Trading based on classification and regression trees

Axel Sundén

Abstract

This thesis investigates whether stock picking based on classification and regression trees can be implemented as a successful algorithmic trading system, if only based on technical analysis. To evaluate the performance of this method a fictional portfolio was constructed from the Stockholm Stock Exchange OMX-30, traded on a five-year period.

By means of implementation, classification of the assets in the portfolio was initially conducted. By using threshold values of the weekly returns and comparison with the index of the portfolio, every asset was classified as either outperforming, neutral or underperforming. With a satisfactory classification, each asset that is considered as outperforming is held over a period of one week and at the end of the period the position is terminated and a rebalancing of the portfolio is made. If no assets are classified as outperforming, the portfolio is liquidated and invested at a risk-free rate, defined as the STIBOR 1 week rate.

When backtesting the model we find that the hit ratio of the overall classification is slightly larger than 50 %. During backtesting over the complete trading period it is found that an immense increase of portfolio value is generated. However, since the model is used in sample no predictive validity outside the range can be made. For this reason, 10-fold cross-validation and resubstitution techniques are employed in order to increase the validity if used in an out-of-sample test. Further, a rolling Sharpe ratio is introduced to evaluate the risk-adjusted returns for both portfolios and it is found that the rebalanced portfolio exhibits greater values.

It is concluded that algorithmic trading based on classification and regression trees can be effective in finding patterns that influence the stock prices and that it can form the foundation for an algorithmic trading system.

Acknowledgements

I would like to thank Karl Hallberg at Nordea for sharing his experience as a quantitative analyst and for his knowledge in algorithmic models. I would also like to thank Professor Boualem Djehiche, my supervisor at the Royal Institute of Technology (KTH), for his efforts and experience of portfolio theory. Further, I would like to thank Gert Engman for providing me with invaluable comments regarding my thesis. Moreover, I would like to thank Johan Obermayer for our discussions regarding mathematical statistics and portfolio theory. Finally, I would like to thank my family for their support as well as my friends at the Royal Institute of Technology for making my time there a great and memorable experience.

Contents

1	Intr	roduction	2									
2	Background											
	2.1	Forecasting of the financial markets	4									
	2.2	Algorithmic Trading	5									
		2.2.1 Classification and Regression Trees and Technical Analysis	7									
3	Theory 1											
	3.1	Data sets	11									
		3.1.1 Other data	12									
	3.2	Method	13									
		3.2.1 Financial data and basic portfolio construction	13									
		3.2.2 Technical Indicators	15									
		3.2.3 Oscillators	16									
		3.2.4 Stochastics	19									
		3.2.5 Indexes	23									
		3.2.6 Indicators	26									
	3.3	Construction of a Tree Classifier	34									
		3.3.1 Growing the tree	36									
		3.3.2 Pruning the tree	40									
	3.4	Construction of Rebalanced Portfolio	42									
	3.5	Sharpe Ratio	43									
4	\mathbf{Res}	ults	45									
	4.1	Performance of model	45									
	4.2	Behavioural Finance	54									
Co	onclu	ision	57									
Bi	ibliog	graphy	58									

Chapter 1 Introduction

Developments in the field of financial mathematics have been great during recent years. Many new areas of research have been introduced and most significantly, the combination of progressions within financial mathematics, aided by tremendous computational power, has made it possible to conduct research that would have been impossible only a few years ago. One specific field that has especially increased is the electronic financial market and it is estimated that today the majority of trading is conducted electronically. One sector within the electronic financial market that has achieved an extra focus is algorithmic trading and methods of forecasting the markets.

Algorithmic trading is a method where a computer is conducting a specific investment instead of a human. As described in the literature, these trading systems implement historical data with respect to well-defined rules, whereas traditional trading only implements a specific strategy [18]. The same underlying technique, as for algorithmic trading, is also applicable for the methods of forecasting financial markets. As computational handling is needed, it can be said that there exists many benefits as well as pitfalls of this quantitative approach to trading and forecasting. First of all, many suggest that quantitative approaches are superior in comparison with traditional approaches. The main reason for this is that the algorithms do not suffer the biases of that humans tend to have, such as escalation bias, self-confidence bias et cetera. Moreover, a quantitative approach can make an investment more efficiently at a lower price thanks to a quicker simultaneous analysis of many factors. On the other hand, a quantitative approach to trading can also be less effective since the judgmental aspect of a human is not directly applied. Furthermore, there might also be issues with the calibration of trading system, resulting in e.g. incorrect timing of the buying and selling of an asset. Generally, the method of conducting investments based on mathematical models has both been praised and condemned as described by Brock et al. (1992) and Fama and Blume (1966).

Since global markets continuously evolve and become more interactive, forecasting of financial markets and trading activity will play a more crucial role in the future. For this reason, the topic of trading and forecasting at an increased pace coupled with improved accuracy will also be playing a more important role ahead. However, this task itself is very complex due to the non-stationary, noisy and deterministically unpredictable nature of the financial markets. This study examines the method called *classification trees* and its ability to forecast stock price movements effectively when only based on technical market data. The method itself will be described, implemented and finally evaluated with the objective of determining if it can yield significantly better returns than simply investing in an index fund.

This thesis is structured as follows: Chapter 2 describes the background of the field of study, classification trees as well as the usage of the methods within the financial markets. Furthermore, this section illustrates the basic structure of the model. Chapter 3 introduces the underlying data of the analysis and how the financial portfolios and technical indicators are constructed as well as the mathematics of classification trees. Chapter 4 consists of the results from the study and the influence of behavioural finance on the results.

Chapter 2

Background

2.1 Forecasting of the financial markets

There exists a lot of research that support the idea of forecasting financial markets. In an early study, conducted by Lo and MacKinlay (1988), it is shown that weekly stock market returns do not follow a random walk process. Employing a specification test based on simple volatility proves this fact and it is suggested that the markets are not completely random. There are some fundamental ideas regarding this behaviour of the markets. First of all, the non-stationary behaviour of financial markets is regarded to be a result of a distribution of the financial time series that is constantly changing. Secondly, the noise of the time series is due to the incomplete information of the financial markets where models do not manage to capture the autocorrelation between historic and future prices. Finally, the paradoxical behaviour of being both deterministic but still unpredictable refers to that the financial markets are in a long perspective deterministic, but random on a short horizon. Besides this study, Keim and Stambaugh (1986) present findings that suggest that there exist some variables that can predict the stock market movements on short term basis and that forecasting in some sense is possible.

Most studies regarding predictability of stock markets focus on well-established markets, some European but especially the US market. Regarding the European markets, Ferson and Harvey (1993) find that returns are to some extent predictable by investigating national equity markets and various global variables. Regarding the US market, Fama and French (1992) investigate the relationship between various fundamental variables and stock returns. Their main finding is that cross-sectional variation in average returns can be captured by book-tomarket equity and size. Rosenberg, Reid, and Lanstein (1985) find that there exists a positive correlation between the ratio of a firm's book value of common equity to its market value and the average returns on stocks. Similar results are also found by Stattman (1980). However, as proposed by Fama and French (1992), many of these factors are only scaled versions of price and may thus be redundant when describing returns, suggesting that technical factors, i.e., factors that only represent information embedded in the stock price, might not capture future returns correctly. This brings up the topic of technical analysis as a tool for predicting the stock market that may or may not be satisfactory.

Technical analysis (TA) is a method for predicting stock returns by finding patterns and relationships in the historic financial time series. This method is based on using primarily stock price and volume data. Different variations of this data are also present within technical analysis such as e.g. highest and lowest stock price as well as derivative prices.

The method of technical analysis itself stems from the *Dow Theory*, established by Charles Dow in the beginning of the nineteenth century and has been both praised and rejected. For instance, according to Fama (1970) the weak form of the efficient market hypothesis states that all information regarding a company is reflected in its stock price. Thus, it should be impossible to forecast the stock prices since all information is already known. However, many empirical studies do not support the weak form of the efficient market hypothesis and rather suggest that the financial markets are only efficient in theory. These findings suggest that technical analysis can be used to capture profitable opportunities when trading.

Evidence of profitable trading, when using technical analysis, is shown by Kwon and Kish (2002). Their findings suggest that technical trading rules are superior when compared with traditional buy-hold (B&H) strategies. This is found by investigating technical trading rules such as momentum, trading volume and moving averages by employing bootstrap, GARCH and t-test methodologies. Similar results are also found by Neftci (1991) when studying trading rules of various algorithms based on technical analysis. Strong support regarding technical analysis is also found by Brock, Lakonishok and Blake (1992) where they implement moving averages and trading range breaks as trading rules. Their findings suggest that stock returns are correlated differently depending on if buy or sell signals are present. Furthermore, and what might be relevant regarding our study, are the findings of Wong et al. (2003) where the timing of entry and exit in the stock market is investigated. Technical indicators such as moving averages and relative strength indexes are employed within their study and it is found that it is possible to generate substantial positive returns when using their approach. Further, Vasiliou et al. (2006) show similar findings by implementing trading techniques based on technical analysis. Consistent with previous studies, their research suggests that trading based on technical analysis can generate large excess returns over buy-hold strategies. Over a one-year (250 days) period, trading based on various indicators generated returns that varied between 36.10 % and 55.65 % while buy-hold strategies only generated returns of 12 %. Hence, since the efficient market hypothesis is contradicted by these empirical studies, it might be possible to use technical analysis to construct efficient trading rules for construction of a trading algorithm based on classification trees.

2.2 Algorithmic Trading

Ever since technical analysis has shown itself to be beneficial, it has been implemented on numerous occasions as algorithmic trading systems. One of the first systems was created by Pruitt and White (1988) and was based on technical indicators such as volume, moving averages and relative strength indexes. As this system generated excess returns, many other trading systems were also implemented, as described by e.g., Brock et al. (1992), Bessembinder and Chan (1995) as well as Ratner and Leal (1999).

As trading systems have been refined to make better stock price predictions, various alternative approaches have been developed. Methods that have gained ground for implementation are parametric models. One of these models is the Generalized Methods of Moments (GMM). This method's success in estimating financial time series was investigated by Hansen and Singleton (1982). Moreover, the use of the GMM models for construction of optimal portfolios was also studied by Brandt (1999). The advantages of this approach and the reason why it is so commonly used in forecasting are, as described by Leung et al. (2001), that one does not have to create assumptions of the distribution of the time series and that the method is very flexible. Other forms of parametric approaches regarding trading systems and forecasts of financial time series can be based on the theory of random walk. This approach tends also to be a relevant benchmark due to its coupling to the efficient market hypothesis.

However, rather than using parametric methods such as the GMM or random walk theory, there exist nonparametric methods that may in several cases prove themselves better. For instance, in a study made by Hill et al. (1996) it is found that non-parametric methods can generate much better results than several traditional statistical models including advanced time series models. The reason for this is first and foremost that traditional parametric methods show bad mapping capabilities for nonlinear relationships, often seen in real-world financial time series. An example of this is shown by Refenes et al. (1994) where traditional statistical approaches of forecasting financial time series quickly becomes unsatisfactory if exposed to nonlinearities within the data sets. For this reason, non-parametric methods such as various neural network and decision tree approaches have been taken into account within the field of forecasting financial markets.

There exist many different non-parametric methods for algorithmic trading. Some of them have gained extra focus, such as artificial neural networks, genetic algorithms and fuzzy logic as described by Chavarnakul and Enke (2009). Another development is the method of chart pattern analysis as made by Lo et al. (2000). In this study, various chart patterns for analysis of technical indicators are introduced as well as employing pattern recognition processes such as non-parametric kernel regression to the financial time series.

As seen in the previous section, non-parametric methods are regarded as superior in comparison to statistical and parametric methods. However, they all have different strengths and weaknesses. For instance, neural network approaches may not exhibit convergence since the optimization of the algorithm is conducted by searching for a minimum along the error surface. During this procedure one may not find a global minima because of getting stuck in a local one [23]. Moreover, the neural network methods often experience problems with overfitting when learning from the data in the sample as well as with the setting of the various parameters that are needed in the algorithm. Finally, the different steps in most of the non-parametric methods are very difficult to interpret. Nevertheless, non-parametric model such as classification trees exhibits most of the advantages of the subsequent non-parametric models, though at the same time they are very simple to interpret and implement in a system.

2.2.1 Classification and Regression Trees and Technical Analysis

As shown, often problems can be solved by using standard pattern recognition. The use of the pattern recognition exist in many disciplines such as computer vision, medicine, manufacturing as well as in finance. Examples of this are:

- Medical diagnosis of a patient based on symptoms and test results, e.g. blood pressure, pulse and temperature
- Credit risk of a customer regarding loans based on e.g. credit history, income and present loans
- Species of a certain flower based on e.g. colour, size and other variables

Tree structured approaches such as classification trees might provide a superior perspective and may also generate several advantages in comparison with other non-parametric methods. Maybe most importantly, a tree-structured alternative provides overwhelming simplicity due to its binary nature. This results in both fast classifications of new data as well as a lucid overview of the model, which can be easily explained. Moreover, the method yields decision boundaries that are complex and the method is equally appropriate for both ordered data and for categorical data or a mixture of them.

An approach based on classification and regression trees exhibits several benefits if compared with the parametric and nonparametric approaches that have already been introduced. As shown by Breiman et al. (1984), who introduced the method, classification trees can be used with any data set given a satisfactory vector of questions. This means that the model can be applied to both categorical data and ordered data, which in its turn has the consequence that one does not have to be bound to a certain category when conducting the research. Secondly, the model is very good in handling nonhomogeneous relationships with respect to conditional information. This has the implication that, when a node is split into two children, the algorithm searches for the next best split within each child independently of each other. Moreover, within standard structures of data, Breiman et al. (1984) also show that the results are invariant under monotone transformations of individual variables, i.e., if x_i is an ordered variable within a specific problem, the optimal split will be the same regardless of transformations such as $x'_i = x^n_i$ are conducted. The biggest advantage of classification trees may be the nonparametric and robustness properties concerning misclassified points and outliers within the data set. The model regards each data point as only one among N data points and thus one does not have to have a priori data structures or distributions.

The classification tree model is a method that uses a *classifier* or a *classification rule* to establish what class a certain object should fall into. For an implementation, it is necessary to construct a classifier based on a historic data to be able to classify future observations. Thus, one already must have a data set where the class of each observation is known, i.e., there exists a class label for each case. For this reason, this method of pattern recognition is known as *supervised learning*. The structure of supervised learning is illustrated by the schematic

Figure 2.1: Illustration of the pattern recognition process



Diagram of the different steps during a statistical pattern recognition method.

diagram in figure 2.1, as proposed by Duda and Hart (1973). The basic process is initiated when an object enters the system. The object has several features that are extracted by a sensor and is compared with predefined classes. When conducting this, the specific object is classified into a certain class membership by using specific classification rules as shown in figure 2.1. The technique should therefore be regarded as a multi-stage decision process with binary decisions at each level. The tree structured nature of the method means that it is constructed by nodes and branches. Each branch is the connection between two adjacent nodes. There always exist more nodes than branches in the structure as there must exist nodes at the end of the final branches. Every split node is called a *parent node* and the two nodes that are produced by the split are called child nodes. Generally, the first node within the tree is called the root node. The final nodes in the tree are known as *terminal nodes* and have an associated class label to them and denotes the final classification of the data. As seen, the terminal nodes are never split into children. The other type of nodes are instead known as nonterminal nodes or internal nodes since they are always split into two children. As an example we introduce a basic classification tree in figure 2.2.

In this example we use two describing features in the example, namely X_1 and X_2 , which together create a feature vector. This can be created if one utilizes ordered variables and fixed-dimensional data. For instance, consider the example in figure 2.2 again. In this case, the procedure partitions the two-dimensional space \mathbb{X} with respect to posed questions of the form $x_i \leq c_n$ where c_n is a constant. The filters at each level are made in a way that the data in each subset of \mathbb{X} becomes more homogenous with respect to its class. However, the way of viewing the recursively partitioning process with rectangles becomes very burdensome when the dimension is increasing. For this reason, the same structure can also be described by a binary tree structure. At each level the same type of



Figure 2.2: Illustration of the pattern recognition process



Figure 2.3: Illustration of the pattern recognition process



Tree based approach of the previous example. As seen, if not $X_1 \leq 0.5$ then one achieves the child t_3 that denotes the classification *. Similarly, if $X_1 \leq 0.5$ holds and not $X_2 \leq 0.5$ one achieves child t_5 and class \diamond . Similarly if $X_1 \leq 0.5$, $X_2 \leq 0.5$ and not $X_1 \leq 0.25$ one achieves class \times . Finally, if $X_1 \leq 0.5$, $X_2 \leq 0.5$ and $X_1 \leq 0.25$ one achieves class +.

filtering process is conducted as in the previous case and the positive response results at each level results in a step to the left child, t_L , and a negative response results in a step to the right, t_R . By doing this, the previous example results in a tree diagram as in figure 2.3.

However, it is apparent that constructing a tree classifier is problematic if not implemented properly. Actually, it turns out to be very dependent on the splitting and the stop-splitting criteria of the data, i.e., when a node is a child or a terminal node. To fully understand how the problems arise and how they can be solved, we must first have an understanding of the preliminary fundamentals. For this reason, data and the features of the data, which will be employed during implementation, as well as the mathematics of classification trees will be described in the upcoming sections.

Chapter 3

Theory

3.1 Data sets

The data that we use in this study comes from 10 randomly picked stocks traded at OMX-30 at the Stockholm Stock Exchange. We have decided to use weekly data that spans from 1 January 2003 until 1 January 2008 for the companies as shown in 3.1. The weekly closing prices, quoted in SEK and provided by Morningstar, Inc. constitutes the historical data set that is employed to calculate stock returns.

Assets:	Abbrevation:	Assets:	Abbrevation:
ABB	ABB	Nordea	NORDEA
AstraZeneca	AZN	Sandvik	SAND
AtlasCopco	ATC	SKA	SKA
Boliden	BOL	SHB	SHB
Electrolux	ELUX	Volvo	VOLVO

Table 3.1: Companies used in this study

The closing prices for the different companies can be seen in figure 3.1 and in figure 3.2. The plots in the figures are not adjusted for splits and extraordinary events as can be seen for e.g., Atlas Copco in mid-2007 and Electrolux in mid-2006. However, these are neglected in this study but adjusted for when calculating the returns and technical indicators by interpolation.



Figure 3.1: Price of asset used in the rebalanced portfolio and in the index portfolio

A plot of the price of assets used in the study.

Figure 3.2: Price of asset used in the rebalanced portfolio and in the index portfolio



A plot of the price of assets used in the study.

3.1.1 Other data

To be able to calculate other factors such as the technical indicators, we use data for the 10 stocks traded on OMX-30 on the Stockholm Stock Exchange for the same period. The data sets used are opening prices, highest price, lowest price, traded volume and time of day. This historical data is also quoted in SEK and provided by Morningstar, Inc. Besides what has already been stipulated, no other modifications of the data has been made with the exception for converting it from a .txt file to a MySQL database. This format has shown itself suitable since all implementations have been conducted using MATLAB (R2007a).

3.2 Method

Let us first of all review the scope of this study. The aim is to investigate whether a profitable portfolio of stocks can be built by using a classification and regression tree approach based on technical analysis. Therefore, our main objectives are creating and evaluating the performance of a general index portfolio compared to a rebalanced portfolio that is based on technical analysis. To be able to do this we must establish a set of technical indicators that describe our securities from a technical analysis point of view. Moreover, we must classify our securities based on whether a particular security performs better or worse than the average. For this reason, we will first of all create a set of rules that indicate the performance of the portfolio assets over time. Secondly, we will create an algorithm based on the classification tree approach that classifies our assets according to this set of rules. When the classification for each new time period is made, we rebalance our portfolio with respect to this classification.

To form a solid foundation for the forthcoming analysis, this chapter will describe the fundamental tools that will be used. First of all we will describe the mathematics regarding financial data and how stock market returns are calculated as well as how an index portfolio is created. Secondly, we will present the different technical indicators that will be used within the classification algorithm. Thirdly, we will introduce the algorithm of classification and regression trees and the mathematics behind it. In this section, we will also explain the different procedures in the algorithm, i.e., growing the tree, pruning the tree and finally how to conduct the classification of new data. Finally, we present the method of constructing the rebalanced portfolio in compliance with Swedish law regarding investment funds.

3.2.1 Financial data and basic portfolio construction

The financial data used regarding returns has the price of an asset as an underlying basis. When calculating the rate of return in the discrete case of a portfolio one has to calculate the rate of return of each asset within the portfolio. Let us take a number of quotations T, where $t = 1, \ldots, T$, of an asset i. Secondly, introduce a vector of returns $r_{i,t}$ that denotes the observation of each random variable r_i . If the price of this specific asset at time t is introduced as S_t and the price of the same asset at time t - 1 is S_{t-1} then the rate of return of this asset is consequently

$$r_{i,t} = \Delta S_t = \frac{S_t - S_{t-1}}{S_{t-1}}$$

If this is expanded into the continuously compounded case and we denote $S_i(t)$ as the time series of the price of the same asset *i* we get,

$$r_{i,t} = \ln S_i(t) - \ln S_i(t-1)$$

This result can help us define the expected return R_i of the same asset and from basic probability theory we thus achieve,

$$R_{i} = E(r_{i}) = \bar{r}_{i} = \frac{1}{T} \sum_{t=1}^{T} r_{i,t}$$

Similarly, let us examine the case where we have portfolio consisting of multiple assets. First of all, let us introduce N assets within the portfolio, where the asset i in the portfolio has return r_i and weight or fraction w_i of the total portfolio and since our budget constraint $\sum_{i=1}^{N} w_i = \mathbf{w}^T \cdot \mathbf{1} = 1$, one achieves the return of the portfolio over period t - 1 to t as,

$$r_p = \sum_{i=1}^N w_i r_i$$

If we introduce a vector consisting of the expected returns of the assets within the portfolio as \mathbf{R} we get the expected return of the portfolio as,

$$R_p = \sum_{i=1}^N w_i R_i = \mathbf{w}^T \cdot \mathbf{R}$$

In the case of rebalancing of an portfolio we may allow *short selling*. When constructing a portfolio we might also set specific restrictions regarding the weights of each asset. By definition short selling represents the act of selling assets that have been borrowed. Thus, the weight of each asset can both be negative or positive is this act is allowed. However, if short selling is restricted we must introduce a restriction according to $\forall i \ w_i \ge 0$. This means in that we may only buy assets and is known as taking a *long* position.

The continuous compounded case, as already introduced, can be expanded into an approximation that can handle small asset returns as seen in equation 3.1. However, to achieve an exact result one has to add an error term ϵ_t .

$$\Delta S_t \approx \ln S_t - \ln S_{t-1} = \ln \left(\frac{S_t}{S_{t-1}}\right) = \epsilon_t \tag{3.1}$$

Now we can introduce an important difference between the price and the rate of return of an asset. To fully understand the difference, we introduce the concept of autoregressive models. Autoregressive models used in this setting are typically the AR(1) process,

$$X_t = c + \varphi X_{t-1} + \epsilon_t$$

As seen, variable X_t depends on the constant c, the variable X_{t-1} and a white noise, i.e., $\epsilon_t \sim N(0, \sigma^2)$. For convenience the constant is normally set as c = 0. Moreover, if the parameter $|\varphi| < 1$ we achieve a process that is stationary. If $\varphi = 1$ we achieve a random walk and if $\varphi > 1$ we get a non-stationary process. A process that is stationary is considered to show mean reversion meaning that prices and returns eventually move back towards the mean. Similarly, if the logarithm is applied to the price of an asset tomorrow we get the logarithm of the price today times the parameter φ with an added error term ϵ_t , i.e., the price of an asset tomorrow is the same as the price today times a factor and then an error variable. The factor within the process is normally $\varphi = 1$ which results in

$$\ln S_t = \ln S_{t-1} + \epsilon_t$$

Thus, it is shown that price data is an autoregressive process that exhibits a random walk. On the contrary, the returns data, as seen from equation 3.1, are stationary and is white noise around zero. Once again, we end up with the fact that returns should theoretically be highly stochastic and stationary, given that the markets would be efficient as in the efficient market hypothesis.

3.2.2 Technical Indicators

In the previous section it is shown that rate of return is stochastic and that the price of assets cannot be forecasted given that the efficient market hypothesis holds. Nevertheless, in the empirical studies that have been introduced it is shown that future stock prices and returns are somewhat correlated with various technical indicators.

Technical indicators can be divided into different groups depending on their nature and name. However, the group membership of some factors can be dual since some of the features of two or more groups might be present and the grouping is not definite. They all include various combinations of the open, high, low or close prices over different time periods and for each new time period a new data point is produced. We have chosen to group the indicators according to the preceding list:

- 1. Oscillators
- 2. Stochastics
- 3. Indexes
- 4. Indicators

It should initially be stressed that many of the technical indicators complement each other. In an interview, the founder of the magazine *Technical Trends*, Arthur A. Merrill, recommends that several indicators should be used simultaneously. In the interview he stipulates that if 40-50 different indicators are used the diversification will be favoured, hence reducing the risk of a emotional attachment [3]. However, even though this does not affect an algorithmic trading system directly, it may affect the implementation when choosing which indicators to use. For this reason, several technical indicators, as seen in table 3.2, are employed in the study and the ones that are the most significant ones during classification will be favoured automatically by the algorithm.

Throughout this section we use some specific variables when calculating the various technical indicators and these are summarized in table 3.3.

Table 3.2: Technical indicators that are used in the study and the specific group that each indicator belongs to.

Name of technical indicator:	Type of technical indicator:
Momentum	Oscillators
Accumulation Distribution Oscillator	Oscillators
Fast Stochastics %K	Stochastics
Median Price	Indicators
Fast Stochastics %D	Stochastics
Negative Volume Index	Indexes
Highest High	Indicators
Positive Volume Index	Indexes
Lowest Low	Indicators
Slow Stochastics %K	Stochastics
Price and Volume Trend	Indicators
Slow Stochastics %D	Stochastics
Accumulation Distribution Line	Oscillators
Acceleration Between Times	Oscillators
Relative Strength Index	Indexes
Bollinger Bands	Indicators
Volume Rate of Change	Indicators
PercentB	Indicators
Price Rate of Change	Indicators
Bandwidth	Indicators
On-Balance Volume	Indicators
Volatility	Stochastics
Chaikin Volatility	Stochastics
William's % R	Stochastics
William's Accumulation Distribution Line	Indicators

Table 3.3: Variables that are used in the technical indicators

Variable:	Details:			
P_h	Price high			
P_l	Price low			
P_o	Price open			
P_c	Price close			
P_t	Price at time t			

3.2.3 Oscillators

Oscillators are technical indicators that fluctuate inside a specific range. This can for instance be fluctuation over a central line, a point or between upper and

lower boundaries. Moreover, these indicators vary over time but cannot trend for longer periods. For this reason, they are tools that are widely used to calculate if a particular asset is overbought or oversold. They are also used to spot the strength and direction of momentum of an asset's movement. The common interpretation of centralized oscillators is that the asset is *bullish* when trading occurs above the centreline of the oscillator, and vice versa when the asset is *bearish*. The technical indicators that belong to this group are Accumulation distribution oscillator, Price momentum and Acceleration between times.

Accumulation Distribution Oscillator

The accumulation distribution oscillator is a technical indicator that takes the opening, closing, high and low prices of the asset into account. This type of oscillator is regarded as normalized since it is divided by the range of the measurement period and treats each new measurement period individually. The interpretation of the accumulation distribution oscillator is as described in Kaufman (1987) as,

$$ADO = \frac{(P_h - P_o) + (P_c - P_l)}{2 \cdot (P_h - P_l)} \cdot 100$$

In the previous expression P_h denotes the high price, P_o is the opening price, P_c is the closing price and P_l is the low price of the period. To visualize the concept the accumulation distribution oscillator of AstraZeneca is plotted in figure 3.3.



Figure 3.3: Plot of accumulation/distribution line as a technical indicator

The concept of accumulation distribution oscillator implemented for AstraZeneca.

Price Momentum

Momentum is a measure that calculates the ratio between closing prices at different time horizons. When we employ the closing price we get,

$$MO_N = \frac{P_{c,t=0}}{P_{c,t=N}}$$

In the expression the variable N denotes the time lag when calculating the difference in price. In this study we use two different time lags, N = 2 and N = 4, two and four weeks. The price momentum for AstraZeneca is shown as an example in figure 3.4. Price momentum can be used in many ways and



Figure 3.4: Plot of price momentum as a technical indicator

The figure shows the price momentum indicator for AstraZeneca. One can clearly see the trend indications.

can also range over both short and long periods depending on the scope of the analysis. One way of using the momentum oscillator is to find specific trends in the financial data such as uptrend or downtrend movements. The general approach is to sell when the indicator has reached its peak and starts to decline, and vice versa. Further, it is assumed that when the indicator has reached very high levels, the price may also reach significantly higher levels in the near future. An advantage of the oscillator is its dual function where it can be used both as a relative strength-ranking tool as well as a momentum indicator for analyzing price movements and internal strength of the stock [13].



Figure 3.5: Plot of acceleration between times as a technical indicator

The plot indicates the acceleration of price momentum over time for AstraZeneca.

Acceleration Between Times

Acceleration between times is the measure of change in momentum between two periods and described by Kaufman (1987) as,

$$ABT_N = MO_{t=0} - MO_{t=N}$$

In our survey we calculate the acceleration between times for the assets with the price momentum as data input, where we employ a period of N = 4 as seen in figure 3.5. The acceleration between times provides us with a clear indication of the trend of the momentum over the past month.

3.2.4 Stochastics

Technical indicators that belong to this group are *Fast stochastics*, *Slow stochastics*, *Chaikin volatility*, *Ex-post volatility* and *Williams %R*. The indicators within this group have in common that they focus on the relationship between the closing price and the price range of stocks over time. The different indicators are very capable of signalling when the market is considered overbought or oversold as well as upward and downward trend patterns.

Fast Stochastics

Fast stochastics is an indicator that can be measured over different time periods. The indicator is divided into type F%K and F%D. As proposed by Achelis

(1995), the first type is calculated as,

$$\%K = \frac{P_c - \min P_{l,t=1,\dots,N}}{\max P_{h,t=1,\dots,N} - \min P_{l,t=1,\dots,N}} \cdot 100$$

In the preceding expression min $P_{l,t=1,...,N}$ and max $P_{h,t=1,...,N}$ denotes the lowest low price during the N last periods respectively highest high during the N last periods. A common value of the parameter is N = 14, though this value can altered depending on the orientation of the study. When calculating the F%K in this study the parameter was set as N = 6. When calculating F%D



Figure 3.6: Plot of fast stochastics as technical indicators

Since fast stochastics is a percentage ratio the y-axis will range between 0 and 100%.

one has to employ a *n*-period simple moving average to the F%K stochastics. The parameter of the simple moving average is often set as n = 3 and this value is employed in this study as well.

Fast stochastics may be used in several ways. One of the most important indications is when F%K and F%D intersect with each other, generating either a bullish or a bearish signal. For this reason the F%D line is called the trigger line, since this is the line that triggers a signal when it intersects with F%K. Further, the movements of the indicators in relation to the price movements may also give clear indications of changes in the trend. The appearance of the two stochastics for AstraZeneca can be seen in figure 3.6.

Slow Stochastics

Slow stochastics are indicators that are very similar to the fast stochastic indicators. The reason for this is that they are calculated from the fast stochastics by employing simple moving averages. Hence, we achieve two different types of slow stochastics, S%K and S%D. The first type, S%K, is calculated by using

Figure 3.7: Plot of slow stochastics as technical indicators



Since slow stochastics is a percentage ratio the y-axis will range between 0 and 100%.

a simple moving average applied to F%K with the parameter set as n = 3. The second type, S%D, is calculated by applying a simple moving average with n = 3 to the S%K instead. By this procedure, we achieve a smoothed out version of the fast stochastics as seen in figure 3.7. This implies that we reduce the number of false intersections between the two lines and that we may achieve a more accurate perspective of the price movements. As seen though, the F%D and S%K are exactly the same. However, the F%D is employed as a trigger line in fast stochastics and this is not the case for S%K in slow stochastics. However, the concept of intersection is important in slow stochastics as well and the crossing of the S%K and S%D may indicate important trend movements. Besides this, other important indications of price movements include the case when the indicators move towards the upper and lower bounds.

Chaikin Volatility

Chaikin volatility is measured by calculating the spread between high and low prices over a predetermined period with an applied *n*-period exponential moving average (EMA) and then measuring the *m*-period rate of change as explained by Achelis (1995),

$$Volatility_{Chaikin} = \frac{\left(HL_{EMA_{n,t=0}} - HL_{EMA_{n,t=m}}\right)}{\left(HL_{EMA_{n,t=0}}\right)} \cdot 100$$
$$HL_{EMA_{n,t}} = EMA_n(P_{h,t} - P_{l,t})$$

The normal appearance of this indicator can be seen for AstraZeneca in figure 3.8 where we have used n = 6 and m = 6.

When using the Chaikin volatility as an indicator, there exist different approaches. One theory is that when prices turn towards the bottom one will





One can clearly see the indications of the Chaikin volatility.

experience a sudden increase in the Chaikin volatility over a short period of time right before hitting the bottom. The same theory stipulates that when getting closer to the top, the volatility will decrease, signifying a bull market that is mature. The fundamental idea of this is that traders focus on terminating positions quickly before hitting the bottom and when being close to the top trades are made on a slower pace. On the other hand, the second approach suggests the opposite explanation. This approach suggests that while being at the top, traders are uncertain of possible outcomes, which results in increased volatility when taking and terminating positions at a higher pace. Similarly, while being at the bottom positions are taken at a slower pace since they exhibit great uncertainty, resulting in a decreased volatility.

Williams %R

Williams % R (percent R) is a stochastic indicator that shows whether a stock is overbought or oversold. The indicator is very similar to the F% K but is defined instead as [1],

$$\%R = \frac{\max P_{h,t=1,...,N} - P_{c,t=1}}{\max P_{h,t=1,...,N} - \min P_{l,t=1,...,N}} \cdot -100$$

Since this is a percentage ratio the plot will be bounded between [0, -100] as seen in figure 3.9. The main indication of % R is when the line is close to the upper and lower bounds, normally around -20 % and -80 %, meaning that the stock is overbought or oversold respectively. One may advantageously combine this indicator with price indicators. The general approach is not to sell an asset before the price decrease of the asset is realized.





William's %R implemented for AstraZeneca. Since the indicator is a negative ratio, the y-axis spans normally from 0 to -100.

Ex-post Volatility

Ex-post volatility, or historical volatility, is calculated as the standard deviation of the historical time series of returns. One must use the continuously compounded logarithmic returns over a specific time period and we choose to measure the volatility over 6 weeks. To calculate the volatility over a time period T, the standard deviation σ_{SD} of the returns is scaled according to,

$$\sigma = \frac{\sigma_{SD}}{\sqrt{T}}$$

The ex-post volatility is measuring the magnitude of the return movements for a given time frame and is therefore sometimes considered as the risk when taking a position. Consequently, a stock with large ex-post volatility may have great differences in return, which might generate large profits, but at the same time large losses. As seen in figure 3.10 where the ex-post volatility of AstraZenca is plotted, the values are ranging between 0 % and 7 %. Nevertheless, the magnitude of volatility can be much more fluctuating and exhibit large values and often depends on the size of the company, traded volume and price.

3.2.5 Indexes

Within this study, we implement the *Relative strength index*, *Positive volume index* and *Negative volume index* as technical indexes.

Figure 3.10: Ex-post volatility of AstraZeneca



Ex-post volatility for AstraZeneca. One can clearly see that there exist periods that have higher magnitude and periods that have lower magnitude of the volatility.

Relative Strength Index

The relative strength index is an indicator that evaluates the difference between the present prices with the most normal price. Since the measure is ratio based it normally take values in [0, 100]. When calculating the relative strength index one may use various time-spans and we choose to calculate our indicator using data for 6 weeks. However, one might also use shorter and longer periods.

The relative strength index is employed to evaluate several aspects of an asset. The preliminary theory is that the maximum value 100 of this indicator signifies that the price has decreased throughout the whole measurement period. On the other hand, the minimum value 0 indicates that the price has increased throughout the whole period. Generally, it is considered that values over 70 and below 30 indicate oversold or overbought assets respectively. Hence, an index rising above 30 would be viewed as a buy-signal. Similarly, a sell-signal is when the index decreases and falls below 70. A popular method of using this indicator is to look for *divergences* between historical prices and the indicator. This means that one tries to find patterns where the relative strength index does not show the same appearance as the price movements, signifying an upcoming price movement in accordance with the relative strength index movements in the near future. The appearance of this index can be seen for AstraZeneca in figure 3.11.

Positive and Negative Volume Index

Technical indicators such as the positive and negative volume indexes indicate whether the volume has increased or decreased over a predetermined time pe-



Figure 3.11: Relative Strength Index of AstraZeneca

Relative strength index for a slightly longer period than 6 weeks is constructed AstraZeneca. One can clearly see the indications of being overbought and oversold.

riod. When high values of the positive volume index are evident it is believed that uninformed and inexperienced traders take positions. Conversely, low values are thought to indicate that well-informed investors take positions. Similarly, the same analogy holds for negative volume index.

To calculate the index one has to use the closing price and the trading volume. When calculating the positive volume index, as in Achelis (1995), one determines if the trading volume has increased between two consecutive trading periods. If the trading volume has not increased, the positive volume index is equal to the positive volume index of the previous time. However, if the trading volume has increased the positive volume index is set as,

$$PVI_t = PVI_{t-1} \cdot \left(1 + \frac{P_{c,t} - P_{c,t-1}}{P_{c,t-1}}\right)$$

Similarly, the previous expression holds for the negative volume index, NVI_t , but with a minor modification. If the trading volume has not decreased, the negative volume index is equal to the value of the previous time. However, if the negative volume index has decreased the value is changed with the same closing price ratio. Since both calculations need an initial value both indexes are given a start value of 100.



Figure 3.12: Positive and Negative Volume Index of AstraZeneca

Positive and negative volume index for AstraZeneca. One can clearly see the indications of being overbought and oversold.

3.2.6 Indicators

This group consist of various indicators that do not fall into the earlier groups. They include calculations based on price and volume at different periods. We have chosen to use indicators such as the *Highest High*, *Lowest Low*, *Median Price*, *Price & Volume Trend*, *On-balance Volume*, *Bollinger Bands*, *Percent B*, *Bandwidth*, *Accumulation Distribution Line* and *William's Accumulation Distribution Line*.

Highest High, Lowest Low and Median Price

These indicators measure the highest high and the lowest low price over a specific time period as well as the median price. In this study we choose to use two measurement periods, 4 and 6 weeks, for the highest high and the lowest low. The appearance of this can be seen for AstraZeneca in figure 3.13.

The indication of highest high and lowest low are sudden changes of the factor. This might be an indication of a changing trend. To achieve good indications of this one might use upper and lower bands that can change over time. When the indicator crosses a band, one should enter the market and when the opposite band is crossed one should exit the market and reverse the position taken at the outset. Another recommendation stipulates that one should take long positions when the previous 4-week high is penetrated. Similarly, when the previous 4-week low is crossed one should instead take short positions.



Figure 3.13: Highest high and lowest low of AstraZeneca

Highest high and lowest low for AstraZeneca. One can clearly see the indications of being overbought and oversold.

The median price on the other hand is the mid-point of the high and low for each period. This line can in many cases be accompanied by a moving average, which might give great insights regarding the stock price movements. In figure 3.14 we see the median price for AstraZeneca with a 6-period modified moving average¹ as well as their difference. When the intersection of the lines occur from above by the moving average with the median price, a buy signal is generated and when the crossing is made from below a sell signal is made. Hence, if the difference between them is negative and then turns positive, a buy signal is at hand and vice versa as can be seen in the lower plot in figure 3.14.

Price and Volume Trend

The price and volume trend is an indicator that takes the volume and closing price into account. The indicator is mainly used to establish the strength of price trends and indicate weak price movements. This is made possible since the indicator takes the percentual change of prices into account and not only that a price change occur in general as presented by Achelis (1995),

$$PVT_t = PVT_{t-1} + volume \cdot \frac{P_{c,t} - P_{c,t-1}}{P_{c,t-1}}$$

The main indication of this factor is the flow of money in and out of a security.

¹Modified moving average is defined as $MA_t = MA_{t-1} + \frac{1}{n}(P_t - (MA_{t-1}))$



Figure 3.14: Median price, Modified Moving Average and their difference for AstraZeneca

Median price and 6-period modified moving average for AstraZeneca. One can clearly see the indications for taking long and short positions.



Figure 3.15: Price and Volume Trend for AstraZeneca

Price and Volume Trend for AstraZeneca. One can clearly see the indications.

Hence, it indicates whether the security is experienced as bullish or bearish. Generally, one is can advantageously use the divergence between the indicator

and the price to conclude if the price will increase.

Volume and Price Rate of Change

The price rate of chance is calculated as the difference between the closing price today and the closing price of a previous time period N,

$$PROC = \left(\frac{P_{c,t=0} - P_{c,t=N}}{P_{c,t=N}}\right) \cdot 100$$

If the price of an asset experiences an increase, the price rate of change will rise, and vice versa. It is said that prices usually evolve in a cyclical manner and this can be captured by the price rate of change at different time periods. In



Figure 3.16: Price rate of change for AstraZeneca

Price rate of change for AstraZeneca. One can clearly see the indications for taking long and short positions.

this study we have chosen to use the price rate of change at 2 and 6 weeks as can be seen in figure 3.16. Moreover, it is possible to get good indications of price movements if the price rate of change cycles are compared with the market cycles.

The volume rate of change is an indicator that is very similar to the price rate of change. It is calculated in the same manner except that ones uses volume instead of price. It is a known fact that specific price chart formations such as breakouts, tops and bottoms can be seen in the volume rate of change chart as a sharp peak.

On-Balance Volume

By relating the volume to the price movements as a momentum indicator, we obtain the on-balance volume. The indicator is calculated by looking at the



Figure 3.17: Volume rate of change for AstraZeneca

Volume rate of change for AstraZeneca. One can clearly see the indications for taking long and short positions.

cumulative volume of a specific asset,

$$\begin{array}{lll} OBV_t = OBV_{t-1} + Volume_t & \text{if} & P_{c,t} > P_{c,t-1} \\ OBV_t = OBV_{t-1} - Volume_t & \text{if} & P_{c,t} < P_{c,t-1} \\ OBV_t = OBV_{t-1} & \text{if} & P_{c,t} = P_{c,t-1} \end{array}$$

As seen, the on-balance volume indicates if volume is flowing in or out of a particular asset depending on the price. Movements will be experienced in the on-balance volume before large price movements are seen. The reason for this is that informed investors take positions ahead of the public.

When analyzing the on-balance volume one introduces three trends: rising, falling and neutral. We are able to see all these type of trends in figure 3.18. When the trend shifts from a rising trend to a falling trend, it is said that short positions should be taken. Similarly, if the trend goes from falling to a rising, long positions should be taken. This is consistent with the assumption that the indicator precedes the price changes. However, to be able to profit from this type of analysis, all positions have to be taken at a fast pace.

Bollinger Bands, %B and Bandwidth

Bollinger bands consist of three lines superimposed on the price chart as seen in figure 3.19. The middle band of the Bollinger bands is a simple moving average of the price and in this study we have chosen to employ a 10-week simple moving average. The upper and the lower bands are constructed by the middle band



On-balance volume for AstraZeneca. One can clearly see the indications for taking long and short positions.

but shifted up respectively down with a number D standard deviations. In this study we conduct this shifting with 2 standard deviations (D=2).



Figure 3.19: Bollinger bands for AstraZeneca

Bollinger bands for AstraZeneca. One can clearly see the indications for taking long and short positions.

Bollinger bands capture the price movements between the upper and the lower bands. As one seen, the bands become wider during periods of higher price volatility and when the volatility is low the bands become tighter. The Bollinger bands have some important features. First of all, when a price penetrates a band, it is considered that the particular trend will continue. Hence, if the upper band is crossed the price will continue to increase and vice versa. Secondly, if the bands are tightening one can expect significant price movements. Moreover, when a new top or bottom is generated outside the bands and accompanied by a top or bottom inside the bands, one can expect a break in the present trend.

To be able to fully capture the information from the Bollinger bands we introduce the indicators *Bandwidth* and %B (percent B), which both derive from the Bollinger bands. The *Bandwidth* is defined as the distance between the upper and the lower Bollinger band whereas %B is defined at time t as,

$$\%B = \frac{P_{c,t} - Bollinger_{lower \ band}}{Bollinger_{upper \ band} - Bollinger_{lower \ band}} \cdot 100$$

The *Bandwidth* helps us to measure how much tighter or wider the bands become and provide us with information as previously stated. In addition, %Bhelps us define the location of the price with respect to the upper and lower band. This measure also helps us in locate where the price crosses the Bollinger bands and to what extent, i.e., a value of %B below 0 meaning a penetration of the lower Bollinger band, a value above 100 meaning a crossing of the upper band. The amount of what is below or above these limits denotes to what extent the price crosses the bands as a percentage of the *Bandwidth*.

Figure 3.20: Bandwidth and Percent B for AstraZeneca



Bandwidth and Percent B for AstraZeneca. One can clearly see the indications that have been described in the text.

Accumulation Distribution Line

The accumulation distribution line is an indicator that is based on the assumption that a price movement is more important if it is accompanied by a large volume and vice versa. For this reason, the indicator includes not only various price data but also volume data and is defined as,

$$ADL = \frac{(P_c - P_l) - (P_h - P_c)}{(P_h - P_l)} \cdot volume$$

The accumulation of an asset is signified by this indicator as an increase, whereas a distribution of the asset is indicated by a decrease. We are able to identify both of these characteristics in figure 3.21. Just as with other indicators we may also regard this indicator in relation to the price movements. For instance, if prices decline at the same time as this indicator moves up, we are very likely to see prices moving up in the near future.

Figure 3.21: Accumulation Distribution Line for AstraZeneca



 $\label{eq:complete} \mbox{Accumulation Distribution Line for AstraZeneca.}$

William's Accumulation Distribution Line

To be able to express this indicator let us first define the *true range low* (TRL) as the lowest among today's low and yesterday's close. Further, let us define *true range high* (TRH) as the highest among today's high and yesterday's close. At this point, define today's accumulation distribution line as, At this point, we

$$\begin{aligned} AD_t &= P_{c,t} - TRL & \text{if} \quad P_{c,t} > P_{c,t-1} \\ AD_t &= P_{c,t} - TRH & \text{if} \quad P_{c,t} < P_{c,t-1} \\ AD_t &= O & \text{if} \quad P_{c,t} = P_{c,t-1} \end{aligned}$$

can define the William's accumulation distribution line as,

$$WADL_t = WADL_{t-1} + AD_t$$

This indicator is mainly used as a tool for finding divergences. The approach is that one should sell a particular asset if the asset makes a new high at the same time as this indicator fail to do so. Similarly, one should buy a particular asset if it makes a new low but William's accumulation distribution line fail to do so. An example of the appearance of this indicator can be seen in figure 3.22.

Figure 3.22: William's Accumulation Distribution Line for AstraZeneca



William's Accumulation Distribution Line for AstraZeneca. One is able to spot the indications for taking long and short positions.

3.3 Construction of a Tree Classifier

Let us define a measurement space X that consists of vectors of data. Each vector **x** in this measurement space consists of various measurements of predefined factors (x_1, x_2, \ldots) . These factors must describe our object in any somehow sense and if predicting the financial markets they can for instance be as in previous sections (e.g., oscillators, stochastics, indexes and indicators).

Our main task is to classify our objects with respect to the data found in each measurement vector $\mathbf{x} \in \mathbb{X}$. Given this, we are about to define our set of various classes as $C = \{1, 2, \ldots, J\}$, i.e., we have J different classes in total. These different classes could be different species of flowers, various medical diagnoses or other relationships depending on what type of classification one intends to carry out. In this study we focus on classifying our data into three performance classes of assets (J = 3): *outperforming, neutral* or *underperforming*. To establish if an asset is outperforming, neutral or underperforming we initially create a set

of rules where the excess rate of return of the rebalanced portfolio in relation to the portfolio index, r, is evaluated against a predetermined threshold, R. The threshold itself is dependent on the risk attitude of the investor and throughout this study we set R = 1.5%. Hence, the set of rules can be stated as:

r	>	R	\Rightarrow	Outperforming
r	\leq	R	\Rightarrow	Neutral
r	<	-R	\Rightarrow	Underperforming

When creating a classification of future data, one must conduct the observations with respect to historic and already classified data (e.g., outperforming, neutral or underperforming). For this reason, a learning set L must be constructed. This set consists of N vectors that are made up by d-dimensional feature vectors \mathbf{x} , accompanied by the correct classification of each feature vector. In a more general approach this means that the learning set of the classification tree is denoted by,

$$L = \{ (\mathbf{x}_1, j_1), \dots, (\mathbf{x}_N, j_N) \}$$

where each data set $(\mathbf{x}_1, j_1), \ldots, (\mathbf{x}_N, j_N)$ is made up by the feature vector $\mathbf{x}_n \in \mathbb{X}$ and the associated class $j_n \in \{1, \ldots, J\}$ where *n* is the index of all objects $n = 1, \ldots, N$. Consequently, in our study we use a measurement matrix (n.b. for *each* time period *t* and *each* asset *m*), where *N* equals the number of cases observed, i.e., our learning period, and *d* equals the number of technical indicators and we end up with a $N \times d$ matrix of measurement data with an $1 \times N$ associated vector of classes.

The classification of the data is conducted in a binary procedure and continues in a repetitive splitting. The first split of the data results in two major descendant and disjoint subsets, X_1 and X_2 , of our measurement space \mathbb{X} , i.e., $X_1 \cup X_2 = \mathbb{X}$ and so forth. This splitting continues recursively and results in more but smaller subsets that are disjoint and more homogenous than the previous set. In a more general approach this recursive classification can be described as a partitioning of our measurement space \mathbb{X} into J disjoint subsets. If the subsets that are created are being denoted A_1, A_2, \ldots, A_j and $\bigcup_j A_j = \mathbb{X}$ then the classifier partitions the data into subsets such that j is the predicted class given $\forall \mathbf{x} \in A_j$.

However, it is difficult to achieve a tree classifier that is not *overfitted* or exhibits *overtraining*. Overfitting and overtraining occurs when the model has overadjusted itself too much to a specific learning set and thus becomes deficient with new situations and data. However, there are many ways of avoiding these challenges and as one concludes the most crucial steps of constructing a good classification tree is to decide when a split should occur and when the splitting should stop. Generally, if the splitting is stopped very early the classification will not be accurate and the classification will be very heterogenous. Similarly, if the splitting continues for a long period of time, eventually the classes will only consist of one object and the system will be overfitted. For this reason it must be stressed that a satisfactory approach to tree construction must be chosen and one clearly understands that the three most important steps in growing a classification tree is:

- 1. How to decide when and how a node should be split
- 2. How to decide when the node splitting should be terminated and a node should be declared terminal
- 3. How to decide what class the terminal node belongs to

When conducting the previous tasks the learning set is the only source of information. For this reason it is crucial not to use more data than necessary and efficient methods must be employed. The general consensus is that it is more difficult to handle the first two steps than deciding what class a terminal node belongs to. Let us for this reason introduce the methods of these steps and illustrate how to grow a tree.

3.3.1 Growing the tree

Tree classifiers can be grown in different ways but they all have one fundamental idea in common and that is that each new subset is purer and more homogenous than the previous set. To illustrate this procedure let us first of all introduce the necessary preliminary theory of Bayesian probability. First of all, define the resubstitution estimate p(j,t) as the probability that an object belonging to class j is at the same time at node t. This can be expressed more explicitly as

$$p(j,t) = \pi(j)\frac{N_j(t)}{N_j}$$

In the preceding expression $N_j(t)$ denotes the number of observations that exist in node t and associated with class j, N(t) is the total amount of cases in the learning sample where $\mathbf{x}_n \in t$ and $\pi(j)$ is the probability that class j will be presented to the classification tree also known as prior probability. Bayesian probability theory now provides us with the conditional probability stating that given that a specific case has already fallen into node t, the probability that the case also belongs to class j is given by

$$p(j|t) = \frac{p(j,t)}{p(t)}$$

where $p(t) = \sum_{j} p(j, t)$ is the marginal probability denoting the probability that any case is in node t.

When conducting the splitting of the data a set of questions Q, generated from the measurement vector, is presented initially to the root node and eventually to child nodes. The set of questions are different depending on what type of data that is used and the set may either be of categorical or ordered nature. More explicitly, introduce the vector of measurements $\mathbf{x} = (x_1, \ldots, x_M)$ and let each measurement either be categorical or ordered. If the measurement x_m is categorical and, assuming that it can take a value $\{k_1, \ldots, k_L\}$, then the set Qposes a question $x_m \in S$ where S is defined over $\{k_1, \ldots, k_L\}$. On the other hand, if the measurement x_m is ordered the set Q includes instead questions such as $x_m \leq h$ where h is defined on $(-\infty, \infty)$. It should be noted at this point that the number of questions within the set Q are finite and that the split always is dependent on one variable. Thus, at each new node, starting at the root node, questions on the preceding form are posed to each node. When the question set Q is introduced to a node, each variable in \mathbf{x} is investigated by the algorithm and the best possible split is selected, making the potential subset more homogenous. This is recursively made for all variables in \mathbf{x} and we achieve the set P denoting the best splits for each factor (x_1, \ldots, x_M) , i.e., the technical indicators in our study. At this point, each best possible split in the set P is compared with all best possible splits and the split that is the superior one and results in the most homogenous subset, is selected.

The performance of the classification tree is highly dependent on the increase in homogeneity of each new subset. For this sake, let us introduce, for any node t in the classification tree T, a measure i(t) that signifies the impurity of the node. In each split s, the data is divided into proportion p_R and p_L depending on if the data is sent to node t_R and t_L as seen in figure 3.23. The proportions are defined as $p_L = p(t_L)/p(t)$ respectively $p_R = p(t_R)/p(t)$ and $p_L + p_R = 1$.

Figure 3.23: Basic structure of splitting a node



The figure displays the splitting procedure. When a split s generates two children, a proportion p_L and p_R of the initial data is sent to the children denoted by t_L and t_R .

Thus, we will enjoy a decrease in impurity per split which is equal to the impurity of the root node subtracted by the proportion of impurity in each child as in the subsequent expression,

$$\Delta i(s,t) = i(t) - p_R i(t_R) - p_L i(t_L)$$

Regarding a complete classification tree, like the small one in figure 3.23, we will have a set of terminal nodes. Let us define this set of terminal nodes as \tilde{T} and now define the overall tree impurity by,

$$I(T) = \sum_{t \in \widetilde{T}} I(t) = \sum_{t \in \widetilde{T}} i(t)p(t)$$

Obviously, when a new split s is made, the new tree T' that is formed results in an impurity,

$$I(T') = \sum_{\tilde{T} - \{t\}} I(t) + I(t_R) + I(t_L)$$

It is now logical that minimizing the overall tree impurity is the same as maximizing the decrease in tree impurity since it is dependent on each split s and the node t,

$$\Delta I(s,t) = I(T) - I(T') = I(t) - I(t_L) - I(t_R)$$

Consequently, we may form a rule for determining when we have reached a terminal node. This is conducted by introducing a threshold $\beta > 0$ that stops the splitting when the decrease in impurity per split has turned sufficiently small,

$$\max_{s \in \mathcal{S}} \Delta I(s, t) < \beta$$

Having established a stopping criterion we must also introduce how each terminal node $t \in \widetilde{T}$ is assigned a class j(t). We conduct this by observing for which class j, that $N_j(t)$ in the terminal node is the greatest and assign this node class j(t), i.e., the maximum value of p(j|t) by means of the plurality rule. Now, given that the object is in node t, let us also assign an expression r(t) for which the object is classified incorrectly as,

$$r(t) = 1 - \max_{j} p(j|t)$$

Further, introduce a penalty of misclassifying an object that belongs to class i into class j as C(i|j) where,

$$\begin{cases} C(i|j) \ge 0, & i \ne j \\ C(i|j) = 0, & i = j \end{cases}$$

When we implement the classification algorithm in this study we set the C(i|j) = 1 if $i \neq j$ and C(i|j) = 0 if i = j. Having this stipulated, we are able to express the expected cost of incorrectly classifying an object of class j as class i when being in node t as,

$$\sum_{j} C(i|j) p(j|t)$$

This in its turn makes it possible to refine the initial class assigning rule and we choose to assign a terminal node class j(t) = i if *i* minimizes the subsequent misclassification cost. Moreover, redefine r(t) as,

$$r(t) = \min_{i} \sum_{j} C(i|j) p(j|t)$$

In the same procedure as for tree impurity for a complete classification tree, we can express the misclassification cost of the complete classification tree as,

$$R(T) = \sum_{t \in \widetilde{T}} R(t) = \sum_{t \in \widetilde{T}} r(t)p(t)$$

This form of splitting and these class assignment rules are used as preliminary basics when implementing our algorithm. However, the presented method must be altered due to some deficiencies. When using this form of R(T), the total cost of classifying incorrectly is decreased for every split and it seems like the results are improving. Moreover, it may be the case that all splits in S return a misclassification cost of zero and that for this reason there does not exist a number of best splits.² However, this is not the case and one would eventually

 $^{^{2}}$ For a proof on this matter please refer to Breiman et al. (1984)

end up with an overfitted system. For this reason, we refer to the findings of Breiman et al. (1984) where they show that it is more effective to create a large complete classification tree that overfits the data and then use various techniques of *pruning* to achieve better generalization to new data. For this reason, we initially create a very large tree and then use the pruning techniques as presented in upcoming sections. Furthermore, what we also take into account are advantageous techniques for splitting nodes and creating the complete classification tree. In some literature, it is argued that the choice among well-recognized splitting rules does not matter. However, according to Salford Systems³ the decision of splitting rule might reduce the error rate by 5-10 % depending on data set [32]. For this reason we choose to evaluate the *Gini Index* as well as *Twoing Rule* splitting rules and implement the superior one in the final algorithm.

Gini Index

The Gini index splitting rule is a very simple but effective method, based on the ideas of the previously mentioned splitting impurity function. Hence, at each node we try to find the split that makes a potential subset more homogenous than the previous set and this is conducted by choosing the split that reduces the impurity of the tree the most. However, the major difference is that this method does not exhibit the deficiencies of the preliminary impurity function.

Let us once again consider the method of assigning a node t with the class j that has the largest posterior probability. The fundamental idea of the Gini index splitting criterion is that we pick a random object from the node t and assign it class i. Consequently, the probability of this class is chosen is the posterior probability p(i|t). In addition, consider the likelihood of this case really belonging to class j, which is the posterior probability p(j|t). The Gini index, which is an impurity measure, is now defined as the sum over the intersection of these probabilities as,

$$i(t) = \sum_{j \neq i} p(j|t)p(i|t) = 1 - \sum_{j} p^2(j|t)$$

By means of implementation, the Gini index tries to conduct a split into two children where the largest or most important class of the parent is separated from the other classes and this is continued in a recursive manner. In this study we only focus on the classes that are the largest and do not consider any classes as more important than others by using customized weights or variable misclassification costs.

Twoing Rule

The Twoing rule is very different in comparison with the previous splitting rule but employs the preliminary ideas of splitting and Bayesian probability. The fundamental technique of the splitting rule is that we initially try, rather than splitting by the most important or largest class, to segment the data into two equally sized subgroups. When conducting this separation, we try to find a general common factor that can be used to separate the initial set into the two

³Market leader in choice modeling and data mining software development

subgroups. When the separation has been made, we search for the split that, as previously shown, results in a maximization of the decrease of the impurity. Having established the splitting that would result in the biggest decrease of impurity, we redefine the selection of the initial separation of the set of classes by conducting yet another maximization. At this point we focus on the initial separation of the learning set into two groups and search for the possible separation that would decrease the impurity of the previous split the most. Hence, at each node the Twoing splitting rule conducts a dual maximization. As a matter of fact, one may regard the splitting at each node as if there only existed two possible classes. By doing this⁴, we may reduce the overall complexity and we can express the impurity as,

$$i(t) = \frac{p_L p_R}{4} \left[\sum_{j} (|p(j|t_L) - p(j|t_R)|) \right]^2$$

3.3.2 Pruning the tree

Recall that it is very ineffective and difficult to create a good classification tree by terminating the splitting procedure after a certain amount of splits. Instead, as suggested by Breiman et al. (1984), it is more effective to grow a very large tree, sometimes until each terminal node is associated with only one object or that all objects belong to the same class, and then use pruning techniques to reduce the size of the tree.

By means of the previous splitting techniques, we construct the maximum tree T_{max} . When this maximum tree is created we strive to prune the tree by using the previous misclassification rates R(T) together with the *complexity* of any subtree, $|\hat{T}|$. The complexity of any tree is the number of terminal nodes in a partially ordered subtree T of the overly large tree T_{max} , i.e., $T \preccurlyeq T_{max}$. By definition, given a *complexity cost* of each terminal node as $\alpha \ge 0$, the cost complexity measure is,

$$R_{\alpha}(T) = R(T) + \alpha |\hat{T}|$$

Consider the alternative that we have a small tree, then obviously we achieve a large R(T). Similarly, if we have a large tree, with only one object per terminal node and class, R(T) = 0 but $R_{\alpha}(T) \neq 0$ due to the complexity of the tree. Hence, we strive to minimize the cost complexity of the tree by searching for subtrees that can be eliminated in a nested partially ordered set of subtrees,

$$T_{max} \succ T_1 \succ T_2 \succ \ldots \succ T_R = \{root\}$$

When deciding which subtree to eliminate, it is a fairly ingenious step to consider finding the weakest connection of a subtree. To find the weakest connection, we introduce a function $g_r(t)$, where we step through all connections in the tree, as^5 ,

$$g_r(t) = \frac{R(t) - R(T_{rt})}{|\hat{T}_{rt}| - 1}$$

 $^{{}^{4}}$ For a proof on this matter please refer to Breiman et al. (1984)

 $^{{}^{5}}$ For a derivation of this expression we refer to Breiman et al. (1984)

In the expression, t denotes an internal node of one of the nested subtrees T_r and T_{rt} denotes its connection to the node t. Hence, the weakest connection in the tree T_r is the one that has the smallest value for $g_r(t)$,

$$g_r(t_r^\star) = \min_{t \in T_r} g_r(t)$$

This pruning is continued recursively until we reach the root node which means that for each new pruning level, a nested subtree is eliminated from the initial tree, $T_{r+1} = T_r - T_{t_r^*}$. Consequently, we will achieve a partially ordered set of subtrees as previously described, starting with the maximum tree and ending with the root node. Further, at each new pruning level, we update the complexity cost as $\alpha_{r+1} = g_r(t_r^*)$ which in its turn results in an increasing sequence $\{\alpha_r\}$ for $r \ge 1$ where $\alpha_1 = 0$. At this point, we introduce a theorem that, under the subsequent conditions, states that the minimal cost complexity tree within the interval $\alpha_r \le \alpha < \alpha_{r+1}$ is,

$$T(\alpha) = T(\alpha_r) = T_r \tag{3.2}$$

We have thus reached the point where we have a sequence of both subtrees and cost complexity parameters. We have also established that the minimal cost complexity tree exists within the stipulated sequence of trees. Nevertheless, we do not know which tree that accounts for the best pruning level and must therefore introduce additional techniques to establish this. Among several possible approaches, we have considered using *Cross-Validation*.

Cross-Validation

The method of cross-validation is a based on a large amount of learning and testing sets that are generated from the original learning sample L. The method is very demanding when it comes to computations; however, it produces results that can be considered more stable when using smaller amounts of data. Very often, other methods of pruning techniques would generate more accurate results on the training set due to underestimation of the misclassification error and overfitting of data. However, empirical results it is shown that smaller trees generate better results than complex ones.

The fundamental idea of cross-validation is that we use as much information from the learning sample L as possible by separating it into K subsets. In this study we use the K = 10, which is also the standard within the literature, known as 10-fold cross-validation. Every subset, which we define as L_k , will be used as a test sample in the algorithm. The rest of the available data, $L^{(k)} = L - L_k$ where $k = 1, \ldots, K$, will be used to create new trees $T_r^{(k)}$ by the previously introduced technique. Since we create sequences of trees and cost complexity parameters for both our initial learning sample L and our K generated samples, we will achieve K + 1 sequences of trees and parameters.

As the superior tree T_r , for any sequence of trees, lies within an interval of complexity costs, as seen in equation 3.2, we must redefine this interval to be able to conduct estimations of the misclassification. For this reason, consider that we instead employ the geometric mean of the interval $\alpha_r^{\star} = \sqrt{\alpha_r \alpha_{r+1}}$, which in its turn makes it possible to get an estimation of the misclassification cost.

Recall that our main objective is that we want to minimize the overall misclassification cost, $R_{\alpha}(T)$, of the tree. To do this, we will use the sequence of generated trees to achieve the trees with same complexity as our original tree. Now consider equation 3.2 again. If we use this theorem we can achieve the estimate of the cost complexity measure for each subtree T_r as,

$$\hat{R}(T_r) = \hat{R}(T(\alpha_r^{\star}))$$

This is explicitly conducted by finding all generated trees $T_r^{(k)}$ where α_r^* , which is associated with the tree T_r , lies within the generated interval $[\alpha_r^{(k)}, \alpha_{r+1}^{(k)}]$. When we have found all these trees we evaluate the associated k:th test set L_k in each generated tree $T_r^{(k)}$ and create a vector that has the same amount of ones as the number of incorrect classifications and zeros as correct classifications. This procedure is made for all subtrees T_r and we set $\hat{R}(T_r)$ as the mean of the vector consisting of zeros and ones. Now, the best-pruned tree is the tree that has the fewest nodes, i.e., greatest value of r and that is within one standard error of the minimum $\hat{R}(T_r)$ for all r. This rule is known as the 1 SE rule.⁶

Having established the rule of optimal pruning, we use this in our algorithm for each time period and for each asset. Generally, the pruning establishes trees that are fairly small with few terminal nodes as seen in the figure 3.24 where the graphs represent the error dependence of the number of terminal nodes.

3.4 Construction of Rebalanced Portfolio

To construct our portfolio, we start by introducing the sequence of weekly unclassified sets of features and achieve a classification based on previous patterns. When the algorithm has classified the assets at each new time period as either *outperforming*, *neutral* or *underperforming* we intend to increase the value of the portfolio by rebalancing the assets based on the classification. As the classification algorithm is implemented with a training period of 20 weeks and the technical indicators use a certain number of weeks of historical data, the first trading positions can be taken after 6 months of recorded data.

When conducting the rebalancing of the portfolio, we decide to only take long positions of the assets that are classified as outperforming. However, we restrict ourselves and do not allow taking short positions in underperforming assets nor keeping positions in neutral assets. Hence, at each new time period all positions are changed with respect to the classification of the new time period. When conducting this we include transaction costs and when buying or selling assets 15 basis points⁷ are deducted from the value of each transaction.

Due to regulations regarding investment funds, which is stipulated by Swedish law⁸, the percentage invested per asset is limited due to risk reduction. However,

 $^{^{6}}$ For a more explicit discussion on this topic we refer to Breiman et al. (1984)

⁷One basis point equals 1/100th of one percent

⁸For a comprehensive discussion please see 'Lag (2004:46) om investeringsfonder'.



Figure 3.24: 10-Fold Cross Validation Calibration of Classification Trees

The figure displays the error for each tree generated for AstraZeneca. As seen in the upper plot the best sized tree is the one that has the fewest terminal nodes but within one standard error of the tree with the lowest cost.

we use a simplification of the law and decide only to invest a maximum of 20 % of the total fund value in an asset that is classified as outperforming. If not enough assets are classified as outperforming, the rest of the fund's value will be invested at the risk free rate, here defined as the mean weekly Stockholm Inter Bank Offered Rate (STIBOR) as seen in figure 3.25.

3.5 Sharpe Ratio

When investing in funds it is common to use risk-adjusted returns. This gives an insight of the risk taken in comparison with the return of a portfolio. To be able to measure this one usually employs the *Sharpe ratio* which is defined as,

$$S = \frac{R - R_f}{\sigma}$$

In the subsequent expression, R denotes the realized return of the portfolio, R_f is the risk-free rate of return and σ is the standard deviation of the difference $R - R_f$. When measuring the Sharpe ratio for our rebalanced and index portfolio we use the mean STIBOR as the risk-free rate of return.

The Sharpe ratio is often used to compare two portfolios to each other. The portfolio that exhibits the highest Sharpe ratio is considered as the superior



Figure 3.25: Accumulation Distribution Line for AstraZeneca

The mean weekly STIBOR during the period 010103 - 010108.

portfolio, since it generates higher returns with respect to risk. Generally, it is considered that the risk taken within an investment should be proportional to the Sharpe ratio. This means that if we have one portfolio that has a Sharpe ratio that is half that of another portfolio, the half of the risk should be taken when investing in the first portfolio. Moreover, a portfolio that has a Sharpe ratio that is positive should be bought, whereas a portfolio that exhibits a negative Sharpe ratio should be held short. Similarly, a portfolio that has a Sharpe ratio of zero should not be considered using since one may instead use the risk-free rate of return [33].

Chapter 4

Results

In this chapter we discuss the results achieved when running the model. Further, we will discuss the impact of behavioural finance on the trading system and the stock market since it seems as if the efficient market hypothesis does not completely hold and this may affect the sector of applicability and validity of the system

4.1 Performance of model

The performance of the classification is very crucial for the overall performance of the trading system. What is most important is the rate of misclassification and especially the number of cases where underperforming assets are classified as outperforming. The reason for this is that these classifications will influence the performance the most since they account for large negative returns. Similarly, we wish to achieve a low proportion of assets that are neutral but classified as outperforming. However, since they only account for small negative returns as well as small positive returns, this type of misclassification is not as important as the previous one. Nevertheless, we wish to achieve an overall high rate of correct classifications as well as of assets that are outperforming and classified as outperforming. When our classification algorithm is executed we achieve the results as seen in table 4.1. When running the algorithm using the Gini index and the Twoing rule, we achieve better results when using the Twoing rule and therefore we use this splitting technique for the empirical results.

As seen, we get a total classification that is slightly better than 50 %. Moreover, the classification of an outperforming asset as outperforming is significantly better than just classifying an asset randomly with a probability of correct classification of $33\frac{1}{3}$ %. Similarly, assets that are neutral are classified as outperforming at quite a high rate but still lower than a random classification. Further, what might be most important is the rate of assets that are underperforming and classified as outperforming. This rate is very low and for this reason we get a low degree of negative returns.

Table 4.1: Hit ratio of classification: total, outperforming/outperforming (O/O), neutral/outperforming (N/O) and underperforming/outperforming (U/O)

Asset	Type:	[%]:	Type:	[%]:	Type:	[%]:	Type:	[%]:
ABB	Total	53.8	0/0	50.0	N/O	44.8	U/O	7.4
AstraZeneca	Total	53.8	0/0	37.5	N/O	21.8	U/O	5.7
AtlasCopco	Total	50.4	O/O	35.8	N/O	20.7	U/O	3.9
Boliden	Total	61.9	0/0	48.2	N/O	33.5	U/O	7.6
Electrolux	Total	42.0	0/0	42.8	N/O	27.5	U/O	8.9
Nordea	Total	53.4	0/0	38.4	N/O	23.0	U/O	9.5
Sandvik	Total	47.5	0/0	37.0	N/O	22.1	U/O	9.4
SKA	Total	48.3	0/0	36.9	N/O	22.2	U/O	9.0
SHB	Total	59.9	0/0	34.0	N/O	19.1	U/O	8.7
Volvo	Total	49.9	O/O	32.4	N/O	18.3	U/O	8.2
Portfolio	Mean	52.0	Mean	39.3	Mean	25.3	Mean	7.8

The performance of the classification results in that outperforming assets are chosen at a somewhat slow pace during the rebalancing and it seems like some of the assets that are close to being neutral are being misclassified at a higher rate. However, since few underperforming assets are being included in the portfolio, the amount of positive returns of the rebalanced portfolio is quite stable. This can be seen in figure 4.1.

As one sees, the amount of positive returns is much larger of the rebalanced portfolio than of the index portfolio. One might say that, by using classification trees, one is able to avoid large negative returns. When examining the lower plot in figure 4.1, we see that returns are both negative and positive with approximately the same distribution. If we conduct a histogram over the index portfolio returns and a QQ-plot we see that they are more or less normally distributed with a slightly fat lower tail as seen in figure 4.2 and figure 4.3.



Figure 4.1: Returns of rebalanced portfolio and index portfolio

A bar plot of the returns of the rebalanced portfolio and the index portfolio. One can clearly see that most returns of the rebalanced portfolio are positive.

As we have shown, the returns of the rebalanced portfolio will not show a normally distributed behaviour that fluctuates around zero. Rather, it will be centred at a positive value with a slightly positive skewness as seen in figure 4.4. A QQ-plot, as seen in figure 4.5, reveals that the lower tail is slightly thinner than a normal distribution and that the upper tail is slightly fatter.

As seen in the figures, the returns of the rebalanced portfolio exhibit a slight positive skewness. This holds since the mass of the distribution is to the left and that the upper tail is much fatter than the lower tail. If we measure the skewness and the kurtosis of the index and the rebalanced portfolio we get the results that can be seen in table 4.2.

Table 4.2: Skewness and Kurtosis

Portfolio:	Skewness:	Kurtosis:
Index	-0.69	5.56
Rebalanced	2.74	11.85

Hence, it is obvious that the value of the rebalanced portfolio is performing better than the index portfolio. If we regard the evolution of the value of both portfolios over time, as seen in figure 4.6, we see that the rebalanced portfolio has an annually return of 30-50 % whereas the index portfolio has an annually return of 0-10 %.

Figure 4.2: Histogram of the index portfolio



Histogram of the index portfolio returns. As seen most returns fluctuate around zero and the tails are a bit heavier than the fitted normal distribution.





When we regard the complete time period as in figure 4.7, we see that the re-

Figure 4.4: Histogram of the rebalanced portfolio



Histogram of the rebalanced portfolio returns. We see that most observations are zero and a larger amount of positive returns than negative.

Figure 4.5: QQ-plot of the rebalanced portfolio



QQ-plot of the rebalanced portfolio returns.

balanced portfolio has reached a value of approximately 700 SEK whereas the index portfolio has only reached a value of approximately 150 SEK.



Figure 4.6: Value of rebalanced portfolio and index portfolio

By appearance, it seems like that the rebalanced portfolio follows the index portfolio's movements, but with a greater efficiency. This is not very surprising since the rebalanced portfolio consists of the same assets as in the index portfolio and great portfolio returns will be included in both the index portfolio as within the rebalanced portfolio at the same time. However, since the index portfolio consists of the complete set of assets and is constant during the whole trading period, significant positive returns from some assets will be erased or smoothened by significant negative returns from other assets. For this reason the index portfolio will not be able to experience the same increase in value as the rebalanced portfolio, which is allowed to experience great returns from individual assets.

In table 4.3 we see the performance of the rebalanced and the index portfolio per month, measured from the first trading week. As seen, the rebalanced portfolio has gained an excess rate of return of more than 400 % if compared with the index portfolio. This return might seem immense, however, the excess returns might at the same time be dramatically reduced if transaction costs are increased.

The figure represents a sequencing of index portfolio and the rebalanced portfolio. As seen the rebalanced portfolio experiences a great increase in value in comparison with the index portfolio.

Figure 4.7: Value of rebalanced portfolio and index portfolio





What also must be considered is that the overall market sentiment over the period has been positive and this is something that significantly effects the performance of the rebalanced portfolio. We are only able to conclude that the algorithm works well in a positive market since it is fairly easy to classify outperforming stocks. On the other hand, just as the rebalanced portfolio might experience great returns, it may also be a subject of great losses. For instance, as the algorithm only learns on historical data sudden great changes in the market will not be acknowledged by the algorithm immediately and only after it has been present for a long period of time. As an example of this, we might consider the large negative return during spring of 2006 where the rebalanced portfolio experiences a great decrease in value. Similarly, we see that during the subsequent trading period of the great return (i.e. 2005-2007) the value of the rebalanced portfolio is very volatile. Not surprisingly, the index portfolio is a bit more volatile as well and the general market sentiment seems eventually to be going in to a negative trend.

Year	Type	Jan	Feb	Mar	Apr	May	Jun
2003	Index:	-	-	-	-	-	100.00
	Rebalanced:	-	-	-	-	-	100.00
	Performance:	-	-	-	-	-	0.00
2004	Index:	115.12	117.27	114.38	112.82	117.71	113.91
	Rebalanced:	161.19	163.96	169.47	169.48	171.48	175.34
	Performance:	40.01	39.82	48.16	50.22	45.68	53.93
2005	Index:	108.25	110.50	117.65	119.35	119.92	122.94
	Rebalanced:	229.13	236.13	251.38	254.54	251.19	269.11
	Performance:	111.66	113.68	113.67	113.27	109.47	118.89
2006	Index:	159.57	160.81	170.58	179.14	189.83	148.10
	Rebalanced:	371.06	423.65	443.03	466.93	501.35	513.33
	Performance:	132.53	163.45	159.72	160.65	164.11	246.61
2007	Index:	178.84	180.17	191.87	188.00	179.81	179.96
	Rebalanced:	627.11	673.54	670.53	681.05	681.71	699.35
	Performance:	250.65	273.83	249.47	262.26	279.13	288.61
Year	Type	Jul	Aug	Sep	Oct	Nov	Dec
2003	Index:	102.22	107.78	112.45	109.84	115.45	112.49
	Rebalanced:	102.81	126.76	138.08	144.18	154.41	154.79
	Performance:	0.58	17.61	22.79	31.27	33.76	37.61
2004	Index:	108.83	107.09	112.82	109.67	112.25	109.75
	Rebalanced:	177.28	185.85	196.07	205.77	214.75	222.03
	Performance:	62.89	73.55	73.79	87.63	91.31	102.31
2005	Index:	128.51	135.13	136.45	141.74	145.72	153.17
	Rebalanced:	279.38	283.34	292.92	307.06	336.60	355.88
	Performance:	117.40	109.69	114.67	116.64	130.99	132.35
2006	Index:	153.30	158.28	166.91	172.81	175.78	167.21
	Rebalanced:	523.52	561.03	569.54	602.04	616.00	624.73
	Performance:	241.49	254.44	241.23	248.39	250.43	273.62
2007	Index:	182.85	161.41	161.84	172.35	156.93	150.57
	Rebalanced:	696.71	681.48	699.68	747.81	728.31	764.05
	Performance:	281.03	322.22	332.31	333.88	364.11	407.44
Total	Index:						146.50
	Rebalanced:						751.09
	Performance:						412.69

Table 4.3: Technical indicators that are used in the study and the specific group that each indicator belongs to.

In order to anticipate the risk-adjusted returns of the portfolios we analyze the Sharpe ratio and the difference in the ratio over a rolling years period as introduced earlier.

From figure 4.8 we can see that the Sharpe ratio of the rebalanced portfolio varies somewhat over time between approximately 0.5 and 1. On the other hand, the Sharpe ratio of the index portfolio is significantly lower and varies over time between approximately 0 and 0.5 and sometimes even negative. This means







Figure 4.9: Bar plot of the excess Sharpe ratio

The bar plot shows the difference in Sharpe ratio between the rebalanced and the index portfolio.

that the index portfolio can at some points be compared with the performance of the risk free rate of return. This is not the case for the rebalanced portfolio and it performs better. Moreover, it is interesting that the large negative returns during spring 2006 is acknowledged by the Sharpe ratio, just as the more volatile period during the later phases of the trading period.

4.2 Behavioural Finance

One could argue that irrational investors are the main reason why patterns are present within price data and why technical analysis might actually work. If investors were rational and based their decisions on fundamental analysis, anomalies would not be present and the stock market would be very different than today. If investigating the most important split factors of the algorithm we get a pie diagram as in figure 4.10 and figure 4.11. The abbreviation that are used in the figures are summarized in the table 4.4.

Name of technical indicator:	Abbreviation:
Momentum 2 Weeks	PMOM2W
Momentum 4 Weeks	PMOM4W
Fast Stochastics %K	STOCHOSCK
Fast Stochastics %D	STOCHOSCD
Highest High 4 Weeks	HHIGH4
Highest High 8 Weeks	HHIGH8
Lowest Low 4 Weeks	LLOW4
Lowest Low 8 Weeks	LLOW8
Price and Volume Trend	PVT
Accumulation Distribution Line	ADLN
Relative Strength Index	RSI
Volume Rate of Change 2 Weeks	VROC2
Volume Rate of Change 4 Weeks	VROC4
Price Rate of Change 2 Weeks	PROC2
Price Rate of Change 6 Weeks	PROC6
On-Balance Volume	OBV
Chaikin Volatility	CHVOL
Accumulation Distribution Oscillator	ADO
Median Price	MPRC
Difference Median Price and Modified MA	MPRCDIFF
Negative Volume Index	NVI
Positive Volume Index	PVI
Slow Stochastics %K	SPCTK
Slow Stochastics %D	SPCTD
Acceleration Between Times	ACC
PercentB	PERCENTB
Bandwidth	BANDWIDTH
Volatility	VOLATILITY
William's Accumulation Distribution Line	WADL
William's %R	WPCTR

Table 4.4: Technical indicators that are employed in the classification tree and their abbreviation.

As seen in the pie diagrams, factors such as Price Rate of Change (PROC), Price Momentum (PMOM) and Volume Rate of Change (VROC) with different time spans

Figure 4.10: Pie chart of the split factors



Pie chart of the most important split factors at the root node.









Pie chart illustrating the most important split factors in the first level of children.

are very common splitting factors. This can be regarded as being related to some of the theories within behavioural finance. The anomalies that might be the most important for this study are *herd behaviour* and *momentum investing*. The herd behaviour suggests that investors often move in sync without any predetermined strategy. This could thus affect different momentum and rate of change indicators since, if the price changes immensely, it is likely that other investors will also take positions in the asset. Moreover, the theory of momentum investing is very likely to be related to the previous herd behaviour. This theory suggests that assets that have had a positive returns for a longer period of time are likely to experience even higher returns in the nearby future which supports the findings of momentum indicators among the most important splitting factors.

It can also be believed that the algorithm may encounter problems when using other sets of data within different price ranges. According to a study conducted by Black (1986) investors choose shares based on an irrational behaviour. It is shown that many investors prefer low priced stock since they believe that high priced stocks can decrease more in value during a drop. Similarly, Brennan and Hughes (1991) conclude that a low priced stocks are negatively correlated to the amount of analyst following the company, a result that is also acknowledged by Bhushan (1989). This may be an explanation of why low-priced assets have irrational and less informed investors and might therefore exhibit complex or non-existing patterns for technical analysis. Since splitting rules must be encountered by the algorithm several times before they become important splitting factors, lower priced stocks might therefore not be suitable for algorithmic trading systems based on classification trees.

Conclusion

This chapter concludes the findings of this study. We discuss possible modifications of the study and recommendations for further research and we also bring up the topic of behavioural finance.

By implementing an algorithmic trading system based on classification trees we are able to conclude that the method is very capable and that the results are promising. By using technical analysis we are able to find patterns that, in some cases, may forecast stock price movements. Since we only use a one-week forecasting period, based on the previous 20 weeks, the rates of change of both price and volume are significant. We also find that the trading system can generate significant positive returns even with trading costs included and our modified regulations regarding the maximum weight allowed per asset in the portfolio taken into account. When running our model we achieve 412% in excess return when compared to our initial index portfolio. Further, our rebalanced portfolio also show a greater Sharpe ratio over time. On the other hand, it is also noted that, if the market changes a lot at a fast pace, the algorithm will most likely not perform as well as when the market is experiencing a constant increase. This is a consequence of that the patterns must have been experienced before and related to a particular movement of the stock. Our results are found during a fairly constant increase of stock prices and we have not examined if the algorithm will perform well in other market conditions. For this reason, in future research it may be relevant to analyze the overall trend of the market coupled with the financial cycles. Moreover, we have also assumed that the financial market exhibits perfect liquidity of stocks, meaning that there always exists a buyer and a seller. In addition, we have also assumed that all trades occur instantaneously without any price movement. In a real-life implementation this would not be the case and this might severely affect the performance of the trading system.

One might also wonder how well the algorithm would be able to capture patterns in future data and in sets of data within other price ranges. The influence of behavioural finance and anomalies in the financial market are obvious and must be accounted for if using the trading system with other data sets. Even though the algorithm is used with pruning techniques that reduce the effect of being in-sample, the system has not yet been implemented in a real-life, out-ofsample environment. Most likely, the performance of the trading system will be reduced, however this is yet to be tested.

Future research could include other types of data, both categorical and ordered as well as fundamental data. It would be interesting to examine the performance of other pruning and splitting techniques as well as bagging or boosting methods such as Random forests. Finally, hybrid models of classification and regression trees, together with other pattern recognition models such as artificial neural networks or genetic algorithms may improve the results significantly.

We also conclude that the impact of behavioural finance on the relationship between stock price and future price movements is important and that it is definitely a topic that needs further research.

Bibliography

- Achelis, S. B., 2000. Technical Analysis From A To Z. 2nd Printing, McGraw-Hill, ISBN 0071363483/9780071363488.
- [2] Bessembinder, H., Chan, K., 1995. The profitability of technical trading rules in the Asian stock markets. Pacific-Basin Finance Journal, 3, pp. 257-284.
- Bauer, R. J., Dahlquist, J. R., 1999. Technical market indicators: analysis & performance. Volume 64. Wiley Trading, John Wiley and Sons, ISBN 0471197211/9780471197218.
- [4] Benington, G. A., Jensen, M. C., 1970. Random Walks and Technical Theories: Some Additional Evidence. Journal of Finance, Vol. 25, No. 2, pp. 469-482.
- [5] Bhushan, R., 1989. Firm characteristics and analyst following. Journal of Accounting and Economics 11, pp. 255-274.
- [6] Black, F., 1986. Noise. The Journal of Finance 41, pp. 529-543.
- [7] Brandt, M. W., 1999. Estimating portfolio and consumption choice: A conditional Euler equations approach. Journal of Finance 54, 1609-1645.
- [8] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J., 1984. Classification And Regression Trees. Taylor & Francis Group, LLC. ISBN 978-0-412-04841-8.
- [9] Brennan, M.J., Hughes, P.J., 1991. Stock Prices and the Supply of Information. The Journal of Finance 46, pp. 1665-1691.
- [10] Brock, W., Lakonishok, J., LeBaron, B., 1992. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. Journal of Finance, Vol. 47, No. 5., pp. 1731-1764.
- [11] Chavarnakul, T., Enke, D., 2009. A hybrid stock trading system for intelligent technical analysis-based equivolume charting. Neurocomputing 72, pp. 3517-3528.
- [12] Chen, A.-S., Leung, M. T., Daouk, H., 2003. Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. Computers and Operations Research, Volume 30, Number 6, pp. 901-923.

- [13] DecisionPoint. http://www.decisionpoint.com, accessed on 15 October 2009.
- [14] Fama, E. F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, v. 25 (2), pp. 383-417.
- [15] Fama, E. F., Blume, M. E., 1966. Filter Rules and Stock Market Trading. Journal of Business 39: pp 226-41.
- [16] Fama, E. F., French, K., 1992. The Cross-Section of Expected Stock Returns. Journal of Finance, 47, pp. 427-465.
- [17] Ferson, W. E., Harvey, C. R., 1993. The Risk and Predictability of International Equity Returns. Review of Financial Studies, 6, pp. 527-566.
- [18] Fusai, G., Roncoroni, A., 2008. Implementing Models in Quantitative Finance: Methods and Cases. Springer Finance. ISBN 978-3-540-22348-1.
- [19] Hansen, L. P., Singleton, K., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50, 1269-1286.
- [20] Hill, T., O'Connor, M., Remus, W., 1996. Neural network models for time series forecasts. Management Science, Vol 42, pp. 1082-1092.
- [21] Kaufman, P. J., 1987. The New Commodity Trading Systems and Methods New York: John Wiley & Sons.
- [22] Keim, D., Stambaugh, R., 1986. Predicting Returns in the Stock and Bond Markets. Journal of Financial Economics, 17, 357-390.
- [23] Kumar, M., Thenmozhi, M., 2006. Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest. Indian Institute of Capital Markets 9th Capital Markets Conference Paper.
- [24] Kwon, K.-Y., Kish, R. J., 2002. Technical trading strategies and return predictability: NYSE. Applied Financial Economics, Volume 12, Number 9 / September 01, pp 639 - 653.
- [25] Lo, A. W., Mamaysky, H. M., Wang, J., 2000. Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. Journal of Finance, 55, pp. 1705-1770.
- [26] Lo, A. W., MacKinlay, A. C., 1988. Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test. The Review of Financial Studies, Vol. 1, No. 1, pp. 41-66.
- [27] Neftci, S. N., 1991. Naive Trading Rules in Financial Markets and Wiener-Kolmogorov Prediction Theory: A Study of "Technical Analysis. The Journal of Business, Vol. 64, No. 4, pp. 549-571.
- [28] Pruitt, S. W., White, R. E., 1988. The CRISMA trading system: who says technical analysis can't beat the market?. Journal of Portfolio Management, pp. 55-58.

- [29] Ratner, M., Leal, R. P. C., 1999. Test of technical trading strategies in the emerging equity markets of Latin America and Asia. Journal of Banking and Finance, 23, pp. 1887-1905.
- [30] Refenes, A. N., Zapranis, A. S., and Francis, G., 1994. Stock Performance Modeling Using Neural Networks: Comparative Study With Regressive Models, Neural Networks, Vol 7(2), 375-388.
- [31] Rosenberg, B., Reid, K., Lanstein, R., 1985. Persuasive evidence of market inefficiency. Journal of Portfolio Management 11, 9-17.
- [32] Salford Systems, Resources, White Papers. Do Splitting Rules Really Matter?. http://salford-systems.com/resources/whitepapers/do-splittingrules-really-matter.html, accessed on 20 October 2009.
- [33] Sharpe, W. F., 1994. The Sharpe Ratio. Stanford University, Reprint from The Journal of Portfolio Management. http://www.stanford.edu/wfsharpe/art/sr/sr.htm, accessed on 23 October 2009.
- [34] Stattman, D., 1980. Book values and stock returns. The Chicago MBA: A Journal of Selected Papers 4, pp. 25-45.
- [35] Tanaka-Yamawaki, M., Tokuoka, S., 2009. Adaptive Use of Technical Indicators for Predicting the Intra-Day Price Movements. Knowledge-Based Intelligent Information and Engineering Systems, Volume 4693, pp. 597-603.
- [36] Vasiliou, D., Eriotis, N., Papathanasiou, S., 2006. How Rewarding is Technical Analysis? Evidence from Athens Stock Exchange. Operational Research. An International Journal. Vol.6 No.2, pp. 85-102.
- [37] Wong, W.-K., Manzur, M., Chew, B.-K., 2003. How rewarding is Technical analysis? Evidence from Singapore stock market. Applied Financial Economics, Volume 13, Number 7, pp 543-551.