lecture notes: Mathematical finance. 2001.

[49] Jin-Lung Lin. Notes on testing causality. *Institute of Economics, Academia Sinica, Department of Economics, National Chengchi University*, 2008.

[50] Robin L Lumsdaine and David H Papell. Multiple trend breaks and the unit-root hypothesis. *Review of economics and Statistics*, 79(2):212–218, 1997.

[51] Jan R Magnus and Heinz Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, 2019.

[52] Benoit Mandelbrot. New methods in statistical economics. *Journal of political economy*, 71(5):421–440, 1963.

[53] Richard Melrose. Functional analysis lecture notes for 18.102. *Cambridge*, 6:2014, 2010.

[54] Attilio Meucci. Review of statistical arbitrage, cointegration, and multivariate ornstein-uhlenbeck. *Cointegration, and Multivariate Ornstein-Uhlenbeck (May 14, 2009)*, 2009.

[55] Anna Mikusheva. Lecture notes: Cointegration. 2007.

[56] Serena Ng and Pierre Perron. Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6):1519–1554, 2001.

[57] Masao Ogaki. Engel's law and cointegration. *Journal of Political Economy*, 100(5):1027–1046, 1992.

[58] Joon Y Park, Sam Ouliaris, and Buhmsoo Choi. Spurious regressions and tests for cointegration. *Cornell University, unpublished*, 1988.

[59] Pierre Perron. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica: journal of the Econometric Society*, pages 1361–1401, 1989.

[60] Pierre Perron and Serena Ng. Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *The Review of Economic Studies*, 63(3):435–463, 1996.

[61] Bernhard Pfaff. *Financial risk modelling and portfolio optimization with R*. John Wiley & Sons, 2016.

[62] Peter CB Phillips. Understanding spurious regressions in econometrics. *Journal of econometrics*, 33(3):311–340, 1986.

[63] Peter CB Phillips. Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–547, 1987.

[64] Peter CB Phillips. Optimal inference in cointegrated systems. *Econometrica: Journal of the Econometric Society*, pages 283–306, 1991.

[65] Peter CB Phillips and Steven N Durlauf. Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53(4):473–495, 1986.

[66] Peter CB Phillips and Sam Ouliaris. Asymptotic properties of residual based tests for cointegration. *Econometrica: journal of the Econometric Society*, pages 165–193, 1990.

[67] Peter CB Phillips and Pierre Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988.

[68] R. G. Pierse. Lecture notes: Granger causality and vector autoregressive models. 2019.

[69] Jim Pitman and Marc Yor. A guide to brownian motion and related stochastic processes. *arXiv preprint arXiv:1802.09679*, 2018.

[70] Said E Said and David A Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.

[71] G William Schwert. Why does stock market volatility change over time? *The journal of finance*, 44(5):1115–1153, 1989.

[72] G William Schwert. Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics*, 20(1):5–17, 2002.

[73] William F Sharpe. The sharpe ratio. *Journal of portfolio management*, 21(1):49–58, 1994.

[74] Steve Shkoller. Lecture notes: Introduction to sobolev spaces. 2014.

[75] Robert H Shumway, David S Stoffer, and David S Stoffer. *Time series analysis and its applications*, volume 3. Springer, 2000.

[76] Christopher A Sims, James H Stock, and Mark W Watson. Inference in linear time series models with some unit roots. *Econometrica: Journal of the Econometric Society*, pages 113–144, 1990.

[77] Bent E Sørensen. 1 granger causality. 2005.

[78] Bent E. Sorensen. Lecture notes: Cointegration. 2019.

[79] James H Stock. Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica: Journal of the Econometric Society*, pages 1035–1056, 1987.

[80] James H Stock and Mark W Watson. Testing for common trends. *Journal of the American statistical Association*, 83(404):1097–1107, 1988.

[81] Rauli Susmel. Lecture notes: Lecture 18 cointegration. 2015.

[82] Ruey S Tsay. *Analysis of financial time series*, volume 543. John wiley & sons, 2005.

[83] Ruey S Tsay. *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons, 2013.

[84] Ruey S Tsay. *An introduction to analysis of financial data with R*. John Wiley & Sons, 2014.

[85] Robert A. van de Geijn. Lecture notes: Notes on vector and matrix norms. 2014.

[86] Jeffrey M Wooldridge. On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of econometrics*, 47(1):5–46, 1991.

[87] Juan Pablo Xandri. Notes on metric spaces. 2015.

[88] Eric Zivot and Donald W K Andrews. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of business & economic statistics*, 20(1):25–44, 2002.

[89] Eric Zivot and Jiahui Wang. *Modeling financial time series with S-Plus®*, volume 191. Springer Science & Business Media, 2007.