Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

NTNU
Norwegian University of
Science and Technology

A bridge between statistical learning and agent based modelling in stock market predictions

Master Thesis

Herman Hellenes

Saturday 6th August, 2016

Supervisors: Prof. Didier Sornette , Prof. Alex Hansen (NTNU)
Tutor: Lucas Fiévet
Chair of Entrepreneurial Risks
Department of Management, Technology, and Economics
ETH Zürich

Abstract

The research related to agent based modelling and statistical learning has increased in the recent years, as alternatives to main stream econometrics. This project combines the three for constructing and evaluating a forecasting experiment of capital markets. The model is applied on the stock indexes S&P500 and FTSE100 in a time period of 15 years. The performance of the model is measured in economic profits, where transaction costs are taken into account. The model provides two methods that each provide an annualized Sharpe ratio. The buy-and-hold strategy on the FTSE100 index is by this measure outperformed to above 1 and 2 BPS in transaction costs, respectively. On the S&P500 index, the model underperforms the buy-and-hold. The model is capable of capturing gains in different market regimes on both indexes.

Acknowledgements

It has been a pleasure to spend six months at the Chair of Entrepreneurial Risks. I sincerely thank Prof. Didier Sornette for accepting me at the group and for always inspiring the people around him. I would like to thank Prof. Alex Hansen at NTNU for being my co-supervisor. A special thanks goes to PhD candidate Lucas Fiévet, who has been invaluable from start to finish.

I wish to thank all my colleagues at the Chair of Entrepreneurial Risks, both for long lasting friendships and an extraordinary academic stimuli. There a special thanks goes to Michael Schatz, Richard Senner, Mattias Franchetto, Dr. Spencer Wheatley, Guilherme Demos, Vincent Giorgis, Ahmed Ahmed and Robert Kuert. At last I would like to express my sincere gratitude to my family for their constant support.

Contents

Contents **iii**

1 Introduction **1**

2 Forecasting Protocol **5**

3 The basis of the ABM - Statistical Learning **9**
3.1 Linking Statistical Learning and EMH 10
3.2 Mapping returns into categories 11
3.3 Decision Tree Approach 12

4 Time Series Decomposition into an ABM **15**
4.1 Decomposing into agents of a lower category 15
4.2 Overconfident Agents 17
4.2.1 Implementing overconfidence 19
4.2.2 Overconfidence calibration 20
4.3 Summary of the ABM 22

5 Search Technology **25**

6 Benchmarking **29**
6.1 Risk-Return Performance Measures 30
6.2 Transaction Costs Analysis 32
6.3 Trading quantiles 32
6.4 Prediction Accuracy 33
6.5 Specification of parameters, dataset and trading restrictions . 35

7 Results **37**
7.1 Gains in UK, losses in U.S. 37
7.2 Market Spectroscopy - multi-agents are promising 42

8 Concluding remarks **49**

A Evolution of the simulated annealing procedure **51**

Glossary **53**

Bibliography **55**

Chapter 1

Introduction

Partly due to increased computation power in recent years, methods with origin in physics and computer science have been suggested for describing financial markets. In this project, two such modelling approaches will be combined for the task of predicting the stock market. The two approaches are a specific agent based model (ABM) and a statistical learning method. Even before giving a short introduction of the two approaches, we want to emphasise that a combined model is proposed for building a bridge between the two. In fact, the ABMs will apply statistical learning for forecasting future returns. Thus the two approaches will be entangled. It is a key objective to compare the pure statistical learning approach to the entangled model. The investigations of the complete model's performance in the stock market is as well of high interest.

Statistical learning has a wide range of applications, as well as a method for forecasting capital markets. Our statistical learning method, and its relation to main stream econometrics, is to a large extent in line with [Fiévet and Sornette, 2016]. We formalize and discuss the method in Chapter 3, with the description of the specific model used by the ABMs. [Atsalakis and Valavanis, 2009] provide a literature survey of more than hundred articles that apply a subset of the statistical learning techniques for stock market forecasting. The results of [Allen and Karjalainen, 1999] report of profitable strategies by applying statistical learning, however they see the profits diminish after transaction costs are taken into account. In this project we require certain precautions in order to apply the model to capital markets. This will be elaborated in Chapter 2. As one of the papers addressing these requirements and report of promising returns, we mention [Creamer and Freund, 2010]. The scarcity of such papers can arguably be due to the inversion of the *file drawer bias*. If an article is presenting insignificant empirical findings, it is often more difficult to publish. The bias might be present also for significant findings in econometrics, where the researchers may be tempted to sell their work to a hedge fund instead of handing it in to a journal. Due to this, [Timmermann and Granger, 2004] argue that there is still a hope of successful forecasting.

As mentioned, the ABMs will take usage of statistical learning for predicting stock returns. Agent based modelling is a technique that focuses on the interactions of microscopic elements, autonomous agents, in order to investigate the collective macroscopic behaviour. This is a *bottom up* approach, in contrast to what is generally used in econometrics and economics. A general introduction to agent based modelling can be found in [Bonabeau, 2002]. An ABM can under the right conditions be a self-organizing system and thus give rise to extreme events. This property is often missing in main stream market models.

> *"ABM is a mindset more than a technology. The ABM mindset consists of describing a system from the perspective of its constituent units"* [Bonabeau, 2002]

Despite the promising properties of agent based modelling, there are also issues. [Richiardi et al., 2003], [Windrum et al., 2007] and [Sornette, 2014] highlight several of them. Many of the ABMs are constructed in order to show stylized facts like asset bubbles and power law distributed returns. However, a wide range of ABMs can produce such results. [Sornette, 2014] argues that the personal preference of the modeller is central in the construction of a model. As well, it is genuinely hard to compare the different models and this makes the approach less robust. Further, the calibration of an ABM to empiric data can be problematic. Often, empirical data is assumed to describe the collective behaviour of the agents. Since agent based modelling is a bottom up approach, it is a major challenge to *reverse engineering* the underlying behaviour of the agents to fit with the data. However there are examples of successful calibrations of ABMs in the context of stock markets. The success is measured by the ability to predict future returns out of the calibration sample. We mention that [Wiesinger et al., 2012], [Satinover and Sornette, 2012a], [Satinover and Sornette, 2012b] and [Andersen and Sornette, 2005] report of well calibrated ABMs. Without actually addressing statistical learning, the three latter papers construct ABMs that can be viewed as close to statistical learning methods. This has in particular triggered our interest of proposing the bridge between the two approaches.

As earlier stated, the reader will in Chapter 2 be introduced to the requirements necessary for applying the model to the capital market. These requirements will be essential for the project as a whole. In Chapter 3 follows the description of the statistical learning method. The ABM applying this method is later introduced in Chapter 4. In order to construct a complete model, a so-called *search technology* is presented in Chapter 5, due to the requirements set in Chapter 2. The different benchmark and analysis methods of the results are described in Chapter 6. The evaluations are discriminated into two categories: efficient market hypothesis evaluation and market spectroscopy. This discrimination will become clear in Chapter 6. The results and related discussions will follow in Chapter 7. Our model shows mixed results in the former of the two categories. However the model is able to gain economic profits in different market regimes. This is a very favourable feature. Through the market spectroscopy we observe no or even negative correlation between prediction accuracy and economic profits. The results indicate the usage of a broad set of forecasting models, due to the fact that there is no obvious logic in which of the parameter sets that lead to high economic profits. Lastly, we provide concluding remarks in Chapter 8. A glossary is provided in order to easily look up central variables and concepts.

Chapter 2

Forecasting Protocol

As promised in the introduction, certain requirements and concepts are needed in order to forecast capital markets. [Fama, 1970] introduced the efficient market hypothesis (EMH) as a cornerstone assumption in finance. The essence of the concept is that historical information is fully reflected in the present price of securities, with no additional information contained in them. A rather simple statement like this has made an enormous impact on how people view the financial markets. An implication of EMH is that returns are not forecastable. The evaluation of our forecasting model will thus be related to challenging the EMH. Per se, the hypothesis is not empirically testable. We will in this chapter establish a protocol in order to do this. According to Fama, one can use a couple of assumptions in order to incorporate a testable framework, however it is still far from a trivial exercise. One first assume that new information will immediately be incorporated in the price. One can argue that modelling of an asset price is in fact modelling of news relevant for the asset. If we denote our information set as Ω_t at time t, we can write the return in excess of expected return as $\eta_{t+1} = r_{t+1} - E[\tilde{r}_{t+1}|\Omega_t]$. We have that η_{t+1} is a fair game with respect to Ω_t if

$$E[\tilde{\eta}_{t+1}|\Omega_t] = 0, \tag{2.1}$$

where the tilde indicates that we deal with a random variable and r_t is the return from a given security at time t. Eq. 2.1 implies that it is not possible to *consistently* obtain *excess risk-weighted returns* in capital markets. This is not preventing investors to be lucky, or even repeatedly, but they can not systematically obtain excess risk-weighted returns. Fama presents two further cases of the fair game model. One is the sub-martingale model, which says that one can not consistently obtain excess risk-weighted returns with respect to the underlying security, i.e. a buy-and-hold strategy. Further assumptions lead to the usage of the more strict random walk theory. This is widely applied in the mathematics of financial derivatives. For further reading and applications we refer to [Wilmott, 2013]. There have been a massive research in evaluating the statistical properties of historical asset prices, letting Ω_t be the price sequence. By looking for excess predictability using this information set, researchers and practitioners then seeks to question the so-called weak form of EMH. An overview of statistical tests are found in [Lim and Brooks, 2011]. Highly influential discoveries of statistical anomalies were reported by for instance [Lo and MacKinlay, 1988] and [Granger and Morgenstern, 1963]. However, it is important to keep in mind that findings of for example a deviation from a random walk, is not necessarily equivalent to a violation of EMH. This is emphasised by for instance [Fama, 1970], [Fama, 1991], [Malkiel, 2003]. It is non-trivial to relate certain statistical anomalies and the possibility of achieving consistent excess risk-weighted returns.

The perhaps most striking evidence of EMH are the reported performance of mutual fund managers. The results show that the managers underperform the market on average (e.g. see [Jensen, 1968] and [Malkiel, 1995]). One could assume that the professionals, if anyone, should be able to outperform a buy-and-hold strategy of a broad market index. According to [Malkiel, 2003], one should favour economic significance, not only look into statistical anomalies. An important argument made by [Grossman and Stiglitz, 1980] is that gathering information and creating advanced strategies are costly operations. They reason that, if such investments did not pay off at all, nobody would bother to do it. In the later review [Fama, 1991], Fama agrees with Grossman and Stiglitz in that there is something to be exploited by increasing the level of sophistication of investment strategies. However, Fama argues that the related costs will on average and over time not be sufficiently high in order to prove the EMH wrong.

More precise and testable definitions of EMH have been proposed by [Jensen, 1978], [Malkiel, 1991] and [Timmermann and Granger, 2004]. Instead of specifying that the available information is reflected in the market, they focus on the possibility of obtaining consistent excess risk-weighted returns. Such definition fits well with the testing of technical trading strategies. According to [Timmermann and Granger, 2004], statistical tests are not sufficient to evaluate EMH, also because of the suppression of the importance of risk-premia and the ignorance of transaction costs, as well as possible dividends and interest rate effects. The two latter will not be relevant for our model. However, this motivate us further to carry out a trading strategy approach in order to benchmark in terms of EMH. A review of technical trading strategies is provided by [Park and Irwin, 2007]. However, it is indeed not straight forward to evaluate a technical trading strategy. This will be described in Chapter 6. We will in this project use the EMH definition of [Timmermann and Granger, 2004], which is the following:

EMH definition: A market is efficient with respect to the information set Ω_t, search technologies S_t, and forecasting models M_t, if it is impossible to make economic profits by trading on the basis of signals produced from forecasting model in M_t defined over predictor variables in the information set Ω_t and selected using a search technology in S_t.

The spirit of this definition is that in order to judge the efficiency of the market, one has to be in the shoes of an investor under realistic conditions. From the EMH definition, predicting prices better than random is not challenging the EMH as long as it is not possible to profit from. We will use this as a key argument for the evaluation of our model. An often missed requirement is the search technology specification, which we reveal in Chapter 5. This is crucial for testing strategies, changing the view from ex post to ex ante. A search technology S_t refers to a predecided way to decide which of the forecasting models we at each timestep will apply. Thus, a search technology is identifying the forecasting models M_t ex ante. One can still investigate how all the forecasting models would have performed ex post, where obviously some models will perform very well, on pure luck. But the ex post analysis of the M_t performance can tell something about the behaviour of the model with respect to its parameters. However, we again emphasize that a forecasting model is not valuable for a market participant if it was only shown to be profitable ex post. We have now made a distinction between the ex post and ex ante view, where both the approaches will be used, but for different purposes.

A common method for testing strategies is the *backtest*. This refers to testing strategies by using existing historic data of the underlying asset. Backtesting will be applied in order to evaluate the model, where the search technology is included. We will in Chapter 3 specify the information set Ω_t and through Chapter 3 and 4 define the forecasting models M_t. The final collection of models are summarized in Chapter 4.3.

Chapter 3

The basis of the ABM - Statistical Learning

In this chapter we introduce the basis of our forecasting models M_t, according to the EMH definition in Chapter 2. All the agents are applying statistical learning to predict future returns. By computer scientists the term machine learning is commonly used. To formalise the concept, this is an generic approach of forecasting an outcome based on an observation set, where the formal problem can be formulated as

$$Y = f(X) + \epsilon, \tag{3.1}$$

where the aim is to find a good estimator $\hat{f}$ for the function f, given observed output Y and input X. With the hypothesis that Y and X have a relation, we can predict the future output through the estimate $\hat{Y} = \hat{f}(X)$. By working with empiric data, there will always be some degree of fluctuating outputs, even with the same inputs. This is due to stochastic effects or even as a result of imprecise measurement methods. Therefore, an error term is needed, denoted as ϵ. This term is incorporating the error arising from such sources. The error term has a zero mean and is independent of X.

Let n be the number of observations in the *training* data. The training data is used for calibrating the learning method. We have for each observation i, one output y_i and d different inputs, so $x_i = (x_{i1}, x_{i2}, .., x_{id})^T$. With those definitions, we can write the training set as $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$. Through the calibration, we will obtain $\hat{f}(x_i) \approx y_i$. We use the principle of Eq. 3.1 to do so, fitting $\hat{f}$ to our training set. However, in order to obtain a successful estimation of f, the model should predict $\hat{f}(x_0) \approx y_0$ when x_0 and y_0 are *not* contained in the training data, but are new arriving data. We will refer to such data as *out-of-sample* data.

There exists a broad range of statistical learning methods, from linear regression to support vector machines to neural networks. The learning methods can often be labelled as either supervised or unsupervised. In an unsupervised method, there is a lack of observables of y_i. In this project, a supervised learning model will be applied, due to having access of observing y_i through market quotes. A central issue with any supervised learning model is the bias-variance trade-off. We can formulate the trade-off through considering the expected squared prediction error, which can be written as

$$\begin{aligned} Err(x) &= E[(Y - \hat{f}(x))^2] \\ &= (E[\hat{f}(x)] - f(x))^2 + E[(\hat{f}(x) - E[\hat{f}(x)])^2] + \sigma_\epsilon^2 \,, \end{aligned} \tag{3.2}$$

where x is not a part of the training set used to estimate f. The variable σ_ϵ is the variance of the error term ϵ. In Eq. 3.2 we can identify the first term as the bias squared, the second the variance, while σ_ϵ^2 remains as an irreducible error term. We can see if our model under- or overfit the training data by looking at the performance of the in-sample and the out-of-sample. The out-of-sample error is high both if the model is under- or overfitted. To discriminate the two, we can conclude with an overfit if the in-sample error is low. If it is high, we have underfitting. Low bias is typically obtained through doing a linear regression on linear data. Methods like decision trees and support vector machines are more flexible, thus increase the possibility of overfitting and we get a high variance. With high variance we mean that $\hat{f}$ would change much if our training data is substituted with another another set measured under the same circumstances. Most often it is not possible to obtain both low variance and low bias for an experiment, so that minimizing the error in Eq. 3.2 is not about tuning both to zero. This is therefore known as the variance-bias trade-off and emerge in numerous other optimization problems. The trade-off is addressed in the next subchapter, when we determine our training data and information set.

3.1 Linking Statistical Learning and EMH

To recapitulate, statistical learning is in essence a method of estimating future events based on previous events. There are in principle no underlying assumptions of the process producing Y. That makes it on the one side look weak, having no underlying theory of how the market works. On the other hand, this is a great advantage. Wrong assumptions of the market behaviour can lead into fatal miscalculations. Thus the errors made by a statistical learning approach are in principle not due to limited beliefs of how the market works. Rather they are consequences of the EMH.

For setting the EMH in context of statistical learning, the definition of the EMH in Chapter 2 is useful. The weak form of the EMH is often restricted to let the information set represent past returns only, having $\Omega_t = \{r_1, \ldots, r_t\}$. We have the freedom to use all past return, or a selection. The bias-variance trade-off implies that our results would be highly biased when using a large feature space[1]. When using all past stock returns, the feature space is enormous and even growing day by day. This motivates us to limit the number of past returns in Ω_t. We let the training set be a rolling window of a constant length. The number of returns used are equal to l_{is}, the *in-sample length*. For each return r_i predicted, let a lag of m previous returns to be the input. We denote m as the *memory length*. This specification will from now on be used as the number of inputs d introduced previously. The dataset that will be used for predicting future returns can then be written on the following form.

$$\mathbb{D}_t := \{(\hat{\Omega}_{t-l_{is}}, r_{t-l_{is}+1}), \ldots, (\hat{\Omega}_{t-1}, r_t)\} \text{ where } \hat{\Omega}_t = \{r_{t-m+1}, \ldots, r_t\} \quad (3.3)$$

The information set Ω_t mentioned in the EMH definition is now defined as all the different returns contained in $\mathbb{D}_t$, meaning $\Omega_t = \{r_{t-m+1-l_{is}}, \ldots, r_t\}$. Before we go further into the details of the statistical learning setup, it is worth mentioning which actions we allow our models to carry out in the market. In this project, we only consider going long through ordinary buying, or going short through one day short selling. The result of a prediction will thus lead to switching between long and short position, or staying in the current.

3.2 Mapping returns into categories

If not in a strict mathematical sense, one can view stock returns as continuous. They are quantitative, which allows for predicting quantitative outputs with a statistical learning method. This is generally called *regression*. When the dataset is stochastic and noisy, the resulting feature space can be very large. However, a trading strategy can still be successful without predicting the exact returns. For example, the ability to predict only up- and downmoves in the stock price can be very profitable. We choose to use discrete outputs for our model. Predicting such outputs are referred to as a *classification* method. As well, we suggest using discrete inputs through a pre-processing that is mapping the stock returns into discrete data. When simplifying the price time series into discrete returns, the information incorporated in the time series are strongly compressed. For example, binary returns are obtained from collapsing the real returns into up and down movements, ignoring magnitude. We compress a real return into one of v

[1]The feature space refer to the space of the collection of all input variables.

discrete categories, namely the mapping $\mathbb{R} \rightarrow \mathbb{N}_v = \mathbb{C}$. With the memory length m and $|\mathbb{C}| = v$ different categories of the returns, $\hat{\Omega}_t$ can be constructed in v^m unique ways.

3.3 Decision Tree Approach

We will apply the *decision tree* approach in this project, a well known approach in the field of statistical learning. This is a non-parametric supervised learning method, meaning that no explicit assumptions of the functional form of f in Eq. 3.1 are necessary. Let the input be an ordered historical sequence of m discrete returns, such that the feature space is of size v^m. This is a classification problem due to predicting outputs that take on a finite set of values, recalling that $Y, X_j \in \{1, 2, \ldots, v\}$. The decision tree approach divides the feature space into v^m hyper-rectangular disjoint sets, $B_1, B_2, \ldots, B_{v^m}$ or lower, depending if one introduce a stopping criteria due to possible overfit. In each set B_j, a probability distribution for the output Y is estimated from the training data. The *decision boundaries* are always parallel to the axes in the feature space, separating the feature space by rectangular boxes.

A decision tree is designed as a structure similar to a flowchart. The tree is build up of connected nodes that in our case are representing either a condition or a value. Each node representing a condition are referred to as *branch nodes*. The branch nodes have a set of subordinated *child nodes*. Depending on the outcome of the condition at a branch node, the decision process moves to the respective child node that corresponds to this condition. The nodes that contain a value have no child nodes. Such nodes are denoted as *leaf nodes*. The tree is build by selecting a *root* node where we from there make recursive binary splitting. The inputs are discriminated into to the disjoint sets B_j, in the end we will have v^m leaf nodes where the respective observation frequencies will be stored. Based on the observation frequency, we have the empiric statistical probability of an observation to belong to a specific leaf. To recapitulate, the observation frequency is simply the average of all inputs from training data which end up in the set B_j. Predictions are then made using the probabilities. The specific algorithm used in this project is an optimised classification and regression tree algorithm (denoted as CART for short) based on [Breiman et al., 1984]. The main tool for implementing the decision trees are to be found in [Pedregosa et al., 2011].

The decision tree approach has typically a high variance and a low bias. One can make the tree finer and finer grained in the calibration of the training data, having an unique path for each configuration of the elements. Recall from the bias-variance trade-off discussion, that when the feature space grows, the chance of over-fitting increases. An approach often taken in statistical learning is to look at the performance of the prediction in the training data versus the out-of-sample data, in order to investigate if either a underfitting or overfitting is present. The length of the training and out-of-sample data can then be varied to obtain a optimal value. This can tell what quality of prediction the model obtains. However, this technique is functioning well if the samples are statistically independent. The samples used for stock predictions are strict sequential in time. As well, the returns in capital markets are not strictly stationary, which makes this approach even less suited. We thus use fixed and limited in-sample length l_{is} and out-of-sample length l_{os}, where the former is varied across simulations. It could be interesting to also vary out-of-sample length l_{os}. That could be done by recursive predictions, adding the predictions to the training data one by one. We will however in this project have the restriction $l_{os} = 1$.

A broader overview of decision trees and their alternatives are presented by [James et al., 2014] and [Loh, 2011]. It is worth mentioning that through the usage of a binary pre-processing of the data, all classification methods give same results. This is because the true decision boundaries in fact will turn out to be the same across methods. In this project, we will in the end use binary mapping as input for the statistical learning algorithm, therefore we do not discuss different methods and choices. For further motivating a binary mapping, in addition to its simplicity, we can argue that market participants tend to be focusing on up- and down moves, not evaluating the values of returns in a linear manner. Such observations have been reported by for instance [Tversky and Kahneman, 1974].

Chapter 4

Time Series Decomposition into an ABM

We will now introduce the ABM. Recall that the ABM is applying statistical learning in order to predict the stock market. This approach allows for several agents that collectively incorporate the market returns. On the most fundamental level, since investors are collectively producing the price time series, one can in theory decompose the time series into the actions of all investors. Of course this is an infeasible task in real capital markets. However, decomposing time series is a reoccurring topic in the literature, where the idea is to decompose the series into a superposition of underlying basis states. For example Fourier decompositions have been used in numerous fields of research to decompose time series, e.g. to distinguish long term price cycles from shorter cycles. Another approach is to apply a bottom up approach, using agent based modelling. We will decompose the time series into K interacting and heterogeneous agents.

4.1 Decomposing into agents of a lower category

We aim to decompose a sequence of $|\mathbb{C}|$ categories into K sequences of $|\mathbb{C}'|$ categories. In principle, $|\mathbb{C}|$ can be both larger, less or equal $|\mathbb{C}'|$. Let elements of the rolling window sequence of Ω_t be compressed into $\mathbb{C}$ categories. Recall the information set Ω_t is defining the set of available returns that can be used for forecasting through statistical learning. Further, let $\mathbb{C} = \{-1, 0, 1\}$. In order to compress real returns into trinary returns, we introduce a noise threshold r_τ. The real returns that are in absolute magnitude smaller than r_τ are set to zero. Thus we filter out noise from the returns, where noise is defined by r_τ.

The information set Ω_t is split into K heterogeneous agents that each get assigned an information set $\Omega_t^{(k)}$. These sets have the same length as Ω_t,

$|\Omega_t| = l_{is} + m$. Let Ω_{tj} represent the j'th ordered element in Ω_t. The elements in the agents' information sets take on binary values, i.e. $|\mathbb{C}'| = 2$. We let $\mathbb{C}' = \{-1, 1\}$ and the following rules apply for the information sets.

$$\Omega_{tj}^{(k)} = \Omega_{tj} \ \forall \ k, \text{ if } \Omega_{tj} \neq 0 \tag{4.1}$$

$$\sum_{k=1}^{K} \Omega_{tj}^{(k)} = 0, \text{ if } \Omega_{tj} = 0 \tag{4.2}$$

which implies the conservation rule

$$\Omega_{tj} = \frac{1}{k} \sum_{k=1}^{K} \Omega_{tj}^{(k)}. \tag{4.3}$$

We immediately see that a consequence of the choice of $|\mathbb{C}'| = 2$ and $\Omega_{tj}^{(k)} \in \{-1, 1\}$, is that we require an even number of agents. In the example of $K = 2$, one can simply write Eq. 4.1 and 4.2 together as

$$\Omega_{tj}^{(1)} = \begin{cases} +\Omega_{tj}^{(2)}, & \text{, if } \Omega_{tj} \neq 0. \\ -\Omega_{tj}^{(2)}, & \text{, if } \Omega_{tj} = 0. \end{cases} \tag{4.4}$$

With K as even, at each point in time where $\Omega_{tj} = 0$ we have $\binom{K}{K/2}$ ways of specifying the agents' information sets. Let the number of zeros contained in Ω_t be denoted as ρ. The number ρ is referred to as the number of *ambiguous* returns. As there are $\binom{K}{K/2}$ combinations for each ambiguous return, we will have $\binom{K}{K/2}^{\rho}$ combinations along the whole Ω_t. Thus, degeneracy is introduced through this decomposition. In the field of computer science, this is referred to as a *compression loss*. Agents have different beliefs about the ambiguous returns and this will through the statistical learning method possibly result in different beliefs of future returns. This makes the agents heterogeneous.

In the case of $r_\tau = 0$, all agents have the same dataset and thus predict the same future returns. We refer to this situation as having a *one-agent model*, due to the fact that we could replace all the identical agents by one of them. This situation will be a key reference point for our forecasting models. The one-agent model is in other words just the application of the statistical learning method without having agents interacting. One can also take a different view, arguing that the ABM is a *multi-agent extension* of the statistical learning method. For evaluating the contribution of the ABM to market forecasting, one can compare the multi-agent model to the one-agent model. The economic profits of the ABM should be improved or equal to the one-agent model in order to report of a promising method. Given the ABM is successful, we are obtaining a turn-around in the principle of causality. To elaborate, there is a implicit "left-right" causality in Eq. 4.3, meaning that the agents are determined by a decomposition of the information set. If the ABM predicts the market better than the one-agent model, i.e. the direct use of Ω_{ij} without a decomposition, one can argue that the causality is turned. This is the essence of what is called *reverse engineering*. In short, a reversed engineering is the calibration of a model to a given environment and successfully reproduce the behaviour of this environment. This is discussed further in the ABM summary in Chapter 4.3. We frame a null hypothesis regarding the multi-agent approach as the following.

Hypothesis 1 H_0 *: The introduction of a multi-agent extension in statistical learning does not result in improved trading performance.*

If the null hypothesis 1 is rejected, one can argue for a successful reverse engineering of the market. The approach for evaluating the null hypothesis will be revealed in presented in Chapter 6.

4.2 Overconfident Agents

We recall that a compression loss is present due to having ρ ambiguous returns in the information set Ω_t. According to the conservation rule in Eq. 4.3, there are many possible configuration of the agents' information sets. The choice will be motivated by *overconfidence*. To clarify, confidence is nothing but the judgement of probability. The concept of overconfidence has been paid much attention in behavioural fiance, with a strong link to psychology. Thus the following statement is worth pondering;

> *"Perhaps the most robust finding in the psychology of judgement is that people are overconfident"* [De Bondt and Thaler, 1995].

An effect related to this is the so-called *illusion of knowledge*. This describes the effect of disagreeing agents that becomes increasingly polarized, when available information can support both sides. [Lord et al., 1979] argue that people are biased by their initial opinion when interpreting relevant empirical information. This implies that people tend to get biased towards their initial opinion by suppressing evidence that disconfirm their belief. As well, Lord et al. state that people *"draw undue support for their initial position from mixed or random empirical findings"*. This is often referred to as the *confirmation bias*. A consequence is that slightly heterogeneous agents that do decisions based on the same set of information will end up increasing the polarization among them. When the information is ambiguous and the predictability is low, [Griffin and Tversky, 1992] argue that experts are even more overconfident than nonexperts. From the psychological experiment by [Wason, 1966], there is even a bias implying a *search* for confirmatory information. A possible explanation for the confirmation bias is that people are averse to ambiguity. From the experiments by [Heath and Tversky, 1991], people facing choices under uncertainty are drawn to choices that make them feel competent. Further discussions on overconfidence and the confirmation bias within in the field of finance are to be found in e.g. [Barber and Odean, 2001] and [Hirshleifer, 2001]. Several theoretical models have been suggested with the aim to incorporate this, see [Daniel et al., 1998], [Scheinkman and Xiong, 2003] and [Odean, 1998]. There, overconfidence is interpreted as a too optimistic estimate of the precision of the information available to an agent. We will interpret overconfident agents as agents with a confirmation bias, leading to the illusion of knowledge.

Since EMH is central in the construction and evaluation of the model, it is worth discussing the above arguments in light of EMH. The EMH's antitheses can in some aspects be the findings from behavioural finance. There have been shown numerous deviations from the assumption that investors are rational. A review of empirical evidence of EMH can be found in [Yen and Lee, 2008]. Since the seminal EMH paper by [Fama, 1970], the theory has specially since the nineties been questioned by the academics rooted in behavioural finance. Both camps have gained major attention, noticing that Fama himself shared the Nobel Price together with his criticiser and behavioural finance scholar Robert Schiller. One can have long discussions related to if behavioural finance and EMH are mutually exclusive or not. However, both Fama and Malkiel (see [Fama, 1991], [Fama, 1998] and [Malkiel, 2003]) are arguing that the evidence presented in behavioural finance are not actually challenging EMH. Many studies show that ex post, somewhere in time and certain asset class, the market was inefficient. Thus there are seldom evidences of an opportunity of profit through exploiting theories from behavioural finance. In the review by Malkiel, he argues that irrational behaviour, overvaluations, undervaluations and other abnormalities, not necessarily implies an inefficient market. As well, behavioural finance is often criticised of not having a usable framework for treating finance, that pointing out biases is only highlighting spurious special cases instead of describing something more fundamental and general. However, that does not make the findings less true. Also due to the definition of EMH in Chapter 2, it is not a direct contradiction between irrational behaviour and EMH. As well, even with some degree of inefficiencies in the markets it is difficult to consistently profit from trading the market. Thus we leave this discussion without making a further comparison of the two schools.

4.2.1 Implementing overconfidence

To recapitulate the confirmation bias described above, the bias can in some situations result in overconfident agents. We will let our agents interpret ambiguous information in a fashion consistent with this, through the conservation laws in Eq. 4.1 and 4.2. Recall that the agents have the same interpretation of large price movements, i.e. movements that are not filtered out as noise. However with spurious information, meaning the returns are below the noise threshold r_τ, overconfidence becomes present. A small or zero movement in the price is indeed useful information for an investor, however it can be distorted by the confirmation bias. We propose that the agents modify their ambiguous return data in order to improve their confidence in predicting future returns. Thus they end up being overconfident with respect to the real dataset. This turns into an optimization problem that will be discussed below. When the agents get sufficiently polarized, they end up predicting different outcomes when using statistical learning on their respective modified datasets $\Omega_t^{(k)}$. In such situations, where the agents have different beliefs of the future returns, the resulting collective prediction will be set as uniform random. This mean, when the agents are sufficiently polarized, the ABM turn into a *noise trader*. With this specification, the ABM predicts like the sum of the overconfident agents if the polarization is low, but as a noise trader when it is high. Since the polarization depend on r_τ in a non-linear but generally increasing fashion, we expect the ABM to carry out more noise trades when the containment of small returns in Ω_t increases. However, it is worth mentioning that the collective prediction of agreeing agents can be different than the prediction of their corresponding one-agent model, i.e. when $r_\tau = 0$. This implies that the ABM is *not* a pure switching mechanism between the corresponding one-agent model and a noise trader.

We will now describe how the agents are calibrated into overconfident states. Whenever $\Omega_{tj} = 0$, let the choice of $\Omega_{tj}^{(k)}$ of each agent be a function of the *internal uniformity* of the agents. The internal uniformity is denoted as $\bar{p}_t$ and represents the average confidence level of the agents at time t. This confidence level refers to how confident the agents are in their predictions. Recalling that the agents use the statistical learning method presented in Chapter 3, the confidence is based on the agents' information set. The minimized $\bar{p}_t$ is simply obtained when there are as many reoccurring $\hat{\Omega}$ as possible in Ω_t, in total over all agents. Thus we have high *uniformity* when the distribution of $\hat{\Omega}$ is uniform and low if it is skewed. Overconfidence occurs when $\bar{p}_t$ of a multi-agent model is lower than for the corresponding one-agent model.

A feature implied by Eq. 4.3 is the following. Within each $\Omega_t^{(k)}$, there are ρ ambiguous returns that must be assigned values to by the ABM. Each assignment depends on all other assignments across $\Omega_t^{(k)}$. As well, all assignments made by an agent are entangled to all the other agents. To obtain a global optimal $\bar{p}_t$, all possible assignments are relevant for the global outcome. To repeat, increasing $\bar{p}_t$ is equivalent to increasing the number of reoccurring $\hat{\Omega}$ in the data. For measuring this, one can consider the distribution of inputs to the statistical learning method $\hat{\Omega}^{(k)} \in \Omega_t^{(k)}$, denoted as $\phi_\Omega^{(k)}$ for each agent k. One can measure the $\bar{p}_t^{(k)}$ in each agent by performing a Pearson's chi-squared test of the distribution $\phi_\Omega^{(k)}$ [Pearson, 1900]. The null hypothesis of the test is that $\phi_\Omega^{(k)}$ is uniform. We simply define the test as

$$\chi^2_{t,(k)} = \sum_{\theta \in \hat{\Omega}^k} \frac{(\theta - u)^2}{u} \tag{4.5}$$

where $u = E[\phi_\Omega^{(k)}]$ given $\phi_\Omega^{(k)}$ would be uniform distributed. In the case of having independent normal distributed data, as assumed in the null hypothesis, the sample statistics χ^2 would follow a chi-squared distribution. The probability of observing χ^2 is determined by comparing to the χ^2 distribution and from this a p-value is obtained. Again, we can observe at maximum v^m different $\hat{\Omega}$ within Ω_t. For the internal uniformity, we simply use the average p-value resulting from the test of each of the agents. Again, the more reoccurring $\hat{\Omega}$ that appears in the respective datasets, the lower the p-value(s) and internal uniformity will become. Thus we have a quantitative measure of overconfidence and internal uniformity through $\bar{p}_t$.

4.2.2 Overconfidence calibration

It can be high number of possible combinations of selecting values for the ambiguous returns, in total $\binom{K}{K/2}^\rho$. However, it is not essential to obtain the exact global minimum of $\bar{p}_t$ in order to introduce overconfidence. We have no reason to believe that the performance will differ fundamentally between the global minimum and a state that is very near. These arguments motivate the usage of a Monte Carlo approach. A pseudocode is provided in Algorithm 1 in addition to the description that now follows. First, one can observe that by arranging the agents' information sets in a two dimensional lattice, this system can be viewed as a Boolean network or spin lattice. In comparison to the Ising model from statistical physics, our model is more restricted with respect to where *flips* can occur. As well, the interactions in our model use global information, i.e. each flip decision depend on all others in the system. With a flip, we refer to the action of switching values between agents at a site where $\Omega_{tj} = 0$. Due to the degeneracy introduced for those sites, a flip will not violate the laws represented by Eq. 4.1, 4.2 and 4.3.

The method of simulated annealing will be applied in order to approximate the global minimum for $\bar{p}_t$. In this context, $\bar{p}_t$ is often referred to as the *internal energy* of the system. $\bar{p}_t$ will be used as the cost function in the optimization process. More formally, we are seeking the global minimum

$$\psi^* \in \Psi^* := \{\psi \in \Psi : \bar{p}_t(\psi) \leq \bar{p}_t(\pi) \; \forall \; \pi \in \Psi\}, \tag{4.6}$$

where $\bar{p}_t(\pi)$ refers to the internal uniformity of the system in state π. Simulated annealing explores the solution space with a varying probability of accepting suboptimal solutions in the short term. This evolutionary approach is a well known technique for approximating a global optimum in the case of a large solution space, where the main concern is to avoid the solution to be a local optimum.

Further, a flip results in moving the system from a state Ψ towards a state Ψ', with respective internal uniformity $\bar{p}_t(\Psi)$ and $\bar{p}_t(\Psi')$. We will apply the flipping procedure in the spirit of the Metropolis-Hastings algorithm, using Boltzmann statistics. The flip conditions are the following.

- A flip is always applied if it results in a lower internal uniformity.
- If a flip is not resulting in a lower internal uniformity, the flip is applied with a probability $\exp(-\beta\Delta\bar{p}_t)$, with $\Delta\bar{p}_t = \bar{p}_t(\Psi') - \bar{p}_t(\Psi)$.

The variable β is the thermodynamic beta, inverse proportional to the temperature T. This convention has its origin in statistical physics and the analogy is widely used outside of the field. One can see that the temperature regulates to what extent the system is randomized. We can write the *acceptance probability* of moving from state Ψ to Ψ' as $A(\Psi \to \Psi') = \min[1, \exp(-\beta\Delta\bar{p}_t)]$, where A then is an appropriate mapping to an interval between zero and one. The algorithm selects flippable sites randomly, for so applying a flip weighted by the acceptance probability. The procedure start with an initial temperature T_0 and the temperature T is then cooled towards zero. For each temperature step ΔT, ξ random sites are sequentially selected and faced with the flip conditions. The number ξ is denoted as number of *optimization loops*. The Monte Carlo approach can in a simple manner tune the extent of optimization, by increasing ξ, T_0 and the number of temperature steps. It is not in the scope of this project to dig too deep into further details and issues arising with simulated annealing. For the interested reader, an overview is to be found in the paper by [Ingber, 1993]. Further, the simulated annealing process is only done for the initial information sets in each asset class, i.e. before the first trade of the ABM is carried out. Recall that when the trading has started, Ω_t is a rolling window that propagate further through the stock return data. For each new flippable site arriving the window, the acceptance probability is reduced to $A_{T=0}$ and only the new site is flipped. Running the full simulated annealing process for each timestep would be very time consuming, thus with a sufficient low internal uniformity we choose to avoid this. In Appendix A we look into how this process evolves through time, compared to using an uniform probability distribution as the acceptance function.

Algorithm 1 Initialization of the agents

procedure Polarize the agents through minimizing $\bar{p}_t$
$t \leftarrow t_0$; // Initialize time
$\Omega_t^{(k)} \leftarrow \Omega_t$; // Split information set into K agents, as in Eq. 4.1, 4.2 and 4.3
$T \leftarrow T_0$; // Temperature initialization
$x \leftarrow \xi$; // Optimization loops initialization
loop: // Monte Carlo optimizer
  $j \leftarrow$ *random element where* $\Omega_{tj} = 0$;
  $x \leftarrow \text{rand}(0,1)$; // Generate a new random number $\in (0,1)$
  if $A(\Psi \rightarrow \Psi') > x$ **then** // Metropolis-Hastings' acceptance probability
    flip $\Omega_{tj}^{(k)} \; \forall \; k$
  $x \leftarrow x - 1$
  if $x \neq 0$ **then**
    goto *loop*.
  else
    $x \leftarrow \xi$ // Reset inner loop
    $T \leftarrow T - \Delta T$ // Cooling
    if $T > 0$ **then**
      goto *loop*.
    else
      end procedure

4.3 Summary of the ABM

We present a short summary of the forecasting models M_t that now are derived. A flowchart description can be found in Figure 4.1. First, Chapter 4.1 goes through the process of decomposing Ω_t into heterogeneous agents. Returns in Ω_t that are below the noise threshold r_τ are viewed as ambiguous, where the agents are required to assign such returns according to Eq. 4.2. The returns above the noise threshold are set according to Eq. 4.1. This implies the conservation rule in Eq. 4.3, meaning the sum of the agents' information sets remains the same as Ω_t. When $r_\tau = 0$, the ABM behaves like a so called one-agent model, since no returns are viewed as ambiguous and the agents are all identical. Due to this, we can view our model a multi-agent extension of the standard statistical learning method. As well, one can view the decomposition of the agents as a reverse engineering of the market into heterogeneous agents with different perception of the same dataset. The heterogeneity is caused by the introduced overconfidence that polarize the agents. The agents do predictions based on the statistical learning method presented in Chapter 3, where the specific method is decision trees. If the agents predict the same outcome for a future return, the collective prediction by the ABM is set as this prediction. However if the agents have different predictions, i.e. the agents are too polarized, the ABM predicts as a noise trader. With the motivation of the confirmation bias presented early in Chapter 4.2, the agents are calibrated such that their initial confirmation bias, that is leading to overconfidence, is near the maximal. The calibration method is a Monte Carlo approach, elaborated in Chapter 4.2.2.

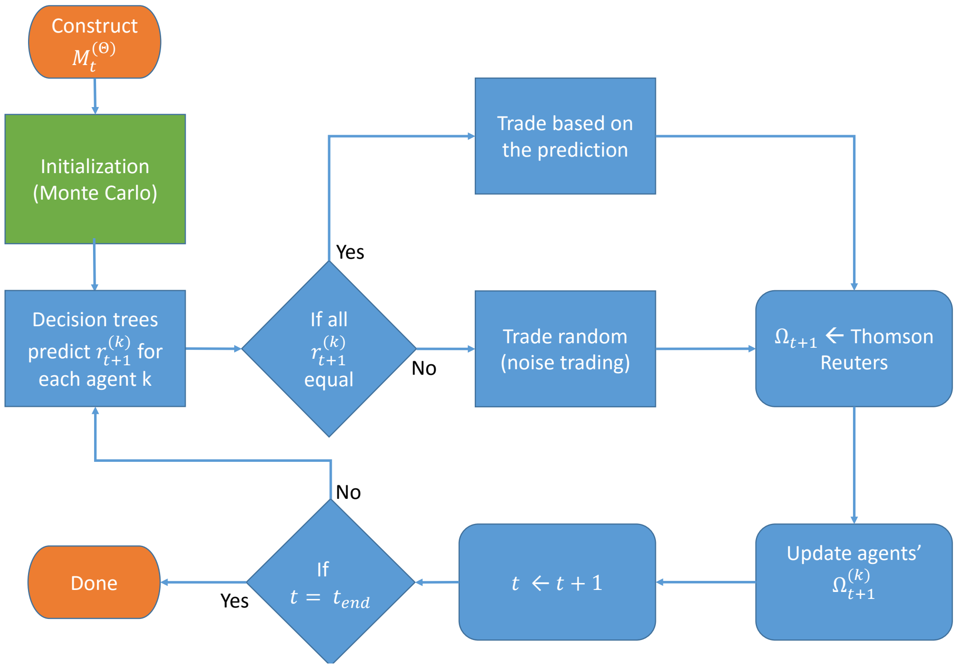

Figure 4.1: The flowchart describes the process of constructing an unique forecasting model $M_t^{(\Theta)}$. A forecasting model is uniquely defined through $\Theta = \{r_\tau, m, l_{is}\}$, the noise threshold, memory length and in-sample length, respectively. The two latter determine the decision trees, while r_τ filter out returns and allow for heterogeneous agents. The initialization is described in Chapter 4.2.2. Then one enter a loop that first predict a future return through decision trees. The upper diamond box raise the condition regarding the collective decision of the agents. The condition discriminate into two trading rules, either noise trading or trade based on the collective decision. The data contained in the information set Ω is provided by Thomson Reuters. The information sets $\Omega_t^{(k)}$ are applied by the decision trees in order to predict future returns. The end of the trading period is here denoted as t_{end}.

To recapitulate, the collection of all forecasting strategies are denoted as

M_t and a single unique strategy is denoted as $M_t^{(\Theta)}$. Through the previous chapters there are only three parameters that are chosen not to be hold fixed. These are the noise threshold r_τ, the memory length m and the in-sample length l_{is}. The two latter parameters where introduced in the former chapter and determining the statistical learning approach. Together, ~~the~~ the three parameters are uniquely determining an individual strategy. Thus we can write $\Theta = \{r_\tau, m, l_{is}\}$. In addition to challenging the EMH, a central question to be answered is the null hypothesis 1. This is questioning the multi-agent model's ability to forecasting stock prices compared to the one-agent model and will be elaborated further in Chapter 6. The next step for addressing the performance of the model is the introduction of a search technology. This will be introduced in the following chapter.

Chapter 5

Search Technology

From the definition of the EMH in Chapter 2, we have required to introduce a search technology S_t. The search technology will decide in each timestep which forecasting model $M_t^{(\Theta)}$ that will be applied for forecasting the next return r_{t+1}. The motivation for a search technology was mentioned in Chapter 2 and is closely related to *data-snooping*. The data-snooping issue is addressed by e.g. [Lo and MacKinlay, 1990] and [Sullivan et al., 1999]. Lo argues the following

> *"Although the likelihood of finding such spurious regularities is usually small (especially if the regularity is a very complex pattern), it increases dramatically with the number of 'searches' conducted on the same set of data"* [Lo, 2007].

The introduction of the search technology aims to mitigate the issue addressed by Lo. There is an infinite number of potential relevant parameters that can be used in a forecasting model. Due to this, [Timmermann and Granger, 2004] argue that all the parameters that have been applied must be revealed for the reader. Not doing so will lead to an enormous bias. We will in the next chapter present all parameters that have been applied in this project.

To recapitulate, highlighting some of many strategies that perform well after a full sweep through the parameter space is not sufficient for satisfying our forecasting experiment. The strategies, i.e. the forecasting models $M_t^{(\Theta)}$, need to be chosen by the search technology *ex ante*. With such a constraint, the approach for testing EMH is close to the test of obtaining consistent excess risk-weighted return through in real time speculating in the market based on Ω_t, M_t and S_t. Since market participants aim to exploit profitable trading strategies, assuming they are on average quite rational, stable statistical patterns in the price time series will have a short lifetime. There is an old saying about EMH:

> *"If any $100 bills are lying around the stock exchanges of the world, they will not be there for long."*

One can thus assume that the search technology should dynamically alternate between strategies. The period where a strategy is capturing positive returns can we expect to decrease with time. This because an increasing amount of financial speculation is carried out with automated trading systems, arguable making the markets more efficient. [Timmermann and Granger, 2004] recommends a *thick modelling* approach, namely *"where a decision based on a combination of outputs of models with statistically similar outputs"*. This is in line with our approach of using the collection M_t and we refer to [Granger and Jeon, 2004] for a broader introduction to thick modelling. The opposite, *thin modelling*, base decisions on a single model. For a changing environment like the financial market, it is thus hard to have a well performing single model through a long period of time. Further, [Timmermann and Granger, 2004] argue that one should not apply a forecasting technique in a point in time where it was not available for the market participants. For an EMH evaluation using historical data, we do a "time-travel" in that sense that we trade with market participants back in time. Thus it is not sensible to apply relatively new techniques like statistical learning and ABMs on data that dates back to e.g. 1930. Since the search technology is emphasized to be an ex ante method, this will lead to a contradiction. For our backtest, we thus only trade in the time period where our techniques were available for the market participants.

There are numerous ways to define a search technology. For example, one can assume that the forecasting model with the highest accuracy will be selected in each timestep. Another possibility is to let the search technology use a Markowitz portfolio optimizer. One can feel tempted to try out many different search technologies for obtaining good results. However, this undermines the whole idea of introducing S_t. To have a large collection of S_t will lead us back to an ex post selection of the best performing method. We limit the number of search technologies to only two. We denote the search technologies as $S_t^{(1)}$ and $S_t^{(2)}$.

Both of our search technologies will be based on the Sharpe ratio. The ratio is defined and discussed further in Chapter 6.1. We let $S_t^{(1)}$ select at each timestep t the model $M_t^{(\Theta)}$ with the highest total Sharpe ratio. Here the total Sharpe ratio refers to a Sharpe ratio that is measured from the point in time where the strategy started trading, until time t. The other search technology, $S_t^{(2)}$, uses a *rolling* Sharpe ratio. Here the Sharpe ratio is measured over the last 100 days, or as many days as possible if we are closer to the starting date of the trading. By construction, the two approaches have their pros and cons. $S_t^{(2)}$ has a shorter response time for selecting models. On the other hand, it might be the case that well performing forecasting models are rather consistent long term, in this case $S_t^{(1)}$ would be a good choice. By applying these two search technologies, we can attempt to evaluate which of the two situations that on average is representable.

We will also pay attention to which of the forecasting models that are used by a search technology. By simply looking at the distribution of r_τ of the $M_t^{(\Theta)}$ picked by each S_t, we can determine which type of ABM that result in high economic profits.

Chapter 6

Benchmarking

As mentioned in the introduction, the weak EMH is stating that it is not possible for an investor to consistently obtain excess risk-adjusted rates of returns, when only using historical returns as information set. There is a concern arising with introducing the term of excess returns. This is because it is relative to a benchmark that is considered to be efficient. As well, the definition of risk is not completely trivial. Therefore we will in Chapter 6.1 introduce different performance measures that have different approaches of the risk evaluation. There is a not a strict consensus in financial literature for benchmarking, treatments of possible biases or measurement methods in general. We attempt to establish a solid approach in this project in order to evaluate our results.

First, we clarify two important issues. It is in principle impossible to detach a trader from the market, since by carrying out trades the trader interferes with the other market participants. All trades influence the market, but as the trading volume grows, so does the market impact. In this project we assume the market impact of the model to be negligible. This assumption is necessary in order to do backtesting. By avoiding this assumption one have to forecast how the market would have reacted to the trades resulting from the model. This would lead us into an absurd scenario. Assuming low volumes traded by the model should be sufficient for avoiding this effect. The second issue is related to the precision of the implementation itself, the *look-ahead* bias. This bias is fairly simple to interpret, however still not easy to mitigate. As the name suggest, this bias is addressing the access of information in a point in time where it should not yet be available for the modeller. We request no such mistakes in our technical implementation of the trading strategies. Procedures addressing this issue is in line with what is described in [Fiévet and Sornette, 2016].

We can distinguish the evaluation of the results by separating them into two main categories. The first is the EMH evaluation, where we take use of the search technology to satisfy the definition of EMH in Chapter 2. The evaluations related to this category are restricted to the contents of Chapter 6.1 and 6.2. The second category is the ex post *market spectroscopy*. The performance obtained by different forecasting models will give us a spectroscopy of the price time series of the traded instrument. This approach has for example been applied by [Zhou et al., 2011]. They explain the term ”spectroscopy” through the analogy of the technique from physics with the same name. More concrete, we can hopefully observe how different strategies capture different aspects of the traded instrument by scanning the parameter space. By being careful, having significant statistics and avoiding data snooping issues, we might gain knowledge of which situations the market can to an extent be predictable and where our models can succeed. Here we again emphasise that this approach build on ex post measurements and should not be confused with the first category. The null hypothesis 1 is comparing the one-agent situation with the multi-agent and will primarily be evaluated through the spectroscopy.

6.1 Risk-Return Performance Measures

The by far most popular risk-return performance measure (RRPM) is the Sharpe ratio. It is defined as the excess return divided by the volatility of the asset price. The volatility is set as the standard deviation. Here the excess return is defined as the mean return of the asset, minus the risk free rate of return. The latter is thus a benchmark rate, often set as the interest rate of a government bond. In this fashion the Sharpe ratio is including both opportunity cost and a risk-return weighting. We will use the Sharpe ratio as our key benchmark, due to its simplicity and preferred measure in the literature. However the Sharpe ratio have several issues that are worth mentioning. It is not entirely correct to call the volatility and risk by the same name. A problem with risk is that it is not an observable, like returns are. The standard deviation is only a precise statistic measure when the time series of returns are produced by a process being parametric and stationary. First, it is doubtful to claim that returns are stationary. Having a parametric process generating returns is as well questionable. Another issue is how the Sharpe ratio favours different types of strategies. Say if a strategy benefits from capturing certain pockets of predictability that causes large upward movements in its returns, but can be rather rare. In this case the Sharpe ratio will be suffering. One can discuss if such a strategy is at all is possible to keep consistent, not being a cause of luck. This question is in principle also relevant for a more stable strategy, due to EMH.

Several alternatives to the Sharpe ratio exist. The Sortino ratio [Sortino and Van Der Meer, 1991] simply exchange the standard deviation in the denominator in the Sharpe ratio with the standard deviation of only the negative returns. The Omega ratio is as well an interesting supplement [Keating and Shadwick, 2002]. This measure rather considers the whole distribution of returns, giving a more nuanced picture of the situation when combined with the Sharpe ratio. By only using mean and standard deviation, potentially a lot of information is lost, specially if a deviation from normality. The Omega ratio does not exclude any higher moments of the return distribution, including kurtosis and skewness. The first step to calculate the Omega ratio is to generate the empirical cumulative distribution function (CDF) of the returns. A loss threshold r_ω has to be chosen exogenous and represents the return that is viewed as being the minimal acceptable one. In fact this is somewhat analogous to the risk free rate used in the Sharpe ratio, but r_ω rather describes the intersection of loss and gain viewed by the investor. The CDF is divided into two areas, gains and losses, where the former is the area above r_ω and the latter is below. Then the Omega ratio $\omega(r_\omega)$ is defined as the probability weighted ratio of gains (G) to losses (L), given a loss threshold r_ω. We can write the ratio as

$$\omega(r_\omega) := \frac{\int_{r_\omega}^{b}(1 - F(r))dr}{\int_{a}^{r_\omega} F(r)} = \frac{G(r_\omega)}{L(r_\omega)}, \tag{6.1}$$

where (a, b) is the interval of returns. For the interested reader, [Kazemi et al., 2004] show that the Omega ratio as well can be viewed in a derivative context.[1]

An extensive overview of different relevant RRPMs is found in [Eling and Schuhmacher, 2007]. They applied 13 different RRPMs for evaluating the return data of 2763 hedge funds. The conclusion is that the ranking of the funds does not change much depending on the measure used. A possible explanation is that the returns investigated are elliptical distributed, which for example is observed by [Lhabitant, 2009] and also in the data presented by [Eling and Schuhmacher, 2007]. The same observations are done for mutual funds, where according to [Guo and Xiao, 2016] there seems not to be any ranking differences. However, in the (white) paper by [Winton Capital Management, 2003] they conclude the opposite and argue for the use of the Omega ratio for achieving a correct ranking of hedge funds. To simplify this problem, we will present both the Sharpe ratio and the Omega ratio for our resulting model. It is out of the scope of this project to investigate in detail the return distribution of all strategies. The risk-free rate of return included the Sharpe ratio will be set to zero, since the ratio will be used for comparing strategies with the same risk-free rate of return. However, the

[1] The Omega ratio is in fact mathematically equal to the expectation value of a call option payoff, divided by the expectation value of a put option payoff. Here the expectations are under the historical probability measure and the underlying asset's strike price can be translated into the loss threshold.

same approach is not as trivial for the thresholds r_ω in the Omega ratio. As discussed by [Winton Capital Management, 2003], the rankings can differ considerably depending on r_ω. This feature has its pros and cons. The advantage is that one can better specify what are considered as profit and loss. On the other hand, the Omega ratio is more complex and intricate for comparison purposes in contrast to the Sharpe ratio. This because r_ω introduces an additional degree of freedom. The risk-free rate of return linearly scales the Sharpe ratio, such that a ranking of strategies applying the Sharpe ratio will be unproblematic. We will for simplicity use the Sharpe ratio for as the key performance measure. The Omega ratio will exclusively be used in order to obtain a more nuanced view of how drawdowns and large returns are impacting the strategies. We consider the thresholds $r_\omega = 0$ and $r_\omega = 1\%$ in the Omega ratio evaluation in Chapter 7.1.

6.2 Transaction Costs Analysis

Transaction costs have a path dependent impact on the returns obtained by a trading strategy. The costs are typically build up as brokerage fees and market bid-ask spread. These costs can vary both between traders and in time. For the EMH evaluation we require the measuring of economic profits, i.e. it is crucial that transaction costs are accounted for. In order to include transaction costs, we first assume the transaction costs for each movement in position to be the same. Meaning that obtaining a long position is as costly as a short position, which often can be achieved even for retail traders when having margin accounts. As well, the transaction costs are assumed to be constant in time. We avoid quoting a specific value to the transaction costs, instead we introduce the *Annualized Sharpe ratio Break Even* (ASBE). This measure represents the cost per transaction that is necessary for the model to obtain the same annualized Sharpe ratio as the corresponding buy-and-hold strategy. The ASBE will be measured in number of basis points (BPS).[2] The number of BPS is subtracted from the model's returns at each point in time where a transaction occurs.

6.3 Trading quantiles

The excess predictability of stock returns does not translate directly into excess returns. To provide a simple example of the difference, consider a strategy that predicts small movements accurately but misses all the large. Thus the strategy can result in having poor economic profits, even though it predicts the market movements better than random. [Fiévet and Sornette, 2016] introduce a measure of the ability to predict the market with resulting

[2]Basis point is a common unit in finance, where one basis point equals 0.01%

positive returns, based on random strategies as a benchmark. One first generate a set of random strategies. The random strategies are generated such that the average change of position will be equal to the strategy being benchmarked. The transaction costs will then be very similar. It is then assumed that the random and non-random strategies can be compared without a further transaction cost discussion. We generate for each trading strategy $M_t^{(\Theta)}$ a set of Φ random strategies generated as described above. Let the compounded returns of the i'th random strategy be denoted as $R_i(t)$. The set of all random strategies is then written as $\mathbb{S}_R = \{R_1(t), ..., R_\Phi(t)\}$. For simplicity, we let $M_t^{(\Theta)}$ represent the compounded returns of the respective forecasting model. The *quantile function* of the strategy $M_t^{(\Theta)}$ is then defined as

$$Q(t) = P(R(t) \leq M_t^{(\Theta)} | R(t) \in \mathbb{S}_R) \tag{6.2}$$

The quantile function contains a lot of information, but we will for each strategy $M_t^{(\Theta)}$ extract the mean only. With the time independent value can we evaluate the average ability of $M_t^{(\Theta)}$ to profit from prediction of future returns, with the random strategies with no predictive power as the benchmark. The average of all the mean quantile functions, where each correspond to a strategy $M_t^{(\Theta)}$, is denoted as $\overline{Q}$. In order to limit the discussion in the next chapter, only $\overline{Q}$ will be presented Chapter 7.2, for each traded instrument.

It is worth mentioning that there exists several other similar methods applying random strategies. One of them is the martingale predictor used by for instance [Satinover and Sornette, 2012a], [Satinover and Sornette, 2012b]. Another is the random strategy used by [Biondo et al., 2013]. The key strength and difference of our approach is that the transaction cost analysis can elegantly be avoided if the above assumption holds.

6.4 Prediction Accuracy

In addition to the quantile function that concentrates on evaluating the trading performance, also the accuracy of the predictions will be evaluated. This is not meant as a performance benchmark per se, but still provides interesting information about the forecasting models. Then one can investigate the relation between excess accuracy and excess return through comparing with the quantile functions. We introduce an accuracy measure through comparing the real and predicted returns. Let the real returns be mapped into $|\mathbb{C}^*|$ categories, while predicted returns are mapped into $|\mathbb{C}'|$ categories. For testing the accuracy of the predicted returns, we start with the following null hypothesis.

Hypothesis 2 *H_0 : the categorical real and predicted returns, mapped into $|\mathbb{C}^*|$ and $|\mathbb{C}'|$ respectively, are statistically independent.*

The values resulting from mapping the predicted and real returns are used for building a contingency table, with the elements as $O_{i,j}$. The table is used in a Pearson's chi-square test of independence [Pearson, 1900], as used in Chapter 4.2, now with Hypothesis 2 as the null hypothesis. We define the test as

$$\chi^2 = \sum_{i=1}^{|\mathbb{C}^*|} \sum_{j=1}^{|\mathbb{C}'|} \frac{(O_{i,j} - \mu_{i,j})^2}{\mu_{i,j}}, \tag{6.3}$$

where $\mu_{i,j} = N p_i p_j$ is the expected frequency of $O_{i,j}$ given the null hypothesis 2. Further we have

$$N = \sum_{i=1}^{|\mathbb{C}^*|} \sum_{j=1}^{|\mathbb{C}'|} O_{i,j}, \quad p_i = \frac{1}{N} \sum_{j=1}^{|\mathbb{C}'|} O_{i,j}, \quad \text{and} \quad p_j = \frac{1}{N} \sum_{i=1}^{|\mathbb{C}'|} O_{i,j} \; .$$

It is common to assume a p-value threshold that can indicate predictive power. Meaning, a strategy that is generating the predicted returns has predictive power if the resulting p-value is below this threshold. The convention is to set this threshold to 0.05. The degrees of freedom of the test is $(|\mathbb{C}'| - 1)(|\mathbb{C}^*| - 1)$. We will apply binary mapping for both predicted and real returns, i.e. $|\mathbb{C}^*| = |\mathbb{C}'| = 2$. In our case we have multiple experiments where we independently carry out statistical tests. Then, some resulting p-values will be very low just by chance, even though the null hypotheses 2 in fact is true. One would expect a fraction of 5% of the tests to be such false positives, meaning type I errors in null hypothesis. For the EMH evaluation, this issue is solved simply by introducing the search strategy, due to having just one experiment. In a forecasting framework, one can argue that this is a more powerful approach than accounting for the false discoveries. However, for looking into the prediction accuracy it is necessary to take the false discoveries into account. According to [Harvey and Liu, 2014], the false discovery rate approach is favourable for the treatment of trading strategies, compared to a family-wise error rate like the Bonferroni procedure. The Benjamini-Hochberg procedure is a good option within the class of false discovery rate approaches [Benjamini and Hochberg, 1995]. A similar approach, which is more appropriate under positive dependence assumptions of the samples, was later provided by [Benjamini and Yekutieli, 2001]. One can argue that our trading strategies are to some extent dependent. But the prediction accuracy is meant first and foremost as an indication of strategy behaviour, not a strict performance measure in it self. Thus do we consider the Benjamini-Hochberg procedure as a sufficient method to account for the type I errors. The procedure functions as the following. Let the resulting p-value for strategy i, with respect its null hypothesis H_i, be denoted as P_i. The ordered sequence of p-values is denoted as $P_1, P_2, \ldots, P_{|\Theta_{r_\tau}|}$, where $|\Theta_{r_\tau}|$ is the number of strategies per noise threshold r_τ. We repeat that $\Theta = \{r_\tau, m, l_{is}\}$. This is uniquely determining a forecasting model and the values for the noise threshold r_τ, memory length m and in-sample length l_{is} are determined in Chapter 6.5. The two latter are specifying the training set of the statistical learning approach, while the former is defined through the ABM that apply statistical learning. Following the Benjamini-Hochberg procedure with the level α, we let z be the largest α in which $P_i \leq \frac{i}{|\Theta_{r_\tau}|}\alpha$. Then we reject all corresponding H_i for $i = 1, 2, \ldots, z$. The maximum acceptable α is set to 0.05.

This analysis is presented in Chapter 7.2 as the final part of the market spectroscopy.

6.5 Specification of parameters, dataset and trading restrictions

In this project we will apply our model to highly liquid stock indexes. Thus it is assumed that all placed orders are possible to carry out. The indexes selected in this project are the U.S. S&P500 and the British FTSE100, where the data is provided by Thomson Reuters. At each prediction done by a forecasting model $M_t^{(\Theta)}$, the whole compounded wealth obtained by the model at time t is invested either into a long or short position. The choice is depending on the prediction of the return at $t+1$. All forecasting models start out with the same initial wealth, where compounded returns of the trades are normalized by this.

Through the last chapters, it has become clear that the essential parameters determining our forecasting models are the noise threshold r_τ, memory length m and in-sample length l_{is}. This is uniquely defined by the set $\Theta = \{r_\tau, m, l_{is}\}$. Below we will provide the specific values used for these parameters. Further, the number of agents K is restricted to 2. Only with two agents the resulting measurements for the ABM differs significantly from the one-agent model, making the two-agent restriction sufficient for now. As well, we avoid to further increase the parameter space. With the framework presented in Chapter 4, the extension to more agents is in principle straight forward.

For an overview of the impact of r_τ on the input dataset, it is worth mentioning the relation between r_τ and the amount of returns being cut off. Stock return distributions are shown to be similar or more heavy tailed than the Gaussian distribution (see for example [Malevergne et al., 2005] and [Pisarenko and Sornette, 2006]). Instead of fitting the whole distribution of returns for more generality, we seek a simple understanding of how the cut-off behaves in our range of noise thresholds. First we look at the dataset consisting of daily returns from the S&P500, between the start of year 2000 until end of 2015. The lowest r_τ applied in the models, $r_\tau = 0.0001$, leads to treating around 1.2% of the returns as noise. The largest noise threshold considered, $r_\tau = 0.001$, corresponds to a 10.6% cutoff. In same time range for the FTSE100, $r_\tau = 0.0001$ to correspond to only 1.0% cut off. The largest threshold, $r_\tau = 0.001$, to corresponds to 9.8%. The set of noise thresholds used in this project is chosen as $r_\tau \in [0.1\%, 0.05\%, 0.025\%, 0.01\%, 0\%]$. The memory lengths used are $m \in [2, 3, 4, 5]$. Lastly, the in-sample lengths are restricted to $l_{is} = 20 + 10b$ where $b \in [1, 2, 3, \ldots, 49]$. The simulations presented in Chapter 7 are exclusively using these sets of parameter values, where we do not exclude any experiments for the reader due to the argumentation in Chapter 5.

The parameters related to the initialization process of the ABM will be fixed through all forecasting models. We will do a short investigation of the annealing impact through preliminary results presented in Appendix A. We settle with the parameters $T_0 = 0.03$, the optimisation loops $\xi = 20$ and having in total 10 equally distanced temperature steps towards zero from T_0.

Closing this chapter, we have finally established the project's approach of forecasting in finance. The key takeaway is famously framed by a legendary scientist.

> *"The first principle is that you must not fool yourself - and you are the easiest person to fool."* (Richard Feynman[3]).

[3]This quote can be found in the book *"Surely You're Joking, Mr. Feynman!": Adventures of a Curious Character*, WW Norton & Company, 2010

Chapter 7

Results

The model presented through the previous chapters give interesting results. Through the discussion in Chapter 2, we did argue for measuring trading performance as economic profits. The Sharpe ratio is chosen as the key performance measure. The performance of the complete model with respect to the EMH evaluation is presented in Chaper 7.1. As stated in the null hypothesis 1, a comparison of the multi-agent model and one-agent model is of interest. This will mainly be discussed through the market spectroscopy in Chapter 7.2. In that chapter, all forecasting models M_t will be discussed, including predictive power, performance measures and the quantile function.

7.1 Gains in UK, losses in U.S.

The forecasting models M_t are applied on two different indexes in the same time period, January 2000 until the end of December 2015. Due to different in-sample lengths in the forecasting models, they will start trading at different points in time. Therefore, the search technologies $S_t^{(1)}$ and $S_t^{(2)}$ start picking models when all models have started trading in the market, which is in January 2002. The risk-return performance measurements are displayed in Table 7.1, without taking into account transaction costs. It is clear that both the search technologies perform well on the FTSE100. On the S&P500, both $S_t^{(1)}$ and $S_t^{(2)}$ have a lower annualized Sharpe ratio than the buy-and-hold strategy. The Omega ratio is also included, where we see that $S_t^{(2)}$ on the S&P500 and the respective buy-and-hold strategy better capture larger returns compared to $S_t^{(1)}$. This is observed by viewing the ranking of the thresholds r_w. By doing this comparison also for the FTSE100, the buy-and-hold strategy is best suited for capturing large returns according to the Omega ratio. Observe in Table 7.1 that the ranking of the strategies are not the same using the Omega ratio instead of the Sharpe ratio. In this situation, the returns obtained by the strategies can witness of non-elliptical bounded distributions.

Table 7.1: Risk-return performance measurements without transaction costs accounted for. Trading period of all strategies is year 2002-2015. The resulting Sharpe ratios have zero risk-free rate of return and are all annualized.

Index	Strategy	Sharpe ratio	Omega ratio	
			r_w=0	r_w=1%
S&P500	Buy-and-hold	0,492	3,704	0,357
	$S_t^{(1)}$	0,394	4,813	0,190
	$S_t^{(2)}$	0,117	5,026	0,309
FTSE100	Buy-and-hold	0,341	3,308	0,343
	$S_t^{(1)}$	0,396	4,967	0,142
	$S_t^{(2)}$	0,526	5,687	0,111

For a proper measure of economic profits, also transaction costs must be included. Table 7.2 displays the search technologies' properties related to transaction costs. The most essential is the ASBE introduced in Chapter 6.2, measured in BPS. Through trading the S&P500, both $S_t^{(1)}$ and $S_t^{(2)}$ are below the break even point without transaction costs accounted for. There is another situation in the FTSE100 trading, where both $S_t^{(1)}$ and $S_t^{(2)}$ have a ASBE above 1 BPS. However, even the ASBE of 2.209 BPS resulting from $S_t^{(2)}$ is probably below the rates most brokers can offer. The performance of $S_t^{(2)}$ is better compared to $S_t^{(1)}$ both before and after accounting for transaction costs. Notice that $S_t^{(2)}$ have more transactions than $S_t^{(1)}$ on the FTSE100, as displayed in Table 7.2. A similar difference is found between the two search technologies trading S&P500, but there $S_t^{(1)}$ has the most transactions. However, the number of transactions are fairly similar for all applied search technologies. A final remark to the Table 7.2 is the ratio of long positions taken by the search technologies. On both indexes the search technologies are close to applying half of the positions long, where the rest as short. On the S&P500, both $S_t^{(1)}$ and $S_t^{(2)}$ have most long positions, where on the FTSE100 the amount of long positions are under half.

The capital gains produced by the search technologies are presented in Figure 7.1 and 7.2, together with the respective buy-and-hold strategies. The gains are measured as Net Asset Value (NAV), where all strategies are starting out with the same. Through time, the compounded returns are added to the NAV. No transaction costs are applied in these figures. Starting with

Table 7.2: Transaction overview. Trading period of all strategies is year 2002-2015. The ASBE is introduced in Chapter 6.2 and measured in BPS. The ASBEs contain zero risk-free rate of return.

Index	Strategy	Transaction Days	Long Positions	ASBE
S&P500	$S_t^{(1)}$	45,57%	53,090%	$< 0{,}00BPS$
	$S_t^{(2)}$	46,54%	50,441%	$< 0{,}00BPS$
FTSE100	$S_t^{(1)}$	46,63%	45,125%	$1{,}055BPS$
	$S_t^{(2)}$	45,58%	46,173%	$2{,}209BPS$

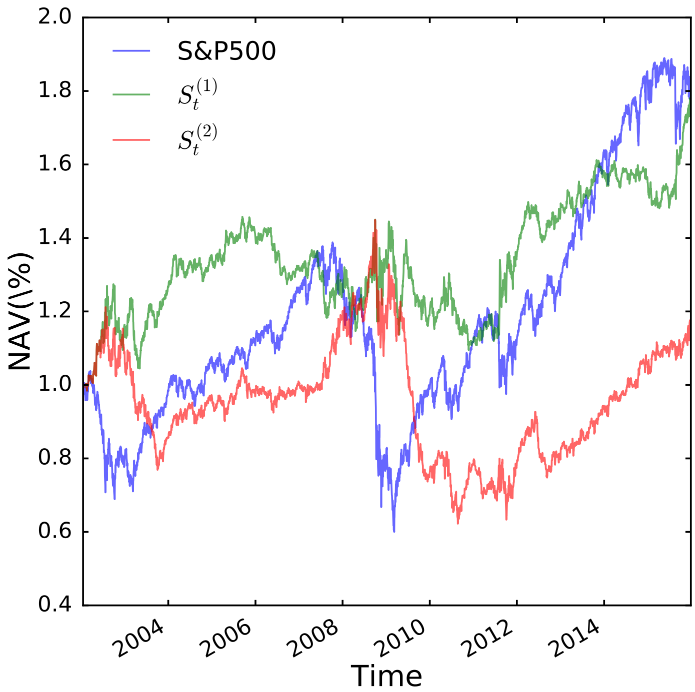

Figure 7.1: Net Asset Value of the search technologies $S_t^{(1)}$ and $S_t^{(2)}$, together with the buy-and-hold strategy, all applied on S&P500. The vertical axis represents the normalized Net Asset Value (NAV), the capital gains measured taking the compounded returns of the strategies. No transaction costs are applied in this figure.

Figure 7.1, the broad picture is that both $S_t^{(1)}$ and $S_t^{(2)}$ perform well in bear markets. This is specially observed during the crashes resulting from the dot-com bubble in 2003 and the financial crisis in 2008-2009. Generally speaking,

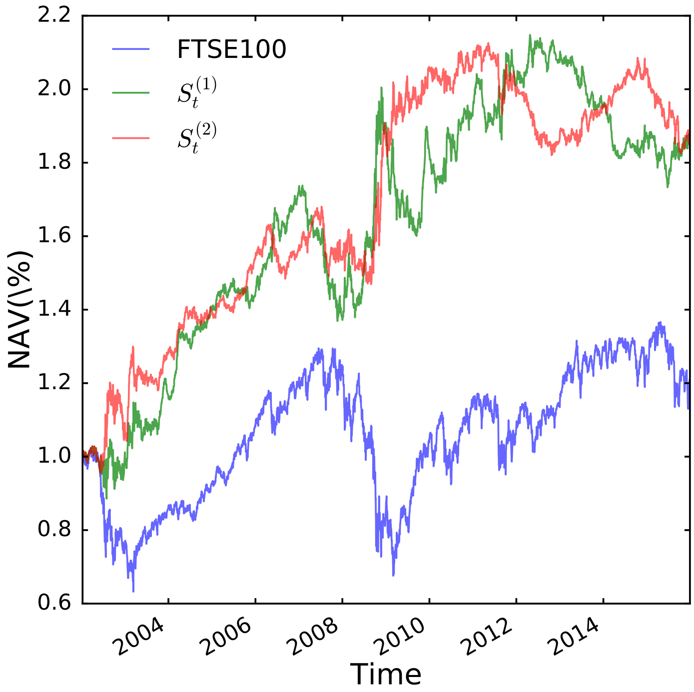

Figure 7.2: Net Asset Value of the search technologies $S_t^{(1)}$ and $S_t^{(2)}$, together with the buy-and-hold strategy, all applied on FTSE100. The vertical axis represents the normalized Net Asset Value (NAV), the capital gains measured taking the compounded returns of the strategies. No transaction costs are applied in this figure.

both strategies underperform buy-and-hold in bull markets, where $S_t^{(1)}$ to a larger extent is capturing upward price movements than $S_t^{(2)}$. We observe a very different situation for the trading of FTSE100, displayed in Figure 7.2. There, both the search technologies are performing well during crashes compared to buy-and-hold. However, they are both on par with the buy-and-hold strategy in bull markets, specially in the time period between the two large crashes. As argued in Chapter 5, the time it would take for a strategy ceasing to capture positive returns can we assume would decrease with time. This since an increasing amount of financial speculation is carried out with automated trading systems. Such an argument is plausible due to observing Figure 7.1 and 7.2. The search technologies are in general not performing well with respect to buy-and-hold in the most recent years.

With the first look, it seems peculiar that the search technologies strongly outperform the buy-and-hold strategy in a market crash. It is indeed a preferable result, however the underlying reason may have a simple explanation. In a dramatic crash, prices have large drops for several days in a row. This would make the forecasting models' training data to be calibrated in favour of short selling. That being said, the in-sample lengths used in the forecasting models are on average rather long, comparing to the crash duration. But in both the largest crashes, on both indexes, the markets were already in a bear or flat state for a period between the market peak and the crash. This time window can give the forecasting models enough time for calibration in order to profit from the crash. It is also worth pointing out that a noise trader will on average perform rather well in the case of a crash, because the amount of short positions will result in extremely large returns. Thus, if the ABM acts like a noise trader, it would most likely perform well. It is a very strong property of a trading strategy to be able to profit in different market regimes. That we clearly see in both search technologies applied on FTSE100 and to some extent by $S_t^{(1)}$ when trading S&P500.

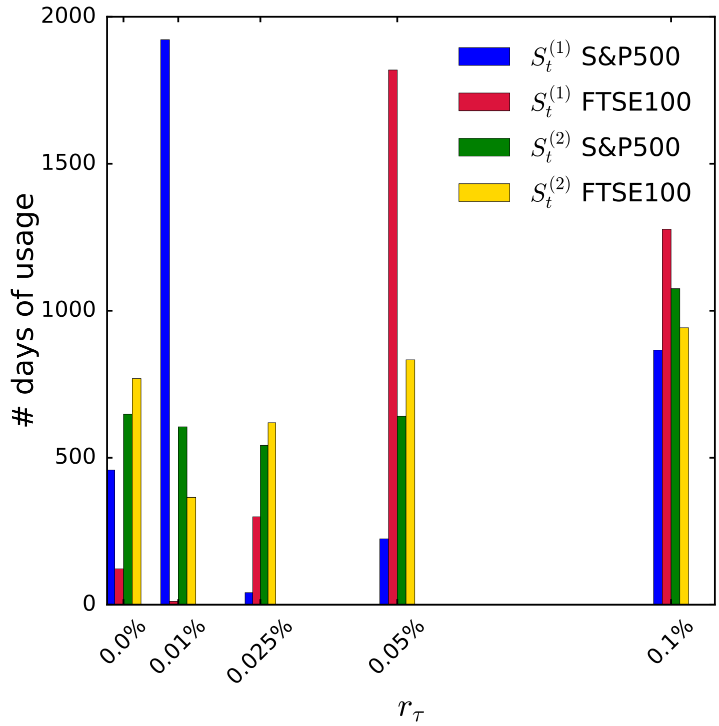

Figure 7.3: The distribution of selected models with respect to the noise threshold r_τ are here displayed. Each bar correspond to a given search technology on a given index. The search technologies $S_t^{(1)}$ and $S_t^{(2)}$ are both applied on FTSE100 and S&P500 in the time period of year 2002-2015.

In light of the null hypothesis 1, there should be a significant fraction of multi-agent models picked by the search technologies. The distributions of selected models with respect to the noise threshold r_τ are shown in Figure 7.3. We observe that both search technologies, on both indexes, are applying models with a wide range of noise thresholds. $S_t^{(2)}$ has a similar distribution on both indexes, almost uniform. On the other hand, the distribution belonging to $S_t^{(1)}$ are on both indexes much more skewed. This could be expected from how the search technologies are defined. $S_t^{(2)}$ picks forecasting models based on a rolling Sharpe ratio, while $S_t^{(1)}$ uses an expanding Sharpe ratio. Thus one can assume that on average, $S_t^{(2)}$ will change forecasting models more frequently than $S_t^{(1)}$, possibly resulting in a less skewed distribution of forecasting models. All in all, it is clear that the multi-agent models are well represented in the search technologies, where the one-agent models in fact are in minority.

7.2 Market Spectroscopy - multi-agents are promising

For a broader view of how the forecasting models behave, we turn to the market spectroscopy. Here all forecasting models M_t are investigated. We repeat that an unique forecasting model is denoted as $M_t^{(\Theta)}$, where $\Theta = \{r_\tau, m, l_{is}\}$. This is uniquely determining a forecasting model and the values for the noise threshold r_τ, memory length m and in-sample length l_{is} are determined in Chapter 6.5. The two latter are specifying the training set of the statistical learning approach, while the former is defined through the ABM that apply statistical learning. M_t is then the collection of all unique forecasting models.

First, Figure 7.4 and 7.5 are displaying the annualized Sharpe ratios of all M_t, for each index respectively. There we pay special attention to how the one-agent models perform compared to the multi-agent models. The one-agent models are clearly not dominating, but neither are the multi-agent models. In fact, when fixing one of the three parameters m, l_{in} or r_τ, none of the resulting forecasting models are consistently dominating. But there are pockets of the parameter space where some forecasting models perform very well in terms of Sharpe ratio. For example, the impact on the resulting Sharpe ratio can be very different depending on the noise thresholds. This motivates the usage of a wide range of parameters, due to the fact that we do not observe any obvious logic in which parameter combination that is superior. The concept of thick modelling discussed in Chapter 5 is in line with this argument. However, through these observations, the null hypothesis 1 can not be rejected. The multi-agent models are contributing to a higher performance, but for certain parameter intervals. Even though the null hypothesis 1 stands, the introduction of multi-agent models can possible contribute positively when applying the search technology if the outperforming models are picked.

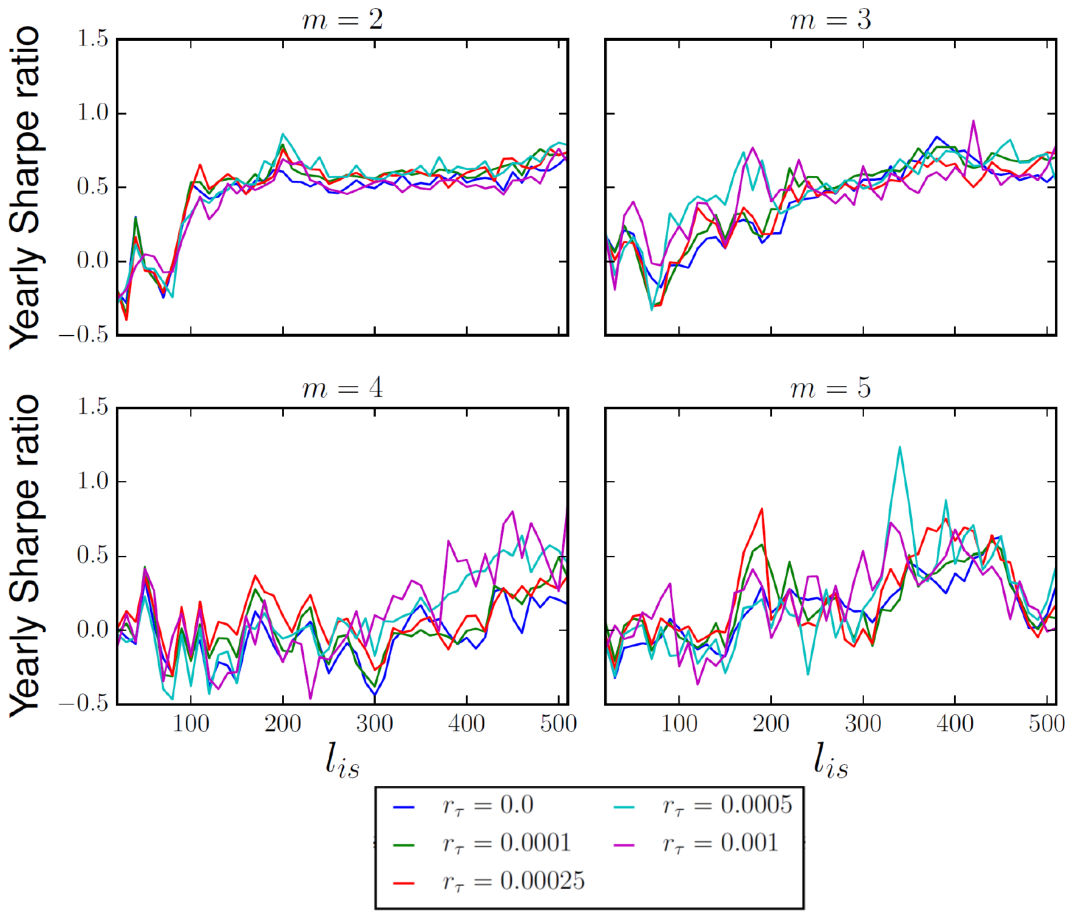

Figure 7.4: All forecasting models M_t are applied on the S&P500 index over the period of year 2000-2015. The resulting yearly, or annualized, Sharpe ratios are displayed, with zero risk-free rate of return. For each memory length m, forecasting models with all the different noise levels are displayed at every in-sample length l_{is}. The in-sample length l_{is} and memory length m are determining the training set of the statistical learning approach. No transaction costs are applied. As a reference, the annualized Sharpe ratio of the buy-and-hold strategy is equal to 0.492 (as stated in Table 7.1).

By the measurements of the quantile function defined in Eq. 6.2, we compare the forecasting models' trading ability with random strategies. The quantile function Q represents the fraction of forecasting models that are performing better than random strategies. As mentioned in Chapter 6.3, only the average quantile function $\overline{Q}$ will be discussed. $\overline{Q}$ is averaged over time and Θ. First, the average quantile function resulting from forecasting the FTSE100 index is equal to 77.57%. For the S&P500 index, $\overline{Q} = 74.25\%$. This tells us that the forecasting models on average perform better than random strategies.

As argued in Chapter 6, there is a non-trivial relationship with the predictability of stock returns and the returns obtained by trading based on predictions. The predictability, or prediction accuracy, is displayed through Figure 7.6 and 7.7. This reports of a general low accuracy. The predictive power is only significant for a fraction of the forecasting models trading the S&P500 index. None are significant on the FTSE100. By comparing with the quantile function one observes that the predictability and performance have no or even negative correlation.

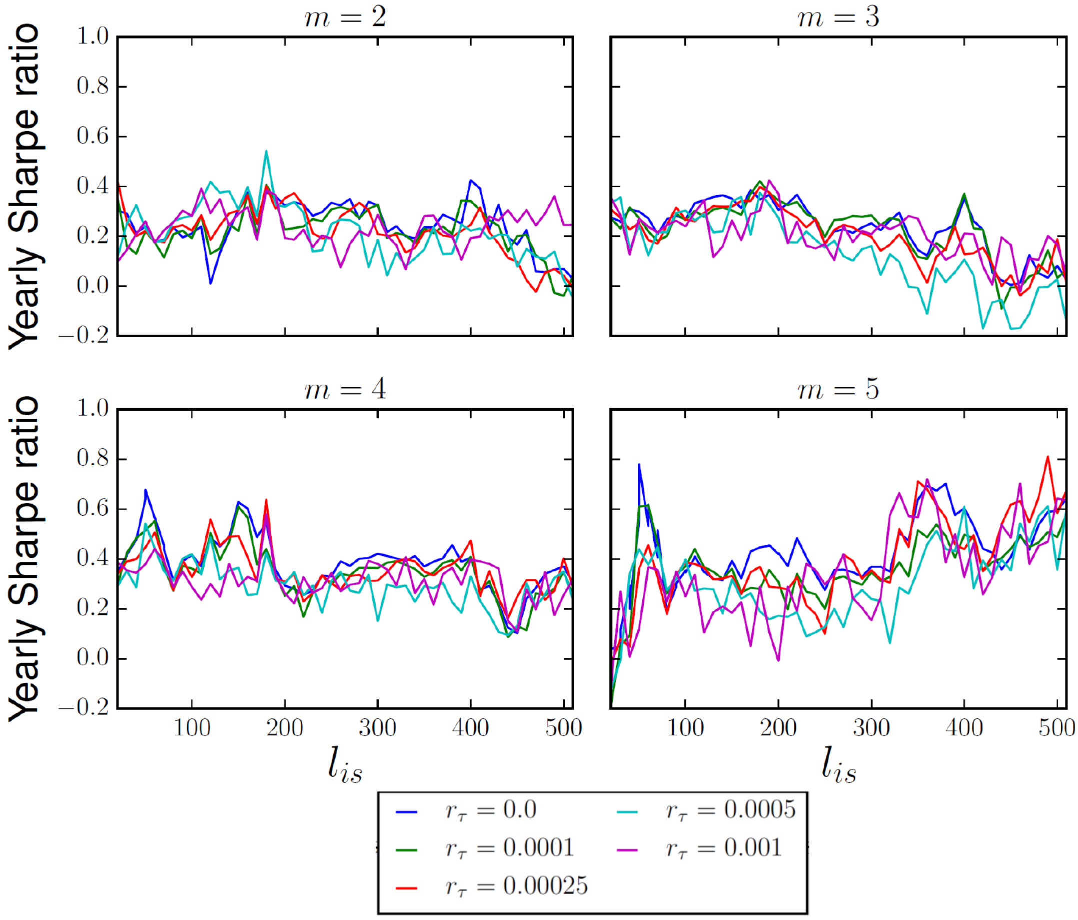

Figure 7.5: All forecasting models M_t are applied on the FTSE100 index over the period of year 2000-2015. The resulting yearly, or annualized, Sharpe ratios are displayed, with zero risk-free rate of return. For each memory length m, forecasting models with all the different noise levels are displayed at every in-sample length l_{is}. The in-sample length l_{is} and memory length m are determining the training set of the statistical learning approach. No transaction costs are applied. As a reference, the annualized Sharpe ratio of the buy-and-hold strategy is equal to 0.341 (as stated in Table 7.1).

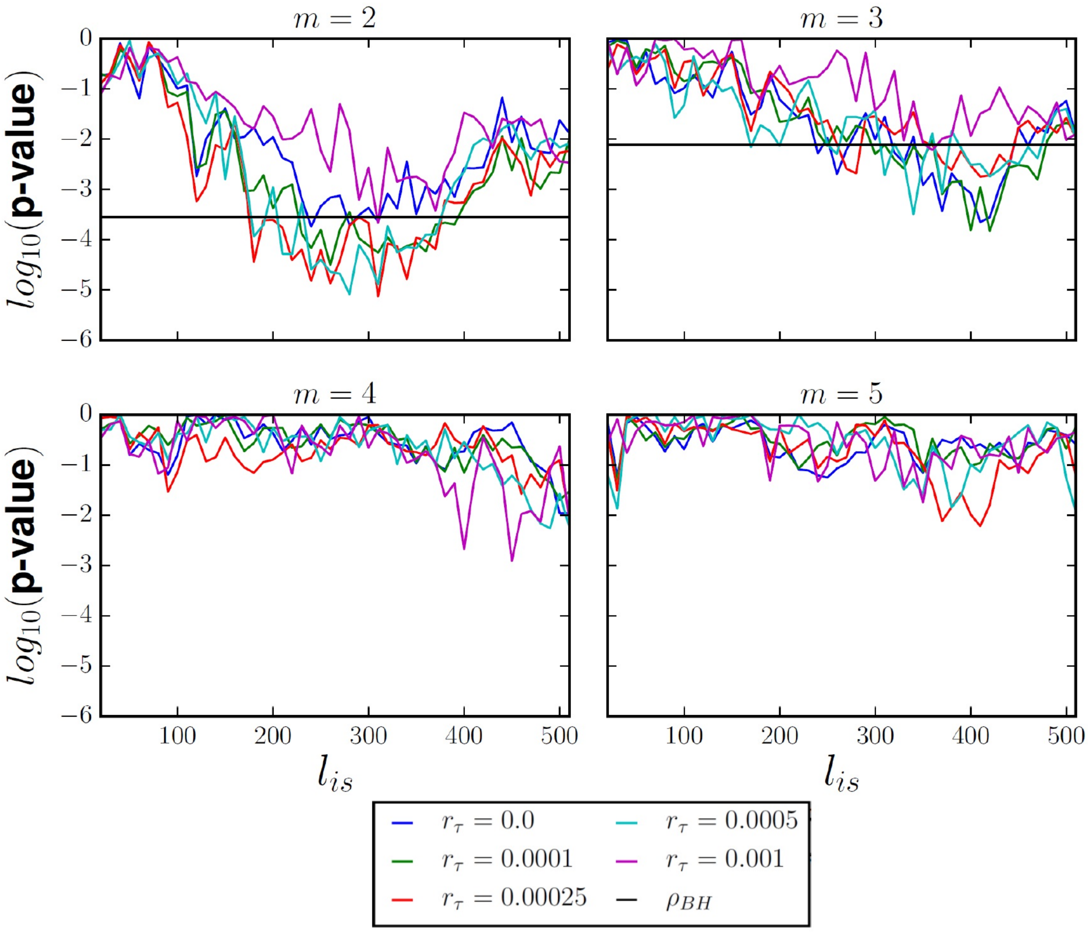

Figure 7.6: The prediction accuracies of all forecasting models M_t are presented in form of p-values. All of M_t are here applied on the S&P500 index in the time span of year 2000-2015. The p-values are defined through to the test in Chapter 6.4. The significance level for the p-values are adjusted according to the Benjamini-Hochberg procedure. The adjusted significance level is visualized in the figure as a horizontal line, denoted as ρ_{BH}. If no line is present, none of the p-values are significant. For each memory length m, forecasting models with all the different noise levels are displayed at every in-sample length l_{is}. The in-sample length l_{is} and memory length m are determining the training set of the statistical learning approach. No transaction costs are applied.

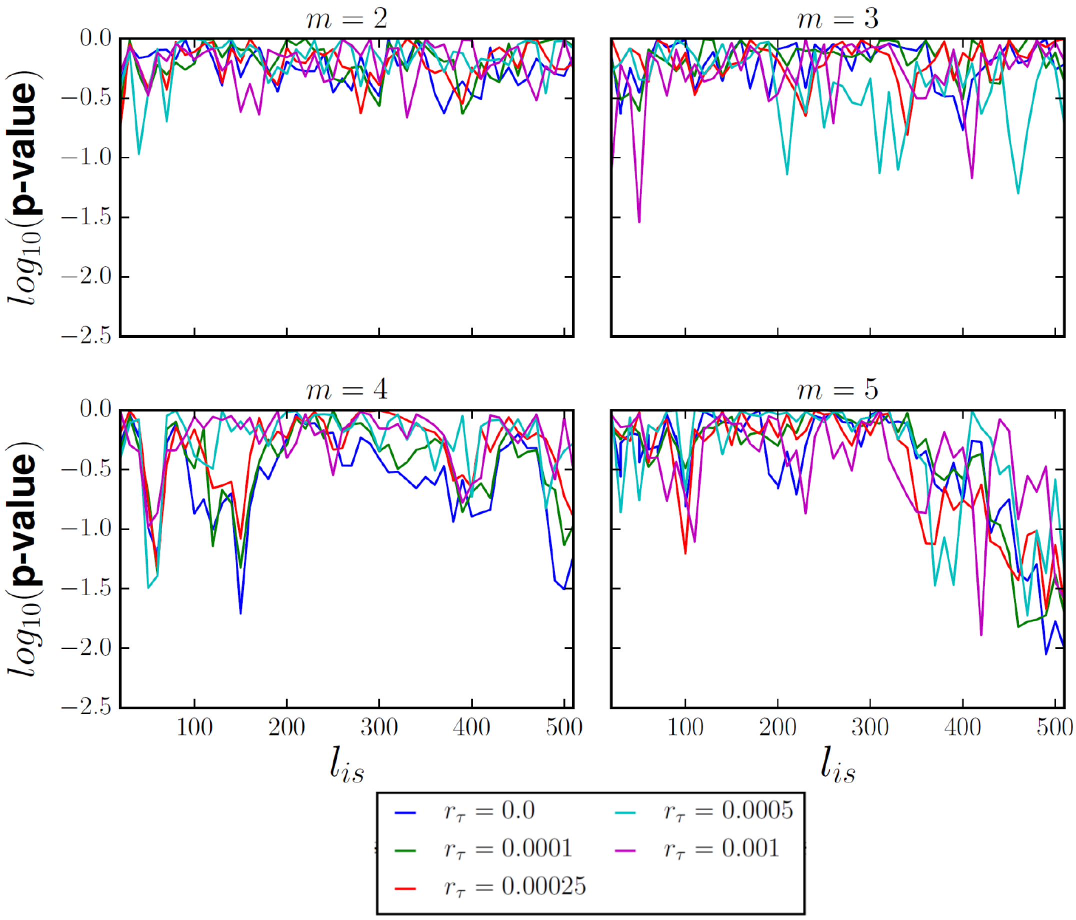

Figure 7.7: The prediction accuracies of all forecasting models M_t are presented in form of p-values. All of M_t are here applied on the FTSE100 index in the time span of year 2000-2015. The p-values are defined through to the test in Chapter 6.4. The significance level for the p-values are adjusted according to the Benjamini-Hochberg procedure. None of the p-values are significant. For each memory length m, forecasting models with all the different noise levels are displayed at every in-sample length l_{is}. The in-sample length l_{is} and memory length m are determining the training set of the statistical learning approach. No transaction costs are applied.

Chapter 8

Concluding remarks

In summary, we have proposed a model for forecasting and trading financial instruments. The model is constructed in order to build a bridge between statistical learning, agent based model and EMH. The EMH definition in Chapter 2 provides an essential framework for implementing the forecasting experiment. The model's performance is measured in economic profits, where transaction costs are taken into account. The experiment take usage of a backtesting methodology, meaning that the model is applied on historical data. This requires a careful description of the methodology, due to numerous possible biases described through the previous chapters.

The model is constructed as an ABM where each agent predicts future returns of a financial instrument with the usage of a statistical learning method. The financial instruments considered are the U.S. S&P500 index and the British FTSE100 index. There exists a broad range of statistical learning methods, however all agents take usage of decision trees. The concept of statistical learning, its relation to EMH and the decision tree approach are described in Chapter 3. Further, the ABM is described in Chapter 4. Motivated by psychology and behavioural finance, the agents are calibrated through a simulated annealing procedure. A summary is then provided in Chapter 4.3. In order to apply the ABM for forecasting and trading, the search technologies in Chapter 5 and benchmarking methodology in Chapter 6 are provided.

Through the EMH evaluation in Chapter 7.1, it is observed that the performances of the models are very different with respect to which index they are trading. There are in total four different experiments, having two search technologies applied on two indexes. The following performance measures refers to the annualized Sharpe ratio. The buy-and-hold strategy's is outperformed by both search technologies on the FTSE100, up to above 1 and 2 BPS, respectively. For the S&P500, both search technologies are outperformed by the buy-and-hold strategy without taken into account transaction costs. In general we are left with no conclusion of which of the two search technologies that is preferred. Going back to questioning the EMH, the conclusion is not straight forward. One can suggest that S&P500 is an efficient market, while FTSE100 is questionable. On the other hand, one can view the situation as having four experiments, where two failed the test while two slightly succeeded. This leaves us with a ambiguous answer. Even though the two indexes are fairly correlated in their trends, the results of applying our model for trading them are very different. However, all experiments except one are able to outperform the buy-and-hold strategy in *different* market regimes. This is a very strong property.

The results from the market spectroscopy are displayed in Chapter 7.2. Through the measured quantile functions, we observe the forecasting models to perform better than random strategies on average. The predictive power, described in Figure 7.6 and 7.6, reports of a general low accuracy. The predictive power is only significant for a small fraction of the forecasting models. By comparing with the resulting quantile function, one can interpret a relationship between accuracy and performance. We observe that the two have no or even negative correlation. Through the Sharpe ratio measurements in Figure 7.4 and 7.5 it is not observed any obvious logic in which parameter range that gives superior performance. This fact motivates the usage of a broad range to be applied in the search technology, in line with thick modelling. As well it is worth mentioning that multi-agent models are not outperformed by the one-agent models consistently, or the other way around. With this we can not reject the null hypothesis 1. It is still possible that the ABMs to some degree are successfully reverse engineering the markets, however it is hard to conclude.

A closer relationship between agent based modelling and statistical learning creates positive synergies, both through the result of our model, but also in terms of bringing the two academic fields closer. We report of mixed results regarding the challenge of EMH through the suggested model. In the context of forecasting in capital markets, it is a necessary to construct the models in the framework of finance. The framework presented through this project can be generalized for backtesting a broad range of trading strategies. The field of agent based modelling has been criticized for not applying a consistent framework for constructing experiments. For forecasting purposes, we suggest a common platform for evaluations and constructions of ABMs.

Appendix A

Evolution of the simulated annealing procedure

The ABM is initialized by a simulated annealing process described in Chapter 4.2.2. For this, the internal uniformity $\bar{p}$ introduced in Chapter 4.2 is a central measure. To get an intuition of how the minimization of $\bar{p}$ evolves through time, we have tracked $\bar{p}$ through the initialization process and further the next year on the S&P500 index. A visualization is provided in Figure A.1. At time zero, the model start trading and one new stock return is arriving its information set each day. Before time zero, the initialization process is applied to the information set of the agents. To recapitulate, $\bar{p}$ is measured from the information sets of the agents. $\bar{p}$ is calculated for each optimization loop of in total ξ loops and temperature step. The temperature cools from T_0 to zero.

For comparing the effect of minimizing $\bar{p}$ during the initializion and during trading, we consider also an uniform random choice for the flips in the ABM. This is displayed as the graph in Figure A.1 starting at time zero. One observes that the internal uniformity is not only on average, but globally much lower through the year with the model in the market when using the simulated annealing procedure compared to a uniform random flipping procedure.

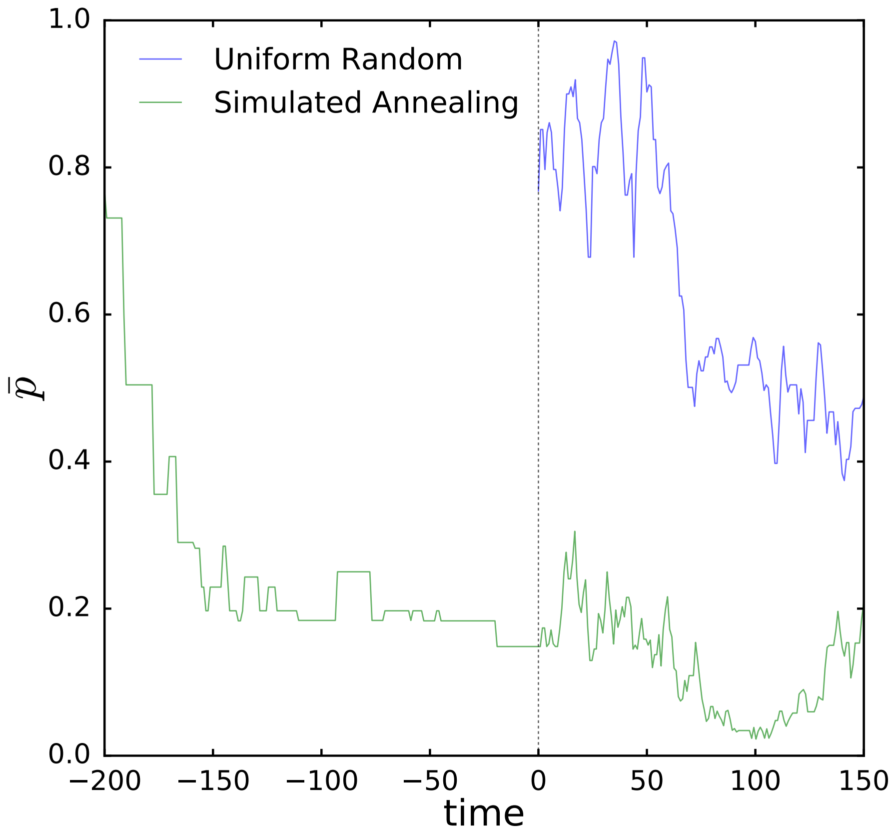

Figure A.1: The plot shows the resulting internal uniformity $\bar{p}$ from the simulated annealing and a uniform random flipping. The prediction period starts at time equal zero, where the model trades the S&P500, from l_{is} days out in the year of 2000 until the last day in the same year. The initialization is made in the time interval before zero. The noise level, r_τ, is set to 0.001, while memory length $m = 2$ and in-sample length $l_{is} = 100$. The initial temperature T_0 equals 0.03 and the number of linear steps to zero temperature is set to 10. The number of optimization loops per temperature step is set as $\xi = 20$.

Glossary

ASBE Annualized Sharpe ratio Break Even. This measure represents the cost per transaction that is necessary for the model to obtain the same annualized Sharpe ratio as the corresponding buy-and-hold strategy. The ASBE is measured in number of basis points (BPS). p. 32

forecasting models M_t is the collection of all forecasting strategies generated through the ABMs. A single unique strategy is denoted as $M_t^{(\Theta)}$, where $\Theta = \{r_\tau, m, l_{is}\}$. p. 7

in-sample length l_{is}, the in-sample length, is the number of returns used in the training set of a statistical learning method. p. 11

information set The information set Ω_t is describing the information available for an market participant or forecasting model at time t. p. 5

internal uniformity $\bar{p}_t$, the internal uniformity, is a measure of the confidence level among the agents at time t. This is minimized with a Monte Carlo technique in order to polarize the agents. p. 19

memory length The memory length m is the number of sequential returns applied as the input in the statistical learning method. p. 11

multi-agent model In contrast to the one-agent model, the ABM takes on the multi-agent model form in the case of $r_\tau \neq 0$. This allow for heterogeneity and possibly different predictions of future returns among the agents. p. 16

noise threshold The real returns that are in absolute magnitude smaller than the noise threshold r_τ, are set to zero. p. 15

one-agent model In the case of $r_\tau = 0$, all agents have the same dataset and thus predict the same future returns. We refer to this situation as having a *one-agent model*, due to the fact that we could replace all the identical agents by one of them. p. 16

return r_t is the return from a given security at time t. p. 5

search technology A search technology in S_t refers to a predecided way to decide which of the forecasting models in M_t that at each timestep will be selected. Two search technologies are applied, $S_t^{(1)}$ and $S_t^{(2)}$, as described in Chapter 5. p. 7

Bibliography

[Allen and Karjalainen, 1999] Allen, F. and Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. *Journal of financial Economics*, 51(2):245–271.

[Andersen and Sornette, 2005] Andersen, J. and Sornette, D. (2005). A mechanism for pockets of predictability in complex adaptive systems. *Europhys. Lett.*, 70(5):697–703.

[Atsalakis and Valavanis, 2009] Atsalakis, G. S. and Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications*, 36(3):5932–5941.

[Barber and Odean, 2001] Barber, B. M. and Odean, T. (2001). The internet and the investor. *The Journal of Economic Perspectives*, 15(1):41–54.

[Benjamini and Hochberg, 1995] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300.

[Benjamini and Yekutieli, 2001] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188.

[Biondo et al., 2013] Biondo, A. E., Pluchino, A., Rapisarda, A., and Helbing, D. (2013). Are random trading strategies more successful than technical ones? *PloS one*, 8(7):e68344.

[Bonabeau, 2002] Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7280–7287.

[Breiman et al., 1984] Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

[Creamer and Freund, 2010] Creamer, G. and Freund, Y. (2010). Automated trading with boosting and expert weighting. *Quantitative Finance*, 10(4):401–420.

[Daniel et al., 1998] Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *the Journal of Finance*, 53(6):1839–1885.

[De Bondt and Thaler, 1995] De Bondt, W. F. and Thaler, R. H. (1995). Financial decision-making in markets and firms: A behavioral perspective. *Handbooks in operations research and management science*, 9:385–410.

[Eling and Schuhmacher, 2007] Eling, M. and Schuhmacher, F. (2007). Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking & Finance*, 31(9):2632–2647.

[Fama, 1970] Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

[Fama, 1991] Fama, E. F. (1991). Efficient capital markets: II. *The Journal of Finance*, 46(5):1575–1617.

[Fama, 1998] Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3):283–306.

[Fiévet and Sornette, 2016] Fiévet, L. and Sornette, D. (2016). The weak efficient market hypothesis in light of statistical learning. *working paper ETH Zurich*.

[Granger and Jeon, 2004] Granger, C. W. and Jeon, Y. (2004). Thick modeling. *Economic Modelling*, 21(2):323–343.

[Granger and Morgenstern, 1963] Granger, C. W. and Morgenstern, O. (1963). Spectral analysis of New York stock market prices. *Kyklos*, 16(1):1–27.

[Griffin and Tversky, 1992] Griffin, D. and Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive psychology*, 24(3):411–435.

[Grossman and Stiglitz, 1980] Grossman, S. and Stiglitz, J. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3):293–408.

[Guo and Xiao, 2016] Guo, B. and Xiao, Y. (2016). A note on why doesn't the choice of performance measure matter? *Finance Research Letters*, 16:248–254.

[Harvey and Liu, 2014] Harvey, C. R. and Liu, Y. (2014). Evaluating trading strategies. *Available at SSRN 2474755*.

[Heath and Tversky, 1991] Heath, C. and Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of risk and uncertainty*, 4(1):5–28.

[Hirshleifer, 2001] Hirshleifer, D. (2001). Investor psychology and asset pricing. *The Journal of Finance*, 56(4):1533–1597.

[Ingber, 1993] Ingber, L. (1993). Simulated annealing: Practice versus theory. *Mathematical and computer modelling*, 18(11):29–57.

[James et al., 2014] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.

[Jensen, 1978] Jensen, M. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, 6(2-3):95–101.

[Jensen, 1968] Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of finance*, 23(2):389–416.

[Kazemi et al., 2004] Kazemi, H., Schneeweis, T., and Gupta, B. (2004). Omega as a performance measure. *Journal of Performance Measurement*, 8:16–25.

[Keating and Shadwick, 2002] Keating, C. and Shadwick, W. F. (2002). A universal performance measure. *Journal of Performance Measurement*, 6(3):59–84.

[Lhabitant, 2009] Lhabitant, F.-S. (2009). *Hedge funds: quantitative insights*. John Wiley & Sons.

[Lim and Brooks, 2011] Lim, K.-P. and Brooks, R. (2011). The evolution of stock market efficiency over time: a survey of the empirical literature. *Journal of Economic Surveys*, 25(1):69–108.

[Lo, 2007] Lo, A. W. (2007). Efficient markets hypothesis. *S.N. Durlauf and L.E. Blume (Eds), The New Palgrave Dictionary of Economics Online, 2nd edn.*

[Lo and MacKinlay, 1988] Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of financial studies*, 1(1):41–66.

[Lo and MacKinlay, 1990] Lo, A. W. and MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. *Review of financial studies*, 3(3):431–467.

[Loh, 2011] Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):14–23.

[Lord et al., 1979] Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11):2098.

[Malevergne et al., 2005] Malevergne, Y., Pisarenko, V., and Sornette, D. (2005). Empirical distributions of stock returns: between the stretched exponential and the power law? *Quantitative Finance*, 5(4):379–401.

[Malkiel, 1991] Malkiel, B. G. (1991). Efficient market hypothesis. In *The World of Economics*, pages 211–218. Springer.

[Malkiel, 1995] Malkiel, B. G. (1995). Returns from investing in equity mutual funds 1971 to 1991. *The Journal of Finance*, 50(2):549–572.

[Malkiel, 2003] Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1):59–82.

[Odean, 1998] Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *The Journal of Finance*, 53(6):1887–1934.

[Park and Irwin, 2007] Park, C.-H. and Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys*, 21(4):786–826.

[Pearson, 1900] Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302):157–175.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct):2825–2830.

[Pisarenko and Sornette, 2006] Pisarenko, V. and Sornette, D. (2006). New statistic for financial return distributions: Power-law or exponential? *Physica A: Statistical Mechanics and its Applications*, 366:387–400.

[Richiardi et al., 2003] Richiardi, M. et al. (2003). The promises and perils of agent-based computational economics. *LABORatorio R. Revelli, Centre for Employment Studies LABORatorio R. Revelli Working Papers Series*, 29.

[Satinover and Sornette, 2012a] Satinover, J. and Sornette, D. (2012a). Cycles, determinism and persistence in agent-based games and financial time-series I. *Quantitative Finance*, 12(7):1051–1064.

[Satinover and Sornette, 2012b] Satinover, J. and Sornette, D. (2012b). Cycles, determinism and persistence in agent-based games and financial time-series II. *Quantitative Finance*, 12(7):1065–1078.

[Scheinkman and Xiong, 2003] Scheinkman, J. A. and Xiong, W. (2003). Overconfidence and speculative bubbles. *Journal of political Economy*, 111(6):1183–1220.

[Sornette, 2014] Sornette, D. (2014). Physics and financial economics (1776–2014): puzzles, Ising and agent-based models. *Reports on Progress in Physics*, 77(6):062001.

[Sortino and Van Der Meer, 1991] Sortino, F. A. and Van Der Meer, R. (1991). Downside risk. *The Journal of Portfolio Management*, 17(4):27–31.

[Sullivan et al., 1999] Sullivan, R., Timmermann, A., and White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *The Journal of Finance*, 54(5):1647–1691.

[Timmermann and Granger, 2004] Timmermann, A. and Granger, C. W. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, 20(1):15 – 27.

[Tversky and Kahneman, 1974] Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

[Wason, 1966] Wason, P. C. (1966). Reasoning. *B. M. Foss, ed.: New Horizons in Psychology*.

[Wiesinger et al., 2012] Wiesinger, J., Sornette, D., and Satinover, J. (2012). Reverse engineering financial markets with majority and minority games using genetic algorithms. *Computational Economics DOI 10.1007/s10614-011-9312-9 (http://ssrn.com/abstract=1553821)*.

[Wilmott, 2013] Wilmott, P. (2013). *Paul Wilmott on quantitative finance*. John Wiley & Sons.

[Windrum et al., 2007] Windrum, P., Fagiolo, G., and Moneta, A. (2007). Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation*, 10(2):8.

[Winton Capital Management, 2003] Winton Capital Management (2003). Assessing CTA Quality with the Omega Performance Measure. *http://www.performance-measurement.org/Winton2003.pdf*.

[Yen and Lee, 2008] Yen, G. and Lee, C. (2008). Efficient market hypothesis (emh): past, present and future. *Review of Pacific Basin Financial Markets and Policies*, 11(02):305–329.

[Zhou et al., 2011] Zhou, W.-X., Mu, G.-H., Chen, W., and Sornette, D. (2011). Investment strategies used as spectroscopy of financial markets reveal new stylized facts. *PloS one*, 6(9):e24391.

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

A bridge between statistical learning and agent based modelling in stock market predictions

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Hellenes	Herman

With my signature I confirm that
- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Zurich, 27.07.2016	Herman Hellenes

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.