The Statistical Analysis about Variation in the Stock Market

Melanie Czerny

Europe University
Faculty of Business Administration
Frankfurt Oder, Germany

Universidad Nacional de Colombia
Faculty of Business Administration
Manizales, Colombia
2019
The Statistical Analysis about Variation in the Stock Market

Melanie Czerny
C1CKJ3M1C

A thesis submitted in fulfilment of the requirements for award of
Master of Business Administration (M. Sc.)
Double Degree
To the
European University Viadrina
And
Universidad Nacional de Colombia

Director: Neil Guerrero Gonzalez
Co-Director: Alberto Antonio Agudelo Aguirre

Europe University
Faculty of Business Administration
Frankfurt Oder, Germany

- 
Universidad Nacional de Colombia
Faculty of Business Administration
Manizales, Colombia
2019
Acknowledgements

Firstly, I would like to thank my supervisor, Professor Neil Guerrero Gonzalez, who supported me not only in my Master’s Thesis but also in all other academic and non-academic aspects of my exchange year in Colombia. With your guidance, I was able to engage with my field of study in a more profound way by applying both new perspectives and innovative thinking.

Secondly, I would like to acknowledge Hernán Dario Perez, a graduate student of electronic engineering, for sharing your knowledge about MATLAB and your area of expertise with me.

Furthermore, I would like to express my sincerest gratitude to Edwin Alexander Urbano Guerrero, a graduate student of civil engineering. Thank you for your unwavering support especially in the final steps of my Master’s Thesis. Your confidence in me and my work helped me to believe in myself through all the obstacles that presented themselves throughout this journey.

In addition, I would like to thank my family for their continuous support throughout all years of my study and for their trust in all decisions I have made. Finally, I would like to thank my friends, who were always there for me and motivated me whenever I doubted myself.

Thank you!
Abstract

The key objective of this paper is to investigate patterns in the German Stock Exchange which allows to make predictions about stock price behaviour. To find these patterns, the stock price data is classified and divided into different segments, by making use of Signal Processing, to be more specific, the Moving Average Filter. The methodology is developed on MATLAB with the company Continental AG, which is part of the DAX. The time period examined is 1998 to 2018 on a daily base. The analysis shows that there are patterns in the stock market and that it can be divided into segments.

Keywords: German Stock Exchange – Signal Processing – Pattern Classification – Prediction
Resumen

El objetivo clave de este trabajo es investigar patrones en la bolsa alemana el cual nos permite hacer predicciones del comportamiento de los precios del mercado. Para encontrar estos patrones, los precios están clasificados y divididos en diferentes segmentos usando el método Signal Processing, más específico el Moving Average Filter. La metodología está desarrollada en MATLAB con los datos de la empresa Continental AG, la cual es parte del DAX. La muestra corresponde a 20 años comprendidos entre 1998 y 2018 con periodos diarios. El análisis muestra que hay patrones en el mercado y que se puede dividir en segmentos de comportamiento similar.

Palabras clave: La Bolsa Alemana – Signal Processing – Clasificación de Patrones – Predicción
Content

1. **INTRODUCTION** ........................................................................................................... 1
   1.1. INTRODUCTION ........................................................................................................... 1
   1.2. STATEMENT OF THE PROBLEM .............................................................................. 2
   1.3. SIGNIFICANCE OF THE THESIS ............................................................................. 2
   1.4. OBJECTIVES .............................................................................................................. 3

2. **THEORETICAL FRAMEWORK** ...................................................................................... 3
   2.1. STOCK MARKET .......................................................................................................... 3
   2.2. STATE OF THE ART .................................................................................................... 12
   2.2.1. LIMITS OF THE CLASSICAL WAY TO ANALYSE STOCK MARKET ......................... 12
   2.2.2. SIGNAL PROCESSING IN STOCK MARKET ............................................................ 13

3. **METHODOLOGY** ......................................................................................................... 16
   3.1. HYPOTHESIS ............................................................................................................... 16
   3.2. DATA BASE ................................................................................................................ 17
   3.3. METHODOLOGY ......................................................................................................... 18

4. **RESULTS** .................................................................................................................... 41

5. **CONCLUSION** ............................................................................................................ 47

6. **REFERENCES** ............................................................................................................. 48
## i. Notations and Acronyms

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Coefficient (for Savitzky-Golay Filter)</td>
</tr>
<tr>
<td>a</td>
<td>Slope of linear function</td>
</tr>
<tr>
<td>b</td>
<td>Coefficient of linear function</td>
</tr>
<tr>
<td>AG</td>
<td>Aktiengesellschaft</td>
</tr>
<tr>
<td>CAPM</td>
<td>Capital Asset Pricing Model</td>
</tr>
<tr>
<td>D</td>
<td>Cash Distributions to investor (Dividend)</td>
</tr>
<tr>
<td>DAX</td>
<td>Deutscher Aktienindex</td>
</tr>
<tr>
<td>e.g.</td>
<td>For example</td>
</tr>
<tr>
<td>h</td>
<td>Coefficient (for Savitzky-Golay Filter)</td>
</tr>
<tr>
<td>IPO</td>
<td>Initial Public Offerings</td>
</tr>
<tr>
<td>i</td>
<td>ith data point</td>
</tr>
<tr>
<td>i.e.</td>
<td>For example</td>
</tr>
<tr>
<td>N</td>
<td>Number of neighbouring points</td>
</tr>
<tr>
<td>NASDAQ</td>
<td>National Association of Securities Dealers</td>
</tr>
<tr>
<td>n.d.</td>
<td>no date</td>
</tr>
<tr>
<td>np</td>
<td>Data Frame</td>
</tr>
<tr>
<td>U.S.</td>
<td>American States</td>
</tr>
<tr>
<td>R</td>
<td>Return</td>
</tr>
<tr>
<td>s</td>
<td>Standard Deviation of a sample</td>
</tr>
<tr>
<td>V</td>
<td>Market Value</td>
</tr>
<tr>
<td>X</td>
<td>Value for X-coordinate</td>
</tr>
<tr>
<td>( \bar{x} )</td>
<td>Mean</td>
</tr>
<tr>
<td>Y</td>
<td>Value for the Y-coordinate</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>Variance</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>( \mu )</td>
<td>Mean</td>
</tr>
<tr>
<td>P(S)</td>
<td>Probability of stability</td>
</tr>
<tr>
<td>P(T)</td>
<td>Probability of tendency</td>
</tr>
<tr>
<td>P(S_n)</td>
<td>Probability of segment n</td>
</tr>
</tbody>
</table>
ii. List of Tables

Table 1. Fundamental Analysis versus Technical Analysis ........................................ 13
Table 2: Example Second Approach ........................................................................ 27
Table 4: Segments and probabilities ......................................................................... 40
Table 5: Frequencies and absolute probabilities ....................................................... 43
Table 6: Conditional Probabilities of Tendency ......................................................... 44
Table 7: Conditional Probabilities of Stability .......................................................... 44
Table 8: All Conditional Probabilities ....................................................................... 45

iii. List of Figures

Figure 1: Utility Function of Risk Averse Investor ..................................................... 7
Figure 2: Systematic and Unsystematic Risk ............................................................... 8
Figure 3: Structure of the graph ............................................................................... 18
Figure 4: Figure: Stock Price of Continental AG, 2018 ........................................... 19
Figure 5: Description of Moving Average Filter 2..................................................... 22
Figure 6: Comparison Stock Price and Smoothed Stock Price of Continental AG, 2018 ................................................................. 23
Figure 7: Calculation First Derivative with n=1 .................................................... 25
Figure 8: Calculation First Derivative with n = 3 ..................................................... 26
Figure 9: Derivative with n=1 and n=3 .................................................................... 26
Figure 10 First Derivative of n=1, n=3, n=5 and n=10 for Continental AG in 2018 ... 27
Figure 11: Derivative Relationship ......................................................................... 28
Figure 12: Segments and Transitions ...................................................................... 29
Figure 13: Stock Price Example .............................................................................. 30
Figure 14: Drastic Change ....................................................................................... 30
Figure 15: First Derivative with Threshold ............................................................... 31
Figure 16: Histogram of First Derivative of Continental AG in 2018 ....................... 31
Figure 17: Gaussian Distribution ............................................................................ 32
Figure 18: Threshold Application ............................................................................ 33
Figure 19: Classification Characteristics for Segmentation ....................................... 35
Figure 20: Threshold constant and weakly increasing (decreasing) ....................... 36
Figure 21: Standard Deviation Calculation with variable mean.......................... 37
Figure 22: Segmentation and Characterisation of Continental AG in 2018 ............ 38
Figure 23: Relationship of Derivative n=1 and N=n ........................................... 41
Figure 24: Segmentation of Continental AG in 2019 ........................................... 47
1. Introduction

1.1. Introduction

On a stock market financial assets such as stocks and bonds are traded, which provides the opportunity to raise capital for companies and provides investment opportunities for market participants. The analysis of financial data is an interesting research topic due to its complex system and its great impacts. Everyday a huge amount of money is traded on the market. Events on the stock market can affect investors with the uncertainty of gains and losses of its investments, such as the loss of their savings and pensions. Furthermore, misdealing and misunderstandings may lead to market crashes and the bankruptcies of companies (Sharkasi, 2006). There are plenty of investigations in this field, interestingly with completely different directions. Some researchers claim that the stock market is unpredictable. They state that there is no autocorrelation in stock data, hence it is following a Random Walk. According to Malkiel (1973, p. 406 – 407) “[n]either fundamental analysis of a stock's firm foundation of value nor technical analysis of the market's propensity for building castles in the air can produce reliably superior results”. Other researchers propose theories and algorithms to analyse stock data, claiming that historical movements of the price do affect stock price development and predictions are possible. The most important question is: “Can tomorrow’s stock price be predicted?” Despite decades of analysis’ the price could not be predicted with a security of 100%. In the past years, electrical engineers got attracted by the difficulty of analysing stock data. Consequently, electrical engineers started researching the “signal” of stock data, using approaches and tools of their field. Nevertheless, the application of electrical engineering instruments on financial stock data combines two fields of high complexity. In this master thesis a methodology is developed, which applies electrical engineering tools, to be more specific Signal Processing, on financial data in a visual way. In this way the complexity of both fields can be simplified, allowing any researcher to follow the investigation and simultaneously invites researchers to investigate further in this topic.
The master thesis starts with bases knowledge about the stock market to provide the foundations needed to understand the research topic. Thereafter, the state-of-the-art is presented. It is followed by chapter 3, which talks about the methodology and the data base of the investigation. Finally, the paper ends with the results and the conclusion.

1.2. Statement of the Problem

Stock price is highly volatile data, which means it has a great risk compared to other investment opportunities. During the last decades researchers analysed the equity market using financial tools with the objective to predict the stock price. Until today stock price movements are not completely predictable, and the possibility of prediction is controversially discussed. Malkiel claims that stock price is following a Random Walk, which means that historical prices do not contain any information about the future direction. Nevertheless, electrical engineering researchers started to investigate in this field, bringing new light by using methods of their field.

By reason of the difficulties in stock price prediction and the new opportunities provided by electrical engineering, I investigate in this master thesis the possibility to predict stock price using Signal Processing methods. The aim is to provide a methodology based on a characterization of the stock price behaviour in time (time series) that allows to make stock price predictions.

1.3. Significance of the Thesis

A stock market provides the opportunity to invest in financial assets, which means buying and selling on the market. Depending on the type of investment, it can be very risky to participate in the stock market, especially for lay people. This investigation can provide more information about stock price behaviour and provide a decision-making tool for investors, thus protect market participants for great losses in their investments.

Furthermore, this work can be seen as the first step towards a more stable stock market, since it provides the foundation of further investigations in a characterization as well as prediction of stock price behaviour.
1.4. Objectives

General Objective

To identify patterns related to variation of the stock price in the equity market and to develop a methodology which allows to make predictive statements about the future price development.

Specific Objectives

I. To implement on any mathematical program at least two strategies based on signal processing to analyse the behaviour of stock.

II. To classify the stock price behaviour into segments of time according to the patterns found.

III. To identify signals that indicate a drastic change in stock price.

IV. To test the performance of the two implemented predictive strategies.

2. Theoretical Framework

2.1. Stock Market

This paper is aimed at researchers of different areas, such as economic and electrical engineering. For this reason, the most important financial aspects are explained to provide the basic knowledge needed to understand this investigation. In the following, the general concepts of the financial market, which build the foundation of this thesis, is explained.

Capital Markets

In the capital market, financial assets, which belong to the group of intangible assets, are traded. The capital market consists of the derivative market, equity market, debt market and foreign exchange market. In the derivative market financial assets such as futures, options and swaps are traded (Rudolph & Schäfer, 2005). As its name indicates, the debt market trades with bonds, such as U.S. Treasury securities (Piljak & Swinkels, 2017). When trading in the
international financial market, investors are faced with the risk of a change in value of different currencies. Therefore, the foreign exchange market provides some instruments to decrease the uncertainty, for instance with forward contracts, future contracts or currency swaps. The equity market is the foundation of this thesis. Therefore, it is explained in detail in the following.

**Equity Market**

From a company’s point of view (the issuer’s point of view) the equity market evolved as a source to raise financial resources which usually are needed for investment purposes. Hence, instead of taking, for example, a bank loan, a firm can sell a part of its ownership to the public (investors) in order to receive the capital, they need. The investors are entitled to the earnings of the company, which is distributed via dividends (De Jong & Rindi, 2009). This means, in return for their invested money they receive a part of the profit the company makes. Dividends are usually provided once a year. However, some corporations provide their dividends throughout the year. From an investor’s point of view the equity market provides an opportunity to invest in intangible assets, thus to avoid the risk that characterises investments in tangible assets, such as depreciation of real estate (Fabozzi & Modigliani, 1992).

**Preferred Stock versus Common Stock**

There are two types of equity securities, which are common stock and preferred stock. One of the main differences between both is related to the degree of priority. Investors holding preferred stock receive dividends before common stockholders receive theirs and also have priority in case of bankruptcy. Moreover, preferred stock holders receive a fixed amount of dividends, while the dividends of common stockholders are variable and uncertain. The corporation can, for instance, decide not to provide dividends to common stockholders and invest this money in the company. Furthermore, only investors holding common stock have voting rights (The Frank J. Fabozzi Series, 2002). Since preferred stock provides a fixed amount of dividends, it shares some characteristics with bonds, thus its value is calculated in a similar way. However, this master thesis will concentrate only on common stocks.
Primary Market versus Secondary Market
When common stock is issued for the first time, it is offered on the so-called primary market, which only deals with new financial assets. If it is also the first time a company offers stock to the public, it is called initial public offerings (IPO) (De Jong & Rindi, 2009). After the initial offer, the stock is traded on the secondary market. Accordingly, on the primary market an investor can buy financial assets directly from the company (e.g. through investment banks), while on the secondary market the assets are being bought from other investors. This paper will investigate only stock that is being dealt on the secondary market.

Dealer or Market Maker
There are different participants on the secondary market. One group of participants, the investors, has already been mentioned. An investor can be any person that wants to buy (or sell if he already holds stock) financial assets. Since there is the possibility that an investor, who wants to sell a certain amount of his assets, does not find a buyer, who wants exactly that asset with the exact amount, the secondary market often needs an intermediary in the market. An intermediary, who is willing to buy as well as to sell the asset and amount the investors want to trade, is called dealer or market maker. A dealer takes position in different stocks, which means he has different financial assets in his inventory. If an investor wants to take position in one of those assets, he can do so immediately by buying from the market maker. However, when selling an asset, the dealer demands a price which is higher than the price that he paid for that asset. This is called bid-ask spread. It can be seen as a cost for immediacy (O'Hara, 1995). A bid-ask spread “consists of the difference between the price at which a market maker is willing to sell the financial asset (i.e., the price it is asking) and the price at which a market maker is willing to buy the financial asset (i.e., the price it is bidding)” (Fabozzi & Modigliani, 1992, p. 6). The difference in price results from the risk the market maker has to take by offering his services. There are two main sources of risks which are the variability in price and the thickness of the market. The first factor is related to the uncertainty of change in price. If the variance of the price is high, the market maker demands a greater bid-ask spread to cover the risk of a decrease in price while holding the asset.
The second factor, thickness of the market, refers to the amount of orders the market maker receives. The higher this number the shorter the time the financial asset is held in the inventory, thus the smaller the risk of a change in price.

**Prices on the Secondary Market**

Before explaining how prices on the secondary market are built, the importance of knowing the price is discussed. The profit an investor makes by buying and selling stocks is called return, which can be calculated with the following formula:

\[ R = \frac{V_1 - V_0 + D}{V_0} \]

Where,
- \( V_1 \) = market value at the end of the interval
- \( V_0 \) = market value at the beginning of the interval
- \( D \) = cash distributions to the investor during the interval

It can be seen that to maximize their return (profit) investors either should try to acquire stocks with high dividend payments or stocks whose price will increase in the future (The Frank J. Fabozzi Series, 2002). The future cash flows are uncertain. The reason for this will be explained later in this chapter. An important piece of information that investors know in advance is the level of uncertainty, which means if there is a possibility that the future cash flow will decrease or increase dramatically or decently. This is the variance of the stock price. Stocks with a great variance are considered risky assets. Investors naturally are risk-averse, which means they would choose stocks with low risk over stocks with high risk if they have the same expected return. Consequently, if they buy the riskier asset, they demand a higher return, which is called risk premium. This concept can also be seen in the utility function of risk averse investors. The utility function is based on two parameters:

\[ U = f(E_w, \sigma_w) \]

These parameters are the expected future wealth and the predicted standard deviation, which is the risk of the asset. Since investors prefer higher expected future wealth \( \left( \frac{dU}{dE_w} > 0 \right) \) and choose an asset with a low standard deviation over an asset with high standard deviation \( \left( \frac{dU}{d\sigma_w} < 0 \right) \) the indifference curve of \( E_w \) and \( \sigma_w \) is upward-sloping (Sharpe, 1964). In figure (1) the utility function is illustrated,
and it is depicted that for risk averse investors the utility of certain values (W1) are higher than the one of uncertain or expected values.

![Utility Function of Risk Averse Investor](image)

*Figure 1: Utility Function of Risk Averse Investor*

For the reasons mentioned above, scientists analyse the market to find investment strategies which maximize the return of the stock. One approach to this is the prediction of stock price or analysis of variation in the equity market.

The prices on the secondary stock market are defined by demand and supply (O'Hara, 1995). An increasing demand of a specific stock increases the price of this stock. But what changes the demand of a stock? Since stock prices are highly volatile data (data with a great variance), many theories and concepts were developed trying to explain what moves the prices in the market.

**Evolution of the Financial Market**

**Classical Approach**

In the classical approach, the price of any financial asset is the present value of expected cash flow, which is also called fair value or fundamental value. In the case of preferred stock, the dividends are fix, which makes it much easier to determine the price. The expected cash flow of common stock is composed of the expected dividends as well as the expected earnings investors gain by selling the stock to another investor or the dealer. The share price therefore reflects the current value of the company as well as the expected growth (Fabozzi & Modigliani, 1992). Therefore, any new information on a company can influence its stock price since it can influence the expected earnings or growth
(Tallmann, 1989). If the earnings of the company are expected to decrease, the dividends will decrease or might become zero, and the price that other investors are willing to pay for this asset will decrease as well. There are various factors or risks that can influence the fundamental value of a company.

**Systematic versus Unsystematic Risk**

Unsystematic risk is also called diversifiable risk, residual risk or company-specific risk. The name company- or firm-specific risk indicates that this risk is unique to a company, for instance, a strike, the outcome of unfavourable litigation or other negative news about the company. Consequently, this risk can be diversified through a portfolio containing different types of stock (Wagner & Lau, 1971). If an investor holds a portfolio with two different stocks (stocks of two different companies) and the stock of one of them faces a dramatical drop in price, he is still left with the other stock whose price might not has changed. In this way market participants are able to reduce risk by diversifying their portfolio. In general, investors try to find stock whose return have a very low correlation to reduce the risk as much as possible.

In contrast to this, the systematic risk is caused by general market or economic conditions, which can affect a whole industry or country, such as inflation (Fabozzi & Modigliani, 1992). Therefore, this risk cannot be diversified. The relationship between systematic, unsystematic risk and the number of stocks held in the portfolio is depicted in the graphic (2).

![Figure 2: Systematic and Unsystematic Risk](image)
Efficient Market Hypothesis
Following the classical approach, Fama established the efficient market hypothesis. According to his theory, if stock prices reflect all available information including insider information, it is called strong form of efficiency. A semi-strong efficiency includes all publicly available information and a weak form of efficiency contains only historical prices. Consequently, no market participant should be able to “beat” the market by using investment strategies to receive higher returns (Fama, 1969). Beat the market means that an investor was able to receive higher returns than he usually should have gotten according to the risk level of the financial asset (Drake & Fabozzi, 2010).
Nevertheless, this theory is criticized by many researchers. Damodaran (2002) proposes another definition of market efficiency, where the market price does not necessarily reflect the real value of the asset at any point in time. The price can be under- or overpriced as long as its nature is random, which means that there is an equal probability for any asset to be under- or overpriced. However, he agrees with Fama that there should not be any successful investment strategy that can find the asset which deviates from its true value.

Asset Pricing Models and the example of CAPM
An asset pricing model that is following Fama’s idea of efficient markets is the Capital Asset Pricing Model (CAPM), which was developed to determine the fair value or fundamental value of financial assets. It assumes that the market is moving to a stable market equilibrium. “The equilibrium price is the price at which quantity demanded equals quantity supplied” (O'Hara, 1995, p. 14). Furthermore, CAPM assumes that any stock has a systematic risk, which is similar to that of the market portfolio. A market portfolio is a portfolio which contains all stock. Since the systematic risk is similar, one can establish a mathematical relationship between the risk of the financial asset and the risk of the market portfolio.

The Capital Market Theory is based on the following assumptions:
- Investors make rational decisions
- Investors make decisions based on the expected return and the variance of returns
- Investors are risk-averse
- The market is perfectly competitive, there is no investor that is able to influence the prices on the market (investors are small compared to the market)
- No transaction costs influence the development of the prices (bid-ask spread, taxes, etc.)

Other asset pricing models, such as the Fama-French three-factor model, are based on similar assumptions of investors and the market. Nevertheless, in the past asset pricing strategies were not able to calculate the market value of stocks (Fabozzi & Modigliani, 1992). An explanation is given by Krause (2001) who claims that the fundamental value of an asset can be seen as the natural price, defined by Smith (1776). And the price that can be observed at the market can be defined as the market price, defined by Smith (1776). Following those assumptions Smith (1776) explains in his work *The Wealth of Nations* that the market price is determined by demand and supply and, therefore, deviates from its fair value. However, in the long run the observed price will move towards its fundamental value.

Nonetheless, the strict assumptions of the fundamental analysis were criticised, and, as a result, behavioural finance was developed. It tries to explain the part of variation, that the fundamental analysis failed to explain.

**Behavioural Finance**

Behavioural finance is a part of behavioural economics. Since the 1970s researchers from the psychology, economic and sociology field have been investigating the rationality in decision-making of individuals through different experiments. They came to the result that human beings deviate from rational choices (Puaschunder, 2018). Thus, one of the main assumptions of classical finance, investors take rational decisions, is said not to be true. The difference between behavioural economics and behavioural finance is that the first “aims to study how individuals make decisions and the way they interact or influence other individuals, organizations, markets, and society (Birnberg and Ganguly, 2012)” (Costa, Carvalho, & Moreira, 2019). “In turn, the field of Behavioral Finance is more focused on the study of errors of judgment and of decision-
making characteristics in financial investments” (Costa, Carvalho, & Moreira, 2019).

Various studies try to explain what factors influence the decision-making process of human beings. The paper of Sadi et al, for example, shows evidence that the offered perceptual errors in the decision-making process has a significant correlation with the personality of the market participant.

**Overreaction Hypothesis and Herd Behaviour**

There are different effects observable on the financial market, based on human behaviour, such as herd behaviour or overreaction. The Overreaction Hypothesis is an example of negative autocorrelation. Autocorrelation means that there is a relationship between the change in share price in one period and another period. Negative autocorrelation occurs, when the price change was positive in the first period and negative in the second one (or the other way around). One example is the overreaction to negative news. The investors expect this new information to affect the new value of the stock much more severely than it does. Usually, a few periods later, the investors are able to evaluate the effect of this negative news properly. Consequently, the price starts to increase again towards its true value. This effect is especially observable when the news contains complex information (Stock, 1990).

Herd Behaviour is an example of positive autocorrelation, which means that there is a relationship between different time periods, where the stock price increases (or decreases). The reason for this is that some investors are observing the actions of other market participants and consider those actions into their decision-making process. If, for example, an investor realises that numerous other stockholders are selling a specific asset, they tend to do the same. Therefore, this effect is called herd behaviour (Oerke, 1999).

To summarize, according to the classical approach the stock price reflects the real value of the company, which is the sum of the future cash flows (Fabozzi & Modigliani, 1992). Behavioural finance claims that the prices on the stock market are affected by many more factors, such as human irrationality which causes a deviation of the real value (Puaschunder, 2018). In this chapter only the basic concepts and some examples were explained. Many papers
investigating the factors that influence stock price. Up until today, there is no researcher who has been able to explain 100% of the variance in stock price.

2.2. State of the Art

2.2.1. Limits of the Classical Way to analyse Stock Market

There exist two different school of thoughts to analyse stock market prices and develop investment strategies, which is the fundamental analysis and the technical one. The fundamental analysis is a numerical approach, usually used for long-term investments since its valuation models consider factors that rather don’t change quickly in time. Examples for valuation models are the Dividend Discount model, Price-to Book Value Ratio or the Discounted Cash Flow model (Wafi, Hassan, & Mabrouk, 2015). Input factors such as profit margins, revenues and return on equity are used to calculate the company’s value and potential growth (Agrawal, Chourasia, & Mittra, 2013). Once calculated the fair price of the share it will be compared with its market value (the price in the stock market) to make a decision on whether invest or not (Wafi, Hassan, & Mabrouk, 2015). To summarize, the fundamental analysis believes that the current share price does not contain all available information, thus can be over or under priced. One of the main criticism of this method is the assumption of market participants being completely rational and never act based on emotions (Achelis, 2001). Several studies and papers, for instance San Miguel, Ryan, & Amaya-Amaya (2004) or Fromlet (2001), in behavioural economics have shown, that investors do not act rationally and proved the existence of psychological effects, such as herd behaviour and exaggeration in the stock market.

Whereas the technical analysis is mainly a visual decision-making tool which uses geometry and pattern recognition to predict share prices (Lo, Mamayski, & Wang, 2000). Achelis (2001) defines technical analysis as a process which analyses historical prices on the stock market to find pattern or trends in the market. This approach assumes that the share price already
contains all information available on the market since it would immediately change if there was new information. Furthermore, it takes human irrationally into account as it is based on investor’s expectation about the price and not the real value. Technical analysts assume that the share price tends to go with the trend and think that human reaction on changes in the stock market are predictable (Agrawal, Chourasia, & Mittra, 2013). However, since it is a visual based approach, it is criticized to be too subjective. The same charts can be interpreted very differently by various persons (Achelis, 2001). In table (1) the main characteristics of the fundamental and technical analysis are summarized:

<table>
<thead>
<tr>
<th></th>
<th>Fundamental analysis</th>
<th>Technical analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Usage</strong></td>
<td>Long term investments</td>
<td>Short term investments</td>
</tr>
<tr>
<td><strong>Strength</strong></td>
<td>Algebraic and numerical, Systematic approach</td>
<td>Visual Tool based on geometry and pattern recognition</td>
</tr>
<tr>
<td><strong>Limits</strong></td>
<td>Assumes investors are rational</td>
<td>Very subjective</td>
</tr>
</tbody>
</table>

*Table 1. Fundamental Analysis versus Technical Analysis*

### 2.2.2. Signal Processing in Stock Market

In the past signal processing was only used in electrical engineering field to analyse information such as sound or images. Thanks to technological development, especially regarding information storage and speed in processing data, signal processing was no longer seen as a tool exclusively used in electrical engineering, hence the application in other fields started. Financial data provides the perfect conditions for the use of signal processing, due to the huge number of information that is available. Referring to the stock market, a high quantity of historical share prices is online and mostly available for free which opens the door to signal processing researchers. There exist several blogs for researchers and also papers, explaining economic theories to electronic engineers to introduce them into the field. Moreover, Darolles, Duvaut and Jay (2013) strongly recommend the use of Signal Processing in Finance in
their book “Multi-Factor Models and Signal Processing Techniques: Application to Quantitative Finance”.

Signal processing is applied in fundamental as well as technical analysis. One example for the use in fundamental analysis are the papers of Wang & Zhang (2013) about the Capital Asset Pricing Model and the Multi-Factor Model. Contrary to the traditional view, they propose that the betas (systematic risk and other risk factors) are only piecewise constant and can have a temporary change due to significant events. They make use of Signal Processing as well as Kalman filter to provide a tool to calculate as well as predict the change in beta and help investors to take decisions.

However, the application of Signal Processing for fundamentalists is limited due to the input factors since it makes it difficult to develop an automatized process. Whereas the knowledge used for technical analysis is the data itself, meaning the price (or return) and time. Once an algorithm is written, there is no need for further human interaction when applying it for example to other data (Agrawal, Chourasia, & Mittra, 2013).

For this reason and because of the critique about the ignorance of emotions and irrational behaviour in the stock market, let to the decision to not do further investigations in this field and rather concentrate on the technical analysis.

There are a variety of examples for the application of Signal Processing using stock data as an input factor in the literature.

Benedetto & Giunta (2013) are analyzing hidden market trends of historical stock quotes to develop an algorithm for financial trading using Signal Processing techniques. Their case study is based on binary options stock trading with one buy or sell order per week, where it showed evidence of asymmetry in the stock market. The signal processing algorithm resulted successfully in 82% of the cases.

Also the paper “Stock market predictive model based on integration of signal processing and artificial neural network (Sharma & Rababaah, 2014) proposes a successful stock index predictive model. As training and testing data the Dow 30 and NASDAQ 100 have been used with an overall prediction accuracy of 96.7%.
Another example is the paper of Atkins, Niranjan and Gerding (2018) who publish in their paper that financial news predict the stock market volatility more accurately than the stock price. For their study they construct machine learning models of Latent Dirichlet Allocation and make use of the simple naïve Bayes classifiers. Using news feed reaches a prediction accuracy of 56%, while using closed stock price reaches only 49%.

Upadhyay, Bandyopadhyay and Dutta (2012) predicted the outperforming stock in the Indian stock market by developing a Multinomial Logistic Regression model. To determine the performance on the stock they used the stock return and variance compared with market return and variance. In the next step the influence of seven financial indicators such as Book Value and Earnings per Share on the stock performance were tested. The classification resulted in an accuracy rate of 56,8%.

There are also other ways to use machine learning approaches in the finance area, such as a comparison of different models to discover the best prediction model like in the paper of Kumar, Dogra, Utreja, & Yadav (2018). They compared Support Vektor Machine, Random Forest, K-Nearest Neighbour, Naïve Bayes and Softmax. Results show that Random Forest algorithm outperforms the others for large dataset and Naive Baysian Classifier has the best prediction for small ones.

Another way of applying machine learning techniques is the forecast of volatility, such as in the paper “An engineering approach to Forecast Volatility of Financial Indices” of Ma, Wong, & Sankar (2007). Techniques such as time series data mining and wavelet transform were used to characterize and predict non-linear time-series, that is to say very volatile data with jumps and drops. The accuracy of this forecast results to be in average over 75% better than other available results.

Furthermore, there are thesis’ and dissertations of the the master or doctor degree in for instance the engineering, computer science and statistics field, researching in the interface of finance and machine learning such as the ones of Larsen (2010), Sharkasi (2006) and Holme (2010). Larsen (2010) develops a prediction model which uses a novel two-layer reasoning approach. In the first layer technical analysis techniques are applied and in the second one machine learning approaches, both combined with money managing strategy tools.
Sharkasi (2006) characterizes international stock markets in his paper using Signal Processing techniques. He discovers that emerging and mature markets deal differently with crashes and events. However, a classification into two segments is too simple and he motivates other researchers to make further investigations in this field. Moreover, he finds evidence of long-term memory in the market as well clockwise transmissions between global stock markets.

In his paper “Statistical Arbitrage: Opportunity spotting for Financial gain in Financial Markets” Holme (2010) provides a zero investment strategy which is able to spot real-time investment opportunities for investors. His developed tool provided between 6% and 44.1% of annual percentage gains. However, he proposes more investigation to assure those outcomes.

All those papers mentioned above show a variety of possibilities for the application of Signal Processing in the finance area, using financial indicators as input factor. The market can be characterized, arbitrage possibilities determined or economic models compared. Furthermore, pattern recognition is used to find trends in the market and predict the stock price or even volatility. All of those papers have successful results. However, the main idea of technical analysis has been ignored. The aim of technical analysis is to provide a visual tool for investors to facilitate their decision making. Even the papers using technical analysis in their investigation, do not provide a visual decision tool as their output result. Therefore, the master thesis aims to close this lack in literature by providing a methodology using both, pattern recognition and visual tools in the research.

3. Methodology

3.1. Hypothesis

I. There exist patterns in the stock price movements on the equity market.
II. The stock price movements can be divided into segments.
III. The future stock price behaviour is predictable.
3.2. Data Base

Data
For this master thesis one of the thirty most successful German companies trading on the Frankfurt Stock Exchange is chosen, which is part of the German stock index called DAX. The origin of DAX goes back to 1988, when the stock index was calculated for the first time with 1000 basis points, which has increased approximately twelve times during the past thirty years (Janßen & Rudolph, 1992). Due to their economical magnitude these companies result to be the most liquid ones on the Frankfurt Stock Exchange. Liquidity is a primer factor when it comes to choose the data base since it reduces the problem of noise. In this case noise can be processed by a delay in time, resulting that the actual price does not reflect the actual value of the stock price, but it shows a delayed value. For developing the methodology one specific company was used which is called Continental AG. It is an automotive manufacturing company specialized in different components of vehicles (Continental, 2019). The National University of Colombia provides access to the software Thomson Reuters Eikon. Therefore, the stock data used in this master thesis are retrieved from this software.

Time Period
To choose an adequate time range, different options were plotted and were visually analysed. The stock price in month, weeks and days were plotted using time periods of ten years, five years, three years, one year and six months. It was observed, that important happenings in the stock market are taking away by analysing in month or weeks. Therefore, the data is analysed in days. Moreover, analysing the days in a time period higher than one year seemed to be too chaotic, since too many important events were happening, whereas six months did not represent enough information about the behaviour of the stock price. Hence, the methodology is developed using daily stock prices within one year. For the development of predictions based on the characterization of the stock price, a data base of twenty years is chosen (1998 – 2018) to accurately analyse the stock price behaviour.
3.3. Methodology

The methodology of this master thesis is developed with MATLAB. MATLAB is a software offered by the U.S. company MathWorks to solve mathematical problems, especially to analyse data, develop algorithms or create models (Mathworks, n.d.). It is mainly used by Engineers, but also by scientists of other fields.

There are two different approaches developed in this master thesis to analyse stock prices in the equity market. Before starting with these approaches, the general structure of the graph is explained. The Frankfurt Stock Exchange in Germany is closed on weekends and holidays (Kirchner, 1999). Hence, the data base does not contain 365, but 254 days. The x-coordinate shows the time in intervals of days and the y-coordinate shows the price in Euros, which can be seen in depiction (3).

As described before, the share price of companies is highly volatile data due to the sheer magnitude of factors that influences it. By plotting the stock price one can easily observe this problem. In graphic (4) the stock price of Continental AG in 2018 for one year is plotted.
For this reason, a tool is needed to clear the data from small fluctuations, which is also called noise, in order to be able to identify trends and significant changes in the data. Thus, a technique of Signal Processing, which is called Signal Smoothing, is used. There are different filters available to smooth data, such as Moving Average Filter or Savitzky-Golay Filter which belong to the category of low-pass filters. As its name indicates, a low-pass filter passes low frequencies and filters those, which are higher than the cut-off frequency. The contrary is a high-pass filter which works contrariwise (MathWorks, n.d.). Hereafter, the models to smooth data are explained, followed by a discussion about the best choice.

Moving Average Filter:
The Moving Average Filter uses the mean of the neighbouring data points to smooth the signal. Firstly, a span needs to be defined, for example 3. MATLAB then uses one data point before and one data point after the one to be smoothed to calculate the average. The result replaces the data point that is aimed to be smoothed. Since there is no neighbour for the end points, they will not be replaced. The higher the amount of neighbouring points are chosen; the smoother gets the data.
This method is given by equation (2) (MathWorks, 2019).

\[ y_s(i) = \frac{1}{2N+1} \left( y(i + N) + y(i + N - 1) + \cdots + y(i - N) \right) \]

Where,
- \( y_s(i) \) = smoothed value for the ith data point
- \( N \) = number of neighbouring data points on either side of \( y_s(i) \)
- \( 2N + 1 \) = span

In this approach all the data points are equally weighted. However, there exist other types of Moving Average Filters, such as the weighted moving average or exponential moving average filter to consider importance of different weights of data points. This is done by adding multiplying factors to each data point (Thinh, Hoang Quan, & Maneetien, 2018).

The main advantage of the Moving Average Filter is the simplicity of application to the data due to its plain algorithm structure (Wang & Zhang, 2016). Apart of its computational efficiency, it is good at rejecting noise which therefore provides a smooth line. Nevertheless, there is a lag between the original and the smoothed share price due to a delay in data. Especially for drastic changes this lag can be significant when analysing data.

**Savitzky-Golay Filter:**

In 1964 Savitzky and Golay published a paper about a filter that uses local least-squares polynomial approximation to smooth data (Schafer, 2011). As in the moving-average filter a data frame needs to be defined, for example 4 data points. Then, the degree of polynomial for this data frame must be chosen. The general equation for the filtering process is the following:

\[ Y_t = \frac{1}{h} \left( \sum_{i=\frac{-np-1}{2}}^{\frac{np-1}{2}} a_i x_{t+i} \right) \]

Where,
- \( np \) = data frame
- \( h \) = coefficient of the filter
- \( a \) = coefficient of the filter

One of the main advantages of this Filter is the obtainment of a universal numerical function (Steinier, Detour, & Deltour, 1972). Using the Savitzky-Golay filter one can overcome the problem of a delay in the smoothed data since it accounts for transient effects. This filter represents the data more closely...
(Hassanpour, 2008), but one must take the risk of overfitting into account. Therefore, it is recommendable to choose a small data frame to be able to fit a low polynomial degree to the data. Furthermore, the Savitzky-Golay filter tends to smooth the data without cutting important minima or maxima (Luo, Ying, & Bai, 2005).

In comparison to the Moving Average Filter the Savitzky-Golay Filter has lots of advantages, such as the account for transient effects. Nevertheless, stock prices movements are highly volatile and behave in a chaotic way. Therefore, it is not useful to try to fit the data in a function. The Moving Average Filter does not use an underlying function to smooth data. Thus, the Moving Average Filter is applied for the Signal Smoothing.

The first step of any of the approaches is smoothing the data using the Moving Average Filter. MATLAB provides a function called *smooth* to realize this process. For the simple Moving Average only two variables need to be defined, the data to be smoothed and the span. The variable of the span does not define directly the number of neighbouring points but the degree of smoothing giving in percentages. If the longitude of the data is e.g. 100 data points and a span of 3% is used, MATLAB will use three data points each time to smooth the graph, one before and one after the data point, that is being smoothed. Since there is no data point before the first one, the stock price for day 1 is not being smoothed and the process starts with the second one. In graphic (5) there is a visual explanation about the Moving Average Filter used by the function *smooth* in MATLAB.
For the Stock Price of Continental AG a smoothing percentage of 1% through a heuristic method is chosen. Regarding the fact that the total amount of data is 254, a span of 2.54 should be used. Since the Moving Average Filter needs an odd number, MATLAB automatically corrects this number to 3 data points (MathWorks, 2019). In the following graph the original stock price and the smoothed stock price of Continental AG is depicted. To visualize the smoothing effect in the best way, only a longitude of one month is presented, which can be seen in graphic (6).
First Approach

In the following the first approach will be explained, which proposes a new method of the application of derivatives. The aim of this approach is to identify interesting patterns in the stock price, which allows to develop a methodology to make stock price predictions. The approach consists of two phases with different steps. According to the discussion before, the first phase is the preparation of the data, which means the application of Signal Smoothing using the Moving Average Filter.

1. Data Preparation
   1.1. Signal Smoothing
   Before starting with this approach, the data will be smoothed according to the Moving Average Filter.

2. Derivative Relationship
   2.1. Calculation of first derivative
After successfully having smoothed the data, the first derivative of the smoothed price is made which will be used in the next steps. The first derivative is the difference between two neighbouring values which gives information about the direction of the graph. The general formula for a derivative is the following (Steward, 2008):

\[ f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \]

Where,
- \( f'(a) \) = the derivative of the function in point \( a \)
- \( f(x) \) = the function in point \( x \)
- \( f(a) \) = the function in point \( a \)
- \( x \) = value \( x \) of the independent variable
- \( a \) = value \( a \) of the independent variable

Since stock price is not continuous but discrete data, only the difference between two neighbouring points have to be calculated. Therefore, formula 3 is transformed to the application of discrete data. Using this finite difference method, there are two ways to calculate the first derivative, which are the progressive and regressive method.

The progressive method takes the difference between the actual value and the next value, which can be seen in formula (5):

\[ (\delta^+ f)(\tilde{x}) = \frac{f(\tilde{x} + N) - f(\tilde{x})}{N} \]

Where,
- \((\delta^+ f)(\tilde{x})\) = derivative of the function \( f \) in the point \((\tilde{x})\) using the progressive method
- \(f(\tilde{x} + N)\) = value of the function \( f \) for the neighbouring point \(\tilde{x} + N\)
- \(f(\tilde{x})\) = value of the function \( f \) for the actual value \( x \) \((\tilde{x})\)
- \(N\) = number of neighbouring points

Whereas, the regressive method is taking the difference between the actual value and the one before:

\[ (\delta^- f)(\tilde{x}) = \frac{f(\tilde{x}) - f(\tilde{x} - N)}{N} \]

Where,
- \((\delta^- f)(\tilde{x})\) = derivative of the function \( f \) in the point \((\tilde{x})\) using the regressive method
- \(f(\tilde{x} - N)\) = value of the function \( f \) for the neighbouring point \(\tilde{x} - N\)
- \(f(\tilde{x})\) = value of the function \( f \) for the actual value \( x \) \((\tilde{x})\)
- \(N\) = Number of neighbouring points

Since the approaches of this master thesis aim to find patterns in the market in order to take decisions about future actions (buy or sell), the regressive method is applied. The usage of this method allows to consider the stock price of the current day to calculate the first derivative, which can be seen graphic (7):
The first step in Derivative Calculation is applying formula 5 to determine the first derivative with the first neighbouring point $f(\hat{x} - 1)$ before the actual value $f(\hat{x})$, using $n = 1$.

In the next step, formula 5 is used again to determine the first derivative. However, this time, $n$ will not be 1, but is chosen to be 3, 5 and 10. Hence, e.g. the third neighbouring point $f(\hat{x} - 3)$ before the actual value $f(\hat{x})$ is used. This process is visually explained in graphic (8):
By using different values for N, which are greater than one, the first derivative will be smoothed. The reason for this lies in the original data, which are volatile stock prices. By applying the Moving Average Filter, a first step of smoothing is done. Nevertheless, the smoothed stock price still has a high variance, which result in a derivative with a high variance. The greater N is being chosen, the greater the degree of smoothing in the first derivative. This process can be seen in the last two graphics. The value for the first derivative in figure (7) is 2, whereas the value for the first derivative in figure (8) is only one. In the following depictions the change of the value can be seen:

Applying this method to the stock price of Continental AG in 2018 results in the derivatives of figure (10), which again are shown in the first month for simplicity reasons.
The regressive method to calculate derivatives require a neighbouring point in the past. Since n=10 uses ten neighbouring points in the past, the derivative can only start in day 10.

2.2. Derivative Relationship
In this step the relationship between the derivative n=1 with N=n (n=3, n=5 and n=10) will be graphically shown. To realize this, the derivative n=1 is plotted in the x-coordinate and the derivatives N=n is plotted in the y-coordinate. If, for instance, the values for the derivative n=1 and n=3 for 5 days are the following:

<table>
<thead>
<tr>
<th>Day</th>
<th>n=1</th>
<th>n=3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>-0.5</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>-1</td>
<td>-1.5</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>2.5</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>-0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 2: Example Second Approach
The corresponding graphic would take this form:

![Diagram](image)

*Figure 11: Derivative Relationship*

**Second Approach**

In the following the second approach will be explained. The aim of this approach is to find pattern in the stock price behaviour which allows a characterisation and segmentation of the data. After having found patterns in the market, this approach tries to forecast which segment follows after the actual segment. Every time a significant change in price happens, the stock price data will be divided into another segment. In this master thesis a drastic change in stock price is called transition.

It is assumed that the stock price data can be divided in the following way:
After each transition follows another segment. To reach the segmentation and prediction of the stock data, the following phases and steps are needed:

1. **Preparation**
   1.1. Smoothing the data
   In the first step the stock price must be smoothed with the Moving Average Filter.

   1.2. Calculation of first derivative
   In the next step the first derivative of the data is calculated, choosing $n$ to be one.

2. **Segmentation**
   2.1. Threshold Calculation
   In the first step of the phase called segmentation, a threshold is developed to divide the stock price data into segments. The threshold will not be applied on the stock price itself, but on the first derivative of the stock price as explained in the following.

   As already mentioned, data in the equity market is highly volatile and therefore, can take any form. Figure (13) shows an example for the development of stock prices during the time:
At first sight this graph might be regarded as “chaotic” without any pattern. On closer examination, however, this graph reveals some important observations as can be seen in depiction (14).

A drastic change in the stock price can be observed, which seems to smooth in the following periods. This observation is very important since it can mean a huge loss for investors. Calculating the first derivative a drop like this can be graphically seen, which is demonstrated in figure (15):
In this depiction it can be observed that the value of the first derivative is much smaller than the other values in the point of the drop. Consequently, the first derivative of stock prices can indicate when the data is changing rapidly. As mentioned before, this happening will be used for the segmentation. To be able to divide the data in segments, the starting and ending points of the segments must be defined in MATLAB. This can be done by using a threshold in the first derivative, which is shown in figure (15). Every time the first derivative crosses this threshold (the red line), the stock price will be divided. In order to find an appropriate threshold to divide the data into segments, the values of the first derivative are plotted in a histogram. The histogram of the first derivative of Continental AG in 2018 is presented in figure (16):
By plotting the first derivative in a histogram one can observe that it follows approximately a Gaussian distribution. The general formula for a single value $x$ is represented by (Bischop, 2006):

$$
N(x|\mu, \sigma^2) = \frac{1}{(2\pi \sigma^2)^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\}
$$

Where,

- $x$ = single real-valued variable
- $\mu$ = mean
- $\sigma^2$ = variance
- $\pi$ = constant
- $1/\sigma^2$ = reciprocal of the variance which is called precision

In the following figure (17) the Gaussian Distribution with $\mu = 0$ and $\sigma = 1$ is depicted:

![Gaussian Distribution](image)

In figure (17) the distribution of data is presented. Approximately 68% of data is within one standard deviation, 95% of data is within two times standard deviation and 99.7% is within three times standard deviation. Based on this rule for Gaussian Distribution, the threshold to divide the stock price data into segments will be whether two times or three times standard deviation. The threshold is not a fix number to provide freedom for adjustments. For some data two times standard deviation might work better than for other data sets due to different values of variances. The application of two times standard deviation means that if the variation in the first derivative is similar to the variation of 95% of the data,
it is still considered as one segment. Is the variation higher than the variation of 95% of the data, it is considered a phase of transition.

2.2. Application of Threshold
Applying the threshold of two (three) times the standard deviation will divide the data as shown below:

As we can see, the data was divided into two segments and a transition part. For analysis reasons, three data points before the drastic change is defined as the start of the transition. To be able to make predictions about the time period when a drastic change will happen, the data points before the actual event have to be analysed. Hence, they must be part of the transition phase. The same is done with the end of the transition to clear the following segment from possible noise. If, for instance, a stable and constant segment follows after the transition, but some data points are still decreasing due to the transition, the segment could wrongly be classified as a decreasing segment. Accordingly, three days before and three days after the green line shown in figure (18) will be relocated to the transition phase. The number of three data points results from testing different values on the stock price of Continental AG.
Furthermore, two corrections for special cases must be done regarding the segmentation of the data. There is the possibility that the first derivative crosses the threshold a couple of times during a short period of time. Therefore, the minimum number of days for one segment must be chose, which is 12 days. A segment that is smaller than 12 days is divided equally and is reallocated to the segment before and the segment after. Another case is the possibility that the value of the first derivative is exactly the threshold, which means the derivative is not crossing but only touching the threshold. Theoretically the transition then would be 7 days, three days before and three days after the data point, and the data point itself. For preventative reasons the transition phase will be decreased to nine days, to avoid noise in the characterisation of the segments.

3. Classification of segments

After dividing the data in different segments, subsegments and phases of transition, a classification of these segments will be realized. In this approach the data will be classified according to its standard deviation and its slope. A segment with a low standard deviation is considered as a stable segment, whereas an instable segment is characterised by its great standard deviation. Furthermore, the tendency of the segment is important. If the average level of the stock price has the tendency to change in time it is whether considered an increasing or decreasing segment. If there is no significant change noticed, the segment is characterised as constant. Moreover, the degree of tendency is considered. If the price, e.g. is only slightly decreasing it is characterised as weakly decreasing, if the slope is very steep it is called strongly decreasing. In the following depiction (19) the different segments of characterising the stock price are presented:
Accordingly, the stock price data can be classified into ten possible segments.

Before the data can be characterised the linear function of each segment is plotted. The formula (8) for linear functions is:

\[ f(x) = ax + b \]

**Tendency:**

The slope of the function is represented by variable a. The threshold between constant and decreasing or increasing is made by using a visual method and then, testing this method on the data base. In figure (20) one can see different slopes with the characterisation of constant, weakly increasing (decreasing) and strongly increasing (decreasing).
The further values of variable \( a \) are developed by using the mean of the slopes of the whole data base as reference, and accordingly allocate values to the different tendencies. Hence, the tendency of a segment will be characterized according to its value \( a \), which can be seen in the following:

\[
\begin{align*}
\text{a} & > 0.2 & \text{strongly increasing} \\
0.2 & \geq a > 0.03 & \text{weakly increasing} \\
0.03 & \geq a \geq -0.03 & \text{constant} \\
-0.2 & \geq a > -0.03 & \text{weakly decreasing} \\
-0.2 & > a & \text{strongly decreasing}
\end{align*}
\]

**Stability:**
To characterize the stability of the segment the standard deviation is used. In the case of a constant segment the general formula (9) for samples can be used:

\[
9. \quad s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

In a stable segment the mean (\( \bar{x} \)) is a constant which is calculated with formula (10):

\[
10. \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

Figure 20: Threshold constant and weakly increasing (decreasing)
In the case of a segment with the tendency to increase or decrease the applied value for the mean is not a constant but is a function, since it needs adjustments. Therefore, the formula for the linear function (8) is used as the mean, which can be seen in figure (21):

![Graph showing standard deviation calculation with variable mean.](image)

*Figure 21: Standard Deviation Calculation with variable mean*

The threshold to divide the data into a stable and an unstable segment is also developed by using a heuristic method. The highest observable value for the standard deviation in the stock data of Continental AG is approximately 2.64. Hence, the threshold 1.32 is used, which is reliable in the data base. Accordingly, the segment is divided in the following way:

\[
\begin{align*}
  s & \leq 1.32 & \text{stable} \\
  s & > 1.32 & \text{unstable}
\end{align*}
\]

Anyway, it is possible that within one segment the characteristic of the stock prices changes, for instance, it is possible that the average value of the stock price changes from constant to decreasing. In this case subsegments of this segments will be developed. In this way, the segments can be analysed according to the classification of segments.
Applying the characterisation of segments on the Continental AG, the result can be graphically seen below in figure (22):

![Figure 22: Segmentation and Characterisation of Continental AG in 2018](image)

For the company Continental AG in 2018, the stock price is divided into five segments, two subsegments and five phases of transition. The year starts with a transition is followed by two subsegments. The first subsegments is classified as unstable & strongly decreasing and the second one as stable & constant. After the next phases of transition follow the segments instable & weakly decreasing, instable & strongly decreasing, stable & weakly decreasing and again stable and weakly decreasing.

4. Prediction

4.1. Application of the methodology on Sample

Applying the methodology on all 20 years of the Continental AG. Each year of the stock price of Continental AG from 1998 to 2018 is divided in segments and transitions. Furthermore, the segments are characterized and classified according to this methodology.

4.2. Development of Probabilities

After applying the methodology to a data base of twenty years, the corresponding probabilities of each year are calculated. This is made by using the historical behaviour of the stock price. Firstly, the frequency of each segment of the past twenty years is calculated. Then the frequencies are converted into
percentages which results to be the probabilities. To aggregate probabilities, the sum rule and the product rule of probabilities are used.

Sum rule:

11. \[ p(X) = \sum_Y p(X,Y) \]

Product rule:

12. \[ p(X,Y) = p(Y|X) p(X) \]

Making use of both rules, the following relationship between conditional probabilities can be developed, which is also called Bayes’ theorem (Bishop, 2006):

\[ p(Y|X) = \frac{p(X|Y) p(Y)}{p(X)} \]

In this study, the sets of events for stability is shown as followed:

\[ S = \{\text{stable, unstable}\} \]

The corresponding probabilities are:

\[ P(S=s) \text{ and } P(S=u) \]

And the sets of events for tendency are:

\[ T= \{\text{strongly increasing, weakly increasing, constant, strongly decreasing, weakly decreasing}\} \]

With the probabilities:

\[ P(T=si), P(T=wi), P(T=c), P(T=sd) \text{ and } P(T=wd) \]

For facility reasons, the probabilities are given in segments. The segments numbered by figure (19) are used. In the table (4) the segments with the corresponding number of segment, their characteristics, the probabilities according to slope and tendency, and the new defined probabilities according to segments are illustrated.
The aim is to make a prediction of the segment that will follow after the actual segment. Hence, the total probabilities of segments are of no interest when predictions are made. The important probabilities are the conditional probabilities. These are given in a table, where the row shows the actual segment and the column shows the probability of a segment if the actual segment has already happened. If, for instance, the value in the row shows the probability for $P(S_1)$ and the column the probability for $P(S_2)$, it means the probability of $P(S_2|S_1)$ is given. Hence, the probability that segment 2 follows after segment 1. The following table is an example how to read the tables of conditional probabilities:

<table>
<thead>
<tr>
<th>$P(S_1)$</th>
<th>$P(S_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(S_1)$</td>
<td>$P(S_2</td>
</tr>
<tr>
<td>$P(S_2)$</td>
<td>$P(S_1</td>
</tr>
</tbody>
</table>

5. Testing the methodology
After having developed the methodology and having gotten probabilities for the segment, the methodology is tested. For this process, the stock price of the company Continental AG for 2019 is divided into segments and characterized according to the developed definitions. Then, the behaviour of the stock price is compared with the probabilities for each segment.

4. Results

First Approach

In the following graphic the result of the first approach is represented. The input data is the first derivative of $n=1$ and the first derivative of $N=n$ ($n=3$, $n=5$ and $n=10$) of the Continental AG in 2018. For simplicity reasons the four quarters of this graphic are numbered.

This figure shows the increasing level of smoothing by choosing a high number for $n$. By applying $n=10$, which are the green points, it can be observed that the data points are getting closer to 0 than the data points of $n=3$ and $n=5$. 
Moreover, the data points tend to increase from negative to positive. That is because if $n=1$ is negative, the corresponding value for $N=n$ also tend to be negative, and the other way around. However, there are varies data points that are not following this rule. This phenomenon of irregularity exists for both cases: if $n=1$ is negative, there are data points of $N=n$ which are positive. This can be observed in quarter I of figure (22). If $n=1$ is positive, there exist $N=n$ data points that take a negative value, which can be seen in the quarter IV. Nevertheless, since the data points are approximately symmetric, so quarter I seems like quarter IV and quarter II looks like quarter III, there cannot be found any observable pattern or outliers that can indicate a characterization of the market or support the predictability of stock price.

**Second Approach**

With the methodology proposed in this master thesis, the stock price of Continental AG from 1998 until 2018 could successfully be divided into segments. Patterns have been found, that allows to classify different parts of the stock price into segments. This research shows, that the stock price can be divided according to its slope and standard deviation into ten different segments, which are stable & constant, stable & strongly increasing, stable & weakly increasing, stable & strongly decreasing, stable & weakly decreasing, unstable & constant, unstable & strongly increasing, unstable & weakly increasing, unstable & strongly decreasing and unstable & weakly decreasing.
Absolut probabilities

In total, 141 segments and subsegments have been found. In the following table (5) the total amounts of each segment and its probability can be seen:

<table>
<thead>
<tr>
<th>Segment</th>
<th>Characteristic</th>
<th>Amount</th>
<th>P(S,T)</th>
<th>P(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stable &amp; constant</td>
<td>18</td>
<td>12.77%</td>
<td>63.1%</td>
</tr>
<tr>
<td>2</td>
<td>Stable &amp; strongly increasing</td>
<td>8</td>
<td>5.67%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Stable &amp; weakly increasing</td>
<td>35</td>
<td>24.82%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Stable &amp; strongly decreasing</td>
<td>8</td>
<td>5.67%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Stable &amp; weakly decreasing</td>
<td>20</td>
<td>14.18%</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Unstable &amp; constant</td>
<td>8</td>
<td>5.67%</td>
<td>36.9%</td>
</tr>
<tr>
<td>7</td>
<td>Unstable &amp; strongly increasing</td>
<td>19</td>
<td>13.48%</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Unstable &amp; weakly increasing</td>
<td>9</td>
<td>6.38%</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Unstable &amp; strongly decreasing</td>
<td>10</td>
<td>7.09%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Unstable &amp; weakly decreasing</td>
<td>6</td>
<td>4.26%</td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td></td>
<td>141</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

The segment, that occurred the most frequent, is segment 3 (stable & weakly decreasing) which makes a quarter of all segments. It is followed by segment 1 (stable & constant), segment 5 (stable & weakly decreasing) and segment 7 (unstable & strongly decreasing). Each of them makes approximately $\frac{1}{8}$ of the segments. The other segments define approximately $\frac{1}{20}$ of the segments. It can be seen that there are more stable segments (63.1%) than instable segments (36.9%). Referred to the tendency, the data base also seems to be stable, since the stable segments with the tendency of constant, weakly increasing and weakly decreasing occurred the most. When the segment is unstable the tendency also seems to be more drastic, since the strong tendencies appear the most. Especially, the tendency of strongly decreasing is observed very often, in instable segments it makes more than $\frac{1}{3}$ of all the tendencies.

Conditional Probabilities

The most interesting part of the results are the conditional probabilities because they give hints about the future stock price behaviour. Hence, it can be used to
predict the stock market. The tables are structured as explained in the chapter before. Before looking at all the segments, the conditional probabilities according to their tendency and their stability are shown in table (6):

Conditional probabilities of tendency:

<table>
<thead>
<tr>
<th></th>
<th>Increasing</th>
<th>Constant</th>
<th>Decreasing</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increasing</td>
<td>42.3%</td>
<td>19.7%</td>
<td>38%</td>
<td>100%</td>
</tr>
<tr>
<td>Constant</td>
<td>46.2%</td>
<td>19.2%</td>
<td>34.6%</td>
<td>100%</td>
</tr>
<tr>
<td>Decreasing</td>
<td>65.9%</td>
<td>15.9%</td>
<td>18.2%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 5: Conditional Probabilities of Tendency

Conditional probability of stability

<table>
<thead>
<tr>
<th>stable</th>
<th>unstable</th>
</tr>
</thead>
<tbody>
<tr>
<td>stable</td>
<td>81%</td>
</tr>
<tr>
<td>unstable</td>
<td>33%</td>
</tr>
</tbody>
</table>

Table 6: Conditional Probabilities of Stability

A very interesting observation is, that if the segment is stable, it also tends to stay stable with a probability of 81%. If it is unstable, there is a probability of 67% that the following segment will be unstable, too. With regard to the tendency, a constant segment is the less probable that occurs after any segment, followed by a decreasing one. The most probable is an increasing segment. Nevertheless, the probability for an increasing segment is the highest after a decreasing one. In two out of three cases, a decreasing segment was followed by an increasing segment in the data base of twenty years. Hence, a good investment strategy is to buy stocks after a decreasing segment. The tendency to increase after an increasing segment and a constant segment does not reach 50%, therefore there is a high risk to lose money or have zero returns on the investment.

The following table (8) shows all conditional probabilities, which means $P(S_n|S_{n-1})$: 
The dashed lines separate the stable from the unstable segments. The probabilities marked in green show opportunities in the market and the red marked numbers show risks. To develop a good investment strategy both probabilities have to be considered. For example, after a segment 3, there is a probability of 28.57\% that a stable & weakly increasing segment follows. Anyway, there is almost the same probability that a stable & weakly decreasing segment follows. It is not recommendable to invest after segment 2, since there is a probability higher than 50\% that a decreasing segment will follow.

**Testing the methodology**

The year 2018 ends with a stable & strongly decreasing segment. According to the calculated probabilities, the year 2019 also has a high probability to start with a stable segment. After decreasing segments there is a probability of 65.9\% that an increasing segment follows. The probabilities according to segments are the following:

- Segment 1: 12.5\%
- Segment 3: 50\%
- Segment 6: 25\%
- Segment 7: 12.5\%
The year 2019 starts with an unstable & strongly increasing segment, which is **segment 7**. The probabilities after segment 7 are the following:

<table>
<thead>
<tr>
<th>Segment 1, 2, 4, 5</th>
<th>5.26%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment 6</td>
<td>21.05%</td>
</tr>
<tr>
<td>Segment 7</td>
<td>26.32%</td>
</tr>
<tr>
<td>Segment 8, 9</td>
<td>15.79%</td>
</tr>
</tbody>
</table>

The second segment of the Continental AG 2019 is **segment 8**. The corresponding probabilities for segment 8 are not very informative, but they are shown anyway:

<table>
<thead>
<tr>
<th>Segment 1, 2, 5, 6, 7, 8, 10</th>
<th>11.11%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment 9</td>
<td>22.22%</td>
</tr>
</tbody>
</table>

After segment 8 follows **segment 6**, which had a probability of 11.11%. The probabilities for the segments after segment 6 are:

<table>
<thead>
<tr>
<th>Segment 3, 5, 8, 10</th>
<th>12.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment 4, 7</td>
<td>25%</td>
</tr>
</tbody>
</table>

Segment 6 is followed by **segment 10**. The corresponding probabilities are:

<table>
<thead>
<tr>
<th>Segment 2</th>
<th>33.33%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment 5</td>
<td>16.67%</td>
</tr>
<tr>
<td>Segment 7</td>
<td>50%</td>
</tr>
</tbody>
</table>

The last segment of the characterisation of the stock price of Continental AG in 2019 is **segment 9**.
We can see that only one time a segment followed, which has a conditional probability of 0%. All the other segments were in the event set of the corresponding segment. Furthermore, it can be observed, that there was a change from stable to unstable in 2018 and 2019. After that, all the following segments were characterised as unstable which is consistent with the absolute probabilities of stability.

5. Conclusion

This research aimed to find patterns in the German Stock Market by using Signal Processing techniques to predict the future stock price behaviour. Based on a quantitative analysis the methodology was developed on a data base of twenty years of the German company Continental AG. This paper illustrates that there are repetitive patterns in the market which can be used to characterise the stock price and divide historical data in segments. There were ten segments used in this investigation with division criteria, which worked for the whole data base.

However, this investigation is facing some limitations. The sample of stock price data needs to be expanded to truly validate the results. Especially, to validate the heuristic methods used in this methodology. A solution could be an algorithm established with machine learning, which improves and cleans the methodology.
from noise automatically. Anyway, it should not be too precise to prevent the effect of overfitting. Furthermore, the data base includes only the most successful, hence the most liquid stocks on the German stock market. Further research is needed to test the validation on smaller companies and foreign stock markets.

Nevertheless, this research is a useful foundation to further investigations. It was possible to combine the finance field with the electrical engineering field in an uncomplicated way, providing useful visual tools for investors as well as researcher of any knowledge level.

6. References


https://www.researchgate.net/publication/258690848_Statistical_trading_using_signal_processing_and_simulation_techniques/link/58b6b569a6fdcc2d14d5dbf6/download


