UNIVERSITY OF SOUTHAMPTON

FACULTY OF BUSINESS, LAW AND ART

BUSINESS SCHOOL

Volume 1 of 1

LEARNING TO BE SMART: CAN HUMANS LEARN TO IMPROVE PROFITABILITY AND RISK CONTROL IN FINANCIAL TRADING?

by

MING-WEI HSU

Thesis for the degree of Doctor of Philosophy

May 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF BUSINESS, LAW AND ART

BUSINESS SCHOOL

Thesis for the degree of Doctor of Philosophy

LEARNING TO BE SMART: CAN HUMANS LEARN TO IMPROVE PROFITABILITY AND RISK CONTROL IN FINANCIAL TRADING?

MING-WEI HSU

Recently, the efficient market hypothesis has faced strong challenges from various fields, and the purpose of this thesis is to provide empirical evidence for the challenges to the efficient market hypothesis from two perspectives. The first one is from the field of machine learning. While an increasing number of machine learning studies report the high accuracy of stock market prediction, this is not consistent with the efficient market hypothesis which suggests that current stock prices discount available information and that it is not possible to obtain systematic returns by exploiting any predictability of prices. As most of the machine learning studies choose relatively simple test settings, I suspect that the reported high accuracy might result from biased performance measurement. That is, the selection of methodological factors is influential on prediction performance in stock markets. To test my conjecture, I run the benchmark with a comprehensive combination of the methodological factors to collect the performance measures under various settings. Next, I analyze the relationship between the prediction performance and the methodological factors. I find the significant influence of the selection of methodological factor on prediction performance, which means that the reported high prediction performance might be biased and my results are not against the prediction of the efficient market hypothesis.

The second challenge is that there is increasing evidence of anomalies in financial markets. This suggests that the underlying rationality principle of the efficient market hypothesis may be flawed. The manner in which individuals learn from experience also remains a matter of debate. The rationality assumption would be justified if individuals follow Bayesian learning, i.e., individuals learn from experience to appropriately adjust their probability estimates and finally make rational and appropriate decisions. To examine the relationship between experience and performance measures, I use linear mixed models to analyze spread trading data. I find that, as individuals gain experience, they increase their degree of risk-taking and realize higher returns. However, these

returns are subject to greater volatility and, as a result, they achieve lower risk-adjusted returns. Since the individuals following Bayesian learning should be able to appropriately update probability estimates conditioned on new information, their decision choices and their risk-adjusted performance should be improved. My results show that individuals fail to follow Bayesian learning. On the other hand, my results can be explained by reinforcement learning, wherein individuals repeat behavior that was rewarding in the past. Traders may try several trading strategies with different levels of risk. Since higher risk generally brings both higher profits and greater losses, traders who undertake riskier strategies will either make higher profits or suffer greater losses. Those traders making a higher profit are reinforced by the riskier strategies and overlook the underlying risk, which leads to lower risk-adjusted performance. Hence, my results cast doubt on the validity of the rationality assumption.

To further explore the degree of rationality with trading data, I propose a method to estimate the degree to which an individual behave like a rational agent, and other behavioral characteristics. The experience weighted attraction (EWA) can be used to estimate the degree of rationality in psychological experiments, but cannot be used with trading data. The reason is that the number of strategies available to decision makers was limited in psychological experiments, but in real-world trading environments, traders have no limits in terms of the strategies they can adopt. We propose a decision-based strategy mapping framework (DSM) to resolve this problem. The DSM is designed to artificially limit the strategy space associated with real-world trading data, by using scenarios. In each scenario, individuals are assumed to have only one decision to make. This allows us to estimate, using data associated with an individual's real-world trading, their behavioral characteristics associated with EWA. Subsequently, we examine the relationship between the estimated behavioral characteristics of traders and their trading behavior and performance. My results suggest that those traders who behave like rational agents tend to trade more actively. However, surprisingly, those traders who are more rational do not achieve superior trading performance.

In conclusion, the findings of this thesis support the efficient market hypothesis that the markets are efficient, at least to the extent to which excess returns cannot be earned with state of the art machine learning techniques. However, the results of learning behavior from individual-level analysis suggest that the rational agent assumption of the efficient market hypothesis is likely to over-simplify individual behavior in the real world.

Table of Contents

Table of Contents i
List of Tables v
List of Figures vii
DECLARATION OF AUTHORSHIP ix
Acknowledgements xi
Chapter 1: Introduction 1

1.1. Background 1
1.2. Challenges to the EMH from Machine Learning (First Paper) 3
1.3. Is experience the mother of risk wisdom? (Second Paper) 4
1.4. Estimating Behavioral Characteristics associated with Learning Models in Financial Markets (Third Paper) 6
1.5. Contribution 7
1.6. Structure of Thesis 8

Chapter 2: Bridging the Divide in Financial Market Forecasting: Machine Learners vs. Financial Economists 9

2.1 Introduction 11
2.2 Related Work 12

2.2.1 The Efficient Market Hypothesis 12
2.2.2 Price Prediction in Financial Markets using Machine Learning 13

2.3 Hypothesis Development 17
2.4 Experimental Design 20

2.4.1 Data, Variables, and Forecasting Horizon 22
2.4.2 Performance Measurement 31

2.5 Empirical results 34

2.5.1 Tests of the Hypotheses of Experimental Factors on Market Predictability 34
2.5.2 Sensitivity Analysis 47

2.6 Discussion 55
2.7 Conclusions 58
2.8 Online Appendix A: Comparing the forecasting ability of commonly employed econometric techniques 62

2.9 Online Appendix B: Identifying the Prediction Method Producing the Best Accuracy and ROI for Each Stock Index 66
2.10 Online APPENDIX C: Analysis of Markets with Intraday Data Available Prior to 2008 71

Chapter 3: Is experience the mother of risk wisdom? 75

3.1 INTRODUCTION 76
3.2 LITERATURE AND HYPOTHESES 78

3.2.1 Experience and Risk-taking Behavior 78
3.2.2 The Effect of Experience on Performance 79
3.2.3 The Relationship between Risk-taking Behavior and Returns Volatility 81

3.3 METHODOLOGY 81

3.3.1 Data 81
3.3.2 Variables 84
3.3.3 Models 91
3.3.4 Controlling Biases 94

3.4 RESULTS 98

3.4.1 The Effect of Experience on Risk-taking Behavior 98
3.4.2 The Effect of Experience on Performance 108
3.4.3 Investment Size 117
3.4.4 The Relationship between the Volatility of Returns and Risk-taking Behavior 119

3.5 DISCUSSION 120

3.5.1 Change of Risk-taking Behavior through Experience 120
3.5.2 Risk-Return Trade-off 121
3.5.3 Bayesian Learning 121
3.5.4 Reinforcement Learning 122
3.5.5 Dynamic interaction of risk-taking behavior, volatility and experience 122
3.5.6 Survivorship Bias 123

3.6 CONCLUSION 123
3.7 APPENDIX: TRADE NUMBER AS EXPERIENCE MEASURE 125

Chapter 4: Estimating Behavioral Characteristics associated with Learning Models in Financial Markets 131

4.1 Introduction 132
4.2 Literature 134
4.2.1 The influence of individual characteristics on trading behavior 134
4.2.2 The influence of experience: learning effect 135
4.2.3 Behavioral models of learning 137
4.2.4 Hypotheses 139
4.3 Methodology 141
4.3.1 Data 141
4.3.2 Learning models 143
4.3.3 Decision-based Strategy Mapping Framework (DSM) 148
4.3.4 Applying Learning Models to Real Markets 150
4.3.5 Variables 151
4.3.6 Testing Hypotheses 153
4.4 Results 154
4.4.1 Scenario/learning model comparison 154
4.4.2 Hypothesis test 156
4.5 Discussion 161
4.6 Conclusion 163
Chapter 5: Conclusion 165
List of References 169

List of Tables

Table 1: Financial Market Forecasting Studies: Design differences 14

Table 2: Summary of the Experimental Setup 21

Table 3: Summary of the Financial Market Data Set...23

Table 4: Covariates Employed in the Financial Time Series Forecasting Models 26

Table 5: Candidate Values for SVM and ANN Meta-Parameters 31

Table 6: Regression Analysis of Predictive Accuracy – ML techniques....................................41

Table 7: Regression Analysis of ROI – ml techniques 43

Table 8: Regression Analysis of Predictive Accuracy – SVM vs. AR 44

Table 9: Regression Analysis of ROI – SVM vs. AR...45

Table 10: REGRESSION ANALYSIS OF SLIDING WINDOW SIZE 49

Table 11: WORLD BANK, IMF and MSCI MARKET CLASSIFICATIONS 51

Table XII Spread Trading Data Summary 83

Table XIII: Risk-taking Behavior Related to Experience 101

Table XIV: Trading Frequency Related to Experience – Sensitivity Analysis..........................103

Table XV Risk-taking Behavior Related to Experience – Survivorship Analysis....................105

Table XVI: Performance Related to Experience..108

Table XVII: Performance Related to Experience – Heckman 2-stage Method 112

Table XVIII: Performance Related to Experience – Sensitivity Analysis 114

Table XIX: Performance related to Experience – Survivorship Analysis 117

Table XX: Volatility of Returns Related with Risk-taking Behavior 119

Table XXI: Risk-taking Behavior Related to Experience (Trade Number)..............................127

Table XXII: Performance Related to Experience (Trade Number) 129

Table 23: Behavioral characteristics estimated by EWA model 139

Table 24: Descriptive summary of data 143

Table 25: Model Fitness in Five Scenarios 154

Table 26: Regression Result – Estimated Parameters of EWA Models on Trading Behavior .. 156

List of Figures

Figure 1: PREDICTION PERFORMANCE IN ACCURACY (LEFT) AND ROI (RIGHT) ACROSS EXPERIMENTAL FACTORS. 36

Figure 2: PREDICTION PERFORMANCE IN ACCURACY ACROSS FINANCIAL MARKETS 37

Figure 3: PREDICTION PERFORMANCE IN ROI ACROSS FINANCIAL MARKETS........38

DECLARATION OF AUTHORSHIP

I, MING-WEI HSU..[please print name]

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

LEARNING TO BE SMART: CAN HUMANS LEARN TO IMPROVE PROFITABILITY AND RISK CONTROL IN FINANCIAL TRADING? ....

....

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as: :
   Hsu, M.W., Lessmann, S., Sung, M.C., Ma, T. and Johnson, J.E., 2016. Bridging the divide in financial market forecasting: machine learners vs. financial economists. *Expert Systems with Applications*, *61*, pp.215-234.

Signed: ....

Date: ....

Acknowledgements

I would like to thank my supervisors, Prof. Johnnie Johnson, Prof. Ming-Chien Sung, and Dr. Teijun Ma, for their patient guidance, encouragement and continuous support of my Ph.D. study. I would also like to thank Prof. Stefan Lessmann for his suggestions and guidance in machine learning techniques.

Profound gratitude goes to Dr. Ian Dawson and Dr. Mee Chi So, who are the members of my upgrade committee and gave me valuable suggestions. Especially, Dr. Ian Dawson provided very detailed opinions and brilliant insights regarding the contribution of my papers.

Special mention goes to Dr. Peter Fraser-Mackenzie and Juan Carlos Moreno-Paredes for their support and suggestions in data analysis.

Finally, I would like to thank the Ministry of Education, Taiwan for providing the sponsorship which allows me to undertake this research.

Chapter 1: Introduction

Eugene F. Fama and Robert J. Shiller were awarded the Nobel prize in economics in 2013 for their contribution to our knowledge of asset pricing. Fama introduced the efficient market hypothesis (Fama, 1970) , and Shiller is one of the most prominent critics of the hypothesis. The fact that two economists with opposing views share the Nobel prize suggests that the debate concerning the EMH is still raging.

In the center of the debate is the assumption of rationality. Evidence from both experimental and empirical research supports the view that individuals can learn from experience. However, the manner in which individuals learn and improve their decision making remains a matter of debate. Specifically, if individuals can learn from experience to appropriately adjust their probability estimates and finally make rational, correct decisions, the rationality assumption would be justified (Charness and Levin, 2003; Chiang et al., 2011) . Experimental results generally question the point of view, and empirical results are relatively few(Roth and Erev, 1995). Hence, it is important to develop empirical evidence concerning the learning process.

The well-known implication of the EMH, that stock prices cannot be accurately predicted, faces challenges from the outside of the traditional finance field: recent machine learning studies report promising directional prediction accuracy. Machine learning techniques may be used to achieve better profit than human traders since traders are subject to irrational behavior which is the nature of human being and cannot be totally avoided. The adoption of machine learning techniques among financial institutions may change the structure of the markets, and hence the investigation of the impact on financial theories is needed.

The aim of this thesis is to examine, via empirical evidence, the nature of learning by traders and the extent to which machine learning techniques may be used to question some of the underlying assumptions of the EMH. First, I run a comprehensive forecasting simulation on the major stock indexes around the world to examine the challenge to efficient market hypothesis posted by the high prediction accuracy of machine learning studies. Second, I examine the learning process of individual traders from a risk perspective and discuss the implication of the results on the two fundamental elements in neoclassical economic theories: rationality and risk aversion. Third, I will apply learning models, which in the past have only been tested with experimental data, on empirical trading data, in order to provide empirical evidence to add to the debate of whether individuals learn to be rational.

1.1. Background

The term ‘efficient market’ was introduced by Fama (1970), and Samuelson (1965) provides theoretical support for the randomness of stock returns. The name of the hypothesis itself caused controversy, since efficiency in economics used to refer to the Pareto optimality, an optimal state of resource allocation. However, the EMH provides an analytical framework for asset pricing in financial markets, and asset pricing models are developed based on the concept.

The first building block of the EMH is rationality. All market participants can be represented by rational agents who make decisions rationally based on all available information to maximize personal welfare. The second building block of the EMH is that all information is free and available to all participants immediately in the market. Thus, prices are determined by rational agents with all information taken into account, and all information is reflected in prices. When a new event happens, the new information becomes available to every rational agent immediately, and the new prices are determined in a very short time. Since all prices reflect all information at all time, the only way to beat the market and obtain excess returns is to foresee new events. Proponents of the EMH argue that it is impossible to systematically foresee new events, and, consequently, it is impossible to predict prices and to obtain excess returns.

The most obvious criticism of the EMH is associated with the assumption of the rational agents. Several behavioral biases, which are departures from rational behavior, are documented, such as overreaction (Bondt and Thaler, 1985), loss aversion (Kahneman and Tversky, 1979; Odean, 1998a), over-confidence (Gervais and Odean, 2001) and disposition (Shefrin and Statman, 1985). The evidence suggests that individual participants in financial markets are not always rational, and the aggregated opinion on prices does not necessarily reflect all information as the EMH predicts. The other assumption of the EMH, that everyone can obtain free and timely information is unrealistic for many market participants. Grossman (1976) and Grossman and Stiglitz (1980) claim that there is no reason to gather information and trade in an efficient market, so such markets will eventually collapse. That is, it is impossible for perfectly efficient markets to exist.

The major implication of the EMH is generally considered to be that prices cannot be predicted in an efficient market since all information should already be reflected in prices. However, several market anomalies, represented by patterns that can help predict prices, have been documented. These include the January effect (Keim, 1983), the weekend effect (French, 1980) and the momentum effect (Jegadeesh and Titman, 1993). Certain characteristics of organizations, such as firm size, price-to-earning and price-to-book ratios, have also been found to be helpful in predicting stock earnings and prices (Fama, 1991). In addition, irrational market phenomena, such as the Internet bubble and the sub-prime financial crisis, post strong challenges to the EMH.

Currently, most advocates of the EMH agree that not all market participants are rational and pricing irregularities can happen for short periods (Malkiel, 2003). It is widely accepted that a perfectly efficient market cannot exist. Otherwise there would be no incentive to trade (Grossman and Stiglitz, 1980). However, it is argued that the EMH still describes the markets quite well, that the surprising phenomenon described above seems to be an exception and that predictable patterns of prices will not persist for a long time. That is, even if prices are partially predictable, there is no systematic approach to obtain excess returns from the market in the long run (Beechey et al., 2000).

However, promising directional prediction accuracy is reported in recent machine learning studies, which may allow the excess return to be earned. I perform a comprehensive simulation with state of the art machine learning techniques in the first paper to clarify whether the high accuracy could be achieved easily. I then examine the rational agent assumption of the EMH in the second paper from a learning perspective. Specifically, I investigate the relationship between experience and performance measures with trading data as a rational agent should follow Bayesian learning – i.e., individuals learn from experience to appropriately adjust their probability estimates and finally make rational and appropriate decisions. In the third paper, I go one step further and propose a method to estimate the rationality level of an individual from trading data. There is a psychological learning model which can be used to estimate the rationality level from laboratory experiments, but this model cannot be used with trading data due to the difference between the laboratory environment and the real world. I propose a systematic approach to resolve this problem.

1.2. Challenges to the EMH from Machine Learning (First Paper)

Despite the advocates of the EMH suggesting that stock prices cannot be accurately predicted systematically, the challenge has been taken up in the Machine Learning literature. In particular, a number of machine learning techniques have been used to try to predict stock trends or prices, including support vector machines (SVMs) and artificial neural networks (ANNs) (George S. Atsalakis and Valavanis, 2009). Recent studies report promising directional prediction accuracy. For example, Pan et al. (2005) obtain 80% accuracy, de Oliveira et al. (2013) report 86% accuracy, and Patel et al. (2015) achieve 90% accuracy in predicting the direction of the stock market. Such high prediction accuracy is likely to enable the creation of excess returns. Clearly, such a result would not be consistent with the EMH implication that no excess return can be obtained in a systematic manner.

To shed light on the reasons for the apparent contradiction between the EMH and the machine learning studies, I conduct a comprehensive benchmark study with all combinations of the factors, which allows me to investigate the influence of factors on market predictability. In this, I explore the influence of five factors which may affect prediction accuracy, including market maturity, prediction horizon, simulation methodology, the usage of technical indicators and machine learning techniques. In particular, I assess the influence of the factors on the predictability of different financial markets and the profitability of model-based trading in these markets using a dataset of 34 financial time series, which cover most major stock markets in the world, on both a daily and hourly basis. To my best knowledge, this is the first study to compare machine learning prediction models across nearly all important stock indexes over the globe on both intraday and daily basis.

While most previous ML studies examine prediction accuracy in one or two financial markets on one prediction horizon, I examine prediction accuracy across a large number of financial markets around the world on both a daily and hourly basis. The large-scale and breadth of the simulation experiments allow me to compare the influence of five key experimental factors on market predictability. This, in turn, allows me to explore the real implications of the ML-based financial market prediction studies for the financial economists' view of market efficiency. Specifically, a bridge between the ML and the financial economics is attempted.

The results of the extensive forecasting experiments suggest that certain factors associated with previous ML studies are influential on the prediction accuracy, including market maturity, forecasting method, prediction horizon and machine learning techniques. I find that prediction accuracy and trading profit is higher in mature markets than that in emerging markets. Daily prediction accuracy is higher than hourly, and SVM generates higher accuracy than ANN. I find that the inclusion of technical indicators has insignificant influence on the prediction accuracy.

My findings provide evidence for the view that stock prices are partly predictable: the prediction accuracy of the ML models is significantly higher than random guesses in my experiments. This is consistent with the previous ML studies that ML techniques, such as SVM and ANN, are useful in predicting financial markets. However, the level of accuracy in my experiments (most are below 60 percent), is substantially lower than that commonly published in previous studies. A possible explanation is that I use naive SVM and ANN models and a basic set of covariates to preserve the reproducibility of my experiments. On the other hand, my results indicate the importance of methodological factors regarding the prediction performance. Certain selections of factors, such as market maturity and prediction horizon, can be significantly influential on prediction accuracy. Hence, previous ML studies may have drawn conclusions from over-optimistic prediction performance and, consequently, may underestimate the degree of market efficiency more generally.

I fail to find evidence inconsistent with the EMH which suggests that stock prices are partly predictable and there is no systematic approach to obtain excess returns. Although the simple trading simulations (which always follow the generated predictions) generally achieve positive return rates, the returns mostly turn negative after transaction cost is taken into account. This suggests that the high prediction accuracy does not guarantee excess returns in financial markets due to the cost of arbitrage (Malkiel, 2003).

1.3. Is experience the mother of risk wisdom? (Second Paper)

Empirical studies show that individuals can learn through time to improve their profitability and reduce their disposition effect (a behavioral bias describing the reluctance to sell stocks in loss) (Feng and Seasholes, 2005; Nicolosi et al., 2009; Seru et al., 2010). However, previous studies have failed to account for the risk accepted by these individuals. The risk return trade-off suggests that higher returns are often accompanied by higher risk (Fama and MacBeth, 1973; Glosten et al., 1993). Consequently, previous results which suggest that individuals learn to improve their performance through time may be misleading. It may simply be that individuals engage in riskier behavior through time and this pushes up their returns. Hence, I examine the extent to which traders learn through time/experience by examining the performance improvement of traders as measured by their Sharpe ratio, a widely used risk adjusted return measure in the economic and finance literature (Sharpe, 1998). To my best knowledge, this is the first study to investigate the effects of learning on investment performance with empirical trading data and individual level analysis, taking account of their volatility of returns and risk-adjusted performance.

The results of this study show that traders increase their returns through experience (measured in terms of the time since their first participation in the market). This result is in line with the literature (Feng and Seasholes, 2005; Nicolosi et al., 2009; Seru et al., 2010). However, I find that the volatility of their returns also increases and their risk adjusted performance decreases as they gain more experience. The results imply that traders improve returns at the expense of taking higher risk and, consequently, obtain lower risk-adjusted performance. This also shows that a single performance measure could lead to misleading conclusions regarding performance improvement.

The rationality assumption of the EMH is based on the notion of ‘Bayesian learning,' and these results cast doubt on Bayesian learning which suggests that through time individuals are better able to estimate probability conditioned on new information and control trading risk. From Chiang et al. (2011)’s point of view, the implication of Bayesian learning is the performance improvement through experience. However, I find that traders have increasing returns and decreasing risk-adjusted performance, which is not consistent with the implication. Thus, the evidence does not lend support Bayesian learning.

On the other hand, reinforcement learning seems to provide a reasonable explanation for my findings. Reinforcement learning suggests that individuals update the likelihood of strategy selection according to the rewards of the previous actions. In this case, traders may use a number of trading strategies with different levels of risk. The riskier strategies generally lead to higher profit and higher volatility. Those traders using riskier strategies are more likely to have higher returns and are reinforced by the riskier strategies. The finding that traders increase their risky behavior through experience provides further support for this explanation.

My findings suggest that traders follow reinforcement learning rather than Bayesian learning, and traders tend to pursue returns while underestimating the underlying risk. The implication is that traders are not aware of the importance of trading risk, and lack of the sense of trading risk could be harmful to both traders and markets. As a result, further intervention from financial authorities is needed to remind traders the potential risk.

I find a notable change of risk-taking behavior is exhibited through experience. Traders intensify their risky behavior when they gain more experience, and this leads to higher volatility of returns. The implication of my results is that risk preference is affected by personal accumulated experience, in addition to total wealth (loss aversion) and prior outcomes (house money effect). Therefore, experience should be accounted for in considering risk preference, since even moderate changes in risk preference can lead to substantial volatility of asset prices compared to the level of the underlying consumption variability (Allen and Gale, 1994; Mehra and Sah, 2002).

1.4. Estimating Behavioral Characteristics associated with Learning Models in Financial Markets (Third Paper)

Individual decision making has been a fundamental question in economic and finance. Recently, emerging empirical evidence suggests that trading behavior is affected by several factors, such as past retunes, personal experience and sophistication (Chiang et al., 2011; Choi et al., 2009; Glaser and Weber, 2009; Kaustia and Knüpfer, 2008; Y.-J. Liu et al., 2010; Nicolosi et al., 2009; Seru et al., 2010; Thaler and Johnson, 1990). On the other hand, trading data from financial markets provides very limited information of individual characteristics, such as age, gender and career, in addition to trading information. However, certain individual behavioral characteristics that are less readily discernable have been shown to affect how individuals learn to adjust their behavior in psychology experiments, which leads to that the effect of these behavioral characteristics has been largely neglected in empirical studies. One of such behavioral characteristics is the extent to which an individual behaves like a rational agent, i.e., the level of rationality, which can be estimated with behavioral models in laboratory experiments. But the estimation of rationality level cannot be easily done with the trading data in the real world. To shed light on the effect of these behavioral characteristics, we develop a methodology which allows us to determine, using empirical data, some important behavioral characteristics of individual traders associated with learning models and we examine to what extent these influence the individual's trading behavior and performance.

We estimate the behavioral characteristics of traders used as parameters in a behavioral learning model (i.e., the experience weighted attraction (EWA) model). Thanks to comprehensive behavioral experiments conducted in the psychology field, behavioral models are proposed to describe how individuals make decisions after receiving feedback from their actions in the past (Camerer and Ho, 1999). The most important behavioral characteristic is the weight of foregone payoff which is used to control how unchosen strategies are reinforced. A strategy with a superior payoff in the past is reinforced and is more likely to be chosen in the future. The weight of foregone payoff is the key to determine the extent to which an individual behaves like a rational agent and can be regarded as the measure of rationality.

Nonetheless, it is problematic to apply behavioral models on real trading data: subjects participating experiments usually select one option among a few alternatives given all relevant information (i.e., strategy space of subjects is limited), while traders in financial markets face an unlimited strategy space and have to make a series of decisions, including the market change direction, the stake size and when to open and close a position, with a number of sources of public market information which may or may not be relevant to the performance of the traders.

Therefore, I propose the decision-based strategy mapping framework (DSM) to help solve this problem of unlimited strategy space. DSM creates the concept of 'possible scenarios.' In each scenario, individuals are assumed to have only one decision to make. For example, in a 'buy/sell' scenario, the traders are assumed to believe the decision of choosing to buy or sell is the one thing that will affect their eventual return. In this way, DSM creates a limited strategy space, which allows me to apply behavioral models on the trading data in financial markets. Subsequently, I am able to assess to what extent a model actually fits the trading data; this also enables me to assess which of the scenarios best represents the factors that traders actually consider when placing their trades. Furthermore, with DSM, I am able to use an individual's trading data to estimate their behavioral characteristics (estimated with EWA) and the manner in which they learn for their previous experience; such as the weight they place on foregone payoffs. I then examine the relation between the behavioral characteristics of individual traders and their behavior and trading performance.

The results show that traders believe that their decisions on stake size are influential on their profit point, i.e., the percentage profit they earn. I also find that traders who put greater weight on foregone payoffs for unchosen strategies (i.e., opportunity costs) tend to place higher stake sizes in average and trade more frequently. As rational economic agents are generally considered to actively consider opportunity costs when making decisions, the results suggest that those traders who behave like rational agents tend to trade more actively. It is important to notice that those traders who are more rational do not achieve better trading performance. In addition, traders who start trading with a preference for lower stake sizes achieve overall better trading performance (higher total profit and lower volatility of returns).

1.5. Contribution

In summary, I conduct comprehensive simulation experiments with machine learning techniques in predicting the change direction of stock index in the first paper, and the major contribution is:

- Providing evidence for that the methodological issues might be the source of the difference between the prevailing view in support of the EMH in the financial economics literature and the high prediction accuracy of stock prices reported in ML studies.
- Providing empirical evidence for the prediction ability of ML techniques, such as SVM and ANN, to recognize the patterns of market anomalies across major financial markets.

In the second paper, I analyze spread trading data on an individual level to examine the change of both behavior and performance through experience from a risk perspective. The major contribution is:

- Risk taking behavior is also affected by accumulated experience, in addition to past returns and total wealth.
- Traders increase returns by taking higher risk, and this leads to lower risk adjusted performance.
- A Higher level of risk taking behavior does not necessarily lead to higher volatility of returns.

In the third paper, I estimate individual behavioral characteristics with empirical trading data by incorporating the proposed DSM framework and the EWA learning model. The major contribution is:

- Proposing DSM to overcome the problem of unlimited strategy space in applying behavioral models to real trading data.
- Providing evidence for the relation between trading performance and behavioral characteristics.

1.6. Structure of Thesis

This thesis is structured as follows. In Chapter 2, I present the result of comprehensive forecasting simulation to clarify the factors leading to the disagreement between the EMH and the machine learning studies. In Chapter 3, I examine the investment performance from a risk perspective when traders gain experience. In Chapter 4, I propose the decision-based strategy mapping framework to estimate unobservable individual characteristics which are shown to be influential on trading behavior and performance. I conclude in Chapter 5.

Chapter 2: Bridging the Divide in Financial Market Forecasting: Machine Learners vs. Financial Economists.

Abstract— Financial time series forecasting is a popular application of machine learning methods. Previous studies report that advanced forecasting methods predict price changes in financial markets with high accuracy and that profit can be made trading on these predictions. However, financial economists point to the informational efficiency of financial markets, which questions price predictability and opportunities for profitable trading. The objective of the paper is to resolve this contradiction. To this end, we undertake an extensive forecasting simulation, based on data from thirty-four financial indices over six years. These simulations confirm that the best machine learning methods produce more accurate forecasts than the best econometric methods. We also examine the methodological factors that impact the predictive accuracy of machine learning forecasting experiments. The results suggest that the predictability of a financial market and the feasibility of profitable model-based trading are significantly influenced by the maturity of the market, the forecasting method employed, the horizon for which it generates predictions and the methodology used to assess the model and simulate model-based trading. We also find evidence against the informational value of indicators from the field of technical analysis. Overall, we confirm that advanced forecasting methods can be used to predict price changes in some financial markets and we discuss whether these results question the prevailing view in the financial economics literature that financial markets are efficient.

Keywords: Financial time series forecasting, market efficiency, machine learning

Chapter 2

Research Highlights

- We perform a comprehensive benchmark in financial time series forecasting
- We determine if the best machine learning methods predict financial time series more accurately than the best econometric methods
- We examine the impact of forecasting methodology on the predictability of financial time series
- We clarify the influence of market maturity, forecast horizon, model simulation methodology, prediction method, and technical indicators on market efficiency

2.1 Introduction

Financial markets facilitate international trade, aggregate, discount and convey information concerning the future prospects of organizations and economies, and are enablers of economic growth. Given their importance, financial markets have attracted much research, including the modeling of market prices. Applications of such prediction models include the management of financial risks (W.-S. Chen & Du, 2009) and the support of investment decisions (Chang et al., 2009; de Oliveira et al., 2013; Huang et al., 2008).

It is common practice to demonstrate the merit of a novel machine learning (ML) model for time series forecasting using financial market data such as stock or commodity prices, currencies, or financial indices. We survey such experiments in our literature review and show that many studies report high levels of forecasting accuracy. Some studies also identify opportunities to trade profitably on model predictions (Chang et al., 2009; de Oliveira et al., 2013; Doeksen et al., 2005; Huang et al., 2008; Huck, 2010; Kara et al., 2011; Schumaker and Chen, 2009). However, few of these studies discuss the implications for market efficiency (see Table 1).

The success of financial time series forecasting demonstrated by ML experiments is surprising given the theory and evidence from the financial economics literature. More specifically, the efficient market hypothesis (EMH) suggests that current stock prices discount available information and that it is not possible to obtain systematic returns by exploiting any predictability of prices (Malkiel, 2003). This theory is widely supported by financial economists (Fama, 1970, 1991). For example, Jensen (1978, p. 96) claims that "there is no other proposition in economics which has more solid empirical evidence supporting it."

Our research is motivated by the need to develop an understanding of the reasons for the disagreement between the EMH and empirical evidence in the ML literature. Given the prominence of the EMH, for example in the design and regulation of financial markets, such understanding is important. In addition, the paper helps fill a research gap, as cross-fertilization between the ML and financial economics literature is limited. For example, the ML community relies on advanced, data-driven forecasting methods in the form of, for example, support vector machines (SVM) or artificial neural networks (ANN). Such methods are rarely considered by financial economists who prefer econometric, often linear methods (Campbell and Thompson, 2008; Fama and French, 1993, 2012). Contradictory findings may thus result from different modeling cultures. In addition, few ML studies discuss their empirical findings in the light of the EMH. The feasibility of predicting a financial market with an accuracy above 80 percent seems to be taken for granted in the ML literature, although such figures suggest significant inefficiency, with important implications for the use of financial markets as effective allocative mechanisms. In summary, the objectives of this study are to: (i) examine the magnitude of disagreement between previous findings of the ML and financial economics literature, (ii) investigate the experimental factors in ML studies which may help explain this discrepancy, and (iii) explore the implications for studies examining financial market efficiency.

To achieve our objectives, we compare the predictive accuracy of the most widely used ML and econometric methods and find that the best ML methods outperform the best econometric methods. In doing so, we extend previous ML studies in financial time series forecasting (de Oliveira et al., 2013; Kara et al., 2011) by performing an independent evaluation of ML methods (i.e., we do not propose a new method, which may bias an evaluation (Hand, 2006)). This enables us to examine the degree to which previous findings in the ML literature (i.e., high forecasting accuracy) generalize to novel experimental conditions. For example, unlike most previous studies, which examine forecast accuracy and resulting trading profitability in *one* financial market, either within a day or across trading days, we examine forecast accuracy and trading profitability across a *large number* of financial markets. We then compare the performance of forecasts of market prices *within a day and across trading days*, which allows us to make three contributions: First, we assess the degree to which factors associated with forecasting methodology (e.g., forecasting method and horizon, etc.) affect predictive accuracy and trading profitability. Second, our results enable us to provide guidance on how to organize benchmarking experiments in financial time series forecasting and to identify the origins of disagreement between the ML and the financial economics literature. Third, we examine the implications that result from ML-based forecasting studies should have on the financial economists' prevailing view of widespread market efficiency. In particular, we attempt to build a bridge between the ML and the financial economics domains. This is important because advances in either domain are rarely reflected in the other. Thus, our study is a first step toward unlocking the potential for collaborative gains.

The remainder of the paper is organized as follows: In the following section, we review related work in the financial and ML literature. Next, we develop six hypotheses concerning how different methodological factors affect predictive accuracy. We then elaborate on our experimental design, before presenting empirical results and discussing their implications. We conclude with a summary of our main findings.

2.2 Related Work

2.2.1 The Efficient Market Hypothesis

It is widely believed that financial markets are efficient in aggregating diverse sources of information concerning an asset's future prospects. In particular, the EMH states that asset prices reflect all relevant information and it is impossible to generate excessive returns through 'more informed' (e.g., model-guided) investment decisions (Fama, 1970). Milder versions of the EMH relax the assumption of prices incorporating all relevant information: semi-strong and weak form efficient markets being ones where prices reflect all publicly available information and historic price information, respectively.

The EMH is based on the belief that market participants make rational decisions and that any mis-pricing will quickly be eliminated by those seeking to gain from these anomalies. However, persistent evidence of irrationality has been identified, such as overreaction (Bondt and Thaler, 1985) and the disposition effect (Dhar and Zhu, 2006; Grinblatt and Keloharju, 2001a; Shefrin and Statman, 1985). Furthermore, persistent market anomalies, such as the January effect (Keim, 1983) and the weekend effect (French, 1980), have been observed and these are difficult to explain in terms of the EMH. In addition, Lo et al. (2000) show that technical analysis can be used to predict stock movements.

Proponents of the EMH argue that market efficiency is a simplification that may not always hold true, but will on most occasions and for most investors. Equally, they argue that anomalies may appear, but will disappear when they become known to the market. In addition, the ability to predict certain market prices does not imply that these can be exploited to earn an excess profit (Malkiel, 2003). Summing up the literature, Fama (1998) indicates that although some studies appear to identify market inefficiency, the EMH cannot be rejected unless three elements are present: endurance - the inefficiency should survive in the long term; homogeneity - the inefficiency should be apparent in the same form across different markets; robustness - the methodology used to demonstrate the inefficiency should be sufficiently robust to provide confidence that the inefficiency exists. It is in this light that we examine the literature which employs ML for financial market prediction.

2.2.2 Price Prediction in Financial Markets using Machine Learning

Despite financial economists' widespread belief in the veracity of the EMH, a large body of ML literature examines the predictability of financial market prices and the profitability of model-based trading (see Table 1). For example, financial market forecasting is a popular application domain to develop new modeling methodologies and to demonstrate their potential. The prevailing approach is to develop dynamic regression models, which predict future market prices on the basis of past movements in those prices and other price time series (e.g., stocks, indices, currencies, etc.).

TABLE 1: FINANCIAL MARKET FORECASTING STUDIES: DESIGN DIFFERENCES

Study	Modeling subject	Market	Forecasting horizon	Prediction Method		Dynamic simulation	Technical indicators	Subject of prediction			Result discussed re EMH
				SVM	ANN			Change direction	Price change	Trading Strategy	
Ormoneit and Neuneier (1996)	Index	German	Hourly		X						
Steiner and Wittkemper (1997)	Stock	German	Daily		X	YES				X	YES
Kuo et al. (2001)	Index	Taiwan	Daily		X		YES			X	
Chen et al. (2003)	Index	Taiwan	Monthly		X		YES			X	
Mittermayer (2004)	Stock	U.S.	Hourly	X						X	
Zhang et al. (2004)	Index	China	Daily		X		YES			X	
Vanstone et al. (2005)	Stock	Australia	Daily		X		YES			X	
Pan et al. (2005)	Index	Australia	Daily		X		YES	X (80%)			
Huang et al. (2005)	Index	Japan	Weekly	X				X (75%)			
Armano et al. (2005)	Index	U.S.	Daily		X		YES			X	
Chen and Ho (2005)	Index	Taiwan	Daily	X			YES		X		
Doeksen et al. (2005)	Stock	U.S.	Daily		X		YES			X	
Bodyanskiy and Popov (2006)	Index	U.S.	Daily		X			X (72%)	X		
Qian and Rasheed (2007)	Index	U.S.	Daily		X	YES	YES	X (65%)			

Hassan et al. (2007)	Stock	U.S. stocks	Daily		X		YES		X		
Tseng et al. (2008)	Index	Taiwan	Daily		X		YES		X		
Huang et al. (2008)	Index	Taiwan, Korea	Daily	X	X		YES	X (80%)			
Schumaker and Chen (2009)	Index	U.S.	Daily	X						X	
Zhang and Wu (2009)	Index	U.S.	Daily		X		YES		X		
Chang et al. (2009)	Stock	Taiwan	Daily		X		YES			X	
Lee (2009)	Index	U.S.	Daily	X			YES	X (87%)			
Tsai and Hsiao (2010)	Stock	Taiwan	Season		X	YES		X (78%)			
Hadavandi et al. (2010)	Stock	U.S.	Daily		X				X		
Huck (2010)	Stock	U.S.	Weekly		X		YES			X	YES
Bollen et al. (2011)	Index	U.S.	Daily		X				X		YES
Kara et al. (2011)	Index	Turkey index	Daily	X	X		YES	X (75%)			
Dai et al. (2012)	Index	Taiwan, Hong Kong, Japan	Daily	X			YES		X		
de Oliveira et al. (2013)	Index	Brazil index	Monthly		X		YES	X (86%)			
Ballings et al. (2015)	Stock	Europe	Year	X	X		YES	X (70%)			
Patel et al. (2015)	Index	India	Daily	X	X		YES	X (90%)			

* Ormoneit and Neuneier (1996) predict hourly volatility.

This table lists the studies using machine learning techniques in stock markets. The modeling subject is the target of the models and can be the price of an individual stock or the index price of a stock market. The market is the region of the stock market. The dynamic simulation is marked with an ‘X’ if the sliding window method is used. The technical indicator is marked with an ‘X’ if the technical indicators are included.

A researcher faces many degrees of freedom in organizing a forecasting study. We believe that the specific choices made may have a sizeable effect on the observed results. Consequently, previous findings related to price predictability are best appraised in the context of the specific experimental conditions under which they were obtained. To that end, we review previous studies and their experimental design in Table 1. The common denominator is that all these studies employ some advanced prediction method to forecast price developments in some financial market. In particular, we concentrate on studies that consider support vector machines (SVM) and/or artificial neural networks (ANN). Due to their ability to recognize patterns in nonlinear, dynamic time series data (Chang et al., 2009; Lee, 2009; Żbikowski, 2015), these methods are especially popular in financial market forecasting. Additional factors, other than the prediction method, considered in Table 1 are the financial instrument that is being forecast (individual stocks and indices), the geographic market, the forecasting horizon, whether the study uses a static or dynamic approach to simulate model-based trading, and whether it uses technical indicators among the covariates. We further motivate these factors when developing our hypotheses in Section 3.

The information in Table 1 suggests that previous studies predominantly seek to forecast financial indices in Asian markets and the Taiwanese market in particular, although US markets also occur frequently. Only six out of 28 studies consider other markets (e.g., Australian, German). Overall, ANN is the most popular prediction method, and only four studies consider both ANN and SVM. One of the interesting findings is that very few studies evaluate prediction models in a dynamic fashion. Rather, the prevailing approach, used in 25 out of 28 previous studies, is to split a financial time series into a training and a hold-out test set. We refer to this approach as ‘static’ because it uses the same prediction model throughout the whole testing period, without updating. A dynamic approach such as sliding-window cross-validation performs model training and evaluation multiple times using smaller chunks of sequentially ordered data (Lessmann et al., 2011).

The merit of technical analysis is hotly debated in the financial economics literature (Fama, 1970; Lesmond et al., 2004; Lo et al., 2000). Advocates of technical analysis stress the predictive ability of technical indicators, whereas opponents refute their value. Nonetheless, technical indicators are widely used in previous ML studies, with 21 out of 28 studies using these indicators, possibly together with other covariates (e.g., financial news).

Table 1 also summarizes previous studies according to whether they: (a) predict the direction of price movements (e.g., rise or fall); (b) predict actual price changes; (c) examine whether investing according to model predictions would produce a profit; (d) discuss their findings in the light of the EMH.

Predicting price direction/size of changes involves forecasting a discrete target variable (i.e., classification)/continuous target variable (i.e., regression). In the former case, Table 1 also reports the directional accuracy, which can be as high as 90 percent (Patel et al., 2015). Clearly, a model that forecasts the direction of price/market movements correctly in 90 out of 100 cases, say days, would facilitate enormous profits, contradicting the view that financial markets are efficient. The directional accuracies shown in Table 1, therefore, illustrate the disagreement between (several) ML forecasting studies and studies in the financial economics literature. If the results of the ML studies are reproducible across markets in the long term, the excess returns will be strong evidence against the EMH and most asset pricing models. However, this implication is discussed in only 3 of 28 studies (see Table 1). Our objective is to identify the factors that explain this disagreement.

2.3 Hypothesis Development

To shed light on the origin of differences in findings concerning market efficiency in the finance and ML literature, we develop a series of hypotheses. These examine the extent to which experimental factors in ML studies affect forecast accuracy and trading profit derived from forecasts and whether the best performing ML technique outperforms the best performing econometric modelling procedures.

The financial economics literature suggests that informational efficiency differs between established and emerging financial markets (Griffin et al., 2010; Ojah and Karemera, 1999); established markets being more efficient and, thus, more difficult to predict. One argument to support this view is regulation. For example, in a highly regulated market, governmental institutions create and enforce detailed rules concerning the release of new information that might affect a company's stock prices. In general, these rules enforce greater and wider disclosure of relevant information. Consequently, capitalizing on private information becomes illegal and less likely.

The relevance of regulation for our study is that selecting a particular financial market for analysis might predetermine the level of predictive accuracy. Specifically, a selection bias toward easier to predict/less efficient markets could explain the high forecasting accuracy reported in the ML literature. However, this is a valid explanation only if predictability and market efficiency do indeed differ with market maturity. To clarify this, we test:

H1a: Predictive accuracy is higher in emerging cf. mature markets, and

H1b: Model-based trading gives higher profits in emerging cf. mature markets.

Another explanation for observing high accuracy/profits in forecasting studies could be the methods used for model assessment. Most ML stock price forecasting studies use a static approach to evaluate model performance (see Table 1), where predictions are generated by a model that is trained with a fixed set of samples. This enables a large number of samples to be included in the training set, which increases the opportunity to recognize price patterns. The static approach is based on the assumption that stock markets exhibit long-term memory (Lo, 1991). However, it disregards the latest price information, which never enters the (single) training set. Sliding-window cross-validation, on the other hand, evaluates a model from multiple origins of a financial time series (i.e., trading periods) and thus uses recent samples for model training. We compare the static approach and sliding-window cross-validation to examine the effect of model assessment methods on model performance.

A static model evaluation paradigm might overestimate the performance of a prediction model. For example, by selecting a single evaluation period (i.e., the test set), the static approach suffers the risk of picking a *lucky sample*. Consider for example the years 2008/2009. Global stock prices all over the world fell significantly in response to the subprime credit crises. In such a period, a naïve forecasting model, which simply predicts price decreases, would display high accuracy, although it does not embody any predictive insight. The risk of over-estimating the value of a model is far less using sliding-window cross-validation, since this repeats the evaluation of a prediction model multiple times. Consequently, we test the following hypotheses:

H2a: Predictive accuracy is higher under a static model evaluation approach, and

H2b: Model-based trading gives higher profits under a static model evaluation regime

The predictive performance of any forecasting model depends largely on the degree to which the covariates are correlated with the target variable. Many financial market forecasting models are auto-regressive models, where the covariates refer to past realizations of the target variable. For example, to predict the closing price of stock X on day *t*, a model might use the closing prices of X at day *t*-1, *t*-2, etc. Indicators from the technical analysis are another type of covariate. They apply additional transformations of prices with the goal of creating more informative variables. Some practitioners and analysts argue that certain transformations of the prices and trading volumes in the past could generate additional information that is helpful in predicting the prices in the future. These transformations are known as technical indicators. The value of technical indicators remains an open question. For example, Lo et al. (2000) find some indicators with practical value. Similarly, Brock et al. (1992) conclude that trading strategies that rely on technical indicators can produce excess returns. However, other studies argue that technical indicators have little predictive value and do not facilitate profitable trading, (e.g., Fama, 1970; Lesmond et al., 2004).

Many ML studies employ technical indicators, but several studies do not, relying rather on raw prices or simple price differences (e.g., closing price – opening price) (see Table 1). Informative covariates play a key role in predictive modeling. The use or omission of technical indicators may, therefore, affect the degree of predictive accuracy considerably. More specifically, the financial economics literature may under-estimate the degree of inefficiency in financial markets because the predictive value of technical indicators in that literature has been underestimated. ML studies commonly use advanced data-driven models that discern nonlinear relationships and complex variable interactions automatically, but these models are not common in financial economics studies. It may be that ML studies, therefore, extract full information value from the technical indicators. To clarify the influence of technical indicators on predictive accuracy and market efficiency, we test:

H3a: Predictive accuracy is higher if a model incorporates technical indicators, and

H3b: Model-based trading gives higher profits if a model incorporates technical indicators

We also examine the degree to which the selection of a specific forecasting horizon affects predictive accuracy. In general, one expects accuracy to decrease when forecasting further ahead (Hyndman and Athanasopoulos, 2014). This view is supported by the EMH, which predicts that time may be needed for stock prices to fully reflect all available information (Fama, 1970); possibly explaining the rising interest in high-frequency trading (Chordia et al., 2013; Menkveld, 2013).

The information reported in Table 1 indicates that previous ML studies in financial time series forecasting predominantly consider forecast horizons of one day or more; possibly because financial data of such granularity is freely available online, whereas high-frequency, intraday data is not available for free. This could indicate that previous forecasting studies, albeit evidencing market inefficiency, have actually underestimated the degree of inefficiency, because they do not consider short forecasting horizons (e.g., intraday forecasting). To examine this proposition, we test the following hypotheses:

H4a: Predictive accuracy is higher for shorter forecast horizons, and

H4b: Model-based trading gives higher profits if forecasting shorter periods into the future

A number of studies have suggested that SVM outperforms ANN in terms of predictive accuracy (e.g., W.-H. Chen, Shih, & Wu, 2006; W. Huang et al., 2005; K. Kim, 2003; Ou & Wang, 2009; Tay & Cao, 2001). Consequently, to examine the effect of the ML forecasting model on predictive accuracy for financial time series and on the assessment of market efficiency, we test the following hypotheses:

H5a: Accuracy is higher for predictions based on SVM (cf. ANN), and

H5b: Model-based trading gives higher profits when using SVM (cf. ANN) for prediction

The financial economics and ML literature examining forecasting accuracy largely rely on econometric and ML methods, respectively. To examine whether the adoption of econometric *vs.* ML methods effects predictive accuracy, we perform a comparison of the commonly employed ML methods (ANN and SVM) with the commonly employed econometric models: autoregressive model (AR), generalized autoregressive conditional heteroscedasticity model (GARCH) and autoregressive integrated moving average model (ARIMA) (Charles, Darné, & Kim, 2011). According to our experiments, reported in online Appendix A, AR outperforms ARIMA and GARCH. We also find that SVM outperforms ANN. In order to assess the effect of the forecasting model (econometric vs. ML) on predictive accuracy for financial time series and on assessment of market efficiency, we test the following hypotheses:

H6a: Accuracy is higher for predictions based on the best performing ML cf. econometric method (i.e., SVM vs. AR), and

H6b: Model-based trading gives higher profits when based on predictions from the best performing ML cf. econometric method (i.e., SVM vs. AR)

The proposed set of hypotheses, enable us to examine the extent to which various factors, such as emerging market data, static evaluation setting, technical indicators, shorter forecasting horizon and selected models (e.g., SVM vs. ANN and econometric vs. ML), influence the accuracy of financial time series forecasting. Integrating such insight with information how ML studies are typically conducted (i.e., Table 1) enables us to explain apparent inconsistencies between the results of the ML and financial economics literature.

2.4 Experimental Design

We investigate the influence of five experimental factors (market maturity, model simulation methodology, covariate composition, forecast horizon and prediction method (SVM vs. ANN; best ML (SVM) vs. best econometric (AR)): see Table 2) on the predictability of price movements in financial markets and the profitability of model-based trading. We test hypotheses based on these experimental factors using a data set of 34 financial time series. Whilst this is not a large sample, it facilitates a reasonably inclusive study since most major markets in the world are examined. In addition, since most ML studies use only a single financial time series, our study is relatively large, and we feel, therefore, that the conclusions can be relied upon.

There is little discussion on the influence of the methodological factors on the prediction performance. We survey both the financial and the machine learning literature to select these methodological factors which are potentially influential on prediction performance. The first factor is market maturity since there is strong evidence from the financial literature showing the difference in market efficiency between the emerging markets and the matured markets and it is difficult to predict the movement of an efficient market (Griffin et al., 2010; Ojah and Karemera, 1999). We examine the influence of the model simulation methodology (static v.s. sliding-window) because of the general belief that stock markets exhibit long-term memory (Lo, 1991) and that the latest price information is often disregarded in the training data set. Our choice of covariate composition is from the long-lasting debate on the effectiveness of technical indicators (see Table 1). We include the forecast horizon factor as the efficient market hypothesis suggests that some time may be needed for prices to fully reflect all available information (Fama, 1970).The machine learning researchers have shown significant interest in the comparison between SVM and ANN (e.g., W.-H. Chen, Shih, & Wu, 2006; W. Huang et al., 2005; K. Kim, 2003; Ou & Wang, 2009; Tay & Cao, 2001); hence, we contribute to the literature by providing the evidence from the major markets.

More specifically, we examine all combinations of factors in a full-factorial setup. To illustrate our procedure, consider one experiment with specific choices for each of the experimental factors. For example, the experimental setup could be such that we use *SVM* to forecast the *FTSE 100* using a forecasting horizon *of one day, excluding technical indicators* from the covariates and measuring forecast accuracy using a *static evaluation approach*. This experiment produces a set of predictions for all periods (i.e., days) of the test set, from which we can calculate forecast accuracy and the profitability of model-based trading. Having completed the first experiment, we change one factor (e.g., incorporate technical indicators among the covariates), and repeat the forecasting simulation. We continue this process for all combinations of factor levels (see Table 2) and all financial time series. This enables us to compare predictability across different experiments and to examine the impact of the experimental factors. The design is similar to that of Gerlein, McGinnity, Belatreche and Coleman's (2016) study in which repeated experiments are performed to clarify the influence of several factors (e.g., the size of sliding window and the number of covariates), on the prediction performance of simple ML classifiers. In the following subsections, we discuss the factors and motivate our choices of individual factor levels.

TABLE 2: SUMMARY OF THE EXPERIMENTAL SETUP

Experimental factor	Factor levels	Hypotheses
Market maturity	High income level vs. middle income level	H1a,b
Model simulation methodology	Sliding-window cross-validation vs. a static evaluation	H2a,b
Covariate composition	Simple price covariates vs. simple price covariates combined with technical indicators	H3a,b
Forecast horizon	Daily vs. hourly	H4a,b
Prediction method	SVM vs ANN	H5a,b
	SVM vs AR	H6a,b

This table lists the levels of all experimental factors and the associated hypotheses.

2.4.1 Data, Variables, and Forecasting Horizon

We obtain our data from TickWrite Data Inc.[1] It comprises time series of financial indices from 34 markets for both established and emerging markets over a 6-year period (2008-14). An exception is the Brazilian market for which we have data for 4 years (2010-14). We included as many markets as possible to cover both mature and less mature markets since one of the aims of the study is to examine the influence of market maturity. It was important for comparability that the same period was employed for each market included in the study. This restricted the sample period because the availability of intraday data is limited and, in many markets, is only available from 2008 onwards. This forced us to choose the data period 2008 to 2014, where intraday data is available for most markets.

Table 3 summarizes the data set. We have focused on predicting national stock indices because these indices are used in the majority of previous ML studies that predict direction of price changes (e.g. Bodyanskiy & Popov, 2006; de Oliveira, Nobre, & Zárate, 2013; C.-J. Huang, Yang, & Chuang, 2008; W. Huang, Nakamori, & Wang, 2005; Pan, Tilakaratne, & Yearwood, 2005; Qian & Rasheed, 2007).

[1] http://www.tickdata.com/

TABLE 3: SUMMARY OF THE FINANCIAL MARKET DATA SET

No.	Economy	Region	Income Level	Index	Code	Start Date	End Date
1	US	America	High	S&P 500	SP	1/2/2008	19/2/2014
2	Canada	America	High	SP TSX Composite Index	TS	1/2/2008	19/2/2014
3	Japan	Asia & Pacific	High	Nikkei 225	NE	1/2/2008	19/2/2014
4	Korea	Asia & Pacific	High	KOSPI 200 Index	KM	1/2/2008	19/2/2014
5	Hong Kong	Asia & Pacific	High	Hang Seng Index	HI	1/2/2008	19/2/2014
6	Singapore	Asia & Pacific	High	Straits Times Index	ST	1/2/2008	19/2/2014
7	China	Asia & Pacific	Middle	ShangHai SE Composite Index	SH	1/2/2008	19/2/2014
8	Malaysia	Asia & Pacific	Middle	FTSE Bursa Malaysia KLCI Index	KL	1/2/2008	19/2/2014
9	Thailand	Asia & Pacific	Middle	Thai Stock Exchange MAI Securities Index	TH	1/2/2008	19/2/2014
10	Indonesia	Asia & Pacific	Middle	Jakarta Composite Index	JC	1/2/2008	19/2/2014
11	France	Europe	High	CAC 40	CF	1/2/2008	19/2/2014
12	UK	Europe	High	FTSE 100	FT	1/2/2008	19/2/2014
13	Italy	Europe	High	FTSE MIB Index	II	1/2/2008	19/2/2014
14	Germany	Europe	High	DAX	DA	1/2/2008	19/2/2014
15	Hungary	Europe	Middle	BUX	BU	1/2/2008	19/2/2014
16	South Africa	Africa	Middle	FTSE/JSE Africa Top40	TO	1/2/2008	19/2/2014
17	Turkey	Middle East	Middle	ISE-100	TU	1/2/2008	19/2/2014
18	Switzerland	Europe	High	Swiss Market Index	SW	1/2/2008	19/2/2014

19	Spain	Europe	High	IBEX 35	IB	1/2/2008	19/2/2014
20	Netherland	Europe	High	AEX	AE	1/2/2008	19/2/2014
21	Belgium	Europe	High	BEL20	BE	1/2/2008	19/2/2014
22	Portugal	Europe	High	PSI-20	PP	1/2/2008	19/2/2014
23	Sweden	Europe	High	OMX ALL-SHARE Stockholm Index	SM	1/2/2008	19/2/2014
24	Norway	Europe	High	OSE All Share Index	OL	1/2/2008	19/2/2014
25	Denmark	Europe	High	OMX Copenhagen Index	KA	1/2/2008	19/2/2014
26	Finland	Europe	High	OMXH25	HE	1/2/2008	19/2/2014
27	Austria	Europe	High	ATX	AT	1/2/2008	19/2/2014
28	Czech	Europe	High	Prague Stock Exchange Index	PS	1/2/2008	19/2/2014
29	Lithuania	Europe	High	OMX Vilnius Index	VI	1/2/2008	19/2/2014
30	Estonia	Europe	High	OMX Tallinn Index	TA	1/2/2008	19/2/2014
31	Latvia	Europe	High	OMX Riga Index	RI	1/2/2008	19/2/2014
32	US	America	High	Dow Jones Industrial Average	DJ	1/2/2008	19/2/2014
33	US	America	High	NASDAQ-100	ND	1/2/2008	19/2/2014
34	Brazil	America	Middle	Brazilian Bovespa Futures	BR	06/01/2010	19/2/2014

This table lists the stock market index used in this study. The income level is from the World Bank. The start date of Brazil is later than other markets due to data availability.

2.4.1.1 Market maturity

To test our first hypotheses (H1a and H1b) related to market maturity, we use the World Bank income level[2] to categorize financial markets (fourth column in Table 3): 26 markets in our sample stem from high income (mature) economies and 8 from middle income (less mature) economies. This approach to categorization is consistent with previous experiments (Choong et al., 2010a; Claessens et al., 2006). However, idiosyncrasies of particular classification systems may introduce bias. To test the robustness of our findings in relation to market maturity, we also employ market maturity classifications provided by the Morgan Stanley Capital Index (MSCI) (Zunino et al., 2009) and the International Monetary Fund (IMF) (Kim and Shamsuddin, 2008).

2.4.1.2 Model simulation methodology

We examine hypotheses related to the model simulation methodology employed. In particular, we compare results from the most common approach adopted, namely a 'static model evaluation' (e.g., de Oliveira et al., 2013), with those from a 'dynamic model evaluation' (e.g., Lessmann et al., 2011). The full details of these approaches are provided in 4.2.2. These models are developed using a fixed number of data points (e.g., days or hours) or 'sliding window size). In comparing results for the static and dynamic approaches, we examine the effect of alternative sliding window sizes of 25, 50 and 100.

2.4.1.3 Model covariates

To test hypotheses related to the effect of technical indicators (H3a and H3b), we consider the seven technical indicators that have been included in at least one study in our literature survey (these, together with the papers in which they appear are shown in Table 4). We develop sixteen covariates by varying parameters of these technical indicators, using popular variations found in the literature (e.g., Bollen et al., 2011). In addition, we create a second set of covariates using the characteristic points of a financial time series, namely, the opening, closing, highest and lowest price and the change of price (from opening to closing) observed in a predefined interval (e.g., 1 day). To test H3a and H3b, we compare the predictive ability of forecasting models incorporating covariates based on the reference values (open, high, low, close and change) to models that incorporate both the reference values *and* the sixteen technical indicator-based covariates. We deliberately employ simple covariates as a benchmark against which to examine the predictive value of the technical indicators because some previous research studies question the value of their informational content (Lesmond et al., 2004).

[2] http://data.worldbank.org/about/country-and-lending-groups

Clearly, these suspicions will be confirmed if technical indicators cannot improve the accuracy of predictions of a model based solely on a simple benchmark.

TABLE 4: COVARIATES EMPLOYED IN THE FINANCIAL TIME SERIES FORECASTING MODELS

	Covariate	Definition	Papers where covariate employed
Price-based covariates	**Opening price in period *t*.**	OI_t	
	Highest price in period *t*	HI_t	
	Lowest price in period *t*	LI_t	
	Closing price in period *t*	CI_t	
Indicators from the technical analysis	**Simple Moving Average (SMA)** SMA is the average closing price in a fixed length moving window. We calculate SMA for 5, 10 and 20 periods.	$SMA_{t,n} = \frac{\sum_{i=0}^{n-1} CI_{t-i}}{n}$	Hassan et al. (2007), Hadavandi et al. (2010), Huck (2010), Bollen et al. (2011)
	Moving Average Convergence / Divergence (MACD) MACD is the difference between a longer and a shorter exponentially weighted moving average (EMA). In general, a buy/sell signal is triggered when the MACD line crosses the zero line, which is usually a nine-day EMA. Our models use MACD and a nine-day EMA as covariates	$EMA_{t,n} = EMA_{t-1,n} + \frac{2(CI_t - EMA_{t-1,n})}{n+1}$ $MACD_{t,n} = EMA_{t,12} - EMA_{t,26}$	Hassan et al. (2007), Bollen et al. (2011)
	Relative Strength Index (RSI) The RSI indicates whether a market is overbought. We calculate RSI for 6, 9 and 14 periods.	$RSI_{t,n} = \frac{\sum_{i=0}^{n-1}(CI_{t-i} - CI_{t-i-1}) \times I\{CI_{t-i} - CI_{t-i-1}\}}{\sum_{i=0}^{n-1} \lvert CI_{t-i} - CI_{t-i-1} \rvert} \times 100$	Huck (2010), Bollen et al. (2011)

Williams %R Williams %R represents the relative position of the current share value in the past n periods. We use n=5 and n=10 periods.	$WilliamsR_{t,n} = \frac{HHI_{t,n} - CI_t}{HHI_{t,n} - LLI_{t,n}} \times -100,$	Hassan et al. (2007), Hadavandi et al. (2010), Bollen et al. (2011)
Accumulation/Distribution Oscillator (ADO) Given some financial time series, the ADO measures the strength of an upward/downward trend.	$ADO_t = \frac{(HI_t - OI_t) + (CI_t - LI_t)}{2 \times (HI_t - LI_t)} \times 100,$	Hadavandi et al. (2010)
Stochastic Oscillator (SO) SO compares the current share price to the values in the past 10 days. We include the %K and %D variables in our models.	$StochasticK_{t,} = \frac{CI_t - LLI_{t,10}}{HHI_{t,10} - LLI_{t,10}}$ $StochasticD_t = \frac{\sum_{i=0}^{2} StochasticK_{t-i}}{3}$	Hassan et al. (2007), Hadavandi et al. (2010)
Bollinger Bands (BB) BB consist of a middle, upper, and lower line. Advocates of BB believe that the trend is likely to revert when share prices approach the upper or lower line. To accommodate BB, we consider three covariates: mid, upper and lower values.	$BollingerMid_t = SMA_{t,20}$ $BollingerUp_t = BollingerMid_t + 2 \times \sigma_{t,20}$ $BollingerLow_t = BollingerMid_t - 2 \times \sigma_{t,20}$	Bollen et al. (2011)

* Our notation is as follows: CI_t is the closing price of a share/index in period t. OI_t, HI_t, and LI_t denote, respectively, the opening, highest, and lowest price in period t. We use *n* to denote the length of the window of a moving average. $I\{x\}$ is an indicator function, which is one if *x* is true and zero otherwise. $HHI_{t,n}$ is the highest share price observed in period *t* – *n* to *t*. Similarly, $LLI_{t,n}$ is the lowest price observed in period *t-n* to *t*. Last, $\sigma_{t,20}$ is the standard deviation of CI_t calculated over the period *t* – 20 to *t*. The literature column lists the studies which adopt the indicator.

2.4.1.4 Forecast horizons

To test hypotheses concerning the influence of different forecast horizons on market predictability and model-based trading profitability (H4a and H4b), we compare forecast horizons of one day and one hour. A horizon of one day is selected as it is the predominant setting in previous research (see Table 1). We selected an alternative horizon of less than one day because a comparison of intraday forecasting models and models that forecast one day or more ahead has, to our best knowledge, not been undertaken. We select one hour for the intraday setting to gain the benefits of a short forecast horizon (which increases the difference between the intraday setting and the setting where models predict one day into the future) whilst reducing the high computational costs associated with performing a large number of comparisons of computationally complex forecasting models for even shorter horizons. To compare experiments with forecast horizons of one day and one hour, we need to align the time periods of forecast model training, validation, and testing. To illustrate this, consider a test set of five trading days. In the daily setting, a model produces five individual predictions of price change direction (one for each day), which we aggregate to estimate the accuracy of the (daily) model. To compare this model on an equal footing to an alternative model that predicts one hour ahead (i.e., in the hourly settings), the latter should predict the same test set of five days. To achieve this, we let the hourly model produce one prediction for every hour for each of the five days in the test set. For example, if a stock exchange opens from 9am to 5pm (i.e., 8 hours), we create 8*5=40 predictions, estimate the accuracy of the hourly model on the basis of these 40 predictions, and compare this accuracy with that of the daily model.

2.4.1.5 Forecasting methods

A large number of ML prediction methods are available (Hastie et al., 2009). However, we focus on the two methods predominantly used in the ML literature on financial time series forecasting, SVM and ANN (Ballings et al., 2015). One of the explanations for their popularity for time series forecasting (cf. other advanced techniques such as bagged or boosted decision trees), is that they are better-suited to handle continuous covariates (Lessmann, Baesens, Seow, & Thomas, 2015), which frequently occur in financial time series forecasting (e.g., prices, price differences or technical indicators).

The SVMs are developed for binary classification problems, and the key concept is to use hyperplanes to define decision boundaries which separate data points of different classes. The idea is to map the original data points from the input space to a high-dimensional feature space such that the classification problem becomes simpler in the feature space. The mapping is done by the kernel function which should be chosen according to the situation. In its basic form, an SVM can be characterized as a regularized linear classifier, which estimates the target label, y, of an observation $\boldsymbol{x}$ by means of a linear function. Let the vector $\boldsymbol{\beta}$ and the scalar β_0 denote the coefficients and the intercept of such a linear function. Then, to develop a SVM, one solves the following mathematical program:

$$\min_{(\beta_0,\boldsymbol{\beta})} \mathcal{L} = \|(\beta_0,\boldsymbol{\beta})\|_2 + \lambda \sum_{i=1}^{n} \max(1 - y_i(\boldsymbol{\beta} \cdot \boldsymbol{x} + \beta_0), 0), \tag{1}$$

where $\mathcal{L}$ is the function of beta and beta0 which is to be minimized.

To prevent an SVM from overfitting the training data, the first term on the right-hand side of (1) penalizes model complexity through minimizing the magnitude of the model coefficients (Hastie et al., 2009). The second term measures the degree to which the model fits the training data accurately and is called the Hinge loss (Hastie et al., 2009). The scalar λ is a meta-parameter of SVM that allows the user to control the trade-off between high model fit and low model complexity (Smola and Schölkopf, 2004).

In our study, the scalar y refers to the change in direction (up or down) of a financial index in a future period (i.e., next day or next hour) and $\boldsymbol{x}$ represents a vector of covariates which we use to predict y (previous values of y, technical indicators, etc.). We estimate the vector of model parameters, $\boldsymbol{\beta}$, during model training. We then compute predictions, $\hat{y}$, as follows:

$$\hat{y} = sign(\boldsymbol{\beta} \cdot \boldsymbol{x} + \beta_0) \qquad (2)$$

SVMs are able to implicitly project the input data into a nonlinear feature space of higher dimension. Creating a linear model in the transformed space is equivalent to creating a nonlinear model in the original input space (Cristianini and Shawe-Taylor, 2000). The nonlinear transformation of the data is (computationally) feasible because the dual program of (1) incorporates the input data only in the form of scalar products. Using a kernel function, SVMs are able to compute the scalar product in the feature space directly (i.e., without actually transforming the data). This technique is known as the kernel trick. A kernel function can transform low dimensional input space to a higher dimensional space, which could convert a non-separable problem to a separable problem. There are many options for a kernel function, such as linear, polynomial and radial basis functions. The choice of a kernel function depends on the features of the problem. For example, the linear kernel function is preferred when the problem is linearly separable and the polynomial kernel function is popular in natural language processing. Due to the nonlinear nature of stock markets, we use the radial basis kernel function which is widely considered to be effective with nonlinear problems. We use the radial-basis-kernel-function, which is the standard kernel in SVM applications (Keerthi and Lin, 2003).

ANNs are another nonlinear prediction method. In general, a feed-forward neural network consists of input, hidden and output layers, where each layer has multiple information processing units called neurons. The neurons of one layer are fully-connected to the neurons of the next layer. The neurons in the input layer are simply the original covariates. That is, we clamp the input covariates on a distinct set of input units. Subsequently, the values are passed through a set of weights to produce the inputs to the next layer of "hidden" units (the hidden layer). These units utilize a nonlinear function, which is usually a sigmoid function, to process their input and to produce their outputs. As many hidden layers can be stacked as necessary, and a feature of deep learning neural network is a large number of hidden layers. The outputs of the neural network are the outputs of the final layer. A learning algorithm is used to calibrate the weights between units to make the output meet some desired requirement. We design our ANNs such that the output has only one neuron, which is the standard setup for regression and binary classification problems (Hastie et al., 2009). In the former case, the output neuron models the value of the continuous target variable and in the latter case the posterior probability of the discrete target variable. The number of neurons in the hidden layer, Z, is a meta-parameter. Let $\boldsymbol{a}_z$ be a vector of weights that connect the input neurons to the z^{th} hidden neuron, b_z a threshold attached to hidden neuron z, and let g^h be a nonlinear function. We can then write the output of the hidden layer as follows:

$$G(\boldsymbol{a}_z, b_z, \boldsymbol{x}) = \begin{bmatrix} g^h\left(b_1 + \sum_{j=1}^{m} a_{1j}x_j\right) \\ \vdots \\ g^h\left(b_z + \sum_{j=1}^{m} a_{zj}x_j\right) \end{bmatrix}_{Z\times 1} . \tag{3}$$

In a similar way, the result of the ANN, $f(\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}_Z, b_Z)$, is:

$$f(\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}_Z, b_Z) = g^o\left(\sum_{z=1}^{Z} \beta_z\, G(\boldsymbol{a}_z, b_z, \boldsymbol{x})\right), \tag{4}$$

where $\boldsymbol{\beta}$ denotes the weight vector that connects the hidden and the output layer, and g^o is the function that transforms the result of the output neuron. We follow Chang et al. (2009) and choose $g^o$and g^h to be logistic functions. This allows us to interpret the output of the neural network as an estimate of the posterior probability of an upward/downward price change.

To determine the parameters of the ANN ($\boldsymbol{a}_z, b_z$, and $\boldsymbol{\beta}$), one minimizes a suitable loss-function using gradient-based methods. Specifically, we create an ANN model through solving (5) using a quasi-Newton algorithm, where λ is once more a regularization parameter to penalize model complexity and prevent overfitting.

$$\min_{\boldsymbol{\beta}} \mathcal{L} = \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i, \boldsymbol{\beta}))^2 + \lambda \|\boldsymbol{\beta}\|_2. \tag{5}$$

ANN and SVM contain two meta-parameters. First, the regularization term λ appears in both models. The second meta-parameter for SVM is a parameter of the radial-basis kernel function, (denoted by γ). For ANN, the second meta-parameter determines the number of neurons in the hidden layer. We employ grid-search to identify suitable values for the meta-parameters of ANN and SVM (Cherkassky and Ma, 2004). In particular, based on recommendations from the literature (Berry and Linoff, 1997;

Xu and Chen, 2008), we define candidate values for each meta-parameter and empirically assess all possible combinations of meta-parameter settings. We select the combination with maximal forecast accuracy on a separate validation sample to compute predictions for an out-of-sample test set. Table 5 reports the candidate settings for meta-parameter values.

TABLE 5: CANDIDATE VALUES FOR SVM AND ANN META-PARAMETERS

ANN	Hidden layer neurons	6, 18, 48
	λ	1, 100, 0, 0.01
SVM	λ	$2^{-10}, 2^{-8}, 2^{-6}, 2^{-4}, 2^{-2}, 1, 2^{2}, 2^{4}, 2^{6}, 2^{8}, 2^{10}$
	γ	$2^{-10}, 2^{-8}, 2^{-6}, 2^{-4}, 2^{-2}, 1, 2^{2}, 2^{4}, 2^{6}, 2^{8}, 2^{10}$

This table lists the parameter values which are examined to find the best combination of parameters in the validation stage.

2.4.2 Performance Measurement

2.4.2.1 Indicators of predictive accuracy

As indicated above, our target variable is ‘direction change’:1 if the index prices increase from one period to the next (e.g., from day *t* to day *t*+1), and 0 otherwise. We assess the predictive accuracy of a forecasting model in terms of the percentage of correctly classified observations (hit rate).

Since the main aim of the paper is to examine the influence of methodological factors on prediction performance, we do not attempt to optimize accuracy. Previous ML studies generally focus on one specific market. However, to achieve our objective, we attempt to forecast *all* markets with a consistent, clearly documented methodology. This facilitates replicability across indices, which we consider an integral part of our research. Consequently, we deliberately do not try to maximize predictive accuracy for an individual index.

We estimate the profitability of trading on a forecasting model’s predictions based on a simple trading strategy, which we call ‘follow prediction.' In particular, we assume that a trader, via futures contracts, buys into the market if the predicted ‘direction change’ equals 1, and sells otherwise. We calculate the return on investment (ROI) in period *t* (roi_t), as follows:

$$roi_t = 1 + \frac{ABS|CI_t - CI_{t+1}|}{CI_t} * PC_t, \tag{6}$$

where CI_t is the closing price of the financial index at period t and PC_t is 1 if the model-predicted 'direction change' in period t+1 is correct, and -1 otherwise. We employ this ROI formula as we use the information available in period t to predict the 'direction change' in period t+1. A period is either one day or one hour, depending on the experimental setting. Consequently, we obtain a profit/loss if the prediction is correct/wrong. The overall ROI over all test periods t = 1, 2, …,T is then calculated as follows:

$$ROI = \prod roi_t \, , t \in T \tag{7}$$

As Shleifer and Vishny (1997) indicate that transaction costs may affect the trading profit in exploiting the anomalies or the recognizable price patterns, we calculate the ROIs with transaction cost accounted in additional analysis.

2.4.2.2 Model evaluation and simulation of model-based training

To test the accuracy of forecasting models, we distinguish two approaches for out-of-sample model evaluation. The standard approach in the literature is to split a financial time series into three non-overlapping sets for model training, validation, and testing (see Table 1). The training set is used to estimate the parameters of the forecasting model (e.g., $\boldsymbol{\beta}$ in (1)). The validation set is used to tune meta-parameters (e.g., the regularization parameter λ in (1)) by means of empirical experimentation. A fully-specified model with fixed meta-parameters is then developed on the union of the training and validation sets (to maximize data utilization) and applied to generate predictions for the observations in the test set. To measure forecast accuracy, the predictions are compared to the actual values of the target variable in the test set. This allows us to compare models in terms of their predictive accuracy on hold-out data. We call this approach a static evaluation because the same model is used to forecast all observations in the test set (de Oliveira et al., 2013; Kara et al., 2011).

To implement the static approach, we split the data points in a financial time series chronologically: 50, 25 and 25 percent for training, validation and testing, respectively. Our data set, which runs from February 2008 to February 2014, contains 1,500 trading days for each of the 34 financial time series. Consequently, by way of example, when developing forecasting models to predict one day ahead, we use the last 25% of trading days (375) for hold-out testing. The models that we eventually compare (e.g., SVM vs. ANN) are trained and validated on the preceding 75% (1,125) of trading days. In particular, we use the first 50% (750) of trading days to train the forecasting models with alternative meta-parameter settings (see Table 5) and assess their accuracy on the following 375 trading days (our validation sample). The validation sample predictions reveal the best meta-parameter setting for a given time-series. We use this setting to estimate a final model using the full 1,125 days sub-sample and use the resulting model to predict the hold-out test set.

We contrast the static model evaluation approach to a dynamic sliding-window cross-validation (Lessmann et al., 2011). Here, we use the same data partitioning as in the static setting. For example, when developing forecasting models to predict one day into the future we again use the first 50% (750) days of a time series to estimate one forecasting model per meta-parameter setting (e.g., 12 settings for ANN). Every meta-parameter setting provides a candidate model. However, we use these candidate models to develop predictions for a single data point, namely for day 751, which is the first day in the validation partition. Next, we shift the training and validation window one period (day) forward. That is, using days 2 to 751 for model training, we estimate one forecast model per meta-parameter setting and let the resulting models predict day 752. We repeat the estimation of models and prediction of one day ahead until we have developed predictions for all days in the validation period (i.e., days 751 to 1,125). We then calculate the accuracy of candidate forecasting models on the validation sample and select the meta-parameters that give maximal accuracy.

This dynamic approach differs from the static approach in that every data point of the validation set is predicted with a different model. In other words, we update forecasting models to incorporate recent information (e.g., we train forecasting models that predict the price ‘direction change’ for trading day 752 on a data set that includes trading day 751).

After completing model selection in the dynamic approach, we collect predictions for the test set. To that end, we proceed as in the model selection step. For example, when comparing SVM and ANN models used to predict one day ahead, for each day of the test set, we estimate one SVM and ANN model. Considering, for example, the first day in the test set (day 1,126), we use the preceding days to train the forecasting models and let them predict trading day 1,126. We then shift the evaluation window one period forward and repeat. The main difference to the model selection stage is that we do not consider all meta-parameter settings. Instead, we estimate one SVM and one ANN model, for which we use the best meta-parameter values as identified in the preceding model selection step. In summary, each data point of the test set is predicted with a different forecasting model, but all models use the same settings for meta-parameters.

If the forecast horizon is not one day but one hour, we proceed in exactly the same way as described. However, the size of the initial data sample is much larger because we now have one data point per opening hour of the stock exchange.

Sliding-window cross-validation offers some degrees of freedom. In particular, when predicting a particular data point (e.g., day), one can either use all previous data points for model training or work with a fixed training window (Lessmann et al., 2011). We opt for the latter approach and use a window size of 50. Using a fixed window size (cf. all available trading days prior to a day *t*) has the advantage that it decreases computational costs. More importantly, assuming that stock returns show no long-term memory (Cheung and Lai, 1995; Lo, 1991), using very old data for training a forecasting model might actually harm its accuracy. However, we acknowledge that the choice of the window size might affect the results of sliding-window cross-validation. Therefore, to secure reliability, we empirically analyze alternative window sizes of 25 and 100 data points.

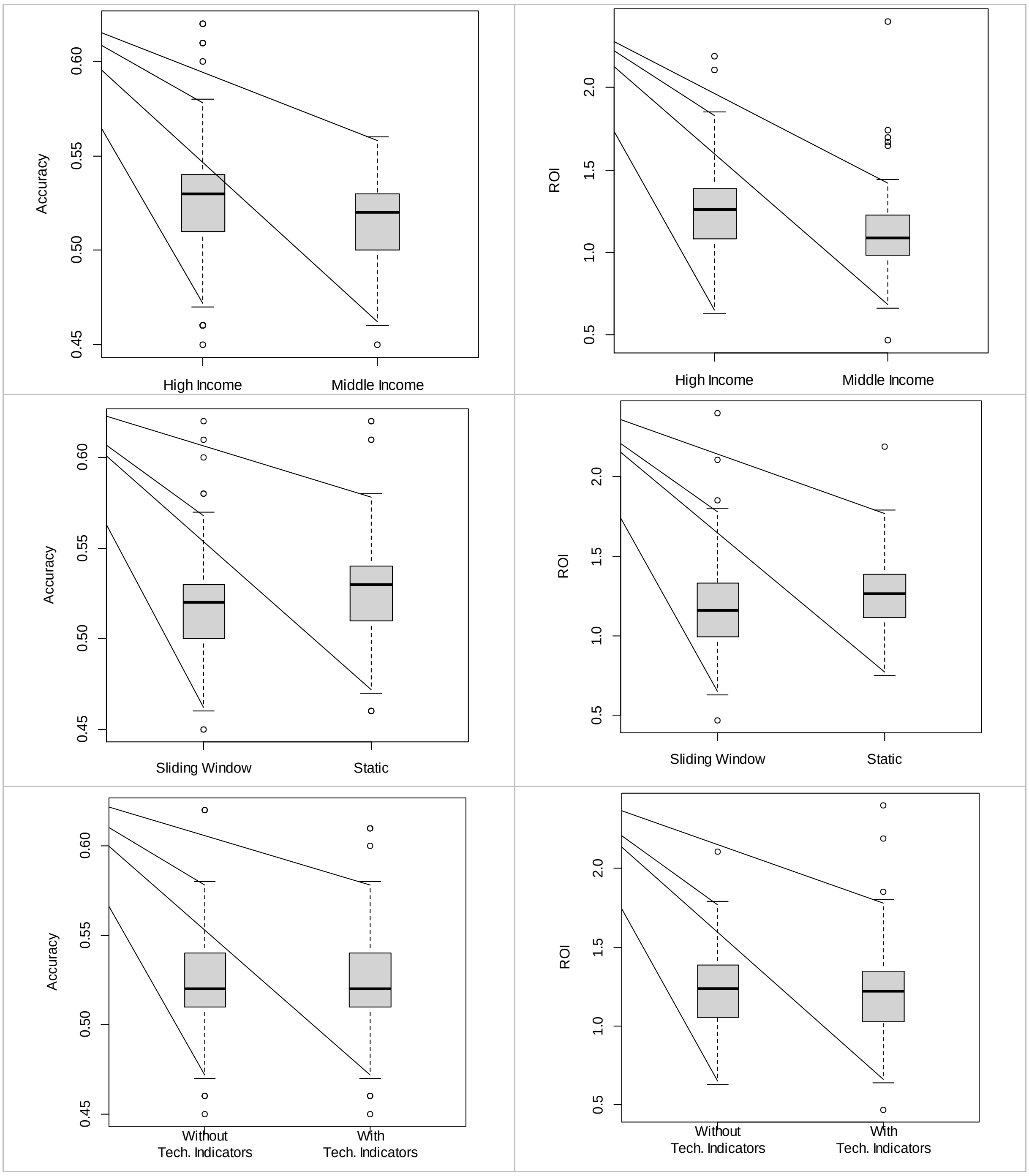
2.5 Empirical results

2.5.1 Tests of the Hypotheses of Experimental Factors on Market Predictability

The experimental design includes five main factors: market maturity, model simulation methodology, covariate composition, forecast horizon and prediction method (SVM vs. ANN; best ML (SVM) vs. best econometric (AR))(see Table 2). Each factor has two alternative settings (e.g., SVM and ANN for the factor prediction method) and these are compared in multiple simulations. That is, we obtain 272 individual prediction results for SVM and ANN (34 financial markets × 2 forecast horizons × 2 covariate compositions × 2 model simulation methodologies), enabling us to show the performance difference between the two prediction methods. We compare prediction results for other factors in a similar way. We provide an overview of the empirical results in Figure 1, which reports the distribution across the experimental factors for predictive accuracy and ROI.

A number of conclusions emerge from the results presented in Figure 1. First, a model-based trading approach produces higher ROI in high (cf. middle) income markets. More specifically, the majority of prediction models produce an ROI greater than one in high income markets; suggesting that profitable trading on model predictions is possible and questioning market efficiency in the corresponding markets. In middle income markets, a larger number of prediction models do not produce a profit, although the median ROI is still greater than one.

We also examine the performance of the forecasting models at the individual market level. We obtain a distribution of prediction performances for each market because we develop multiple models to examine the other experimental factors: forecast horizon, prediction model, etc. These results are displayed in Figures 2 and 3.


Accuracy
0.45
0.50
0.55
0.60
High Income
Middle Income
ROI
0.5
1.0
1.5
2.0
Sliding Window
Static
Without
Tech. Indicators
With
Tech. Indicators

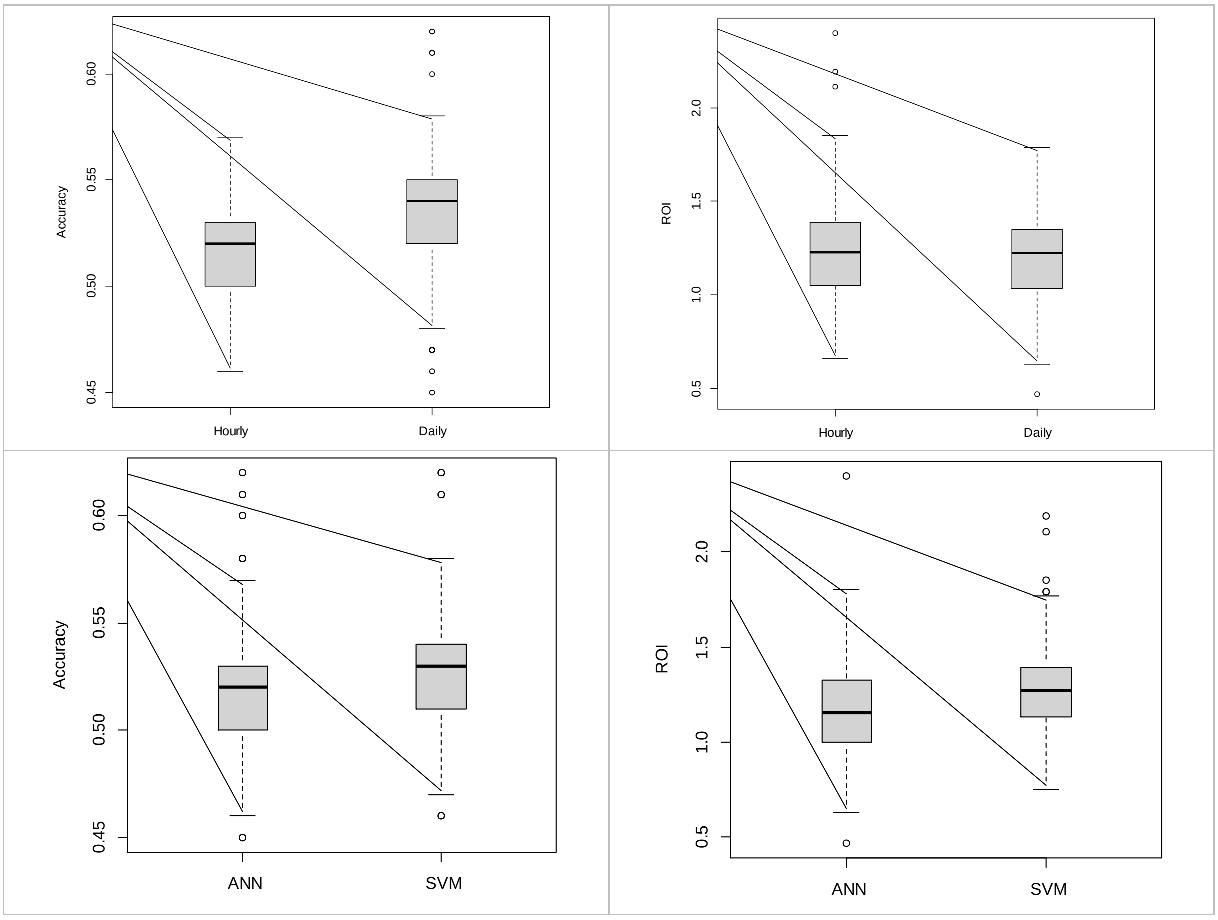


Figure 1: PREDICTION PERFORMANCE IN ACCURACY (LEFT) AND ROI (RIGHT) ACROSS EXPERIMENTAL FACTORS.

Each boxplot shows the performance (accuracy or ROI) of all experiments in terms of one methodological factor. For example, the first boxplot in the left column represents the accuracy in terms of market maturity. The left box shows the accuracy in the high income markets (more mature) and the right box displays the accuracy in the middle income markets (less mature).

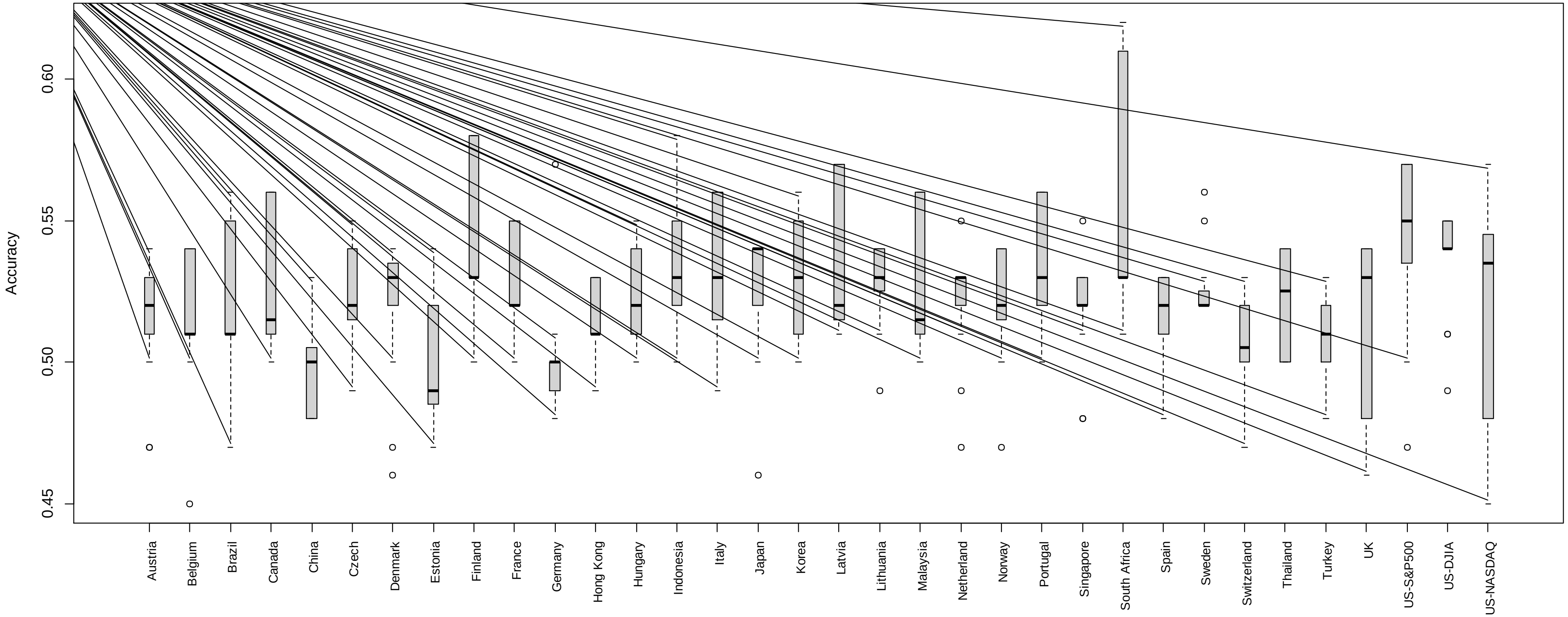


Figure 2: PREDICTION PERFORMANCE IN ACCURACY ACROSS FINANCIAL MARKETS

Each boxplot shows the performance (accuracy) of all experiments in the market.

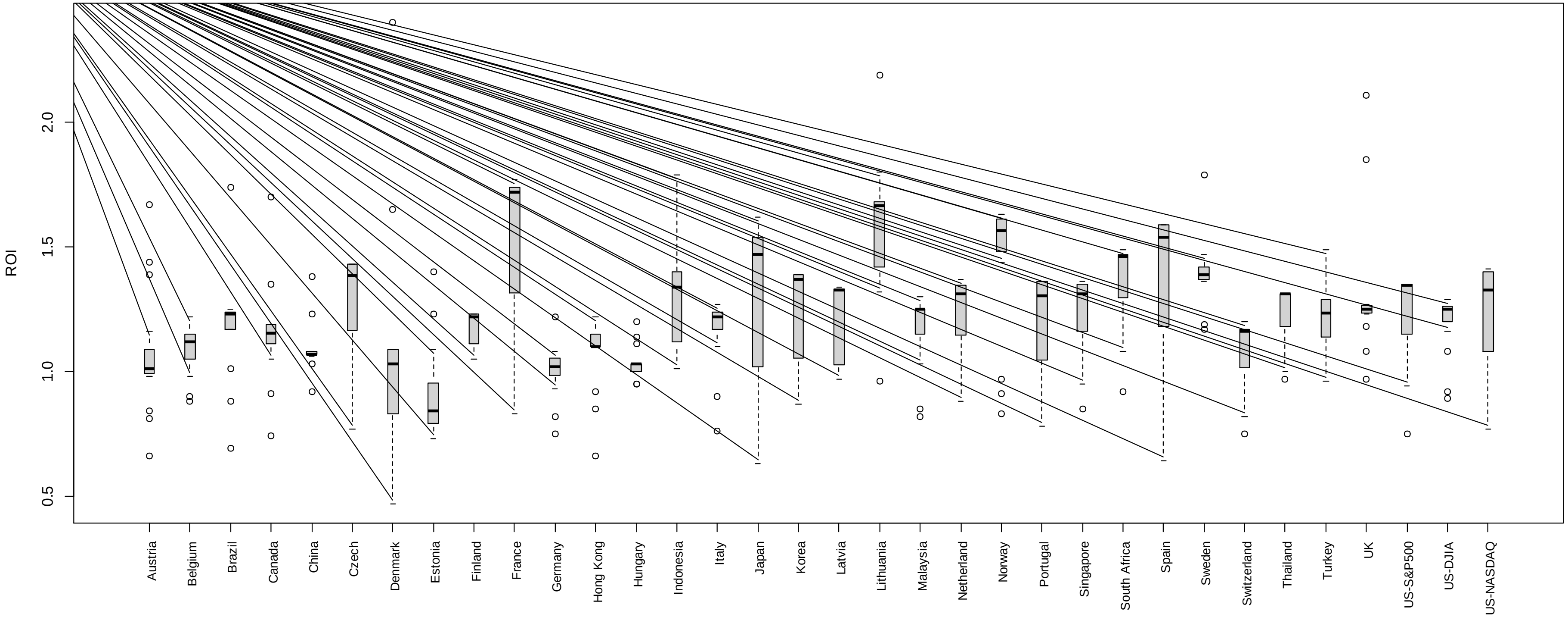


Figure 3: PREDICTION PERFORMANCE IN ROI ACROSS FINANCIAL MARKETS

Each boxplot shows the performance (ROI) of all experiments in the market.

Figures 2 and 3 evidence substantial variation in predictability across markets. This is true for the comparison across high and medium income markets but also for comparisons between markets within these groups (e.g., the Finnish and German markets belong to the high income group but display large differences in their spread in predictive accuracy). Results at an individual market level also suggest that profitable model-based trading is feasible based on the covariates we employ (see Figure 3) since the median ROI is considerably greater than one for most markets. Examples of exceptions include markets in Austria, Denmark, Estonia, Hungary, and Germany.

In summary, the main finding from the results displayed in Figures 2 and 3 is that the selection of the financial market exerts a major influence on the observed level of predictability and thus market efficiency. Our analysis of the literature reveals that previous ML studies predominantly rely on data from a single market. The wide variation that we observe in market efficiency between markets suggests that this focus on a single market risks over- or under-estimating the potential for prediction across all financial markets.

The results presented in Figure 1 indicate that a sliding-window methodology typically indicates lower levels of accuracy and ROI than a static methodology. This suggests that the cross-validation method employed affects prediction performance and should be considered in future comparisons. The static approach, which prevails in prior research (see Table 1), benefits from using a larger number of samples for model training compared to the sliding-window cross-validation. The dynamic approach benefits from including the most recent samples. Therefore, our results indicate that including recent price information does not improve stock market prediction more than incorporating a larger number of samples. This supports the view that financial markets exhibit long-term memory (Lo, 1991).

The results presented in Figure 1 suggest that technical indicators for financial market modeling do not offer many advantages over simple covariates, as the accuracy and ROI distributions of the two settings are similar. This is surprising since we deliberately select simple covariates in the form of reference prices (open, high, low, close) as a benchmark for the technical indicators. For example, we do not consider price differences or moving averages, which have been used in prior research (see Table 1). Therefore, our results provide strong evidence against using technical indicators for financial time series forecasting.

The results displayed in Figure 1 suggest that predictive accuracy is lower when forecasting price movements one hour (cf. one day) ahead but the ROI distribution does not display much difference between the two settings. However, it is noteworthy that both daily and hourly forecasting horizons produce, on average, a positive return (i.e., ROI > 1).

Our results support previous findings that SVM performs better than ANN in financial market modeling (Kim, 2003; Ou and Wang, 2009; Tay and Cao, 2001). The median accuracy and the median

ROI are higher for SVM compared to ANN. In addition, the distribution of ROI displays less spread for SVM, implying greater stability of SVM-based prediction models. The lower quartile of the ROI distribution is higher for SVM (cf. ANN), and the ROI distribution of ANN shows an outlier value where ROI drops below 50%, whereas SVM do not suffer from such large losses. SVM can thus be considered more robust. One possible explanation is that there is tremendous noise in stock market data and ANN has been shown to be inferior on noisy data (Kim, 2003). Also, in our simulations, ANN takes considerably higher computing resources than SVM does, which decreases the probability of finding the best parameters during the validation. While the previous studies compare SVM and ANN with one or two daily stock market data, we extend the scope by including both daily and hourly data across 34 stock markets which comprise most major markets in the world. Given the large scope of our simulation and the consistency between our findings and the previous studies (Kim, 2003; Ou and Wang, 2009; Tay and Cao, 2001), the superior performance of SVM for financial time series forecasting is convincing. This implies that SVM should be considered as an important benchmark for assessing novel models and methods.

To formally test our hypotheses, we examine the statistical significance of observed mean differences across factor levels. We achieve this using regression analysis. In particular, we estimate the following regression model to explain predictive accuracy:

$$\text{Accuracy} = \alpha + \beta_{\text{MI}} MI + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVM + \varepsilon, \tag{8}$$

where *MI, ST, T, D* and *SVM* are dummy variables, taking the value 1 for middle income markets, a static model simulation, when using technical indicators alongside basic price covariates (open, high, low, close and change), a daily forecast horizon and SVM, respectively, and taking the value 0 otherwise (i.e. for high income markets, a dynamic model simulation, when only using basic price covariates, an hourly forecast horizon and ANN). This model is also used to explain ROI.

To formally test our hypotheses related to the relative performance of, what we show to be, the best ML and econometric models (i.e., SVM and AR, respectively), we estimate the following regression model to explain predictive accuracy:

$$\text{Accuracy} = \alpha + \beta_{\text{MI}} MI + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVMA + \varepsilon, \tag{9}$$

where *MI*, *ST*, *T*, *D* and *SVMA* are dummy variables, taking the value 1 for middle income markets, a static model simulation, when using technical indicators alongside basic price covariates, a daily forecast horizon and the best performing ML model (SVM) prediction, respectively, and 0 otherwise (the reference model is AR). This model is also used to explain ROI.

If we expect a regression model to go through the origin, we can remove the intercept term from the model without introducing bias in estimating coefficients. In the case of prediction accuracy, *going* *through the origin* means the average accuracy is 0%. However, since the accuracy of randomly guessing the price direction is around 50%, the accuracy of our models are unlikely to be around 0%. Similarly, the regression line for ROI is also unlikely to go through the origin, unless our models generate huge negative returns and lose all the original capital. Consequently, we expect that the regression lines for both accuracy and ROI will not go through the origin. Hence, we include the intercept terms to avoid potential bias in estimating the parameters of the regression models (Brooks, 2014). Tables 6 and 7 summarize the results of the regression analyses for accuracy and ROI, respectively, related to the comparison of the two commonly employed ML methods (SVM and ANN). Tables 8 and 9 summarize the results of the regression analyses for accuracy and ROI, respectively, related to the comparison of the best performing ML and econometric methods (SVM and AR).

TABLE 6: REGRESSION ANALYSIS OF PREDICTIVE ACCURACY – ML TECHNIQUES

Predictive accuracy	Estimated Coefficient	Std. Error	t value	p value	Cohen's d	partial η^2
(Intercept)*	0.5081	0.0023	221.4	**< 10^{-16}**		
Market maturity (MI)	-0.0103	0.0023	-4.389	**< 10^{-4}**	0.4016	0.0346
Model simulation methodology (ST)	0.0086	0.002	4.319	**< 10^{-4}**	0.3350	0.0335
Covariate composition (T)	-0.0001	0.002	-0.037	0.971	0.0028	0.0000
Forecast horizon (D)	0.0174	0.002	8.748	**< 10^{-16}**	0.7105	0.1245
Prediction method (SVM)	0.0106	0.002	5.315	**< 10^{-6}**	0.4154	0.0499
Residual standard error	0.02323	df	538			
R^2	0.2096	Adjusted R^2	0.2023			
F-statistic	28.54	(on 5 and 538 DF)				
p-value	< 2.2^{-16}					

* Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; T: Technical indicators included amongst the covariates; D: Forecast horizon of one day; SVM: Prediction method employed is SVM

TABLE 7: REGRESSION ANALYSIS OF ROI – ML TECHNIQUES

ROI	Estimated Coefficient	Std. Error	t value	p value	Cohen's d	partial η^2
(Intercept)*	1.1764	0.02228	52.8	**$< 10^{-16}$**		
Market maturity (MI)	-0.1496	0.02280	-6.564	**$< 10^{-8}$**	0.6264	0.0741
Model simulation methodology (ST)	0.1021	0.01934	6.015	**$< 10^{-6}$**	0.4220	0.0492
Covariate composition (T)	-0.0168	0.01934	-0.867	0.3864	0.0679	0.0014
Forecast horizon (D)	-0.0442	0.01934	-2.285	**0.0227**	0.1795	0.0096
Prediction method (SVM)	0.1163	0.01934	5.277	**$< 10^{-8}$**	0.4842	0.0630
Residual standard error	0.2255	df	538			
R^2	0.1737	Adjusted R^2	0.166			
F-statistic	22.62	(on 5 and 538 DF)				
p-value	$< 2.2^{-16}$					

* Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; T: Technical indicators included amongst the covariates; D: Forecast horizon of one day; SVM: Prediction method employed is SVM

TABLE 8: REGRESSION ANALYSIS OF PREDICTIVE ACCURACY – SVM VS. AR

Predictive accuracy	Estimated Coefficient	Std. Error	t value	p value	Cohen's d	partial η^2
(Intercept)*	0.5067	0.0026	192.11	**< 10^{-16}**		
Market maturity (MI)	-0.0094	0.0027	-3.4973	**0.0005**	0.3680	0.0295
Model simulation methodology (ST)	0.0055	0.0023	2.4191	**0.016**	0.2146	0.0143
Covariate composition (T)	-0.0003	0.0028	-0.1049	0.9165	0.2394	0.00002
Forecast horizon (D)	0.0183	0.0023	7.9853	**< 10^{-13}**	0.7528	0.1369
Prediction method (SVMA)	0.0129	0.0028	4.6147	**< 10^{-5}**	0.5068	0.0503
Residual standard error	0.0231	df	403			
R^2	0.2142	Adjusted R^2	0.2045			
F-statistic	21.92	(on 4 and 403 DF)				
p-value	< 10^{-16}					

*Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; T: Technical indicators included amongst the covariates; D: Forecast horizon of one day; SVMA: Prediction method employed is SVM and the reference model is AR.

TABLE 9: REGRESSION ANALYSIS OF ROI – SVM VS. AR

ROI	Estimated Coefficient	Std. Error	t value	p value	Cohen's d	partial η^2
(Intercept)*	1.1576	0.0242	47.832	**< 10^{-16}**		
Market maturity (MI)	-0.186	0.0248	-7.5092	**< 10^{-12}**	0.8183	0.1230
Model simulation methodology (ST)	0.0718	0.021	3.4181	**0.0007**	0.3018	0.0282
Covariate composition (T)	-0.0245	0.0257	-0.9515	0.3419	0.2402	0.0022
Forecast horizon (D)	-0.0466	0.021	-2.2188	**0.0271**	0.1946	0.0121
Prediction method (SVMA)	0.1638	0.0257	6.3665	**< 10^{-9}**	0.6598	0.0916
Residual standard error	0.2122	df	403			
R^2	0.2301	Adjusted R^2	0.2206			
F-statistic	29.82	(on 4 and 403 DF)				
p-value	< 10^{-16}					

*Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; T: Technical indicators included amongst the covariates; D: Forecast horizon of one day; SVM: Prediction method employed is SVM, and the reference model is AR.

The F-statistics and their corresponding p-values displayed in Tables 6-9 confirm the statistical significance of the four regressions. The adjusted R^2 values suggest that the independent variables explain about 20 (21) percent and 17 (23) percent of the observed variation in predictive accuracy and ROI, respectively (The numbers in brackets relating to the regressions incorporating the best performing ML and econometric methods (SVM and AR). These R^2 values may appear rather low. However, it is important to remember that the development of prices in financial markets is driven by a multitude of factors, many of which are not considered in this study. As a consequence, the performance of the prediction models exhibits substantial unexplained variation. However, it is noteworthy that R^2 is higher in the regression of predictive accuracy. This emphasizes the subtle difference between predicting market movements using an ML model and trading profitably on model-predictions. Our results show that the degree to which the latter is possible depends even more on external factors other than the prediction method (and on other factors considered in this study). As we discuss below (see Section 6), this has important implications for conclusions regarding market efficiency.

We first examine results related to experiments simply incorporating ML forecasting models (SVM and ANN).The results presented in Table 6 and Table 7 demonstrate that the effect of all experimental factors (other than covariate composition) on predictive accuracy and ROI is statistically significant. This confirms the relevance of the chosen factors and that the potential of sophisticated technical indicators to help predict financial market movements (over that achievable via simple reference-price based covariates) might be limited. Note that we also run experiments in the markets in which data are available for longer periods (8 ~ 30 years). The corresponding results, which are available in online Appendix C, agree with those presented in the main part of the paper.

In interpreting the other results of the regression, it is important to remember the reference model that forms the basis of the comparison. Given our coding of the independent variables, the reference model is an ANN model that predicts price movements in high income markets using simple reference prices as covariates one hour ahead and is evaluated using sliding-window cross-validation. Considering the regression coefficient of the intercept in the accuracy regression model (Table 6), such a model produces a directional accuracy of 51 percent. Assuming that the distribution of hourly price movements is roughly balanced, this is only a little better than random. However, we observe in the ROI regression (Table 7) that the seemingly small improvement over a random model is enough to produce a sizeable profit, with an ROI of about 18 percent. The regression coefficient of market maturity (-0.1496) indicates that profitability erodes in middle income markets. Similarly, ROI decreases with forecast horizon. Keeping everything else constant, predicting one day into the future reduces ROI approximately 4 percentage points (regression coefficient of -0.0442). However, predictive accuracy increases in a setting with a daily forecast horizon (regression coefficient 0.0174 in Table 6). The results also confirm that a static model simulation methodology (cf. sliding-window cross-validation) and using SVM instead of ANN both significantly increase accuracy and ROI. It is worth noting that the significance of the intercept terms in both the regressions reported in Tables 6 and 7 supports our decision to include these terms. It is worth noticing the large t-value for the intercept (221.4). Since the t-value, which can be considered as a measure of the precision for the estimation of the coefficient, is calculated as the estimated coefficient divided by the standard error, the large value results from the small value of the standard error (0.0023) which means that the range of the estimated coefficients is relatively small.

To complement the analysis of the five experimental factors, Table 6 and Table 7 also include measures of effect size. In particular, Cohen's d captures the mean difference between groups in standard deviation units. As a rule of thumb, values of 0.2, 0.5, and 0.8 indicate small, medium, and large effect sizes, respectively (Cohen, 1969). The Cohen's d results displayed in Tables 6 and 7 indicate that market maturity, prediction model (SVM vs. ANN), and model simulation methodology have a medium effect on predictive accuracy and ROI. Effect sizes for the factor covariate composition are small, which supports the lower levels of significance that we observe for this factor. Finally, the forecast horizon has a medium to large effect on predictive accuracy but only a small effect on ROI. For completeness, we include an additional measure of effect size in Table 6 and Table 7, the amount of variation that is explained by an individual factor – commonly referred to as partial η^2 in analysis of variance. Overall, the results are in line with those related to the Cohen's d.

We now turn to our analysis of the regressions incorporating the best performing ML and best performing econometric models. As discussed above, the results displayed in Tables 6 and 7, show that SVM outperforms ANN. This result is line with that reported in the literature (Chen et al., 2006; Huang et al., 2005; Kim, 2003; Ou and Wang, 2009; Tay and Cao, 2001). Our experiments also show that AR outperforms ARIMA and GARCH (see online Appendix A). To formally test hypotheses 6a and 6b, we compare SVM and AR by examining the statistical significance of observed mean differences across factor levels in our estimations based on equation (9).

The results presented in Table 8 and Table 9 demonstrate that SVM significantly outperforms AR in terms of accuracy and ROI, providing support for Hypotheses 6a and 6b. It is important to note that the effects of all experimental factors (with the exception of covariate composition) on predictive accuracy and ROI are statistically significant. The significance and direction of the effect of the experimental factors on accuracy and ROI are in line with the results discussed above for the regressions based on the two most commonly employed ML techniques (i.e., equation (8)). This confirms that most of the experimental factors have a significant effect on the predictive accuracy and ROI achievable using both ML and econometric methods.

2.5.2 Sensitivity Analysis

To confirm the robustness of our conclusions, we perform additional experiments related to different sliding window sizes and different systems for classifying market maturity.

2.5.2.1 Analysis of Sliding Window Size

In sliding-window cross-validation, the size of the training window is a potentially important parameter that might affect the observed level of predictive accuracy. Thus, to investigate whether the above conclusions are robust toward alternative settings, we perform experiments with sliding window size 25 and 100. We estimate the following regression model to explain predictive accuracy:

$$\text{Accuracy} = \alpha + \beta_{\text{MI}} MI + \beta_{\text{SW25}} SW25 + \beta_{\text{SW50}} SW50 + \beta_{\text{SW100}} SW100 + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVM + \varepsilon, \tag{10}$$

where *MI*, *SW25, SW50, SW100, T, D* and *SVM* are dummy variables, taking the value 1 for middle income markets, when the sliding window size is 25, 50 and 100, when using technical indicators alongside basic price covariates, a daily forecast horizon and SVM prediction, respectively, and 0 otherwise. Recall that the data to estimate these equations also includes the results of (static) forecasting models that are not subject to a sliding window evaluation. Therefore, using three dummies to encode the three settings for sliding window size is appropriate and does not create linear dependence among covariates. This model is also used to explain ROI.

The results of regression (10) are shown in Table 10. For both accuracy and ROI, all of the coefficients of *SW25*, *SW50* and *SW 100* are negative, indicating that accuracy and ROI are significantly higher under a static model evaluation compared to using a dynamic sliding-window cross-validation. For both accuracy and ROI, the coefficients of sliding window size 100 are very close to that of sliding window size 50, whereas the coefficient of sliding window size 25 is much lower than the others. Hence, increasing the size of the sliding window appears to improve prediction performance, but the marginal benefit appears to decrease. We take the small differences between the results for window sizes of 50 and 100 as evidence that the choice of 50 is appropriate and does not affect the conclusions of the comparison of alternative model evaluation regimes (sliding-window vs. static).

Table 10: REGRESSION ANALYSIS OF SLIDING WINDOW SIZE

	Predictive accuracy				ROI			
	Estimated Coefficient	Std. Error	t value	p value	Estimated Coefficient	Std. Error	t value	p value
(Intercept)*	0.5148	0.0019	273.52	**<10^{-16}**	1.2286	0.0176	69.678	**<10^{-16}**
Market maturity (MI)	-0.0086	0.0016	-5.265	**<10^{-6}**	-0.1328	0.0154	-8.636	**<10^{-16}**
Sliding Window Size 25 (SW25)	-0.0114	0.0020	-5.769	**<10^{-7}**	-0.1224	0.0184	-6.632	**<10^{-8}**
Sliding Window Size 50 (SW50)	-0.0086	0.0020	-4.369	**0.0001**	-0.1021	0.0184	-5.532	**<10^{-7}**
Sliding Window Size 100 (SW100)	-0.0083	0.0020	-4.201	**<10^{-4}**	-0.1029	0.0184	-5.578	**<10^{-7}**
Covariate composition (T)	-0.0006	0.0014	-0.422	**0.673**	-0.0072	0.0130	-0.548	**0.5837**
Forecast horizon (D)	0.0145	0.0014	10.403	**<10^{-16}**	-0.0405	0.0130	-3.101	**0.0020**
Prediction method (SVM)	0.0170	0.0014	12.225	**<10^{-16}**	0.1948	0.0130	14.930	**<10^{-16}**
Residual standard error	0.0230	df	1080		0.2151	df	1080	
R^2	0.2301				0.2507			
Adjusted R^2	0.2251				0.2459			
F-statistic	46.11	(on 7 and 1080 DF)			51.63	(on 7 and 1080 DF)		
p-value	**<10^{-16}**				**<10^{-16}**			

***** Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; SW 25,50,100: Sliding window size 25,50 and 100, respectively; T: Technical indicators included amongst the covariates; D: Forecast horizon of one day; SVM: Prediction method employed is SVM

2.5.2.2 Analysis of Different Market Maturity Classification Methods

Our earlier results show that the market maturity is an influential methodological factor regarding prediction performance. However, there are several stock market maturity classification methods. For example, Claessens et al. (2006) and Choong et al. (2010) classify stock market development based on the income level data of World Bank (as used in our main analysis), Zunino et al. (2009) use the Morgan Stanley Capital Index (MSCI), and Kim and Shamsuddin (2008) adopt the International Monetary Fund (IMF) classification.

To examine whether the adoption of different classification methods could lead to a different conclusion regarding the impact of market maturity, we estimate the following regression models to explain predictive accuracy:

$$\text{Accuracy} = \alpha + \beta_{\text{IMFE}} IMFE + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVM + \varepsilon, \tag{11}$$

$$\text{Accuracy} = \alpha + \beta_{\text{MSCIE}} MSCIE + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{SVM}} SVM + \varepsilon, \tag{12}$$

where *IMFE*, *MSCIE*, *ST*, *T*, *D* and *SVM* are dummy variables, taking the value 1 for IMF emerging markets, MSCI emerging or frontier markets, a static model simulation, when using technical indicators alongside basic price covariates (open, high, low, close and change), a daily forecast horizon and SVM prediction, respectively, and 0 otherwise. That is, the only difference between the main analysis and this additional analysis is the definition of the market maturity variable. A market can be classified as a matured market in one method but an emerging market in another method. For example, Lithuania and Latvia are in the high income group of the World Bank and are classified as emerging markets by the IMF. Table 11 provides details of the manner in which different markets are classified using the World Bank, the IMF and the MSCI classification systems. These models are also used to explain ROI.

Table 12 and Table 13 summarize the results of the regression analyses for accuracy and ROI, respectively. All the estimated coefficients of IMFE and MSCIE are negative and statistically significant. That is, predictive accuracy and the profitability of model-based trading are higher in the markets with greater (cf. lower) maturity. This confirms the conclusion we drew based on the World Bank classification system.

Table 11: WORLD BANK, IMF and MSCI MARKET CLASSIFICATIONS

World Bank		IMF		MSCI	
High Income	Middle Income	Advanced	Emerging	Advanced	Emerging or Frontier
US, Canada, Japan, Korea, Hong Kong, Singapore, France, UK, Italy, Germany, Switzerland, Spain, Netherland, Belgium, Portugal, Sweden, Norway, Denmark, Finland, Austria, Czech, Lithuania, Estonia, Latvia	Brazil, , China, Malaysia, Thailand, Indonesia, Hungary, South Africa, Turkey	US, Canada, Japan, Korea, Hong Kong, Singapore, France, UK, Italy, Germany, Switzerland, Spain, Netherland, Belgium, Portugal, Sweden, Norway, Denmark, Finland, Austria, Czech, Estonia	Brazil, China, Malaysia, Thailand, Indonesia, Hungary, South Africa, Turkey, Lithuania, Latvia	US, Canada, Japan, Hong Kong, Singapore, France, UK, Italy, Germany, Switzerland, Spain, Netherland, Belgium, Portugal, Sweden, Norway, Denmark, Finland, Austria	Brazil, China, Korea, Malaysia, Thailand, Indonesia, Hungary, South Africa, Turkey, Czech, Lithuania, Estonia

This table lists the countries of each market maturity classification standard.

Table 12: REGRESSION ANALYSIS OF PREDICTIVE ACCURACY WITH IMF AND MSCI MARKET CLASSIFICATIONS

Predictive accuracy	Eq. (14)				Eq. (16)			
	Estimated Coefficient	Std. Error	t value	p value	Estimated Coefficient	Std. Error	t value	p value
(Intercept)	0.5088	0.0023	220.69	**<10^{-16}**	0.5118	0.0023	226.20	**<10^{-16}**
IMF (IMFE)	-0.0109	0.0022	-5.023	**<10^{-6}**				
MSCI (MSCIE)					-0.0163	0.002	-8.284	**<10^{-15}**
Model simulation methodology (ST)	0.0086	0.002	4.342	**<10^{-4}**	0.0086	0.002	4.506	**<10^{-5}**
Covariate composition (T)	-0.00007	0.002	-0.037	0.97	-0.00007	0.002	-0.039	0.969
Forecast horizon (D)	0.0174	0.002	8.795	< **2^{-16}**	0.0174	0.002	9.128	**<10^{-16}**
Prediction method (SVM)	0.0106	0.002	5.344	**<10^{-6}**	0.0106	0.002	5.546	**<10^{-7}**
Residual standard error	0.02311				0.02227			
R^2	0.218				0.274			
Adjusted R^2	0.2108				0.2672			
F-statistic	30	(on 5 and 538 DF)			40.6	(on 5 and 538 DF)		
p-value	**<10^{-15}**				**<10^{-15}**			

***** Base model indicated by values of experimental factors given in brackets: IMFE: Emerging market as classified by IMF classification; MSCIE: Emerging market as classified by MSCI; ST; Static simulation methodology; T: Technical indicators included amongst the covariates; D: Forecast horizon of one day; SVM: Prediction method employed is SVM

Table 13: REGRESSION ANALYSIS OF ROI WITH IMF AND MSCI MARKET CLASSIFICATIONs

ROI	Eq. (15)				Eq. (17)			
	Estimated Coefficient	Std. Error	t value	p value	Estimated Coefficient	Std. Error	t value	p value
(Intercept)*	1.1749	0.0228	51.543	**$<10^{-16}$**	1.1965	0.0227	52.685	**$<10^{-16}$**
IMF Emerging (IMFE)	-0.1146	0.0215	-5.331	**$<10^{-6}$**				
MSCI Emerging Frontier (MSCIE)					-0.1448	0.0197	-7.342	**$<10^{-12}$**
Model simulation methodology (ST)	0.1021	0.0196	5.210	**$<10^{-6}$**	0.1021	0.0192	5.326	**$<10^{-6}$**
Covariate composition (T)	-0.0168	0.0196	-0.856	0.3925	-0.0168	0.0192	-0.875	0.3820
Forecast horizon (D)	-0.0442	0.0196	-2.256	**0.0245**	-0.0442	0.0192	-2.306	**0.0215**
Prediction method (SVM)	0.1163	0.0196	5.939	**$<10^{-8}$**	0.1163	0.0192	6.071	**$<10^{-8}$**
Residual standard error	0.2284				0.2235			
R^2	0.1523				0.1888			
Adjusted R^2	0.1444				0.1813			
F-statistic	19.33	(on 5 and 538 DF)			25.04	(on 5 and 538 DF)		
p-value	**$<10^{-15}$**				**$<10^{-15}$**			

***** Base model indicated by values of experimental factors given in brackets: IMFE: Emerging market as classified by IMF classification; MSCIE: Emerging market as classified by MSCI; ST; Static simulation methodology; T: Technical indicators included amongst the covariates; D: Forecast horizon of one day; SVM: Prediction method employed is SVM

2.5.3 The effect of transaction cost on ROI

Shleifer and Vishny (1997) suggest that the costs of arbitrage, such as transaction costs, may allow mispricing to occur.;thus the prices exhibit co-movements or recognizable patterns. But the transaction cost may prevent arbitrageurs from exploiting these co-movements. To examine the effect of transaction cost on ROI, we calculate the ROI with the transaction cost as 0.1%. This is a conservative value given that Shleifer and Vishny (1997) use 0.5%. The results are presented in Figure 4. We present the results regarding the prediction horizon factor as the results of other factors are similar to what we presented earlier. We can see that the ROIs from most simulations are lower than 1, which means no profit is earned. With our trading simulation, a trade is placed every period according to the generated prediction from the model. It is not surprising that the ROIs in the hourly setting are much lower than those in the daily setting as there are eight or more opening hours in one trading day. We do not find evidence against the prediction of the EMH, and our results show that it is important to account for transaction cost in discussing excess returns.

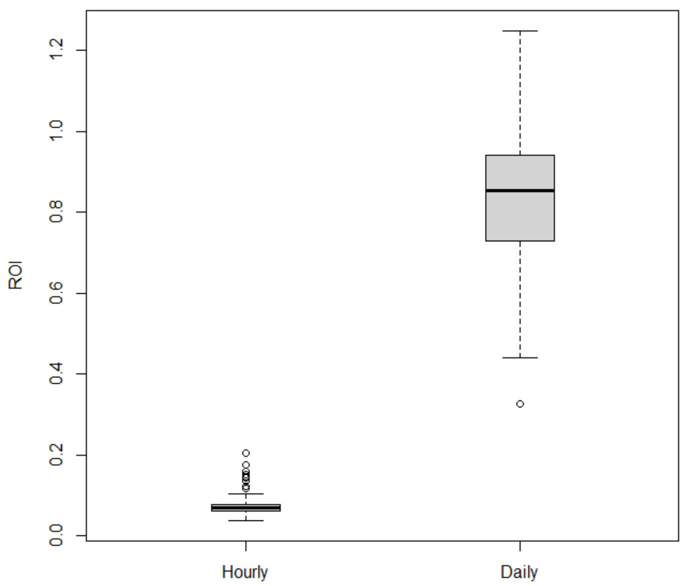


Figure 4: PREDICTION PERFORMANCE IN ROI ACROSS DAILY AND HOURLY DATA

The boxplot shows the ROI of all experiments in terms of daily and hourly data. The transaction cost is 0.1% for each trade, and there is one trade for each period. In other words, there is one trade every day in daily data and one trade every hour in hourly data.

2.6 Discussion

We first discuss our results relating to the factors that influence the prediction accuracy and the profitability of trading on the forecasts based on the most commonly employed ML techniques (SVM and ANN). These results demonstrate that the selection of the data source (i.e., the financial market being forecast) significantly and substantially influences the observed level of prediction accuracy and the profitability of trading on the forecasts. The impact on the profitability of trading on the forecasts is particularly large (Cohen's d equal to 0.6264). However, we cannot accept hypotheses *H1a* and *H1b* because the direction of the influence differs from expectations. Specifically, we find that predictive accuracy and the profitability of model-based trading are higher in financial markets with high (cf. medium) income levels. This is surprising since regulations and rules related to information disclosure should be further developed in such markets. On the other hand, the markets in the high- and medium income groups differ in many ways. It may be that the middle-income markets we consider are sufficiently regulated to rule out the effect of information disclosure policies on market predictability. An analysis of the predictability of low-income markets might give further insights into which other factors might govern the relationship between market income-level and predictability, and thus explain the result observed. Unfortunately, intraday time series data of sufficient length for markets with low income levels is not available from our data provider (TickWrite Data Inc.[3]). Consequently, a focused comparison of high *vs*. low income markets with forecasting horizons of one day or above might be a fruitful avenue for future research. An important implication of our findings is that the financial market selected for a forecasting study has a significant effect on the resulting accuracy of the forecasts. This implies that results which have been observed for one market may not be generalized to a different market. Drawing conclusions from findings derived from a single market, as is undertaken in the majority of existing ML studies, may mask the true degree of informational efficiency in financial markets. Rather, researchers are well-advised to test novel forecasting methods on multiple financial time series. This will offer a clearer picture of the relative advantages of competing methods and enhance the generalizability of empirical studies.

Our results offer support for hypotheses *H2a* and *H2b*, that predictive accuracy and the profitability from model-based trading are higher when evaluating forecasting models using a static train/test set approach (cf. a dynamic approach). The results suggest that the selection of the cross validation method is influential on prediction performance. This suggests that there are considerable risks in drawing conclusions regarding market efficiency based on the results of a single cross validation method.

[3] http://www.tickdata.com/

The results suggest that the prediction performance of ML techniques are more sensitive to the size of samples than to the use of more recent price information. ML techniques make predictions by recognizing patterns. One potential price pattern is long-term memory, which describes stock price behavior with autoregressive models (Greene and Fielitz, 1977; Lo, 1991; Mandelbrot, 1971). Our results suggest that the size of the training sample is a key ingredient of prediction performance. In our experimental settings, there are over 700 samples in the training set for a static approach. The width of a sliding window is fixed at 50, so the training set is always the latest 50 samples before the predicted sample. Hence, the higher accuracy we observe from the static approach implies that the size of the training set is at least as important as recent price information in predicting stock prices. On the other hand, doubling the size of the training window to 100 observations, as we have done in the sensitivity analysis, has not increased predictive accuracy substantially over the 50 observation window size. This suggests that the trade-off between the recency and size of data in financial time series forecasting experiments is indeed complex and would benefit from future research. However, from a practical point of view, ignoring the possibility and necessity to update a prediction model, the static approach appears naïve and is very unlikely to be adopted. For example, consider an experiment where stock prices are forecast one day into the future, and we have a financial time series of four years. Assuming there are roughly 350 trading days per year, this gives 1400 data points. A 70:30 train/test set split will reserve 420 of these data points (i.e., trading days) as a hold-out test set. Such an evaluation implies that a trader uses her stock prediction model for more than a year without updating. We argue that this is not realistic. More importantly, a static model methodology is vulnerable to selecting a test set that does not represent the overall population well (i.e., lucky sample effect).

Our results lead us to reject *H3a* and *H3b*. In particular, our results suggest that the performance of a forecasting model that includes technical indicators is not significantly better than one that uses basic reference prices as covariates. We find that even advanced nonlinear prediction methods such as SVM and ANN, which are able to discern complex relationships among covariates and the target variable, are unable to distill predictive information from technical indicators beyond that contained in basic price covariates. Consequently, our results further support previous criticism of technical indicators (Fama, 1970; Lesmond et al., 2004) and cast doubt on their value for predictive modeling.

Our results also suggest that the length of the forecast horizon has a significant effect on predictive accuracy and the profitability from model-based trading. However, the forecast horizon appears to have a different effect on the predictability (cf. profitability) of model-based trading. In particular, we find evidence in favor of *H4b*, in that the profitability of model-based trading is higher for an hourly (cf. daily) forecast horizon. On the other hand, predictive accuracy is significantly higher if forecasting price movements a day (cf. hour) into the future. Furthermore, the size of the forecast horizon effect on the profitability of model-based trading is small (Cohen’s d equal to 0.1795), whereas the forecast horizon effect on predictive accuracy is actually the largest effect observed in the study (Cohen’s d equal to 0.7105). We suggest that two effects maybe important to explain these results. First, a market that is more volatile is ceteris paribus harder to predict. Moreover, the (weak form) EMH suggests that stock prices can take some time to reflect all information. Consequently, volatility is higher in short-term forecasting, which, in turn, suggests lower predictive accuracy. However, in terms of the profitability of model-based trading, another effect comes into play. Second, forecasting shorter horizons creates more opportunities to trade on model predictions. We can only trade one daily forecast in a day, but multiple hourly forecasts. Furthermore, hourly forecasts and trading allow an investor to exploit the variation of prices within a day. In summary, the volatility argument predicts less accuracy from model-based trading for a forecasting horizon of one hour (cf. one day). However, the feasibility of higher trading frequency and the opportunity to capitalize on intraday price movements suggests that ROI might be higher in the intraday setting. Consequently, the positive effect on profitability for higher frequency trading may be larger than the negative influence of volatility. Given that predictive accuracy is not directly linked to trading frequency and intraday price movements, it is certainly plausible that we observe higher accuracy for a daily forecast horizon. There, higher intraday volatility is the only relevant effect, and this effect has a negative influence on accuracy.

It is important to note that the results observed in the forecast horizon comparison have implications for market efficiency. In particular, the review of previous forecasting studies indicates a bias toward forecast horizons of one day or above in the ML literature. Many such studies report very high accuracies (see Table 1), which, in the light of the EMH, is surprising. However, the results observed in this study demonstrate that predictive accuracy and trading profit are not perfectly correlated in financial time series forecasting. Market efficiency is only questioned if empirical evidence suggests that trading on model-based predictions produces excessive returns. Hence, observing high accuracy contradicts the EMH only if the corresponding prediction models facilitate profitable trading. We observe that this is not necessarily the case. More specifically, we find a forecast horizon of one day, the setting predominantly considered in previous ML research, to be associated with higher accuracy. However, we demonstrate that this forecast horizon is associated with less model-based trading profit than a shorter horizon of one hour. Furthermore, we find that the profit of the models decreases significantly when transaction cost is considered, which suggests that the high prediction accuracy does not guarantee excess returns in financial markets. Consequently, the disagreement between the financial economics and the ML literature concerning market efficiency might be less than the results of published studies suggest, and we do not find evidence against the prediction of the EMH.

Our results also support *H5a* and *H5b* that SVM predicts price movements more accurately than ANN and also produce higher trading profits. Furthermore, Cohen’s d suggests that the effect is large (0.4154 for accuracy and 0.4842 for ROI). This finding is interesting given the recent history of the application of SVM and ANN. The former gained much popularity in early 2000, leading to a partial replacement of standard feedforward ANN. However, with the advent of extreme learning machines (Huang et al., 2006) and, more recently, deep neural networks (Schmidhuber, 2015), attention has shifted back to neural prediction models. It seems likely that such advanced types of neural networks will be applied in the domain of financial time series forecasting. An implication of our results is that studies which aim to investigate the potential of these and advanced prediction methods for financial time series forecasting should routinely compare them with SVM since SVM predict price movements with high accuracy. In other words, SVM represents an important benchmark for assessing the marginal utility of new methods.

Our results also support *H6a* and *H6b* that the best ML model (which we show to be SVM) predicts price movements more accurately than the best econometric method (which we show to be AR) and also produces higher trading profits. Furthermore, Cohen's d suggests that the effect is sizeable (0.51 for accuracy and 0.66 for ROI). These results are in line with the literature that ML techniques outperform econometric methods (Donaldson and Kamstra, 1999; Pai and Lin, 2005). Our results, therefore, show that ML techniques, such as SVM and ANN, are useful techniques for detecting market anomalies. The conventional approach in the financial economics literature (Fama, 1970; Fama and French, 1993) is to use autocorrelation and linear regression models to examine the relation between explanatory factors and stock prices. (e.g., Keim, 1983; French, 1980). ML methods work in a different way. They are trained to recognize patterns in a data-driven manner and do not require human intervention. We find that such an approach facilitates profitable model-based trading, even when using models with fairly naive covariates. Both ANN and SVM achieve an ROI greater than one in most of our settings and the reference setting in particular (see Figures 1 and 3). Clearly, studies that scrutinize the degree to which financial markets are efficient must ensure that a modeling method is employed that fully exploits all predictive information contained in the covariates. Our results indicate that ML methods are well suited for this task and, thus, deserve a place in the financial economists' toolbox.

2.7 Conclusions

The EMH predicts that excess returns cannot be earned in a systematic way, by, for example, model-based trading. However, many ML-based financial time series forecasting studies seem to find ways to anticipate market developments with surprisingly high accuracy. The direction of price movements in these studies is often predicted with 80 percent accuracy and above. Some studies also report that their models facilitate profitable trading (Bitvai and Cohn, 2014). We set out to clarify the origins of the apparent contradiction between the ML and EMH literature. To that end, we perform an extensive forecasting benchmark in which we use two established ML methods to predict price movements in most major stock markets, and we compare, what we show to be, the best ML and econometric models, for predicting these movements. This study, to our best knowledge, is the first to compare intraday and daily ML and econometric prediction models across most major markets.

We find that the maturity of a financial market, the prediction method, the horizon for which it generates forecasts, and the methodology to simulate model-based trading all have a significant effect on market predictability and the feasibility of profitable model-based trading. Consequently, decisions that forecasting studies have taken with respect to these factors can help explain the results observed. This is not true for our last experimental factor, covariate composition since we find that popular technical indicators are no more predictive than basic reference prices.

Overall, we do not find overwhelming evidence which contradicts the EMH, since the results cannot be said to pass Fama's (1998) tests of endurance, homogeneity, and robustness; the EMH acknowledges that stock prices are partly predictable in the short run in some markets. Most of the predictive accuracies we observe are well below 60 percent. However, the level of accuracy we observe is substantially lower than that commonly published in previous research, and it is acknowledged that we might have increased the levels of accuracy by including more sophisticated covariates. Consequently, we would suggest that inefficiency may exist in some markets.

However, the EMH is a theory related to general market behavior and acknowledges that anomalies can occur at certain times and in certain situations. In this context, the importance of our results lies in the insight they provide into the methodological issues which might explain the difference between the prevailing view in support of the EMH in the financial economics literature and the high accuracy in predicting financial market prices achieved in ML studies. Clearly, our results suggest that the econometric models generally employed in the financial economics literature may have led to overly pessimistic views of the degree to which financial price series can be predicted. It is clear from our results that practitioners interested in predicting financial time series are well-advised to consider ML techniques in their arsenal of methods. In fact, to aid this process, we show in Table B.1 in online Appendix B, the experimental settings that have given the best results for each individual market.

Furthermore, our analyses help to provide a realistic estimate of the potential and limitations of ML techniques for financial index forecasting. In particular, they suggest that the following factors may have given a false sense of the degree to which the results of ML studies contradict the EMH: First, there may be a bias toward the specific markets studied, with an emphasis on those that might be easier to predict. Second, there may have been a focus on less suitable model evaluation methodologies (i.e., a static approach as opposed to sliding-window cross-validation), which are prone to give optimistic estimates of model accuracy. Third, the focus on forecast horizons of one day and above in prior research may have led to a false impression of the predictability of financial markets. In summary, our study provides some evidence that it would be unwise to draw conclusions regarding the degree to which financial markets, in general, are inefficient from the accuracy figures reported in some previous ML financial time series forecasting studies.

The main implication of the EMH is that, in an informationally efficient market, it is not possible to obtain systematic, excessive returns from trading the predictions of a forecasting model. The results of our study indicate an imperfect link between the predictive accuracy of a forecasting model and the profitability of trading on the model's forecasts. For example, we show that predicting the direction of price movements one day into the future produces higher accuracy compared to predicting prices one hour into the future, but produces a lower return from model-based trading. Consequently, through concentrating on forecast horizons of one day and above, the predictive accuracy that previous studies observe tends to be higher than what would be observed if examining shorter, intraday, forecast horizons. However, our results suggest that even then this does not necessarily imply that the corresponding models would facilitate profitable trading. For example, we find a forecast horizon of one day to be associated with relatively lower ROI and the returns observed have not taken account of transaction costs. Furthermore, most models fail to generate profit after transaction cost is considered. This makes our results consistent with the implications of the EMH (Malkiel, 2003).

Our study has focused on predicting national stock indices because these indices are used in the majority of previous ML studies that predict the direction of price changes. However, to confirm the robustness of our findings, it would be useful to extend the analyses to the prediction of stock prices associated with a variety of industry sectors. It would also be useful to incorporate data from a wider range of middle income markets and to employ higher frequency data, once sufficient data becomes available. It should also be noted that due to data availability issues outlined in section 4.1, our main analyses were conducted using data from 2008-14, a period which covered two financial crises, the US subprime mortgage crisis, and the European sovereign debt crisis. Clearly, these unexpected events may have influenced our results. However, additional analyses, reported in online Appendix C, for the 13 stock indices for which we were able to obtain a longer time series of data, produced similar conclusions to the earlier analyses. This gives some comfort that our results are robust to unexpected events such as financial crises. It would be valuable if future studies which employ longer time series of data could confirm this view.

Our results for these hypotheses hold important implications for the machine learning literature since the influence of the methodological factors is severely underestimated in predicting financial markets. The lack of awareness of the methodological factors leads to the biased performance of the prediction model, and the preference of positive results might also play a role in the over-optimistic prediction performance of the published studies. Specifically, many machine learning studies claim to develop novel prediction models which can achieve very high accuracy, i.e. 80% to 90%, but these models are usually tested with only a simple setting, such as daily data in a single market. In other words, these studies may suffer from the over-fitting hazard as well as the lack of general applicability. Since our results show that the change of one methodological factor could affect the prediction performance significantly, we argue that the proposed prediction models should be tested in multiple settings in future machine learning studies.

Our results also have implications for the financial economics literature. In particular, we provide empirical evidence for the ability of advanced ML techniques in the form of SVM and ANN to detect market anomalies across many major financial markets. ML methods are rarely employed by financial economists. However, these techniques can capture complex nonlinear interactions in a financial data set and approximate their relationship to a target variable. This is probably the reason we find evidence that the best ML method (SVM) outperforms the best econometric method (AR) when predicting financial prices. Consequently, our results suggest that ML methods offer the prospect of studying informational efficiency and, thus, providing financial economists with new insights concerning the manner in which financial markets employ information.

2.8 Online Appendix A: Comparing the forecasting ability of commonly employed econometric techniques

We examine the forecasting ability of three of the most widely used econometric methods for forecasting financial prices, namely, AR, ARIMA, GARCH (Charles, Darné, & Kim, 2011). ARIMA is also one of the most widely used benchmarking methods in machine learning studies (George S Atsalakis and Valavanis, 2009). We follow Pai and Lin's (2005) settings for ARIMA and, hence, we do not include technical indicators. We follow Awartani & Corradi (2005) by using GARCH(1, 1) and we employ the AR(1) model with the innovation distribution defined as Gaussian with constant variance. Econometric methods are generally employed to predict the change in the value of an index. However, to compare the results with those studies employing ML techniques, we convert the predictions of change in value to predictions of change in direction.

To achieve these objectives, we estimate the following regression model for predictive accuracy:

$$\text{Accuracy} = \alpha + \beta_{\text{MI}} MI + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{ARIMA}} ARIMA + \beta_{\text{GARCH}} GARCH + \varepsilon, \qquad (13)$$

where *MI*, *ST*, *T*, *D*, *ARIMA* and *GARCH* are dummy variables, taking the value 1 for middle income markets, a static model simulation, when using technical indicators alongside basic price covariates, a daily forecast horizon, ARIMA prediction and GARCH prediction, respectively, and 0 otherwise. That is, AR is the base model. This model is also used to explain ROI.

The results relating to predictive accuracy and ROI are shown in Tables A.1 and A.2, respectively. The estimated coefficients of ARIMA (-0.0162) and GARCH (-0.0072) are negative and statistically significant in the regression reported in Table A.1, indicating that AR outperforms ARIMA and GARCH in terms of accuracy. The coefficient of ARIMA is negative (-0.072) and statistically significant, and the coefficient of GARCH is not statistically significant in the regression reported in Table A.2. These indicate that in terms of ROI, AR outperforms ARIMA and there is no difference between AR and GARCH. Taking both accuracy and ROI, into account, AR out-performs GARCH and ARIMA. Hence, we compare AR with SVM in our main analysis.

TABLE A.1: REGRESSION ANALYSIS OF PREDICTIVE ACCURACY

Predictive accuracy	Estimated Coefficient	Std. Error	t value	p value
(Intercept)*	0.5081	0.0028	181.21	**< 10^{-16}**
Market maturity (MI)	0.0034	0.0029	1.1702	0.2426
Model simulation	0.0114	0.0024	4.6725	**<10^{-5}**

methodology (ST)				
Forecast horizon (D)	0.0037	0.0024	1.5307	0.1266
Prediction method (ARIMA)	-0.0162	0.003	-5.4513	**< 10^{-7}**
Prediction method (GARCH)	-0.0072	0.003	-2.4173	**0.0161**
Residual standard error	0.02458	df	402	
R^2	0.1211	Adjusted R^2	0.1102	
F-statistic	11.08	(on 5 and 402 DF)		
p-value	**< 10^{-9}**			

* Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; D: Forecast horizon of one day; ARIMA: Prediction method employed is ARIMA; GARCH: Prediction method employed is GARCH.

TABLE A.2: REGRESSION ANALYSIS OF ROI

ROI	Estimated Coefficient	Std. Error	t value	p value
(Intercept)*	1.0537	0.0268	39.337	**< 10^{-16}**
Market maturity (MI)	-0.0023	0.0274	-0.0851	0.9322
Model simulation methodology (ST)	0.1791	0.0233	7.7007	**< 10^{-12}**
Forecast horizon (D)	-0.0324	0.0233	-1.3934	0.1643
Prediction method (ARIMA)	-0.072	0.0285	-2.5276	**0.0119**
Prediction method (GARCH)	0.0007	0.0285	0.0258	0.9794
Residual standard error	0.2348	df	402	
R^2	0.148	Adjusted R^2	0.1374	
F-statistic	13.97	(on 5 and 402 DF)		

p-value	**< 10^{-11}**

* Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; D: Forecast horizon of one day; ARIMA: Prediction method employed is ARIMA; GARCH: Prediction method employed is GARCH.

We compare the prediction performance of the most commonly employed econometrics methods (ARIMA, GARCH and AR: Charles, Darné, & Kim, 2011) with that of ANN and SVM. We adopt ARIMA, GARCH and AR among the well-known econometrics methods. We employ the same settings for the econometrics models indicated above, namely GARCH(1, 1) and AR(1) with the innovation distribution defined as Gaussian with constant variance. Overall, we obtain 136 simulation results for each of the prediction models (AR, ARIMA, GARCH, ANN, SVM), resulting in 680 simulation results in total. We estimate the following regression model to explain predictive accuracy:

$$\begin{aligned}\text{Accuracy} = \alpha + \beta_{\text{MI}}MI + \beta_{\text{ST}}ST + \beta_{\text{D}}D + \beta_{\text{ANN}}ANN + \beta_{\text{AR}}AR \\ + \beta_{\text{ARIMA}}ARIMA + \beta_{\text{GARCH}}GARCH + \varepsilon,\end{aligned} \tag{14}$$

where *MI, ST, T, D, ANN, AR, ARIMA and GARCH* are dummy variables, taking the value 1 for middle income markets, a static model simulation, a daily forecast horizon, ANN prediction, AR prediction, ARIMA prediction, GARCH prediction (with SVM as reference prediction), respectively, and 0 otherwise. This model is also used to explain ROI.

The results are displayed in Table A.3. In the regressions associated with both accuracy and ROI, all the coefficients of *ANN*, *AR*, *ARIMA* and *GARCH* are negative, indicating that SVM, the reference model, outperforms the other models. In addition, the coefficients of ANN are higher than those of *AR*, *ARIMA* and *GARCH* for both accuracy and ROI. Consequently, our results are consistent with the literature, that suggests machine learning techniques outperform traditional econometric methods when predicting stock prices (Pai and Lin, 2005).

Table A.3: REGRESSION ANALYSIS OF PREDICTION TECHNIQUES

	Predictive accuracy				ROI			
	Estimated Coefficient	Std. Error	t value	p value	Estimated Coefficient	Std. Error	t value	p value
(Intercept)*	0.5197	0.0025	204.438	**< 10^{-15}**	1.2439	0.0242	51.379	**< 10^{-15}**
Market maturity (MI)	-0.0016	0.0022	-0.720	0.4718	-0.0582	0.0211	-2.756	**0.006**
Model simulation	0.0102	0.0019	5.426	**< 10^{-7}**	0.1470	0.0180	8.207	**< 10^{-14}**

methodology (ST)								
Forecast horizon (D)	0.0097	0.0019	5.176	**<10^{-6}**	-0.0268	0.0180	-1.498	0.135
ANN	-0.0108	0.0030	-3.634	**0.0004**	-0.1240	0.0283	-4.380	**<10^{-4}**
AR	-0.0129	0.0030	-4.351	**<10^{-4}**	-0.1638	0.0283	-5.785	**< 10^{-7}**
ARIMA	-0.0292	0.0030	-9.816	**< 10^{-15}**	-0.2358	0.0283	-8.326	**< 10^{-15}**
GARCH	-0.0201	0.0030	-6.774	**< 10^{-10}**	-0.1631	0.0283	-5.759	**< 10^{-7}**
Residual standard error	0.0245	df	672		0.2335	df	672	
R^2	0.1961				0.1847			
Adjusted R^2	0.1877				0.1762			
F-statistic	23.42	(on 7 and 672 DF)			21.75	(on 7 and 672 DF)		
p-value	**< 10^{-15}**				**< 10^{-15}**			

* Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; D: Forecast horizon of one day; ANN: Prediction method employed is ANN; AR: Prediction method employed is AR; ARIMA: Prediction method employed is ARIMA; GARCH: Prediction method employed is GARCH.

2.9 Online Appendix B: Identifying the Prediction Method Producing the Best Accuracy and ROI for Each Stock Index

Table B.1 displays the setting of the experiments with the best performance for each stock index. As we discuss in the results section in the main paper, certain settings are positively influential on performance. For example, daily prediction appears more often than hourly prediction in the following table, indicating that daily predictions are generally more accurate and produce higher ROI than hourly predictions. Similarly, the static setting appears more often than the sliding window setting, indicating that predictions produced using the static setting are generally more accurate and produce higher ROI than predictions derived from a sliding window setting. In Table B.1, mature markets are more often those that display higher accuracy figures; e.g., 0.58 in France and 0.62 in Denmark. It is worth noting that our analysis, discussed in the main paper, suggests that SVM outperforms ANN significantly in terms of accuracy. However, ANN achieves the highest accuracy in more markets than SVM. We can explain the contrast by looking at the ANN and SVM boxplot in Figure 1: the average accuracy of SVM is higher, but there are more outliers for ANN, i.e., SVM performs better than ANN on average, but ANN occasionally achieves good performances across the combination of settings.

TABLE B.1: SETTING OF BEST ACCURACY AND ROI FOR EACH STOCK INDEX

		Setting of Highest Accuracy					Setting of Highest ROI				
Market	Index	Accuracy	Model	Method	Tech. Indicators	Horizon	ROI	Model	Method	Tech. Indicators	Horizon
Netherland	AEX	0.56	ANN	Static	No	Daily	1.3	SVM	Static	No	Hourly
Austria	ATX	0.56	SVM	Static	No	Daily	1.79	SVM	Static	No	Hourly
Belgium	BEL20	0.55	SVM	Static	Yes	Daily	1.37	ANN	Static	No	Daily
Brazil	Brazilian Bovespa Futures	0.54	ANN	Sliding Window	No	Daily	1.4	ANN	Sliding Window	No	Daily
Hungary	BUX	0.53	SVM	Sliding Window	Yes	Hourly	1.38	SVM	Sliding Window	No	Hourly
France	CAC 40	0.58	SVM	Static	No	Daily	1.79	SVM	Static	No	Daily
Germany	DAX	0.56	ANN	Static	Yes	Daily	1.39	ANN	Static	No	Daily
US	Dow Jones Industrial Average	0.55	ANN	Static	No	Daily	1.29	SVM	Static	No	Hourly
UK	FTSE 100	0.56	ANN	Static	No	Daily	1.27	SVM	Static	Yes	Hourly
Finland	OMXH25	0.53	ANN	Static	No	Hourly	1.59	ANN	Static	No	Hourly

Hong Kong	Hang Seng Index	0.53	ANN	Static	No	Daily	1.22	ANN	Sliding Window	Yes	Daily
Spain	IBEX 35	0.54	ANN	Static	No	Daily	2.19	SVM	Static	Yes	Hourly
Italy	FTSE MIB Index	0.54	ANN	Static	No	Daily	1.62	ANN	Static	Yes	Hourly
Indonesia	Jakarta Composite Index	0.56	ANN	Static	No	Daily	1.7	ANN	Sliding Window	No	Hourly
Denmark	OMX Copenhagen Index	0.62	ANN	Static	No	Daily	1.49	ANN	Static	No	Daily
Malaysia	FTSE Bursa Malaysia KLCI Index	0.54	ANN	Static	No	Daily	1.22	ANN	Sliding Window	Yes	Hourly
Korea	KOSPI 200 Index	0.57	SVM	Static	Yes	Daily	1.22	SVM	Static	Yes	Daily
US	NASDAQ-100	0.57	ANN	Sliding Window	No	Hourly	1.41	ANN	Static	No	Daily
Japan	Nikkei 225	0.55	ANN	Static	No	Daily	1.77	ANN	Static	Yes	Hourly
Norway	OSE All Share Index	0.55	ANN	Sliding Window	No	Daily	1.36	ANN	Static	Yes	Hourly
Portugal	PSI-20	0.54	ANN	Static	No	Daily	1.63	ANN	Static	Yes	Hourly
Czech	Prague Stock Exchange Index	0.53	SVM	Static	Yes	Daily	1.2	SVM	Static	Yes	Daily
Latvia	OMX Riga Index	0.54	ANN	Static	No	Daily	2.11	SVM	Sliding Window	No	Hourly
China	ShangHai SE Composite	0.54	SVM	Static	No	Daily	1.67	SVM	Static	No	Daily

	Index										
Sweden	OMX ALL-SHARE Stockholm Index	0.56	ANN	Static	No	Daily	1.36	ANN	Static	No	Hourly
US	S&P 500	0.57	ANN	Static	No	Daily	1.35	ANN	Static	No	Daily
Singapore	Straits Times Index	0.55	ANN	Sliding Window	No	Daily	1.2	ANN	Sliding Window	Yes	Daily
Switzerland	Swiss Market Index	0.57	ANN	Static	No	Daily	1.34	ANN	Static	No	Daily
Estonia	OMX Tallinn Index	0.53	ANN	Sliding Window	No	Daily	1.49	SVM	Static	No	Hourly
Thailand	Thai Stock Exchange MAI Securities Index	0.56	ANN	Static	No	Daily	1.74	ANN	Sliding Window	No	Hourly
South Africa	FTSE/JSE Africa Top40	0.55	ANN	Static	No	Daily	1.43	ANN	Static	No	Hourly
Canada	SP TSX Composite Index	0.58	ANN	Static	No	Daily	1.23	ANN	Static	No	Daily
Turkey	ISE-100	0.54	ANN	Static	No	Daily	2.4	ANN	Sliding Window	Yes	Hourly
Lithuania	OMX Vilnius Index	0.54	ANN	Static	No	Daily	1.31	ANN	Static	No	Daily

This table lists the setting of the experiments with the best performance (accuracy and ROI) for each stock index.

2.10 Online APPENDIX C: Analysis of Markets with Intraday Data Available Prior to 2008

In selecting the sample period for the main analysis, we wanted to maximize the number of markets we could use with sufficient daily and intraday data to test the hypotheses. However, the availability of intraday data was limited and, for many markets, was only available from 2008 onwards. To include as many markets as possible, we choose the period 2008 to 2014, where intraday data is available for most markets. The selection of the sample period may introduce sample selection bias. In order to examine whether this bias affected our conclusion, we conducted additional experiments with the thirteen stock indexes for which we could access intraday data for longer periods (listed in table C.1).

Table C.1 : STOCK INDICES WITH INTRADAY DATA AVAILABLE BEFORE 2008

NO.	Economy	World Bank Income Level	Index	Start Date	End Date
1	US	High	S&P 500	1/2/1983	19/2/2014
2	Japan	High	Nikkei 225	1/7/2003	19/2/2014
3	Korea	High	KOSPI 200 Index	1/2/2004	19/2/2014
4	Hong Kong	High	Hang Seng Index	1/12/2006	19/2/2014
5	France	High	CAC 40	1/7/2003	19/2/2014
6	UK	High	FTSE 100	1/7/2003	19/2/2014
7	Italy	High	FTSE MIB Index	1/7/2003	19/2/2014
8	Germany	High	DAX	1/7/2003	19/2/2014
9	Hungary	Middle	BUX	1/7/2003	19/2/2014
10	Switzerland	High	Swiss Market Index	1/7/2003	19/2/2014
11	Spain	High	IBEX 35	1/7/2003	19/2/2014
12	US	High	Dow Jones	1/4/1993	19/2/2014

			Industrial Average		
13	US	High	NASDAQ-100	2/1/1997	19/2/2014

This table lists the stock indexes for which the intraday data is available for longer periods.

We estimate regressions based on equations (8) to explain prediction accuracy and ROI, and the results are presented in Table C.2. The results relating to prediction accuracy are consistent with the results for the larger group of markets for which data was available for 2008-14 (reported in Table 6). We observe some differences in terms of the results relating to ROI compared to those for the larger group of markets for which data was available for 2008-14 (reported in Table 7). In particular, the coefficient of market maturity (MI), which is negative and significant at the 5% level in the results reported in Table 7, is still negative but the p-value increases to 0.0851. This may arise because only one stock index of the 13 indices included in this additional experiment is from a middle income market, i.e., a less mature market; the sample size may, therefore, be insufficient to observe a significant difference. In addition, the coefficient of forecast horizon (D) is positive in these additional experiments and is not statistically significant. As we discussed earlier, this can be regarded as evidence of an imperfect link between predictive accuracy and profitability. The remaining factors show similar results to those reported in Table 7. Overall, the result from the markets with earlier intraday data is in line with our earlier conclusion and suggests that selection bias did not influence the results from the main analyses reported in the paper.

Table C.2: REGRESSION ANALYSIS OF THE MARKETS WITH INTRADAY DATA BEFORE 2008

	Predictive accuracy				ROI			
	Estimated Coefficient	Std. Error	t value	p value	Estimated Coefficient	Std. Error	t value	p value
(Intercept)*	0.5047	0.0026	197.01	**< 10^{-15}**	1.0772	0.0541	19.901	**< 10^{-15}**
Market maturity (MI)	-0.0186	0.0043	-4.360	**<10^{-4}**	-0.1559	0.0901	-1.730	0.0851
Model simulation methodology (ST)	0.0047	0.0022	2.073	**0.0394**	0.1345	0.0480	2.802	**0.0056**
Covariate composition (T)	0.0001	0.0022	0.042	0.9663	-0.0149	0.0480	-0.310	0.7566
Forecast horizon (D)	0.0074	0.0022	3.258	**0.0013**	0.0216	0.0480	0.451	0.6528
Prediction method (SVM)	0.0080	0.0022	3.512	**0.0005**	0.1486	0.0480	3.094	**0.0023**
Residual standard error	0.0164	df	202		0.3462	df	202	
R^2	0.1863				0.0930			
Adjusted R^2	0.1662				0.0706			
F-statistic	9.251	(on 5 and 202 DF)			4.143	(on 5 and 202 DF)		
p-value	**<10^{-7}**				0.0013			

* Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; T: Technical indicators included amongst the covariates; D: Forecast horizon of one day; SVM: Prediction method employed is SVM

Chapter 3: Is experience the mother of risk wisdom?

Abstract— We examine the degree to which individuals modify their financial risk-taking behavior and learn to improve their decision performance as a result of gaining experience of a decision-making task. In particular, we examine whether increased experience tends to result in improvements in risk-adjusted (Sharpe ratio) performance. We achieve our objective by analyzing 6,287,477 trades of 27,868 individual UK spread-traders over a 10 year period, using linear mixed models. We find that, as individuals gain trading experience, they increase their degree of risk-taking and make higher returns. However, their returns are subject to greater volatility and, as a result, this leads to them achieving lower risk-adjusted returns. The conclusion holds after accounting for selection bias and survivorship bias. We discuss the implications for operators and regulators in financial markets and explore the implications for decision makers more generally.

Keywords: Learning; decision making; risk, performance

3.1 INTRODUCTION

The risk-taking behavior of individuals is arguably over-simplified in the financial economic theory. The equity premium puzzle is one of the scenarios which the financial economic theory fails to explain. In Thaler's (2010) example, a dollar invested in U.S. Treasury bills in 1926 would be worth about $14, while a dollar invested in U.S. stocks on the same date would be worth more than $2,000. Mehra and Prescott (1985) show that this huge return difference cannot be explained by the difference of risk alone. Some progress has been made with behavioral insights. Benartzi and Thaler (1995) argue that this puzzle can be understood with a combination of psychological concepts: loss aversion, which describes the tendency to weigh losses more heavily than gains, and mental accounting, which is the implicit method deployed to evaluate financial outcomes. Barberis, Huang and Santos (2001) develop an equilibrium model to explain the equity premium puzzle by adding the house money effect which captures the tendency to take higher risk after making a profit. This shows the need for more empirical evidence of the risk-taking behavior of individuals.

Recent empirical research has emphasized the influence of the outcomes of an individual's previous decisions on their subsequent risk-taking behavior (Choi et al., 2009; Glaser and Weber, 2009; Kaustia and Knüpfer, 2008; Y.-J. Liu et al., 2010; Thaler and Johnson, 1990). There is also strong laboratory evidence that individuals are likely to change their risk-taking behavior as they gain experience of a decision-making task (Camerer and Ho, 1999; Charness and Levin, 2003; Roth and Erev, 1995). However, there is little evidence that confirms these findings in real-world environments. Empirical evidence does support the view that an individual's accumulated experience improves some aspects of financial decisions; notably leading to higher returns and reduced behavioral bias (Chiang et al., 2011;

Feng and Seasholes, 2005; Nicolosi et al., 2009; Seru et al., 2010). However, the impact of an individual's experience on how they handle risk has been neglected; specifically, the impact of increased experience on the volatility of their returns and their risk-adjusted returns. This study aims to fill this gap.

We first examine the degree to which individuals change their risk-taking behavior as they gain experience and find that, in general, risk-taking behavior tends to increase. We explore these research questions within the context of trading in financial markets and are able to use a range of measures to assess the degree of risk taken by individuals (i.e. longer holding times, higher investment size and trading frequency). Our results hold across all these measures of risk-taking.

Second, we examine the impact of an individual's accumulated experience on their risk-related performance, measured in terms of the volatility of their returns and their risk-adjusted returns (Sharpe ratios: Sharpe, 1998). It is generally acknowledged that higher returns are accompanied by higher risk (Fama and MacBeth, 1973; Glosten et al., 1993; Markowitz, 1952). Consequently, it is our contention that previous studies, by focusing on assessing performance simply in terms of returns, may have produced misleading conclusions concerning the degree to which individuals learn as a result of experience. Our results suggest that traders *do not* learn to improve risk-adjusted performance. Rather, whilst experienced traders make higher profits, they suffer higher volatility of returns and decreases in risk-adjusted performance.

Finally, we observe the dynamic interaction of the three elements: experience, risk-taking behavior, and volatility of returns. We find that as traders gain experience, they increase the size of their investments and their trading frequency and also tend to hold positions for longer. Our results also show that longer holding times and higher investment size lead to higher volatility of returns. This provides a reasonable explanation for that traders increase the volatility of returns when they gain in experience. It is worth mentioning that trading frequency is negatively related with the volatility of returns. The positive relationship between trading risk and the level of risk-taking behavior is often assumed to exist. Our results show that this should not be taken for granted.

In sum, our study provides important insights into the degree to which an individual's risk-taking behavior and their risk-adjusted performance changes as a result of the accumulated experience of a decision-making task.

The remainder of the paper is structured as follows. In section 2, we discuss the literature that explores the degree to which individuals change their risk-related behavior and their decision performance in the light of experience. We describe the data and the methodology in section 3. In section 4, we present our results. We discuss the results in section 5 and conclude in section 6.

3.2 LITERATURE AND HYPOTHESES

3.2.1 Experience and Risk-taking Behavior

There is growing evidence that an individual's risk attitude may change as a result of their past experiences. For example, Thaler and Johnson (1990) document the house money effect, whereby prior gains encourage individuals to increase their risk-seeking behavior (confirmed in the context of trading by Liu et al., 2010). Other studies support the view that past experience influences subsequent risk-taking. For example, Kaustia and Knüpfer (2008) find that Finnish IPO investors with higher past returns were more likely to subscribe to the next IPO than those with lower past returns. Similarly, Choi et al. (2009) show that individuals who have achieved higher returns or lower return variance from their 401(k) retirement funds, which are workplace savings plans allowing employees to decide the portion of their salary before taxes which are saved in a specified investment fund in a tax-advantaged way, in the past, tend to increase their rate of saving in this fund. In addition, Glaser and Weber (2009) found that those individual investors who made higher past returns, tend to increase the level of their trading activities, i.e., portfolio turnover and number of transactions, and take greater risk, (e.g., buying higher risk stock and holding a less diversified portfolio).

As indicated above, previous research has demonstrated that the degree of risk an individual is prepared to take changes under different circumstances, e.g., via the house money effect. However, there is a dearth of research examining the effect of experience on risk-taking behavior. An exception is Chiang et al.'s (2011) finding that experienced IPO auctioneers take greater risk and bid aggressively. Consequently, we test the following hypotheses: *As an individual gains experience of a decision-making task they increase their risk-taking, in terms of: how long they are prepared to expose themselves to uncertainty (holding time)* (H1a), *the size of the stake they are prepared to put at risk* (H1b) *and how frequently they are prepared to engage in this risk taking activity* (H1c).

3.2.2 The Effect of Experience on Performance

Some studies show that experienced individuals demonstrate less behavioral bias. For example, List (2004) found that those with more experience were less prone to overvalue goods merely because of ownership (the 'endowment effect') and Dhar and Zhu (2006), Liu et al. (2010) and Gloede and Menkhoff (2011) report that trading experience reduces the tendency to, respectively, sell winning investments and to hold onto losing investments (the 'disposition effect'), to evaluate the risk of a new event in isolation without taking other risks into account and to demonstrate overconfidence.

Some studies report that greater trading experience is associated with higher investment returns (e.g., Feng and Seasholes, 2005; Seru et al., 2010; Nicolosi et al., 2009 and Linnainmaa, 2011). By contrast, Chiang et al. (2011) find that the returns of IPO investors decrease as they participate in more auctions. An important feature of the existing literature is that learning is measured by improvements in returns. However, the risk-return trade-off is a fundamental concept of decision-making in general and financial decision making in particular. For example, a principle of modern portfolio theory is that individuals should seek to maximize returns while holding risk constant or to minimize the risk while holding returns constant (Markowitz, 1952; Sharpe, 1964). Consequently, a risk-adjusted measure of performance is widely used to compare investments (e.g., stocks vs. bonds). However, the current literature examining the degree to which experience helps to improve performance pays little attention to the risk perspective of performance.

To fill this gap, we examine to what extent greater risk is taken by individuals as they gain in experience in a financially-based decision making task (displayed by greater volatility in their returns) and to what extent they improve their performance if the greater risk they assume is taken into consideration. To achieve this, we test the following hypotheses: *As an individual gains in experience of a financially-based decision making task they achieve higher returns* (H2a), *their returns display greater volatility* (H2b) *and they achieve higher risk-adjusted returns* (H2c).

3.2.3 The Relationship between Risk-taking Behavior and Returns Volatility

As indicated in 2.2, when testing H2b, we intend to use the volatility of an individual's financial returns as a proxy for the degree of aggression displayed in their risk-taking behavior. This assumption is often made (e.g., Chiang et al., 2011; Y.-J. Liu et al., 2010). However, there is little empirical evidence supporting the assumption, so we test the following hypotheses: *The volatility of an individual's returns increases when the time they expose themselves to uncertainty increases (H3a), the size of the stake they are prepared to put at risk increases (H3b), and when they engage in the risk-taking activity more frequently (H3c).*

3.3 METHODOLOGY

3.3.1 Data

The data used in this study was collected from a large spread-trading broker based in the U.K. Spread trading has developed rapidly in the U.K. since the 1990s, due to its relatively low transaction costs, the ease of access it provides to retail investors to international markets and because profits which accrue are tax free (Brady and Ramyar, 2006; Paton and Williams, 2005). Traders in spread-trading markets can decide to either buy or sell the market (e.g., a financial index). The profit and loss of each trade depend on the amount invested and how many points the index rises or falls in the direction that the trader predicted. For example, if a trader believes that the FTSE 100 will rise, they might 'buy the index' at, say, £50 per point. If the FTSE 100 rises 20 points and the trade is closed, the trader then makes £50 x 20 =

£1,000 profit. If the FTSE 100 falls 10 points and the trade is closed, the trader's loss is £50 x 10 = £500. Alternatively, traders can 'sell' the market, in which case profits/losses are made if the market falls/rises.

Spread trading data offers a number of advantages over traditional stock market data for the purpose of determining the degree and manner in which individuals alter their risk-taking behavior as they gain experience. In particular, spread trading is short-term trading, and on average spread traders make 3 or 4 round trades (open and close) within one hour (Gulthawatvichai et al., 2013). Consequently, all returns are realized and no estimation of gains is required. By contrast, shares purchased from conventional stock markets are normally held for a longer term with dividends paid in the future which often requires researchers to estimate the return of the stock purchased as they are not definite until sold. This process may lead to bias. For example, Seru et al. (2010) and Nicolosi et al. (2009) measure stock returns over 20- and 30-day periods following the purchase, respectively, and Barber and Odean (2002) assume that all trades occur on the last day of the month in estimating monthly returns. However, all trades in our dataset are closed, so the returns are realized and definite. Using such short-term trading data also reduces the possibility that any observed change in risk-taking behavior arises from changes in an individual's personal circumstances between trades.

The data contains a total of 6,287,477 trades made by 27,868 individual spread traders [24,554 (88%) and 3,314 (12%) are male and female traders, respectively] between October 2003 and March 2013. Each record in the database contains the following information relating to a closed trade: an individual's identification number, the time stamp, and price when each of the trades was opened and closed, whether the trade was to buy or sell the market and the amount invested. Variables, such as the measures of experience, risk and Sharpe ratios are calculated on a trade by trade basis.

Table XII summarizes the descriptive statistics relating to the data. The mean number of trades placed by a trader (300.4) is significantly higher than the third quartile (190), suggesting that a relatively small number of traders place a relatively high number of trades. The distribution of the time between a trader's first and last trade in the dataset is also right-skewed (median: 185 mean: 454.7, third quartile: 658), suggesting that some traders continue to trade much longer than others. The results also suggest that only around 25% of traders make profits and half of the trades being closed within 11.7 minutes of their opening, and over 75% of all trades being closed within 1 hour, indicating the generally short–term nature of spread-trading. Consequently, to avoid the tendency to draw overly-optimistic conclusions from the 'survivors' who place more trades or stay in the market longer than others, there is a need to control for survivorship bias (the detailed method will be discussed later).

Table XII Spread Trading Data Summary

	Mean	**1st Q**	**Median**	**3rd Q**
Panel A: Trader (27,868)				
Age	41	31	40	51
Total number of trades	226.7	10	44	180
Actively trading period (days)	436.7	23	179	637.2
Total Profit[1]	-67.6	-65	-22	0
Total investment size[2]	1323	16	83	433
Mean number weekly trades	763	1	3.5	9.3
Panel B: Trade (6,287,477)				

Holding time(minutes)	356.2	3.2	12.2	54.3
Investment size	4.5	1	1	4
Profit	-1.4	-7	1	6

[1]A trader's total profit is the sum of the profit of all trades and is defined as $\sum R_{ik} * ST_{ik}$, where R_{ik} is the number of points won or lost and ST_{ik} is the size of the investment of the i^{th} trade of trader k.

[2]A trader's total investment size is the sum of the investment size of all trades and is defined as $\sum ST_{ik}$ for trader k.

3.3.2 Variables

Experience

In order to make a profit, spread traders predict the market trend after certain events, such as breaking news and unexpected financial announcements. Spread traders need to observe and infer the relationship between the sources of market information and the market fluctuation. Spread traders may learn by actively participating the markets and observing the outcomes of their trades. If this is the case, the accumulative trade number is a good proxy variable for the experience. Spread traders may also gain experience by observing the events and recognizing the patterns of market scenarios. In this case, the length of time spread traders have stayed in markets represents the experience they have accumulated. Both trading time and trade number are used as the proxy variables of experience in the literature (Chiang et al., 2011; Feng and Seasholes, 2005; Nicolosi et al., 2009; Seru et al., 2010). However, the trading time and trade number are linearly correlated (r = 0.44) in our spread trading data. Therefore, if we include both trade number and trading time in our regression models, we may suffer multicollinearity problem which can lead to bias in estimating and interpreting coefficients of the experience measurements. Consequently, it is appropriate to develop our models using trading time and trade number, separately. Since their respective results are similar, we have chosen to present the results of models using trading time in the main content of the paper and include that using trade number in the Appendix.

In defining the experience of a trader, we measure the length of time they have traded with the spread trading company. Specifically, the experience of trader k associated with their i^{th} trade (E_{ik}) is defined as follows (Seru et al., 2010):

$$E_{ik} = t_{ik} - t_{1k}, k \in N^+; i \in N^+ \quad (1)$$

where t_{1k} and t_{ik} represent the point of time when trader k opened their first and i^{th}trade, respectively. $N^+$is a set of non-negative integers. Chiang et al. (2011) take a logarithmic transform of the experience variable, while Feng and Seasholes (2005) and Nicolosi et al. (2009) use the linear term. To decide whether the logarithmic transform should be used, we use the Cox test to compare the two sets of the regression models which are used to test our hypotheses. We use the logarithmic term in the first set of regression models and the linear term in the second one. The aim of the Cox test is to compare two non-nested regression models. The basic idea behind the Cox test is that, if the first regression model has a better set of independent variables, there should be no explanatory value from the fit of the independent variables from the second regression model to the fitted values from the first regression model. On the contrary, if we find explanatory value from it, the first regression model does not have a better set of independent variables. Therefore, the Cox test is to regress the fitted values of the first regression model on the second regression model, and vice versa. The results show that the regression models with the logarithmic term are slightly better than those with the linear term. The reason might be that the logarithmic transform can capture the human learning characteristic of learning faster in early periods and then slowing down in later periods. Hence, we take a logarithmic transform of the experience variable in our regression models.

Return

We define the return, R_{ik}, as the number of points won or lost on trade *i* of trader k.

Volatility of returns

We measure volatility by the variance of returns (defined in terms of points won/lost on a particular trade). Specifically, the overall risk run to the point of closing the i^{th} trade for individual k, V_{ik}, is defined as the variance of returns from their first trade up to their i^{th} trade:

$$V_{ik} = variance(R_{jk}), j \in [1, i]. \tag{2}$$

Sharpe ratio

The Sharpe ratio (Sharpe, 1998) is commonly used to measure a trader's risk-adjusted performance. The accumulated Sharpe ratio S_{ik} of trader k from their first trade up to the point of closing their i^{th} trade is calculated as follows:

$$S_{ik} = \frac{mean(R_{jk})}{\sqrt{variance(R_{jk})}}, j \in [1, i]. \tag{3}$$

Holding time

Since spread-trading markets are highly volatile (Chordia et al., 2001), it is believed that longer holding times are associated with higher risks as the trader exposes themselves to uncertainty for longer. The holding time associated with the i^{th} trade of trader k, H_{ik}, is defined as the number of seconds between the opening and closing of the i^{th} trade:

$$H_{ik} = t'_{ik} - t_{ik,} \tag{9}$$

where t_{ik} and t'_{ik} represent the times when trader k opens and closes their i^{th} trade, respectively.

Investment size

Investment size reveals an important aspect of risk-taking behavior (Fehr-Duda et al., 2010), and higher investment size is often regarded as taking higher risk (Liu et al., 2010). We denote the size of the investment in the i^{th} trade by trader k by ST_{ik}. This may be a 'long' or 'short' trade.

Trading frequency

Similar to Seru et al. (2010), who examine the number of trades during a pre-defined period, we measure trading frequency, TF_{ik}, as the number of trades in the seven-day period prior to trader k made the i^{th} trade t_{ik}.

Control variables

Demographic variables

Since risk-taking has been found to be greater amongst the young (e.g., Greenwood and Nagel, 2009), we control for a trader's age A_k. We also control for gender G_k, which take the value 1 for male, and 0 otherwise, because males are generally regarded as taking more risk (Bernasek and Shwiff, 2001; Jianakoplos and Bernasek, 1998).

Market volatility variable

Market volatility is controlled as a significant pricing factor to account for the effect on investment performance and behavior (Adrian and Rosenberg, 2008). Our data contains detailed individual trading records from four markets: FTSE 100, DAX 30, Euro-dollar and sterling-dollar exchange rates. We measure their daily market volatility by calculating the variance of the index values at one minute interval. That is, for a given market and a given date, we collect the index values minute by minute and calculate their variance for each trading day. All trades placed in the given market on the given date are associated with the daily market volatility. To make the values comparable between markets, we standardize the daily volatility MV for each market as follows:

$$MV = \frac{X - mean(X)}{\sqrt{variance(X)}}, \qquad (10)$$

where X is the raw value of market volatility.

The corresponding market volatility of the i^{th} trade of trader k is denoted as MV_{ik}.

Disposition Effect

We control for disposition effect (DE), which is a type of behavioral bias and is the tendency for a trader to realize gains faster than losses (Dhar and Zhu, 2006; Odean, 1998b). This affects investment behavior and performance (Seru et al., 2010). Calculating a reliable measure for the DE requires a sufficiently large number of trades. Previous studies, using traditional stock market data where investors may hold positions over several months, usually measure an investor's DE over a long period, such as a year (e.g., Seru et al., 2010). However, due to the much shorter average holding time of most trades placed in spread trading markets, we calculate the DE on a monthly basis. In determining the DE, Odean (1998b) and Dhar and Zhu (2006) calculate 'realized gains' and 'realized losses' and 'paper gains' and 'paper losses' for an investor at the time a stock was sold. The realized gain for a given investor increases by one if a sale is profitable, otherwise 'realized loss' increases by one. Other stocks remaining in the investor's portfolio at the time of the sale contribute to the paper gain or paper loss. For any stock with a prevailing market price exceeding/less than the price at which the stock was bought, the paper gains/losses increase by one. Odean (1998b) then calculate the proportion of gains realized (PGR) and proportion of losses realized (PLR) in over the period (say 1 year), as follows:

$$PGR = \frac{Realized\ Gain}{Realized\ Gain + Paper\ Gain} \tag{11}$$

$$PLR = \frac{Realized\ Loss}{Realized\ Loss + Paper\ Loss}. \tag{12}$$

In this study, we calculate realized gain and realized loss in the same way as Odean (1998b) and Dhar and Zhu (2006) (i.e., the count for realized gain/loss increases by one when a trade is closed in profit/loss). However, we need to modify the method of counting paper gains/losses as spread traders often have only one position opened and the portfolio approach adopted by Odean (1998a) concept is not appropriate. We use Fraser-Mackenzie et al. (2013)'s method to count the paper gains/losses associated with a particular trade. Since the holding time of a trade is generally short (the median is 11.7 minutes), we count the number of minutes a trade is in profit/loss as the paper gains/losses associated with that trade. That is, we assume that spread traders need to make a series of hold/close decisions every minute after a position is opened. We add one to paper gain/loss if a position in profit/loss is held for one minute. We then sum up the realized gain/loss and paper gain/loss and calculate their respective PGR and PLR on a monthly basis. Monthly DE, D_{ik}, associated with the i^{th} trade of trader k is then defined as the difference between PGR and PLR:

$$DE = PGR - PLR. \qquad (13)$$

A positive DE suggests a greater tendency to realize gains than losses.

Accumulated Profit (Account Balance)

The prospect theory predicts that individuals are risk-seeking in loss and risk-averse in profit (Kahneman and Tversky, 1979). Since the reference point used to determine whether the current position is in profit or in loss is usually the initially account balance, which is 0 in our data, we add the accumulated profit (account balance) to the regression models to control the effect of accumulated profit (or loss) on risk-taking behavior. To control the effect of accumulated profit (or loss) on risk-taking behavior, we add the accumulated profit to the regression models. The accumulated profit is calculated from the first trade each trader made with the spread trading company since individuals usually use the initial balance as a reference point in calculating profit/loss. Consequently, we include the accumulated profit (or loss) of trader *k* at the point of closing their $i-1^{th}$ trade, AP_{ik}, in the regressions, as follows:

$$AP_{ik} = \sum R_{jk} * ST_{jk}, j \in [1, i-1], \qquad (14)$$

where R_{jk} and ST_{jk} are the return and investment size of the j^{th} trade of trader k.

Recent Profit

Recent returns have been shown to influence risk-taking behavior (Choi et al., 2009; Glaser and Weber, 2009; Kaustia and Knüpfer, 2008; Liu et al., 2010; Thaler and Johnson, 1990). For example, the house effect suggests that individuals tend to take higher risks after gaining profit as prior gains might cushion the potential loss (Thaler and Johnson, 1990). Consequently, to control for this, we include the profit of the *i-1*-th trade of trader k in the regression models, denoted LTP_{ik}:

$$LTP_{ik} = R_{i-1\ k} * ST_{i-1\ k}, i > 1. \tag{15}$$

Similarly, we control for accumulated profit of all trades made by trader *k* in the week prior to their i^{th} trade (LWP_{ik}).

3.3.3 Models

3.3.3.1 Linear Mixed Model

The spread trading data is essentially panel data, as each trader can place multiple trades. A traditional pooled ordinary least squares (OLS) regression model is not appropriate for the analysis of the panel data. For example, if the traders who are able to learn from their experience place much more trades than others, the estimation will be influenced by the trades of these traders. This will lead to biased results which over-estimate the learning ability of the traders. To account for trader heterogeneity, linear mixed models (LMMs) are employed to control the unobserved trader-specific characteristics (see Seru et al. 2010).

Let Y represent the dependent variables (i.e., returns, risk and Sharpe ratios), E a trader's experience and C the control variables discussed in the previous section. If Z denotes the trade level covariates, and we use

$$Z = \begin{bmatrix} E \\ C \end{bmatrix}, \tag{16}$$

We develop our models in two stages (Cnaan et al., 1997). We let $k = 1, \dots \mathrm{N}$ index the individual traders. The first stage (trade level) model is of the following form:

$$\mathrm{y_{ik}} = \beta_\mathrm{k} \mathrm{z_{ik}} + \mathrm{e_{ik}}, \quad \mathrm{k} = 1, \dots \mathrm{N} \tag{17}$$

where $\mathrm{y_{ik}}$ represents a dependent variable (i.e., returns, risk or Sharpe ratios) of the i^{th}trade of trader k, $\mathrm{z_{ik}}$ is the covariate vector of the i^{th} trade of trader k, $\mathrm{e_{ik}}$ is zero mean error term and β_k are the regression coefficients for trader k, and N is the number of traders in our data.

In the second stage (trader level), the β_k are regarded as dependent variables, and the mean of the β_k depends on trader level characteristics. Let $\mathrm{a}'_{\mathrm{E,k}}$ and $\mathrm{a}'_{\mathrm{C,k}}$ denote the vector of trader level characteristics affecting the coefficients of experience (E) and the control variables (C), respectively. We assume that both slope and intercepts of traders can vary, so we use

$$\mathrm{a}'_{\mathrm{E,k}} = (1, \mathrm{k}), \tag{18}$$

where k is used as the unique ID for each trader. The regression model we estimate is given by:

$$\beta_\mathrm{k} = \mathrm{A_k}\alpha + \mathrm{b_k}, \tag{19}$$

where

$$\mathrm{A_i} = \begin{bmatrix} \mathrm{a}'_{\mathrm{E,k}} & 0 \\ 0 & \mathrm{a}'_{\mathrm{C,k}} \end{bmatrix}$$

$$\alpha' = (\alpha'_\mathrm{E},\ \alpha'_\mathrm{C}).$$

and α'_E and α'_C are regression parameter vectors. Combining equation (17) and (19) we have

$$\mathrm{Y_k = Z_k\beta_k + e_k = Z_kA_k\alpha + Z_kb_k + e_k} \tag{20}$$

There are two types of regression parameters in equation (20). The α's are 'fixed effects' and help to test our hypotheses. The $\mathrm{b_k}$'s, are 'random effects' (independent random variable with zero mean) and are often considered as error terms. We use the method proposed by Pinheiro et al. (2007) to estimate the regression coefficients.

3.3.3.2 Fixed Effect Regression Model

In order to test our first set of hypotheses, we estimate the regression model represented by Eq.12 for different dependent variables, separately. Specifically, we test our hypothesis that as traders gain in experience they increase their risk-taking, in terms of holding time (H1a) using the following regression models:

$$\begin{aligned} H_{ik} = \alpha + \beta E_{ik} + \beta^{ST} ST_{ik} + \beta^{TF} TF_{ik} + \beta^{A} A_k + \beta^{G} G_k + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} \\ + \beta^{AP} AP_{ik} + \beta^{LTP} LTP_{ik} + \beta^{LWP} LWP_{ik} + \varepsilon_{ik}, \end{aligned} \tag{21}$$

where for trader k, H_{ik} is the holding time of their i^{th} trade, E_{ik} is the measure of their experience at the opening of the i^{th} trade, A_k is their age, G_k is their gender, MV_{ik} is the market volatility, D_{ik} is an estimate of the DE, AP_{ik} is the accumulated profit up to the time of opening the ith trade, LTP_{ik} is the profit/loss secured by trader k on their i-1th trade and LWP_{ik} is the profit/loss secured by trader k during the week prior to opening trade i. We use ε_{ik} to denote the regression error term, and α, $\beta, \beta^{H}, \beta^{ST}$, β^{TF}, β^{A}, β^{G}, β^{MV} and β^{D} are determined by parameter estimation. We test the hypothesis H1b and H1c by changing the dependent variable to the investment size (ST_{ik}) and the trading frequency (TF_{ik}).

We then test our second set of hypotheses that a*s traders gain in experience they achieve higher returns* (H2a), *their returns display greater volatility* (H2b) *and they achieve higher* *risk-adjusted returns* (H2c) as measured by the Sharpe ratio (Sharpe, 1966), by estimating the following regression models:

$$R_{ik} = \alpha + \beta E_{ik} + \beta^{H} H_{ik} + \beta^{ST} ST_{ik} + \beta^{TF} TF_{ik} + \beta^{A} A_{k} + \beta^{G} G_{k} + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} + \varepsilon_{ik}, \quad (22)$$

where for trader k, R_{ik} is the return and V_{ik} is the variance of returns, associated with their i^{th} trade and S_{ik} is their accumulated Sharpe ratio up to the point of closing the i^{th} trade. We test the hypothesis H2b and H2c by changing the dependent variable to the variance of returns (V_{ik}) and the Sharpe ratio (S_{ik}).

We finally test our third set of hypotheses that t*he volatility of an individual's returns increases when the time they expose themselves to uncertainty increases (H3a), the size of the stake they are prepared to put at risk increases (H3b) and when the y engage in the risk taking activity more frequently (H3c)* using the following regression model:

$$V_{ik}^{p} = \alpha + \beta^{H} H_{ik}^{p} + \beta^{ST} ST_{ik}^{p} + \beta^{TF} TF_{ik}^{p} + \beta^{A} A_{k} + \beta^{G} G_{k} + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} + \varepsilon_{ik}, \quad (23)$$

where for trader k, V_{ik}^{p} is the variance of returns, accumulated from their trades placed during the period of p days before their i^{th} trade, E_{ik} is the measure of their experience at the opening of the i^{th} trade, H_{ik}^{p} , ST_{ik}^{p} and TF_{ik}^{p} are the total holding time, total investment size and total trade number during the period of p days before their i^{th} trade. We select multiple values of p, such as 1, 7, 14 and 30, to take the short-term trading characteristic of spread trading into account.

3.3.4 Controlling Biases

There are two potential sources of bias which could affect our analysis. The first one is selection bias which results from the sample selection approach failing to ensure the representativeness of the obtained sample. In particular, the first trade of all traders is excluded in our regression models containing the volatility of returns and the Sharpe values. The reason is that we need at least two trades to calculate the volatility, and we only have one trade when we calculate the performance measures for the first trade. Therefore, we cannot calculate the volatility of returns for the first trade of all traders, and these are treated as missing values. Similarly, we cannot calculate the Sharpe ratio for a trader's first trade since we need the volatility of returns to calculate the Sharpe ratios. Consequently, we exclude certain trades and form a non-randomly selected sub-sample from all the trades, which may introduce selection bias to our results. The second one is survivorship bias which leads to over-optimistic results as the traders who fail to survive may be excluded from the analysis. Specifically, the traders who are able to learn to improve their performance through experience may survive longer in the market and dominate the results of the analysis.

3.3.4.1 Selection bias

We follow Seru et al.'s (2010) approach and use the Heckman two-stage method to account for the potential selection bias (Heckman, 1976). The selection bias arise as we only have limited information from a non-random sub-sample, and the Heckman two-stage method is widely used to correct for non-randomly selected samples (Toomet and Henningsen, 2008). The key insight of this method is that this bias correction problem can be approached as an omitted variables problem. Hence, the problem of sample selection bias can be solved by estimating the omitted variable.

The common sample selection problem is modeled as follows: the observation equation is given by:

$$y_i = \beta x_i' + \varepsilon_i \tag{23}$$

where y_i is the dependent variable, x_i' are the regressors, β is the regression coefficient and ε_i is a normally distributed error term. However, some of the $y_i$observations (Sharp ratios in our case) are missing, so a direct estimation with ordinary linear regression will be biased. Heckman (1976)'s solution consists of two stages. All observations are included in the first stage, and only the observations with y_i not missing are selected to the second stage. The first stage regression is used to predict selection:

$$z_i^* = rw_i' + u_i, \tag{24}$$

where z_i^* is 1 if the i^{th}observation is selected, otherwise 0, w_i' are the regressors, r is the regression coefficient and u_i is an error term. The error terms are assumed to follow a bivariate normal distribution:

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & \sigma^2 \end{pmatrix}\right).$$

The observed dependence between y_i and z_i^* can be written as:

$$E(y_i | y_i \textit{ is observed}) = E(y_i | z_i^* > 0).$$

According to Toomet and Henningsen (2008) the observation equation can be re-written as:

$$\begin{aligned} y_i &= x_i'\beta + \varepsilon_i = x_i'\beta + E[\varepsilon_i | z_i^* > 0] + \eta_i = x_i'\beta + E[\varepsilon_i | u_i > -w_i' r] + \eta_i \\ &\equiv x_i'\beta + \rho\sigma\lambda(w_i' r) + \eta_i, \end{aligned} \tag{25}$$

where$\lambda(\alpha) = \varphi(\alpha)/\Phi(\alpha)$is commonly termed the inverse Mill's ratio, $\varphi(\alpha)$ and $\Phi(\alpha)$ are standard normal density and cumulative distribution functions, and η is an independent disturbance term. $\rho\sigma$ is unknown and can be estimated by the ordinary least-squares method (OLS). The selection bias can be adjusted by including the inverse Mill's ratio in the second stage (Heckman, 1976).

We develop our model in a setting similar to Seru et al. (2010): a trade is selected if this is not a trader's first trade, and the Sharpe ratio is calculated and assigned to that trade. The selection condition comes from the fact that we cannot calculate the Sharpe ratio if only one trade is placed. In the first stage, the independent variables are

$$\mathrm{w}' = (\mathrm{E}, \mathrm{TradeNO}, \mathrm{APr}), \tag{26}$$

where E is the experience measure, $TradeNO$ is the number of trades which have been placed and APr is the accumulated profit of trader *k* at the time trade *i* is placed.

We apply the Heckman two-stage method with those regression models having volatility or Sharpe ratios as dependent variables to control for the potential selection bias introduced by our measures of volatility and the Sharpe ratio.

3.3.4.2 Traders Surviving Shorter/Longer

The second potential bias may arise because traders who learn to improve their performance through experience may survive longer in the market and place more trades than those who do not learn. Nicolosi et al. (2009) tackle this so called 'survivorship bias' by only examining the individuals who trade in both the first and the second half of the period covered by their data. However, using this approach, traders with short active periods may be selected (e.g., a trader is selected if they start trading from the end of the first half period and stops at the beginning of the second half period). To avoid this scenario and to account for trader-specific characteristics, we split all traders into two groups: those stay active in the spread-trading market for shorter or longer than or equal to the median of a trader's active period in our data (i.e., 179 days). We define a trader's active period as the time from their first trade to their last trade in the dataset. We then run LMMs with dependent variables of returns, volatility of returns and Sharpe ratios (see model (17)) for both groups of traders, separately.

3.3.4.3 Sensitivity Analysis

In calculating the variables of trading frequency, volatility and Sharpe ratio, we need to specify a fixed time period. To examine the robustness of our results, we calculate those variables with different time periods and check if the original conclusion still holds. In particular, when controlling for trading frequency, in addition to the seven-day period employed in the main analysis, we also use a 14 and 30 day period. Furthermore, we calculate the volatility of returns and Sharpe ratio achieved by trader *k* up to the time of trade *i* from their first trade to the trade *i*. We also use periods of 60, 90, 120 and 180 days to check the robustness of our results.

3.4 RESULTS

3.4.1 The Effect of Experience on Risk-taking Behavior

To examine the impact of experience on risk-taking, we estimate the LMMs summarized by equation (21). The results are presented in

Table XIII and the coefficients for all three measures of risk-taking, holding time, investment size and trading frequency are positive and significant (i.e., 1004.2, $p < 0.001$; 0.0412, $p < 0.001$; 6.591, $p < 0.001$, respectively). These results provide support for H1a, H1b and H1c, namely, that their risk-taking behavior increases as traders' experience increases. Chiang et al. (2011) found that individuals tend to trade in a more aggressive fashion when they gain experience and our results support this conclusion.

To examine the robustness of our results, we calculate the trading frequency using different time intervals, (i.e., 14 and 30 days in addition to the original interval of 7 days). The results are presented in

Table XIV. These results confirm that trading frequency tends to increase as a trader's experience increases.

Table XIII: Risk-taking Behavior Related to Experience

Risk taking measure:	**Holding time**	**Investment size**	**Trading frequency**
Intercept	22976 ***	4.2137 ***	17.319 ***
Experience	2559.1 ***	0.0236 ***	0.8673 ***
Controls			
Holding Time		-0.0000003 ***	-0.000002 ***
Investment size	-204.4 ***		0.1823 ***
Trading Frequency	-33.473 ***	0.0042 ***	
Accumulated Profit	0.1795 ***	0.00009 ***	0.0002 ***
Last Trade Profit	-11.555 ***	0.000004	-0.002 ***
Last Week Profit	-0.3138 ***	0.0002 ***	0.0011 ***
Age	14.493	-0.0188 ***	-0.0725 ***
Gender	2570.9	0.5118 **	-2.1145 ***
Market Volatility	-0.2911 ***	-0.00002 ***	0.0009 ***
Disposition	0.0384 ***	-0.000002 ***	-0.00003 ***

This table presents results for regressions of the form

$$H_{ik} = \alpha + \beta E_{ik} + \beta^{ST} ST_{ik} + \beta^{TF} TF_{ik} + \beta^{A} A_{k} + \beta^{G} G_{k} + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} + \beta^{AP} AP_{ik} + \beta^{LTP} LTP_{ik} + \beta^{LWP} LWP_{ik} + \varepsilon_{ik},$$

where for trader k, H_{ik} is the holding time of theiri^{th} trade, ST_{ik} is the investment size of their i^{th} trade and TF_{ik} is the trading frequency of the i^{th} trade, E_{ik} is the measure of their experience at the opening of the i^{th} trade, A_k is the age, G_k is the gender, $MV_{ik}$is the market volatility, D_{ik} is the estimate of disposition effect, AP_{ik} is the accumulated profit, LTP_{ik} is the last trade profit and LWP_{ik} is the last week profit. We use $\varepsilon_{ik}$to denote the regression error term, and α, β, β^H, β^{ST}, β^{TF}, β^A, β^G, β^{MV} and β^D are determined by parameter estimation. Data employed are for the period 2003 to 2013. ***, ** and * denote significance at 0.1%, 1% and 5% levels, respectively.

Table XIV: Trading Frequency Related to Experience – Sensitivity Analysis

	Trading Frequency (14 days)	**Trading Frequency (30 days)**
Intercept	19.984 ***	10.187 ***
Experience	3.3562 ***	9.3768 ***
Controls		
Holding Time	-0.000003 ***	-0.000005 ***
Investment size	0.2901 ***	0.3796 ***
Accumulated Profit	0.0004 ***	0.0008 ***
Last Trade Profit	-0.0022 ***	-0.0020 ***
Last Week Profit	0.001 ***	0.0003 ***
Age	-0.1021 ***	-0.1587 ***
Gender	-4.1491 ***	-8.2532 ***
Market Volatility	0.0013 ***	0.0016 ***
Disposition	-0.00004 ***	-0.000002 **

This table presents results for regressions of the form

$$TF_{ik} = \alpha + \beta E_{ik} + \beta^{H} H_{ik} + \beta^{ST} ST_{ik} + \beta^{A} A_{k} + \beta^{G} G_{k} + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} + \beta^{AP} AP_{ik} + \beta^{LTP} LTP_{ik} + \beta^{LWP} LWP_{ik} + \varepsilon_{ik},$$

where for trader k, TF_{ik} is the trading frequency of the i^{th} trade in a period of 14 and 30 days, H_{ik} is the holding time of theiri^{th} trade, ST_{ik} is the investment size of their i^{th} trade, E_{ik} is the measure of their experience at the opening of the i^{th} trade, A_k is the age, G_k is the gender, $MV_{ik}$is the market volatility, D_{ik} is the estimate of disposition effect, AP_{ik} is the accumulated profit, LTP_{ik} is the last trade profit and LWP_{ik} is the last week profit. We use $\varepsilon_{ik}$to denote the regression error term, and α, $\beta, \beta^{H}, \beta^{ST}$, $\beta^{A}, \beta^{G}, \beta^{MV}$ and β^{D} are determined by parameter estimation. Data are from the period 2003 to 2013. ***, ** and * denote significance at 0.1%, 1% and 5% levels, respectively.

3.4.1.1 Controlling Survival Bias

To account for potential survival bias, we estimate the LMMs using equation (21) for the trades of those that have active trading periods (a) shorter and (b) greater than (or equal) to 179 days. The results are presented in Table XV. For those traders with shorter active trading periods, the three risk-taking behaviors are again positively related with experience (16393, $p < 0.001$; 3.9138, $p < 0.001$; 6.591, $p < 0.001$). We also find that traders with longer active trading periods also increase their risk-taking measured in terms of holding time (2688.4, $p < 0.001$) and investment size (0.0215, $p < 0.001$) as they gain in experience. However, for this group, their trading frequency decreases with experience (-0.1995, $p < 0.001$).

Overall, these results largely support our conjecture that traders increase risk-taking behavior as they gain in experience. However, traders with longer active trading periods appear to reduce their trading frequency.

Table XV Risk-taking Behavior Related to Experience – Survivorship Analysis

	Surviving Shorter			Surviving Longer		
	Holding Time	**Investment Size**	**Trading Frequency**	**Holding Time**	**Investment Size**	**Trading Frequency**
Intercept	16393 ***	3.9138 ***	12.486 ***	32288 ***	4.4769 ***	23.438 ***
Experience	1004.2 ***	0.0412 ***	6.591 ***	2688.4 ***	0.0215 ***	-0.1995 ***
Controls						
Holding Time		-0.0000004 ***	-0.000002 **		-0.0000003 ***	-0.000002 ***
Investment size	-52.09 ***		0.381 ***	-231.51 ***		0.145 ***
Trading Frequency	-2.2994	0.005 ***		-43.402 ***	0.0039 ***	

Accumulated Profit	0.2798 ***	0.0003 ***	0.0029 ***	0.1823 ***	0.00008 ***	0.0002 ***
Last Trade Profit	-6.7449 ***	0.00007	-0.0074 ***	-11.855 ***	-0.000009	-0.0013 ***
Last Week Profit	-0.5147 ***	-0.00006 ***	0.004 ***	-0.3055 ***	0.0002 ***	0.0007 ***
Age	130.48	-0.0105	-0.1968 ***	-15.85	-0.0266 ***	-0.0747 ***
Gender	1324.6	0.2494	0.0713	814.87	0.7039 **	-3.2912 ***
Market Volatility	-0.0745	-0.00005 ***	0.0005 ***	-0.3063 ***	-0.00002 ***	0.0009 ***
Disposition	0.0136 ***	-0.0000009 ***	-0.00003 ***	0.0402 ***	-0.000002 ***	-0.00002 ***

This table presents results for regressions of the form in two groups of traders that stay active in the spread-trading market for shorter and longer periods

$$H_{ik} = \alpha + \beta E_{ik} + \beta^{ST} ST_{ik} + \beta^{TF} TF_{ik} + \beta^{A} A_k + \beta^{G} G_k + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} + \beta^{AP} AP_{ik} + \beta^{LTP} LTP_{ik} + \beta^{LWP} LWP_{ik} + \varepsilon_{ik},$$

where for trader $\boldsymbol{k}$, $\boldsymbol{H_{ik}}$ is the holding time of their$\boldsymbol{i^{th}}$ trade, $\boldsymbol{ST_{ik}}$ is the investment size of their $\boldsymbol{i^{th}}$ trade and $\boldsymbol{TF_{ik}}$ is the trading frequency of the $\boldsymbol{i^{th}}$ trade, $\boldsymbol{E_{ik}}$ is the measure of their experience at the opening of the $\boldsymbol{i^{th}}$ trade, $\boldsymbol{A_k}$ is the age, $\boldsymbol{G_k}$ is the gender, $\boldsymbol{MV_{ik}}$is the market volatility, $\boldsymbol{D_{ik}}$ is the estimate of disposition effect, $\boldsymbol{AP_{ik}}$ is the accumulated profit, $\boldsymbol{LTP_{ik}}$ is the last trade profit and $\boldsymbol{LWP_{ik}}$ is the last week profit. We use $\boldsymbol{\varepsilon_{ik}}$to denote the regression error term, and $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\beta^{H}}$, $\boldsymbol{\beta^{ST}}$, $\boldsymbol{\beta^{TF}}$, $\boldsymbol{\beta^{A}}$, $\boldsymbol{\beta^{G}}$, $\boldsymbol{\beta^{MV}}$ and $\boldsymbol{\beta^{D}}$ are determined by parameter estimation. We define a trader's active period as the time from their first trade to the last of their trades in the dataset. On this basis, we find that the median of a trader's active period is 179 days. Consequently, we split traders into those that have active periods shorter and longer than or equal to 179 days. Data are from the period 2003 to 2013. ***, ** and * denote significance at 0.1%, 1% and 5% levels, respectively.

3.4.2 The Effect of Experience on Performance

The results of estimating the LMMs using equation (17) are presented in Table XVI and show that more experienced traders make higher returns but take greater risks and achieve lower Sharpe ratios. In particular, the coefficients of experience for the models with return and the volatility of returns as the dependent variable are positive and significant, respectively (0.0471, $p < 0.001$; 78.715, $p < 0.001$), indicating a positive relationship between experience of a trader and both their expected returns and their return volatility, thus, supporting H2a and H2b. However, the results of estimating the model indicate a significant negative relationship between a trader's experience and the Sharpe ratio they achieve (-0.0131, $p < 0.001$). This result does not support H2c, suggesting that traders' risk-adjusted performance *declines* as they gain in experience.

Table XVI: Performance Related to Experience

	Return	**Volatility**	**Sharpe Ratio**
Intercept	0.1076	6252.7 ***	0.0106
Experience	0.0471 ***	78.715 ***	-0.0131 ***
Controls			
Holding Time	-0.00003 ***	0.0017 ***	-0.0004 ***
Investment Size	-0.0119 ***	-2.1632 **	0.000000008 ***
Trading Frequency	-0.0007 ***	-0.8019 ***	0.0003 ***
Age	-0.008	-13.511	-0.002 **
Gender	-0.0348	-4233.3 ***	-0.0338

Market Volatility	-0.0002 ***	0.0215 ***	0.0000001
Disposition	-0.000003 ***	-0.0001	-0.00000007 ***

This table presents results for regressions of the form

$$R_{ik} = \alpha + \beta E_{ik} + \beta^{H} H_{ik} + \beta^{ST} ST_{ik} + \beta^{TF} TF_{ik} + \beta^{A} A_{k} + \beta^{G} G_{k} + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} + \varepsilon_{ik},$$

where for trader k, R_{ik} is the return of their i^{th}trade, V_{ik} is the variance of returns associated with trader k's i^{th} trade and S_{ik} is trader k's accumulated Sharpe ratio up to the point of closing the i^{th} trade, H_{ik} is the holding time of theiri^{th} trade, ST_{ik} is the investment size of their i^{th} trade and TF_{ik} is the trading frequency of the i^{th} trade, E_{ik} is the measure of their experience at the opening of the i^{th} trade, A_k is the age, G_k is the gender, $MV_{ik}$is the market volatility, and D_{ik} is the estimate of disposition effect. We use $\varepsilon_{ik}$to denote the regression error term, and α, β, β^{H}, β^{ST}, β^{TF}, β^{A}, β^{G}, β^{MV} and β^{D} are determined by parameter estimation. Data are from the period 2003 to 2013. ***, ** and * denote significance at 0.1%, 1% and 5 levels %, respectively.

Some previous studies have also found that traders with greater experience achieve higher returns (Feng and Seasholes, 2005; Nicolosi et al., 2009; Seru et al., 2010), and they argue that this arises because traders learn from experience. However, we find that more experienced traders take higher risks resulting in a lower risk-adjusted performance.

3.4.2.1 Controlling Selection Bias: Heckman Two-Stage Method

The results of conducting the Heckman two-stage method are presented in Table XVII. The selection bias is shown to be a legitimate concern in respect of the Sharpe ratio, because $\rho > 0$, indicating that the unobserved values tend to be lower than those which are observed. That is, the risk-adjusted performance of the unobserved trades is lower compared to those trades which we can observe and use to calculate the risk-adjusted performance. After controlling for selection bias in the second stage, we can see that experience is negatively related with Sharpe ratio (-0.0091, $p < 0.001$). Although the magnitude of the coefficient is smaller than that estimated in the earlier analysis, it still remains negative and significant, confirming that lower Sharp ratios are generally obtained by traders with greater experience. Similarly, the volatility of returns is positively related with experience (337.7, $p < 0.001$), which is in line with our earlier results that as traders gain experience, their returns are subject to greater volatility.

To test the robustness of our conclusions, when estimating the LMMs using equation (17) with the volatility of the returns and the Sharpe ratio as the dependent variable, we also calculate the returns volatility and Sharpe ratio using 60, 90, 120 and 180 day periods, respectively. The results all confirm our conclusions that a trader's returns volatility and

Sharpe ratios (presented in Table VII) are positively and negatively, respectively, related to a trader's experience.

Table XVII: Performance Related to Experience – Heckman 2-stage Method

	Volatility		Sharpe	
	First Stage (Selected In Sample)	**Second Stage**	**First Stage** (Selected In Sample)	**Second Stage**
Intercept	1.775 ***	886.7 ***	1.775 ***	0.0322 ***
Experience	-0.0001 ***	337.7 ***	-0.0001 ***	-0.0091 ***
TradeNO	-0.0055 ***		-0.0006 ***	
Accumulated Profit	0.0000004 ***		0.0000004 ***	
Holding Time		0.0064 ***		0.00000001 ***
Investment Size		2.934 **		0.0008 ***
Trading Frequency		-3.402 ***		0.0001 ***
Age		-10.27 ***		-0.0009 ***
Gender		-101.4 *		-0.0039 ***
Market Volatility		0.0925 ***		0.000001 ***
Disposition		-0.0022 ***		-0.0000001 ***
R^2		0.0035		0.0024
Adjusted R^2		0.0035		0.0023

Inverse Mills Ratio	-1898.4 ***	0.0155 ***
Σ	32508	0.5687
ρ	-0.0584	0.0273

This table presents results from Heckman 2 stage method. The regression model in the first stage is

$$w' = (\mathrm{E}, \mathrm{TradeNO}, \mathrm{APr}),$$

where E is the experience measure, $TradeNO$ is the number of trades which have been placed and APr are the accumulated profit of the trader at the time the trade is placed. The regression model in the second stage is

$$V_{ik} = \alpha + \beta E_{ik} + \beta^{H} H_{ik} + \beta^{ST} ST_{ik} + \beta^{TF} TF_{ik} + \beta^{A} A_{k} + \beta^{G} G_{k} + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} + \varepsilon_{ik},$$

where for trader k, V_{ik} is the variance of returns associated with trader k's i^{th} trade, and S_{ik} is trader k's accumulated Sharpe ratio up to the point of closing the i^{th} trade, H_{ik} is the holding time of their i^{th} trade, ST_{ik} is the investment size of their i^{th} trade and TF_{ik} is the trading frequency of the i^{th} trade, E_{ik} is the measure of their experience at the opening of the i^{th} trade, A_k is the age, G_k is the gender, $MV_{ik}$is the market volatility, and D_{ik} is the estimate of disposition effect. We use $\varepsilon_{ik}$to denote the regression error term, and α, $\beta, \beta^{H}, \beta^{ST}, \beta^{TF}, \beta^{A}, \beta^{G}, \beta^{MV}$ and β^{D} are determined by parameter estimation. Data are from the period 2003 to 2013. ***, ** and * denote significance at 0.1%, 1% and 5% levels, respectively.

Table XVIII: Performance Related to Experience – Sensitivity Analysis

	Volatility (180)	**Volatility (120)**	**Volatility (90)**	**Volatility (60)**	**Sharpe (180)**	**Sharpe (120)**	**Sharpe (90)**	**Sharpe (60)**
Intercept	6355.3 ***	6508.2 ***	6409.0 ***	7810.1 ***	0.0219	0.0211	0.0302	0.068
Experience	68.546 ***	64.697 ***	76.839 ***	36.440 *	-0.0078 ***	-0.0057 ***	-0.0039 ***	-0.0017 ***
Controls								
Holding Time	0.0034 ***	0.0041 ***	0.0047 ***	0.0055 ***	-0.000000010 ***	-0.00000001 ***	-0.00000002 ***	-0.00000003 ***
Investment Size	-2.9344	-0.4747	-1.2054	-1.5878	0.0004 ***	0.0005 ***	0.0005 ***	0.0005 ***
Trading Frequency	-1.1466 ***	-1.1306 ***	-1.2044 ***	-1.1761 ***	0.00006 ***	0.00005 ***	0.00005 ***	0.00006 ***
Age	-22.425	-24.928	-25.001	-34.101 **	-0.0018 *	-0.0018 *	-0.0019 *	-0.0026 **
Gender	-3859.1 ***	-4062.5 ***	-4139.0 ***	-5125.2 ***	-0.04	-0.0339	-0.0393	-0.0413

Market Volatility	0.0407 ***	0.0580 ***	0.0667 ***	0.0715 ***	0.0000002 **	0.00000008	0.000000005	-0.0000001
Disposition	-0.0001	0.0002	0.0005	-0.0002	-0.0000001 ***	-0.0000001 ***	-0.0000002 ***	-0.0000002 ***

This table presents results for regressions of the form

$$V_{ik} = \alpha + \beta E_{ik} + \beta^{H} H_{ik} + \beta^{ST} ST_{ik} + \beta^{TF} TF_{ik} + \beta^{A} A_{k} + \beta^{G} G_{k} + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} + \varepsilon_{ik},$$

where for trader k, V_{ik} is the variance of returns associated with trader k's i^{th} trade in a period of 180, 120, 90 and 60 days, and S_{ik} is trader k's Sharpe ratio up to the point of closing the i^{th} trade in a period of 180, 120, 90 and 60 days, H_{ik} is the holding time of their i^{th} trade, ST_{ik} is the investment size of their i^{th} trade and TF_{ik} is the trading frequency of the i^{th} trade, E_{ik} is the measure of their experience at the opening of the i^{th} trade, A_k is the age, G_k is the gender, $MV_{ik}$is the market volatility, and D_{ik} is the estimate of disposition effect. We use $\varepsilon_{ik}$to denote the regression error term, and α, $\beta, \beta^{H}, \beta^{ST}, \beta^{TF}, \beta^{A}, \beta^{G}, \beta^{MV}$ and β^{D} are determined by parameter estimation. Data are from the period 2003 to 2013. ***, ** and * denote significance at 0.1%, 1% and 5% levels, respectively.

3.4.2.2 Controlling Survivorship Bias

The results of estimating the LMMs using equation (17) for the trades of those that have (a) shorter or (b) longer than or equal to 179 days active trading periods are presented in Table XIX. The results indicate that, as in the main analysis, traders in both groups take more risks and their risk adjusted performance reduces as they gain in experience. In particular, the coefficients of the volatility and Sharp ratio when estimating equation (17) are significant and positive/negative, respectively (volatility coefficient for longer and shorter trading period groups: 77.922, $p < 0.001$; 74.109, $p < 0.001$; Sharp ratio coefficient for shorter and longer trading period groups: -0.0275, $p < 0.001$; -0.0108, $p < 0.001$). In addition, as in the main analysis, for traders in the longer trading period group, the coefficient for returns in the model represented by equation (17) is positive and significant (0.0495, $p < 0.001$); suggesting that they achieve increased returns as they gain in experience. However, traders with shorter active trading periods, return is negatively related with experience (-0.4008, $p < 0.001$). It may well be that these individuals cease trading because they fail to learn to improve their returns and, as a result, have little incentive to continue trading.

Overall, the results largely support hypotheses 2a and 2b, namely, that as traders gain in experience they make greater returns but these returns are subject to higher volatility. Equally, the results do not appear to support H2c, indicating that as traders gain experience their risk adjusted performance falls.

Table XIX: Performance related to Experience – Survivorship Analysis

	Return	Volatility	Sharpe Ratio
Surviving Shorter			
Intercept	0.9632	878.23	0.0355
Experience	-0.4008 ***	74.109 ***	-0.0275 ***
Controls			
Holding Time	-0.00005 ***	0.0029 ***	-0.0000002 ***
	-0.0121 ***	-0.2023	0.0002
Trading Frequency	-0.000005	-0.2537	0.00003 ***
Age	-0.0414 **	4.3560	-0.0035 **

3.4.3 Investment Size

Gender	-0.6160	-537.57	-0.0502
Market Volatility	-0.0002 ***	0.0073	-0.0000001
Disposition	-0.000003 ***	0.0005	-0.0000002 ***
Surviving Longer			
Intercept	0.7089	5539.8 ***	0.0704 **
Experience	0.0495 ***	77.922 ***	-0.0108 ***
Controls			
Holding Time	-0.00003 ***	0.0017 ***	-0.000000004 ***
Investment Size	-0.0127 ***	-2.5004 **	0.0003***
Trading Frequency	-0.0009 ***	-1.0005 ***	0.00006 ***
Age	-0.0083	-12.999	-0.0013 ***
Gender	-0.1816	-3319.6 ***	-0.0433
Market Volatility	-0.0002 ***	0.0232 ***	-0.0000001 **
Disposition	-0.000003 ***	-0.0002	-0.00000006 ***

This table presents results for regressions of the form in two groups of traders that stay active in the spread-trading market for shorter and longer periods

$$R_{ik} = \alpha + \beta E_{ik} + \beta^H H_{ik} + \beta^{ST} ST_{ik} + \beta^{TF} TF_{ik} + \beta^A A_k + \beta^G G_k + \beta^{MV} MV_{ik} + \beta^D D_{ik} + \varepsilon_{ik},$$

where for trader k, R_{ik} is the return of theiri^{th} trade, MV_{ik} is the variance of returns associated with trader k's i^{th} trade and S_{ik} is trader k's accumulated Sharpe ratio up to the point of closing the i^{th} trade, H_{ik} is the holding time of theiri^{th} trade, ST_{ik} is the investment size of their i^{th} trade and TF_{ik} is the trading frequency of the i^{th} trade, E_{ik} is the measure of their experience at the opening of the i^{th} trade, A_k is the age, G_k is the gender, $V_{ik}$is the market volatility, and D_{ik} is the estimate of disposition effect. We use $\varepsilon_{ik}$to denote the regression error term, and α, $\beta, \beta^H, \beta^{ST}, \beta^{TF}, \beta^A, \beta^G, \beta^{MV}$ and β^D are determined by parameter estimation. We define a trader's active period as the time from their first trade to the last of their trades in the dataset. On this basis, we find that the median of a trader's active period is 179 days. Consequently, we split traders into those that have active periods shorter and longer than or equal to 179 days. Data are from the period 2003 to 2013. ***, ** and * denote significance at 0.1%, 1% and 5% levels, respectively.

3.4.4 The Relationship between the Volatility of Returns and Risk-taking Behavior

The relation between the volatility of returns and risk-taking behavior is revealed by estimating the LMMs using equation (18) and examining the coefficients of holding time, investment size and trading frequency in

Table XX. We test the robustness of our conclusions, by including returns volatility and risk-taking behavior measured across 1, 7, 14 and 30 day periods. The positive coefficients of holding time (0.0043, $p < 0.001$; 0.0019, $p < 0.001$; 0.0013, $p < 0.001$; 0.0007, $p < 0.001$) suggest that volatility is positively related with the length of time a position is held open. This is as one would expect, as the longer the position is held open, the greater is the market risk to which the trader is exposed (supporting H3a). Similarly, the significant positive coefficients of investment size (0.2301, $p < 0.05$; 0.1226, $p < 0.001$; 0.0888, $p < 0.001$; 0.0406, $p < 0.001$) suggest that as the amount they are prepared to invest increases the volatility of a trader's returns (measured in terms of index points won/lost) increases. The result implies that traders are exposed to higher trading risk when they put a larger amount of money at risk, and hypothesis 3b is supported. The significant negative coefficients of trading frequency (-12.834, $p < 0.001$; -5.5314, $p < 0.001$; -3.964, $p < 0.001$; -2.3659, $p < 0.001$) indicate that as a trader's frequency of trading increases, the degree of risk taken decreases (i.e. the volatility of returns decreases); leading us to reject H3c.

Table XX: Volatility of Returns Related with Risk-taking Behavior

	Volatility of Returns			
	1 day	7 days	14 days	30 days
Intercept	5241.6 ***	6367.3 ***	6095.9 ***	5974.2 ***
Holding Time	0.0043 ***	0.0019 ***	0.0013 ***	0.0007 ***
Investment Size	0.2301 *	0.1226 ***	0.0888 ***	0.0406 **
Trading Frequency	-12.834 ***	-5.5314 ***	-3.964 ***	-2.3659 ***
Age	-12.073	-17.141	-19.36	-14.704
Gender	-4578.4 ***	-5362.7 ***	-5055.9 ***	-4654.6 ***
Market Volatility	0.1498 ***	0.1166 ***	0.1032 ***	0.0723 ***
Disposition	0.0013 **	0.0011 ***	0.0014 ***	0.0015 ***

This table presents results for regressions of the form:

$$V_{ik}^{p} = \alpha + \beta^{H} H_{ik}^{p} + \beta^{ST} ST_{ik}^{p} + \beta^{TF} TF_{ik}^{p} + \beta^{A} A_{k} + \beta^{G} G_{k} + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} + \varepsilon_{ik}, \qquad (24)$$

where for trader k, V_{ik}^{p} is the variance of returns, accumulated from their trades placed during the period of p days before their i^{th} trade, E_{ik} is the measure of their experience at the opening of the i^{th} trade, H_{ik}^{p}, ST_{ik}^{p} and TF_{ik}^{p} are the total holding time, total investment size and total trade number during the period of p days before their i^{th} trade. The values of p are 1, 7, 14, and 30. ***, ** and * denote significance at 0.1%, 1% and 5% levels, respectively.

3.5 DISCUSSION

3.5.1 Change of Risk-taking Behavior through Experience

We find evidence that traders exhibit significant behavioral change regarding risk-taking as they gain experience. In particular, traders hold positions for longer and increase investment amount, which leads to higher volatility of returns. This result suggests that traders' level of risk-taking increases as they gain experience. As shown in Table XII, the majority of traders in our data lose money and, consequently, the increased risk-seeking behavior is consistent with the prediction of prospect theory that individuals are risk seeking in loss. This can lead them to become subject to the escalation of commitment, and our results suggest that this escalation comes in the form of the increased length of time they hold a trade.

We also find that as traders gain in experience they increase trading frequency. However, we found that trading frequency is negatively related to the volatility of returns (which is often used as the measure of trading risk). A possible explanation is that when trading frequently they avoid volatile environments. The overall effect is that as traders gain experience they increase their degree of risk-taking (in terms of trading frequency) but they appear to adjust their related behaviors in order to control the additional risk, resulting in their returns showing lower volatility. This can be regarded as a form of risk homeostasis, also known as risk compensation, which refers to the behavior that individuals take actions to decrease potential risk while facing risky situations (Wilde, 1982). This result indicates the complexity of the interaction between different types of risk-taking behavior.

The greater risk taken by traders as they gain experience in terms of holding time and investment size could be explained by overconfidence. There is much evidence that individuals attribute to themselves greater influence on positive outcomes than their actions merit and have inflated estimates of their true ability (Gervais and Odean, 2001; Odean, 1998a). It is possible that as traders become more experienced in trading and more familiar with the environment (e.g., with trading rules, operations of trading platforms), they become overconfident and, as a result, take higher risk. Furthermore, it has been shown that individuals become more overconfident when feedback on decisions is inconclusive or uncertain (Griffin and Tversky, 1992). Since trading generates uncertainty in returns, the trader may attribute positive outcomes to their own ability to predict the future and negative outcomes to failures of the market (Gervais and Odean, 2001). Consequently, the more they trade, the greater their level of overconfidence.

Our results imply that changes in risk-taking behavior do not simply arise as a result of changes in total wealth (i.e., increasing risk taking arising from loss aversion) and prior outcomes (house money effect) but from personal accumulated experience. It has been shown that small fluctuations in investors' attitudes towards risk could increase the volatility of equity prices (Mehra and Sah, 2002). Hence, our results are important as they demonstrate that it is important to take accumulated experience into account when considering the degree of risk an individual decision maker is likely to take.

3.5.2 Risk-Return Trade-off

Our results concerning the impact of trading experience on returns are consistent with the existing literature, namely, that traders make higher returns as they gain more experience (Feng and Seasholes, 2005; Nicolosi et al., 2009; Seru et al., 2010). Earlier studies have regarded this as the evidence that traders learn to improve their performance. However, importantly, our results also show that these higher returns are achieved at the expense of higher risk and that this additional risk causes risk-adjusted performance to decline. These results suggest that traders 'learn' through the experience of trading that they can achieve higher expected returns by taking more risk but, they appear to fail to learn how to handle risk effectively, leading to a lower risk adjusted performance.

3.5.3 Bayesian Learning

One of the fundamental assumptions of neoclassical economic theories is that individuals are rational and attempt to maximize expected utility (Ackert and Deaves, 2009). This assumption has been challenged by a number of observed behavioral biases, e.g. disposition effect (Dhar and Zhu, 2006; Grinblatt and Keloharju, 2001b; Shefrin and Statman, 1985), overconfidence (Barber and Odean, 2001), overreaction (Bondt and Thaler, 1985) and loss aversion (Kahneman and Tversky, 1979). Despite these behavioral biases, it is possible that individuals can learn from experience to improve financial decision-making to the point where the assumptions of neoclassical theories hold. There is strong evidence that individuals are likely to change their behavior with experience (Camerer and Ho, 1999; Charness and Levin, 2003; Roth and Erev, 1995). However, the manner and degree to which they change their behavior remain a matter of debate. In particular, the rationality assumption would be justified if individuals can learn from experience to appropriately adjust their probability estimates and finally make rational, appropriate decisions; a process referred to as Bayesian learning (Charness and Levin, 2003; Chiang et al., 2011).

In response to the challenges from the evidence of irrational behavior, the advocates of rationality often claim that individuals follow Bayesian learning to update their beliefs, i.e., the probability distribution of events, and achieve rational decision making over time. One way to test the conjecture is to examine whether individuals improve performance through experience. For example, Chiang et al. (2011) argue that Bayesian learners should be able to achieve higher returns by learning from experience. However, such an approach fails to take account of the risk perspective in performance measurement, e.g., higher returns may be obtained at the expense of higher volatility (Fama and MacBeth, 1973; Glosten et al., 1993). We argue that both returns and risk-adjusted performance should be considered in measuring the performance improvement of learners. Since Bayesian learners can appropriately update probability estimates conditioned on new information, they should be able to improve their decision choices, enabling them to not only increase returns but also to improve their risk-adjusted performance. Our findings, that traders' risk-adjusted performance decreases the longer they trade, provides strong support for the view that Bayesian learning does not take place, certainly amongst spread traders (Chiang et al., 2011).

3.5.4 Reinforcement Learning

We argue that reinforcement learning, wherein individuals repeat behavior that was rewarding in the past provides a reasonable explanation for our findings. Reinforcement learning describes a process to update the probability of selecting a strategy based on prior outcomes; the strategies with better outcomes in the past having a higher probability of being selected in the future. Psychological evidence suggests that reinforcement learning explains many of the dynamics of human behavior demonstrated in experiments (Roth and Erev, 1995). We would argue that our study provides field study support for this type of learning amongst traders. In particular, individuals may try trading several strategies with different levels of risk. Since higher risk generally brings both higher profits and greater losses, traders who undertake riskier strategies will either make higher profits or suffer greater losses. Those traders making a higher profit are reinforced by the riskier strategies, while those suffering higher losses are likely to quit the market. We also find that as traders gain in experience they tend to take greater risk. It might, therefore, be argued that these individuals are simply reinforced by risk-taking behavior accompanied with high return and high volatility.

3.5.5 Dynamic interaction of risk-taking behavior, volatility and experience

Most empirical studies focus on the influence of one factor, e.g., prior outcomes, on risk-taking behavior (Choi et al., 2009; Glaser and Weber, 2009; Kaustia and Knüpfer, 2008; Y.-J. Liu et al., 2010; Thaler and Johnson, 1990). We examine the dynamic interaction between experience, risk-taking behavior and returns volatility by considering together the results outlined in sections 4.1, 4.2 and 4.3. In particular, these earlier results suggested that as traders gain more experience they tend to hold their trades for longer and to experience higher returns volatility. In addition, we found that there is a positive relationship between returns volatility and risk-taking behavior, i.e., holding time and investment size. That is, our results suggest that as traders gain experience they increase their risk-taking behavior, in terms of holding trades for longer and of higher investment size and it is this which leads to higher volatility of returns.

Furthermore, our results demonstrate that as traders gain in experience they increase their trading frequency. We also found that returns volatility and trading frequency are negatively related. Consequently, the increase in returns volatility experienced by traders as they gain in experience cannot be explained by their increasing trading frequency.

3.5.6 Survivorship Bias

Comparing the results of the two groups those that stay active for shorter and longer periods in the markets, we find that most traders exhibit similar tendency regarding risk-taking behavior and risk-related performance measures. Both groups increase risk-taking behavior, achieve high volatility of returns and lower Sharpe ratios. However, the difference regarding returns between the two groups is not negligible. Those traders staying active for shorter periods decrease their returns when they gain experience. As individuals can easily perceive the downward trend of returns, leaving the markets seems a reasonable choice for those traders who keep losing more money or making less profit. In addition, those traders staying active for longer periods appear to reduce their trading frequency, which is negatively related with returns, as they place more trades. On the other hand, we find that the risk-taking behavior is significantly related with the volatility of returns among the groups of staying for longer periods, but this does not hold for the other group. One possible explanation is that those traders surviving longer have a better understanding on the effect of their behavior and, hence, is more influential on their volatility of returns. Therefore, our findings show that survival bias is influential on the results regarding returns.

3.6 CONCLUSION

We analyze 6,287,477 trades from 27,868 spread traders over a 10 year period. Our results show that traders make higher returns as they gain experience, which is consistent with the findings of Seru et al. (2010), Feng and Seasholes (2005) and Nicolosi et al. (2009). However, our results also show that traders take greater risk with accumulating experience, resulting in lower risk-adjusted performance. It appears, therefore, that traders only improve returns by taking greater risk. The results are consistent with traders' risk-taking behavior being reinforced by the accompanying higher returns. This leads them to take greater and greater risk without fully understanding of the degree of risk they are taking. This conclusion still holds after we take selection bias and survivorship bias into account.

This, to our best knowledge, is the first study to examine the effects of learning through experience on traders in real-world financial markets, taking account of their risk-taking behavior, and their risk-adjusted performance. The results lead us to question one of the important assumptions underlying the neoclassical approach, namely, that individuals learn, following the Bayesian rule, i.e., they appropriately update prior probability estimations conditional on new information or events. Rather, our results suggest that traders learn by reinforcement, choosing strategies which led to ‘better’ outcomes in the past. It appears that in adopting this approach, traders may over-weight the value of returns in assessing what the 'better' outcomes are and this leads them to under-assess the underlying risk.

It is generally believed that individuals need to be able to observe feedback from their actions in order for learning to take place. It is straightforward for traders to observe the returns of their trades. However, it is far less straightforward for them to observe their volatility of returns and their risk adjusted performance. This may explain the different learning effects observed for returns (a positive relationship with experience) and risk relevant performance (a negative relationship with experience). Since the latter is not easy to observe, traders may take little effective actions to improve those measures or to control trading risk. One solution to alleviate this situation is to provide risk-relevant performance measures on trading platforms, such as web sites and mobile applications. It is important to notice that trading platforms may be reluctant to provide their clients risk relevant information since the profit of trading platforms is mainly from commission fee and cautious clients do not always place many trades. Therefore, policy intervention might be needed.

If our results are mirrored in future studies examining the effects of learning on risk adjusted performance in other financial markets, they have important implications for the efficiency of financial markets. In particular, they suggest that it is unlikely that individual traders will learn the lessons of excessive risk-taking from previous periods of excessive exuberance in financial markets which led to bubbles and eventual crashes. As a result, our findings point to the need for intervention by government or financial authorities to adopt measures which make clearer the risk involved in particular assets or investment strategies and/or to provide incentives for traders to focus more on *risk-adjusted* performance.

3.7 APPENDIX: TRADE NUMBER AS EXPERIENCE MEASURE

We use the number of trade as the measure of experience to test H1 and H2.

In particular, the experience of trader k associated with their i^{th} trade (E_{ik}), is defined as:

$$E_{ik} = i, k \in N^{+}; i \in N^{+} \tag{25}$$

where $N^{+}$is a set of non-negative integers.

We estimate the LMMs (21) to examine the impact of experience on risk-taking behavior. The results are presented in Table XXI and are consistent with the results which we obtain by using trading time as experience measure. The trader experience is positively related with all three measures of risk taking behavior: holding time, investment size and trading frequency (i.e. 2186.9, p < 0.001; 0.1194, p < 0.001; 8.0182, p < 0.001, respectively). These results provide support for H1a, H1b and H1, namely that the risk-taking behavior of traders increases as they gain experience.

We present the results of estimating the LMMs using equations (22) in Table XXII. The results show that more experienced traders make lower returns and achieve lower Sharpe ratios. In particular, the coefficients of experience and Sharpe ratio are negative and significant in (22) , respectively (-0.1257, p < 0.001; -0.0119, p < 0.001), while the coefficient of volatility is not significant. Hence, the results do not provide support for H2a, H2b and H2c.

Regarding H2c, the results are consistent with that we obtain by using trading time as experience measure: traders achieve lower Sharpe ratio when they gain experience, no matter we measure experience with trading time or with trade number. That is, as we conclude earlier, traders fail to learn to improve their risk-adjusted performance through experience.

We find that traders make lower returns when they place more trades, which is consistent with Chiang et al. (2011)'s findings, while Seru et al. (2010) and Nicolosi et al. (2009) show opposing results that traders make higher returns when they place more trades. On the other hand, our earlier results show that traders make higher returns when they stay in the market for a longer time, which is consistent with Seru et al. (2010)'s and Nicolosi et al. (2009)'s findings. That is, the current evidence is consistent in supporting that traders make higher returns when they stay in the market for a longer time, but the results regarding trader number are mixed. This suggests that the learning approaches used by traders are influential on the returns. In particular, traders who observe the patterns of market events tend to make higher returns when they stay in markets for a longer time, while traders who learn by actively participating the markets and observing the outcomes of their trades tend to make lower returns when they place more trades.

Table XXI: Risk-taking Behavior Related to Experience (Trade Number)

Risk Taking Measure:	Holding Time	Investment Size	Trading Frequency
Intercept	21584 ***	3.9191 ***	-6.1537 ***
Experience	2186.9 ***	0.1194 ***	8.0182 ***
Controls			
Holding Time		-0.0000003 ***	-0.000002 ***
Investment size	-206.08 ***		0.1696 ***
Trading Frequency	-36.79 ***	0.004 ***	
Accumulated Profit	0.1687 ***	0.00009 ***	0.0004 ***
Last Trade Profit	-11.563 ***	0.000004	-0.0019 ***
Last Week Profit	-0.3003 ***	0.0002 ***	0.0011 ***
Age	64.354	-0.0193 ***	-0.1291 ***
Gender	3274.1	0.5139 **	-2.1016 ***
Market Volatility	-0.291 ***	-0.00002 ***	0.0009 ***
Disposition	0.0391 ***	-0.000002 ***	-0.00003 ***

This table presents results for regressions of the form

$$H_{ik} = \alpha + \beta E_{ik} + \beta^{ST} ST_{ik} + \beta^{TF} TF_{ik} + \beta^{A} A_k + \beta^{G} G_k + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} + \beta^{AP} AP_{ik} + \beta^{LTP} LTP_{ik} + \beta^{LWP} LWP_{ik} + \varepsilon_{ik},$$

where for trader k, H_{ik} is the holding time of theiri^{th} trade, ST_{ik} is the investment size of their i^{th} trade and TF_{ik} is the trading frequency of the i^{th} trade, E_{ik} is the measure of their experience at the opening of the i^{th} trade, A_k is the age, G_k is the gender, $MV_{ik}$is the market volatility, D_{ik} is the estimate of disposition effect, AP_{ik} is the accumulated profit, LTP_{ik} is the last trade profit and LWP_{ik} is the last week profit. We use $\varepsilon_{ik}$to denote the regression error term, and α, $\beta, \beta^{H}, \beta^{ST}, \beta^{TF}, \beta^{A}, \beta^{G}, \beta^{MV}$ and β^{D}

are determined by parameter estimation. Data employed are for the period 2003 to 2013. ***, ** and * denote significance at 0.1%, 1% and 5%, respectively.

Table XXII: Performance Related to Experience (Trade Number)

	Return	Volatility	Sharpe Ratio
Intercept	1.8654 ***	6391.3 ***	0.0214
Experience	-0.1257 ***	-3.151	-0.0119 ***
Controls			
Holding Time	-0.00003 ***	0.0017 ***	-0.000000008 ***
Investment Size	-0.0134 ***	-2.1313 *	0.0003 ***
Trading Frequency	-0.0009 ***	-0.7406 ***	0.00006 ***
Age	-0.0058	-12.075	-0.0022 *
Gender	-0.0093	-4192.9 ***	-0.0382
Market Volatility	-0.0002 ***	0.0209 ***	0.0000001
Disposition	-0.00001 ***	0.0012 ***	-0.00000007 ***

This table presents results for regressions of the form

$$R_{ik} = \alpha + \beta E_{ik} + \beta^{H} H_{ik} + \beta^{ST} ST_{ik} + \beta^{TF} TF_{ik} + \beta^{A} A_k + \beta^{G} G_k + \beta^{MV} MV_{ik} + \beta^{D} D_{ik} + \varepsilon_{ik},$$

where for trader k, R_{ik} is the return of their i^{th} trade, V_{ik} is the variance of returns associated with trader k's i^{th} trade and S_{ik} is trader k's accumulated Sharpe ratio up to the point of closing the i^{th} trade, H_{ik} is the holding time of their i^{th} trade, ST_{ik} is the investment size of their i^{th} trade and TF_{ik} is the trading frequency of the i^{th} trade, E_{ik} is the measure of their experience at the opening of the i^{th} trade, A_k is the age, G_k is the gender, MV_{ik} is the market volatility, and D_{ik} is the estimate of disposition effect. We use ε_{ik} to denote the regression error term, and α, $\beta, \beta^{H}, \beta^{ST}, \beta^{TF}, \beta^{A}, \beta^{G}, \beta^{MV}$ and β^{D} are determined by parameter estimation. Data are from the period 2003 to 2013. ***, ** and * denote significance at 0.1%, 1% and 5%, respectively.

Chapter 4: Estimating Behavioral Characteristics associated with Learning Models in Financial Markets

Abstract

We develop a methodology which allows us to examine the manner and extent to which important behavioral characteristics of individual traders associated with learning models influence the individual's trading behavior and performance. We estimate traders' behavioral characteristics associated with the experience weighted attraction (EWA) behavioral learning model. EWA was developed on the basis of psychological experiments where the number of strategies available to decision makers was limited. The problem of applying the EWA model is that in real-world trading environments traders have no limits in terms of the strategies they can adopt. We propose a decision-based strategy mapping framework (DSM) to resolve this problem. DSM is designed to artificially limit the strategy space associated with real-world trading data, by using scenarios. In each scenario, individuals are assumed to have only one decision to make. This allows us to estimate, using data associated with an individual's real-world trading their behavioral characteristics associated with EWA. Subsequently, we examine the relationship between the estimated behavioral characteristics of traders and their trading behavior and performance. We find that traders who put greater weight on foregone payoffs for unchosen strategies (i.e., opportunity costs) tend to place higher mean stake sizes and trade more frequently. As rational economic agents are generally considered to actively consider opportunity costs when making decisions, our results suggest that those traders who behave like rational agents tend to trade more actively. However, surprisingly, those traders who are more rational do not achieve superior trading performance.

4.1 Introduction

Empirical studies show the influence of an individual's demographic characteristics on their financial trading behavior. For example, men have been shown to be more risk seeking than women in making financial decisions, such as retirement planning and stock investment (Bernasek and Shwiff, 2001; Jianakoplos and Bernasek, 1998, 1998; Pålsson, 1996). In addition, Goetzmann and Kumar (2008) and Greenwood and Nagel (2009) find that young investors tend to hold less-diversified portfolios and expose themselves to higher risk than older investors. These conclusions are made possible because financial trading data often contains individual measurable characteristics, such as age and gender. However, individual behavioral characteristics which have been shown to affect how individuals learn to adjust their behavior are less readily discernable. As a result, the effect of these behavioral characteristics has been largely neglected in empirical studies. To shed light on the effect of these behavioral characteristics, we develop a methodology which allows us to determine, using empirical data, some important behavioral characteristics of individual traders associated with learning models and we examine to what extent these influence the individual's trading behavior and performance.

In particular, we estimate the behavioral characteristics of traders used as parameters in a behavioral learning model (i.e. the experience weighted attraction (EWA) model), which describes how individuals make decisions after receiving feedback from their past actions. Individuals receive a payoff as a return after choosing one strategy from a number of strategies repeatedly in the experimental setting, and the EWA describes the likelihood of choosing the strategies. The EWA identifies the weight of foregone payoffs as the most important behavioral characteristic. This is used to control how unchosen strategies are reinforced. Specifically, a strategy which achieved a superior payoff in the past is reinforced and is more likely to be chosen in the future. For example, if a trader has two available strategies (say buy and sell), then if the better payoffs are achieved when choosing the 'buy' strategy the trader is more likely the trader to choose a 'buy' strategy in the future. However, while the selected strategy is reinforced according to the payoff achieved, EWA also allows unchosen strategies to be reinforced (e.g., as a result of a trader observing the payoff that would have been achieved had that strategy been chosen). The degree foregone payoffs influence an agent's behavior is regarded as a measure of the agent's economic rationality. In particular, more rational economic agents are seen as those whose actions are more driven by accounting for the opportunity (Charness and Levin, 2003; Payzan-LeNestour and Bossaerts, 2014).

EWA was developed on the basis of psychological experiments ((Camerer and Ho, 1999). However, applying behavioral learning models to real-world trading data is problematic as trading in the real-world differs to the controlled settings in which these models were developed. In particular, subjects participating in experiments usually select one option from a few alternatives given all relevant information, i.e., the strategy space of subjects is limited. However, traders in financial markets face an unlimited strategy space and need to make a series of judgments and decisions, including the likely direction in which the market is heading direction, what stake/investment size to choose and when to open and close a position. They must make these judgments and decisions facing a number of sources of public market information, which may or may not be relevant.

We propose a decision-based strategy mapping framework (DSM) to enable us to develop an EWA model for traders based on their real-world trading data. DSM creates the concept of 'possible scenarios.' In each scenario, individuals are assumed to have only one decision to make. For example, in a ‘stake size’ scenario, traders are assumed to believe that the stake size they select when they place a trade is *the one factor* that will affect their eventual return. Similarly, in a ‘buy/sell’ scenario, the traders are assumed to believe the decision of choosing to buy or sell is the one thing that will affect their eventual return. In this way, DSM creates a limited strategy space, which allows us to develop behavioral learning models based on real-world trading data. Subsequently, we are able to assess to what extent the behavioral model we develop actually fits the trading data; i.e., thus, enabling us to assess which of the scenarios best represents the factors that traders consider when placing their trades. This allows us to use an individual’s trading data to estimate their behavioral characteristics and the manner in which they learn for their previous experience; such as the weight they place on foregone payoffs. We are then able to examine the relation between the behavioral characteristics of individual traders and their behavior and trading performance.

We find that traders think that their decisions on stake size are influential in terms of the percentage profit they achieve. In addition, we find that those traders with higher mean stake sizes who trade more frequently tend to put greater weight on foregone payoffs for unchosen strategies (i.e., opportunity costs). Rational economic agents are generally considered to actively consider opportunity costs when making decisions. Consequently, our results suggest that those traders who behave like rational agents tend to trade more actively. An interesting finding is that those traders who are more rational do not achieve better trading performance. In addition, traders who start trading with a preference for lower stake sizes achieve overall better trading performance (higher total profit and lower volatility of returns).

In this study, we employed data from the spread trading market to undertake our analysis since it offers a number of advantages over data from conventional financial markets. In particular, the short-term nature of spread trading enables us to ensure that all trades which we examine have been closed, thereby ensuring that all returns are realized and no estimation of gains is required. In addition, since 75% of the trades are closed within two hours and the median holding time is around 25 minutes, fewer changes occur in the personal status of the trader and fewer changes arise in the environment than is the case for stock market investment. Furthermore, spread trading provides prompt and deterministic feedback concerning the decisions of traders, and this is important to help traders learn from experience and for the application of behavioral learning models.

This is, to our best knowledge, the first study to use empirical trading data to provide evidence that the behavioral characteristics of individuals employed in learning models are significantly related with their trading behavior and performance. More importantly, the DSM framework provides a novel approach to estimate the behavioral characteristics, such as the degree of rationality, which cannot be observed and measured directly from the empirical data. Also, the DSM framework can be used with other behavioral models, in addition to the EWA model, to estimate further characteristics depending on the research topics in in future research.

The remainder of the paper is structured as follows. In section two, we discuss the literature associated with the influence of individual characteristics on trading behavior and on the degree to which traders learn from their previous experience to change their behavior. We also briefly examine the literature introducing the three learning models which we examine in the paper. We describe the data and the methodology in section three. In section three, we present our results. We discuss the results in section five, and this is followed by the conclusion.

4.2 Literature

4.2.1 The influence of individual characteristics on trading behavior

A large body of literature provides evidence of the relationship between trading behavior and demographic characteristics, such as age, gender, marriage status, and wealth. For example, males have been shown to be more risk seeking when making financial decisions when their entire portfolio of assets, including real assets, stock, and bond investment, is considered (Bajtelsmit et al., 1999; Bernasek and Shwiff, 2001; Pålsson, 1996). In addition, it has been shown that risk aversion associated with investment increases with age (Pålsson, 1996). For example, Goetzmann and Kumar (2008) show that younger investors do not tend to diversify their portfolios to the same extent as older investors and Greenwood and Nagel (2009) found that younger mutual fund managers take higher risks by having more inclination to chase trends and to heavily invest in technology stocks.

Marriage status, education level, income level, wealth, and occupation have also all been shown to be associated with differences in investment behavior. For example, Jianakoplos and Bernasek (1998) find that single women are more risk averse than single men and married couples. Goetzmann and Kumar (2008) show that the level of portfolio diversification is lower among low-income and less-educated investors. Similarly, Hartog et al. (2002) suggest that the level of risk aversion decreases as incomes and wealth increase. They also provide evidence that entrepreneurs are more risk seeking than employees and public sector employees are more risk averse than those in private sectors.

Individual behavioral characteristics cannot be directly observed from survey and trading data, without further analysis. However, these characteristics may also affect trading behavior. For example, the disposition effect describes the reluctance of investors to realize losses (Dhar and Zhu, 2006; Odean, 1998b) and cannot be measured directly from trading records. Rather the disposition effect is estimated using a formula which requires one to measure differences between the tendency to realize gains and losses. The discovery of the disposition effect is important since investors suffering higher disposition effect are less likely to make rational decisions and as a result generally display inferior performance (Seru et al., 2010).

We aim to explore other behavioral characteristics of individual traders which are used as parameters of the EWA model. Given that individuals receive a payoff as a return and have a limited number of strategies to choose repeatedly, the EWA model describes the process to update the attractions of the strategies, which is positively influential on the likelihood of choosing a strategy, in the individuals' minds based on the payoffs. These parameters are estimated with empirical trading data and represent how traders learn from their experience. For example, the parameter of the weight on foregone payoffs represents the degree to which an individual's future behavior is affected by the payoffs they would have secured on unchosen strategies in the past (opportunity costs) in adjusting their future behavior. We explain and discuss the details of the parameters of the EWA model in the methodology section. To our best knowledge, this is the first attempt to estimate the behavioral characteristics with empirical trading data and investigate the relationship between the behavioral characteristics and the trading behavior/performance.

4.2.2 The influence of experience: learning effect

There is a rapidly growing literature exploring the degree to which individual investors change their behavior and performance through time. Some studies have examined the influence which past performance (e.g., returns and volatility) has on subsequent investment behavior (Choi et al., 2009; Glaser and Weber, 2009; Kaustia and Knüpfer, 2008; Thaler and Johnson, 1990). Others investigate the influence accumulated experience (e.g., number of trades, the time trading markets) has on behavioral changes and on performance (Dhar and Zhu, 2006; Feng and Seasholes, 2005; Gloede and Menkhoff, 2011; Linnainmaa, 2011; List, 2004; Nicolosi et al., 2009; Seru et al., 2010).

4.2.2.1 The influence of past performance on future behavior

Past returns and their volatility have been shown to be influential on individual's subsequent financial behavior. For example, Thaler and Johnson (1990) found that a loss is more tolerable after earlier gains (the 'house money effect') and individuals are more inclined to take more risk after prior gains. Similarly, empirical studies show that previously good performance arising from trading encourages individuals to engage in the greater trading activity. For example, Kaustia and Knüpfer (2008) found that IPO investors in the Finnish stock market who had higher past returns were more likely to subscribe to the next IPO than those with lower past returns. Similarly, Choi et al. (2009) show that individuals with experienced higher returns or lower volatility of returns from their 401(k) retirement fund are more likely to increase their 401(k) saving rate. Equally, Glaser and Weber (2009) found that higher past returns led individual investors to increase their level of trading activities, (i.e., portfolio turnover and number of transactions). They also found, in line with Liu et al. (2010), that higher past returns led investors to take greater risk (e.g., buying higher-risk stocks and holding a less diversified portfolio). Further evidence that previous experience influences future behavior was presented by Malmendier and Nagel (2009), who found that individuals who have experienced high inflation periods tend to predict higher inflation rates in the future and are more prone to borrowing and not to take nominally fixed-rate investments.

4.2.2.2 The influence of accumulated experience on performance

Empirical evidence suggests that experienced traders, who place more trades or stay active in markets for longer periods, suffer less behavioral bias and achieve better trading performance. It is argued that these traders accumulate knowledge when they trade or remain materially interested in these markets and this enables them to improve their subsequent performance. For example, List (2004) found that inexperienced individuals are more prone to the endowment effect, the tendency to overvalue goods merely because of ownership. Dhar and Zhu (2006), Feng and Seasholes (2005) and Seru et al. (2010) showed that the disposition effect is alleviated when traders have more experience, and Liu et al. (2010) showed that trading experience mitigates narrow framing, the tendency to evaluate the risk of a new event in isolation without taking other risks into account. Similarly, Gloede and Menkhoff (2011) found that financial professionals with more working experience exhibited less overconfidence.

There are several studies reporting a significant positive relationship between investment returns and experience (e.g., Feng and Seasholes (2005), Seru et al. (2010) Nicolosi et al. (2009), Linnainmaa (2011)). In contrast to the above studies showing that experience leads to better returns and lower behavioral biases, Chiang et al. (2011) found that the returns of IPO investors decreased the more auctions in which they participated. A possible reason for the inconsistent results is the nature of IPO auctions which provides a limited trading opportunity (i.e., opportunity to learn and to apply one's experience) compared with general stock market investment.

In summary, experience has been shown to influence individual investment behavior.

4.2.3 Behavioral models of learning

Psychologists and experimental economists have developed several behavioral models, based on the results of laboratory experiments, to explain the process by which individuals learn to improve their performance through time. We briefly introduce these below:

4.2.3.1 Learning approaches: reinforcement and belief learning

There are two approaches which have been developed to describe individual learning behavior: belief learning and reinforcement learning. Belief learning, which is consistent with the rational assumptions underlying the neoclassical economic theories, suggests that individuals learn (i.e., update their beliefs concerning the probability of events) by following the Bayesian rule (Charness and Levin, 2003). That is, individuals update prior probability estimations conditional on new information or events. On the other hand, reinforcement learning suggests that individuals make decisions based on previous outcomes. In particular, strategies which have led to better outcomes in the past are more likely to be chosen.

The results of laboratory experiments suggest that reinforcement learning can explain most of the dynamics of human behavior (Charness and Levin, 2003; Roth and Erev, 1995), but that some individuals behave in a manner consistent with Bayesian learning. For example, Bruhin et al. (2010) showed that around twenty percent of subjects behaved in a manner similar to rational agents in attempting to maximize expected values. Payzan-LeNestour and Bossaerts (2014) reported that subjects followed belief learning in executing investment tasks if they were told explicitly that the investment values would change regularly during the experiment, but their behavior was consistent with reinforcement learning when such instructions were not given.

4.2.3.2 Experience-Weighted Attraction Model

The experience-weighted attraction (EWA) model has been developed to integrate the features of the major learning models discussed above (Camerer and Ho, 1999). The EWA describes the learning process by which individuals choose strategies based on prior results. The subjects in the EWA model experiments have a limited number of strategies to choose. After choosing a strategy, a subject receives a payoff as a return. It is assumed that the attractions of the strategies in the subjects' minds are then updated based on the payoffs. The strategies with higher attraction are more likely to be chosen in the future. The core of the EWA model is the rules applied to update these attractions.

Camerer and Ho (1999) indicate that both belief-based models and choice reinforcement models can be considered to be special cases of the EWA model, by controlling how strategies are reinforced. The first unique feature of the EWA model is that it controls how strategies which are not selected are reinforced with the parameter of the weight on the foregone payoff. The reinforcement model does not reinforce unchosen strategy at all, while the belief model treats chosen and unchosen strategies equally. The EWA model takes the middle ground with this parameter. The weight on foregone payoff also indicates the degree to which an individual resembles a rational economic agent since the belief models assume individuals update probability estimation of events by following the Bayesian rule (i.e., taking into account all events including both chosen and unchosen strategies). This is an approach which is consistent with the rational assumptions underlying neoclassical economic theories (Charness and Levin, 2003; Payzan-LeNestour and Bossaerts, 2014).

Another unique feature of the EWA model is that it includes the initial attractions of the different strategies as a parameter to be estimated from data. Reinforcement models have no concerns about initial attractions, while belief models assign initial attractions as expected payoffs. Initial attractions represent the prior belief and personal preference for strategies. For example, traders who have a high initial attraction for a 'lower stake size' strategy prefer to place small stake sizes in their first trades.

A key difference between reinforcement and belief models is the extent to which attractions either average or cumulate. Reinforcement models calculate the cumulative attractions from the first trade in evaluating strategies, while belief models consider the average attractions. For example, if a trader chooses a strategy three times and receives three payoffs (say, 1, 2 and 9), the cumulative attraction is derived from the cumulative payoff of 1+2+9=12 and the average attraction is from the mean of the payoffs, 4. We skip the formulas transforming payoffs to attractions in this example for simple illustration (these are discussed in section 3.2.1). It is clear that both the cumulative and average attractions could be valuable in evaluating strategies, and the EWA model mixes them by including the parameter of the extent to which attractions are either average or cumulatively-based.

The EWA model also describes the phenomena that individuals may forget earlier payoffs and beliefs. This is achieved by including the depreciation rate of previous attractions. This captures the decreasing weight placed on payoffs from the earlier trades compared to more recent trades. Similarly, the more historical experience of individuals is discounted through the depreciation rate of experience. The EWA model also includes a parameter for prior experience, i.e., the number of periods/trades experienced before the data is collected. The parameter of the sensitivity to attractions allows traders to have a different response to the same level of attractions in the EWA model.

In sum, the EWA model incorporates both reinforcement learning and rational belief-based learning into one model, and the EWA model allows further comparison and parameter estimation.

We estimate the value of the EWA parameters from the trading data of each trader. That is, we find out the best values of the parameters of the EWA model that enable it to describe best the strategy selection of a specific trader from the trades placed by that trader. For this specific trader, these values represent how strategy selection is formed. For example, if the weight on the foregone payoffs of a trader is estimated to be zero, we know that this trader is only affected by the payoffs from the selected strategies. Similarly, a high initial attraction for the low stake size strategy could be a sign of caution, since that trader would prefer to place small stakes in early trades rather than investing boldly. Consequently, the estimated parameters of the EWA model represent the learning behavioral characteristics of individuals in forming strategies to cope with changing environment. Such characteristics cannot be observed directly from demographic and trading data, and can only be developed by estimating learning models. Table 23 shows the characteristics estimated with EWA model in this study.

Table 23: Behavioral characteristics estimated by EWA model

Notation	Behavioral characteristic
A	Initial attraction of strategies
N	Initial experience
ϕ	Depreciation rate of previous attraction
ρ	Depreciation rate of previous experience
δ	Weight of foregone payoffs
λ	Sensitivity of players to attractions
k	The extent to which attractions are averaged or cumulated

This table lists the parameters of the EWA model. Each parameter represents a behavioral characteristic which captures a feature of individual behavior in the learning process.

4.2.4 Hypotheses

As discussed above, there is clear evidence that trading behavior is influenced by demographic characteristics, such as gender and age (Bernasek and Shwiff, 2001; Goetzmann and Kumar, 2008; Greenwood and Nagel, 2009; Jianakoplos and Bernasek, 1998; Pålsson, 1996). We estimate the behavioral characteristics of a trader using the EWA model with the trading data of the trader. Since the behavioral characteristics, such as the weight of foregone payoffs, initial attractions of strategies and the depreciation rate of previous attractions, describe how the experience of a trader affects the current decision making process, we believe that behavioral characteristics are also likely to influence trading behavior. This view is reinforced by the fact that it has been shown that different individuals change their behavior in different ways based on their experience (Chiang et al., 2011; Seru et al., 2010; Thaler and Johnson, 1990). We examine the effect of the behavioral variables (as shown in Table 23), on the specific aspects of trading behavior which are widely investigated in the literature (Glaser and Weber, 2009; Y.-J. Liu et al., 2010) (i.e. stake size, holding time and trading frequency). We examine this conjecture by testing the following hypotheses:

H1a: The relation between the behavioral characteristics in the learning models of a trader and their average stake size is significant.

H1b: The relation between the behavioral characteristics in the learning models of a trader and their average holding time is significant.

H1c: The relation between the behavioral characteristics in the learning models of a trader and their average trading frequency is significant.

Recent studies have examined the relationship between trading performance and the accumulated experience of individuals. The results suggest that individuals can learn from experience to improve their profit (Nicolosi et al., 2009; Seru et al., 2010). We believe that a range of individual behavioral characteristics might affect the manner and degree to which traders learn. In addition, we believe that these are likely to affect their subsequent trading performance because these behavioral characteristics are estimated with the EWA model which captures human learning. We examine this conjecture by testing the following hypothesis:

H2a: The relation between the behavioral characteristics in the learning models of a trader and their total profit performance is significant.

H2b: The relation between the behavioral characteristics in the learning models of a trader and their volatility of returns is significant.

H2c: The relation between the behavioral characteristics in the learning models of a trader and their Sharpe ratio is significant.

Disposition effect is a type of behavioral bias and indicates the tendency for a trader to realize gains faster than losses (Odean, 1998b). Recent studies show that traders can reduce their disposition effect and improve their trading performance when they accumulate trading experience (Dhar and Zhu, 2006; Feng and Seasholes, 2005; Seru et al., 2010). As the behavioral characteristics which we estimate with the EWA model represent the way individuals' learn from experience, we examine the relation between disposition effect, a commonly behavioral bias, and the behavioral characteristics with the following hypothesis:

H3: The relation between the behavioral characteristics in the learning models of a trader and their disposition effect is significant.

4.3 Methodology

4.3.1 Data

The data used in this study was collected from a large spread-trading exchange platform based in the UK Since the 1990s, spread trading has developed quickly in the UK as an important derivative market, due to the relatively low transaction costs, the tax-free status of gains and the ease of access to international markets (Brady and Ramyar, 2006; Paton and Williams, 2005). Traders can either buy or sell the market (e.g., an index) based on their individual prediction of likely market movements. The investment amount and how many points the index rises or falls determines the profit and loss of each trade. For example, a trader might 'buy the index' at, say, £5 per point if he predicts that the FTSE 100 will rise. If the FTSE 100 has risen 20 points when the trader closes his trade, the profit of the trade is £100 (£5 x 20), but if the FTSE 100 has fallen 10 points when the trade is closed, the trader makes a £50 loss. (£5 x 10) Spread traders also can 'sell' the market, in which case profits are made if the market falls.

Compared with traditional stock markets, spread trading data offers three advantages when using learning models to estimate the behavioral characteristics of traders. First, spread trading is short-term, and most trades are closed within one hour (Gulthawatvichai et al., 2013). Hence, all returns are realized, and no estimation of gains is required. By contrast, individuals generally buy shares for the long term, and their value is tied up with potential future dividends. Returns of stock purchases are not definite until they are sold and researchers, therefore, often need to estimate the return of a stock purchase. These estimates can be subject to error. For example, Seru et al. (2010)'s and Nicolosi et al. (2009)'s estimate stock returns over 20-and 30-day periods, respectively, from the date of purchase by assuming all shares are sold in the end of the periods, Barber and Odean (2002) make even stronger assumptions in estimating monthly returns, namely, that all trades occur on the last day of the month. By contrast, we only examine spread trades that have been closed and the returns realized, with no estimation required. Second, there are fewer external factors affecting the decisions of spread traders. In particular, since most spread trades are closed in a very short period, the personal situation of a trader and the economic and political environment are likely to be invariant during the holding time. By contrast, due to the greater time periods involved, stock market investors may sell stocks due to an emergent need for cash or as a result of a change in the economic and political environment. The relative lack of external factors in spread trading markets means that the trader's view of the likely future direction of the market is the primary focus of their decision. Third, thanks to the short-term nature of spread trading, traders receive feedback concerning the decisions they have made in a timely and unambiguous fashion. This is an important component for effective learning to take place (Skinner, 2014)and, based on their experience, traders can refine their subsequent trading decisions. However, the feedback received by investors in stocks is more implicit, since stocks are often held for several months and values of stocks always rely on evaluations concerning what may happen in the future. This lack of immediate, unambiguous feedback can hinder the learning process (Skinner, 2014). Our aim is to estimate traders' behavioral characteristics based on the manner in which they learn to change their behavior (using the EWA learning model). The less ambiguous route to learning in the spread trading market makes the employment of this data particularly valuable.

We examine all the 59,927 closed trades of 1,005 individual spread traders with a large spread-trading broker between October 2003 and March 2013. Since the accuracy of parameter estimation is highly related with the number of data points, we only include those traders with more than the median number of trades (46). This cutoff is used as learning experiments discussed in Camerer and Ho (1999)'s study involve subjects making more than 40 choices.

We collected the following information relating to each closed trade: an individual's identification number, the times the trade was opened and closed, the opening and closing prices, whether the trader bought or sold the market and the amount invested.

Descriptive statistics relating to the data are displayed in Table 24. The distribution of a trader's active period (i.e., the time between their first and last trade in the data period), is right-skewed (median: 306.8 days, mean: 556.4 days, third quartile: 846.9 days), suggesting that some traders continue to trade considerably longer than others. The descriptive statistics also indicate that only a minority of traders (14%) make profits and that the opening and closing of an individual trade takes place in a short time interval: half of the trades being closed within 25.2 minutes of their opening, and over 75% of all trades being closed within 3 hours.

Table 24: Descriptive summary of data

	Mean	**1st Qu**	**Median**	**3rd Qu**
Panel A (1005 traders)				
Age	42.8	33.0	41.0	51.0
Traders' Total Number Trades	59.6	52	59.0	66.0
Traders' Active Period (day)	556.4	88.7	306.8	846.9
Traders' Total Profit/Loss	-1026	-639	-234	-82
Total stake size	226	64	86	185
Panel B (59927 trades)				
Holding time (minute)	1111.2	5.7	25.2	166.9
Stake size	4.1	1.0	1.0	3.0
Profit point	-1.0	-12	1.0	9.0

This table presents a descriptive summary of the data. There are 908 (90.3%) male traders.

4.3.2 Learning models

As shown in section 2, there is strong evidence that an individual's behavior is affected by their previous personal experiences and performance. Learning models describe how individuals respond to their experience, and we intend to use what has been shown to be the most comprehensive of these models to discern the behavioral characteristics of traders. In particular, we employ the experience weighted attraction (EWA) model that combines the features of the two major learning approaches, reinforcement learning and belief learning. This has been shown to provides better explanatory and predictive power than either of its two constituent learning approaches alone (Camerer and Ho, 1999).

The nonlinear interaction of parameters in the EWA model is the reason why EWA, as a model of human learning, is potentially superior in capturing behavioral characteristics of individuals to the linear methods widely used in the literature.

4.3.2.1 Experience Weighted Attraction (EWA) Model

4.3.2.1.1 EWA model notation

In order model the behavior of spread traders using the EWA model, it is necessary to make the assumption that the payoff of one trader is not affected by the strategies of other traders. This is a reasonable assumption since the movement of underlying markets on which the payoffs to spread traders is based largely on the independent actions of spread traders.

Individual traders are indexed by i and each trader i has m strategies. A 'strategy' could be holding a position for a short time or putting a high stake on a trade. That is, the strategy space of trader i is $S_i = \{s_i^1, s_i^2, \ldots, s_i^m\}$. The actual strategy used by trader i in the t^th^ trade is $s_i(t)$. The payoff of trader i from the t^th^ trade is $\pi_i(s_i(t))$. We modify the payoff function from the original version of EWA by not including the strategies of other traders, since we assume that the strategy of one trader do not affect the payoff of another trader, as discussed above. Each strategy $s_i^j, j \in \{1, \ldots, m\}$, for trader i after the t^th^ trade, is assigned a numerical attraction, $A_i^j(t)$, and strategies with a higher attraction are associated with a higher probability of being chosen. $N(t)$ is the experience of trader i after the t^th^ trade and is one of the core variables which are updated for each trade. If learning takes place in accordance with the EWA model then $N(t)$ is updated for each trade:

$$N(t) = \rho \cdot N(t-1) + 1, t \geq 1 \tag{26}$$

The parameter ρ is a discount factor representing how much we depreciate the impact of previous experience.

The other core variable which we update for each trade is the level of attraction, $A_i^j(t)$, which depends on the payoff yielded by a strategy. The hypothetical payoffs that unchosen strategies would have yielded are weighted with a parameter δ. And the actual payoffs which are earned from the chosen strategy $s_i(t)$ are weighted by $1 - \delta$, which makes a total weight of 1. Hence, the weighted payoff can be written as $\left[\delta + (1-\delta) \cdot I(s_i^j, s_i(t))\right] \cdot \pi_i(s_i(t))$, while the indicator function $I(x, y)$ equals 1 if $x = y$ and 0 otherwise.

To update the level of attraction, $A_i^j(t)$, we sum up a depreciated, experience-weighted previous attraction $A_i^j(t-1)$ and the weighted payoff of the t^th^ trade, and subsequently normalize with the updated experience weight:

$$A_i^j(t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(t-1) + \left[\delta + (1-\delta) \cdot I(s_i^j, s_i(t))\right] \cdot \pi_i(s_i(t))}{N(t)} \tag{27}$$

,where ϕ is a discount factor depreciating previous attraction.

We follow Camerer and Ho (1999) and use the logit function to transform the level of attraction to the probability of choosing strategies. In particular, the probability of choosing strategy j for the t+1th trade of trader i is given by:

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m} e^{\lambda \cdot A_i^k(t)}} \tag{28}$$

The parameter λ represents the sensitivity of traders to attractions. This could vary considerably among traders due to both economic and psychological reasons, such as personal wealth, motivation in trading and perception of awards.

There are six parameters to be estimated: $A_i^j(0)$, $N(0)$, φ, ρ, δ and λ. $A_i^1(0)$, also shorted as A, is the initial attraction of the first strategy for trader i. $N(0)$ is the estimated prior experience which a trader already had before the data was collected, ϕ is the depreciation rate of previous attraction, and ρ is the depreciation rate of previous experience. δ is the weight of unchosen strategies and λ is the sensitivity to attractions.

4.3.2.1.2 EWA parameter interpretation

Weight of foregone payoffs, δ

The most important feature of the EWA model is the parameter, δ, which is used to control how unchosen strategies are reinforced. In the reinforcement model, the chosen strategy is reinforced according to the payoff, and unchosen strategies are not reinforced. The EWA model reinforces the unchosen strategies based multiplying δ by the payoff which would have yielded. That is, reinforcement models do not reinforce unchosen strategies (i.e., δ =0), while the EWA model reinforces unchosen strategies based on the payoffs of these unchosen strategies multiplied by δ. If δ is greater than zero, unchosen strategies are reinforced to a certain extent.

The δ parameter demonstrates clearly the difference among EWA, reinforcement and belief models in capturing the law of actual effect and the law of simulated effect. These are the two basic principles of learning. Many learning experiments, mainly with animal subjects, show that the chosen strategies with superior payoffs are more likely to be chosen subsequently. This is called the law of actual effect as the rewards for actual choices affect subsequent choices. The law of simulated effect states that subsequent choices could be affected by foregone rewards. The unchosen strategy may be considered as simulated successes if high payoffs would have been yielded had a particular strategy been chosen. Simulated successes then could reinforce unchosen strategies and increase the probability of their being chosen in the future. The experiments supporting this simulated effect principle have mostly been conducted with human subjects.

Hence, the key to distinguishing different learning models is the empirical explanation power of the law of actual effect and the law of simulated effect. The reinforcement model claims that only actual payoffs are influential (δ = 0), while belief models consider actual and simulated effects equally (δ = 1). The EWA model allows more flexibility and determines δ based on the data.

Furthermore, the parameter δ can be considered as a proxy measure of the extent to which a trader behaves like a rational economic agent. This is the case since the core concept of belief models is that probability estimates of events are updated by following the Bayesian rule, which is consistent with the rational assumptions underlying the neoclassical economic theories (Charness and Levin, 2003; Payzan-LeNestour and Bossaerts, 2014).

Depreciation rate of previous attraction and experience, ϕ, ρ and k

The parameter ϕ depreciates the level of attraction, $A_i^j(t)$ of a payoff. This parameter captures the decreasing weight placed on more historical trades (cf. to more recent trades). Similarly, the parameter ρ depreciates the experience measure N(t). Consequently, these two parameters are used to describe the cognitive phenomena that in a changing environment individuals discount old experience and forget prior beliefs.

To understand how ϕ and ρ control the decay in the strength of prior beliefs, we consider the numerator and the denominator of the equation (27) in terms of reinforcement and belief models, respectively. The numerator is a running total of depreciated attraction: $\phi \cdot N(t-1) \cdot A_i^j(t-1) + \left[\delta + (1-\delta) \cdot I(s_i^j, s_i(t))\right] \cdot \pi_i(s_i(t))$. The denominator represents accumulated and depreciated experience: $\rho \cdot N(t-1) + 1$. In reinforcement models, the attractions are only affected by payoffs, not by experience; i.e. ρ = 0 and the denominator is always one. Consequently, reinforcement models monitor the running total in the numerator while the experience (the denominator) is never updated. On the other hand, belief models, which require ϕ = ρ, also monitor the running total of attraction (in the numerator) but divide by the accumulated experience (represented by the depreciated number of trades). Consequently, in evaluating strategies, the reinforcement models consider the cumulative (and depreciated) attractions from the first trade, while belief models look at the average (and depreciated) attractions. Both cumulative and average attractions could be helpful in evaluating strategies, and the EWA model mixes them by normalizing depreciated cumulative attraction by depreciated experience.

Camerer et al. (2002) use the k notation to make it clearer that the key difference is the extent to which attractions either average or cumulate: $k = 1 - \frac{\rho}{\phi}$. The equation (26) can be rewritten as:

$$N(t) = \rho \cdot N(t-1) + 1 = (1-k) \cdot \phi \cdot N(t-1) + 1, t \geq 1 \tag{29}$$

When k = 1, the denominator of the equation (27) (i.e., $N(t)$) is always one as we discuss above for reinforcement models. In this case, attractions of strategies, which are updated with the equation (27), can cumulate without any boundary. Consequently, the accumulated attraction of a strategy could be much larger than that generated from the highest payoff. When k = 0, the denominator of the equation (27) is the depreciated number of trades (representing the accumulated experience). Therefore, attractions are weighted averages of lagged attractions and cannot grow outside the bounds of the payoffs.

Initial attractions $A_i^1(0)$ and $N(0)$,

The term $A_i^j(0)$ represents the initial attraction of strategies. Belief models require the initial attraction of strategies to be derived from prior beliefs, while EWA model leaves them as parameters to be estimated based on data. The parameter $N(0)$ is the prior experience and represents the strength of the initial attractions. A small $N(0)$ means that the effect of the initial attractions disappear rapidly, while the effect of the initial attractions persists for a long time if $N(0)$ is large.

Sensitivity to attractions, λ

The parameter λ represents the sensitivity to attractions. Traders could have a different response to the same level of attractions due to personal status, which affects the probability of choosing strategies

4.3.2.2 Reinforcement model

In order to estimate traders' behavior using the reinforcement model we estimate the EWA model with δ=0, ρ=0 and N(0)=1 (Camerer and Ho,(1999) . The EWA model works like choice reinforcement models when δ=0, ρ=0 and N(0)=1, where N(0) represents the initial experience and setting N(0) as 1 indicates the lack of prior experience.

4.3.2.3 Belief model

In order to estimate traders' behavior using the belief model, we estimate the EWA model with δ=1, ρ= ϕ (Camerer and Ho,(1999).

4.3.2.4 Parameter Estimation

In order to estimate the parameters of the EWA model, we use the log-likelihood function from Camerer and Ho (1999) with minor modification since we estimate the parameters for each individual trader rather than treating all traders collectively as a representative agent:

$$LL(A_i(0), N_i(0), \varphi_i, \rho_i, \delta_i, \lambda_i) = \sum_{t=1}^{T} \ln\left(\sum_{j=1}^{m} I\left(s_i^j, s_i(t)\right) \cdot P_i^j(t)\right), \tag{30}$$

where $P_i^j(t)$ is the probability of choosing strategy j for the t^{th} trade of trader i, $s_i(t)$ is the actual strategy chosen by trader i for the t^{th} trade, $s_i^j(t)$ is the predicted strategy for the t^{th} trade of trader i, and $I(s_i^j, s_i(t))$ equals to 1 if the model prediction is correct, which means $s_i^j = s_i(t)$, and 0 otherwise. In this study, we limit the number of available strategies to two, so m is always two. The basic idea behind this log-likelihood function is to add up the predicted probability of those correct predictions for a given set of parameters. We use zero instead of $\ln(0)$ in rare cases. We use the maximum-likelihood estimation method to search for the parameters which maximize the LL function. We tried a variety of starting points to avoid converging to local optima.

4.3.3 Decision-based Strategy Mapping Framework (DSM)

The aim of DSM is to solve the unlimited strategy problem in applying learning models to empirical trading data. The basic idea is to assume that in one *scenario* all traders believe that only one (or more) type of *decision* is relevant to their trading outcomes. Using our data we are later able to assess to what extent this is a valid assumption.

This has the advantage of enabling us to determine the behavioral strategies based on real-world trading data, and this is achieved by assuming that the traders face a limited number of strategies (i.e., limited strategy space) under certain assumptions. We can have as many scenarios as we want to cover all potential *decisions* given the available data.

A *decision* is the basic unit of the modeling process and could be a decision concerning, for example, stake size or holding time. The level of a *decision* is the number of available options for a given *decision* (usually ≥ 2). Hence, the number of available *strategies* in one *scenario* depends on the number of the relevant *decisions* and the level of the *decisions*. For example, if all traders are assumed to believe that the only relevant *decision* is whether to buy or sell, they only have two possible *strategies* available in the *scenario*: buy or sell. Consequently, the number of available strategies is limited under certain assumptions. In one scenario, the combination of decisions forms the strategies. The level of a decision, *b*, is the number of available options for one decision. The number of relevant decisions in one scenario is d. The number of available strategies is b^d. Hence, the strategy space of trader i is

$$S_i = \{s_i^1, s_i^2, s_i^3, \dots, s_i^m\}, \text{ where } m = b^d.$$

For example, in a scenario in which the only relevant decision is to buy or sell, the number of available strategies is $b^d = 2^1$, where b is 2 because there are two options: buy and sell, and d is 1 as there is only one relevant decision in this scenario.

4.3.3.1 Decisions

We examine five types of decision made by traders concerning an individual trade: the stake size to invest, whether to buy or sell, the holding time, length of time traders decide to hold positions showing a profit (profit time) and length of time traders decide to hold positions showing a loss (loss time). In order to demonstrate this methodology, we assume that b = 2 for all decisions faced by the spread traders in our study.

In the decision relating to whether to take a short or long position the strategies naturally fall into two: whether to buy or sell. For those decisions involving time and money, some conversion is necessary. In particular, we convert all possibilities into two options by classifying them into two groups based on the median for a given trader. For example, a decision by trader i concerning stake size for a given trade is regarded as a choice of whether to invest more or less than the median of the stake size of all trades placed by trader i. In this manner, an infinite number of possible strategies can be transformed into a limited, tractable number. This enables us to develop some understanding of the behavioral strategies used by traders and allows us to use the data from all traders (since the approach used guarantees that all available strategies which we examine are available to all traders, i.e., each trader can decide whether to stake more or less than their median amount on a given trade) .

For example, a trader is assumed to make a decision on a given trade of whether to invest above or below their median stake size. Let us assume that they decide to stake more than the median level and this trade resulted in profit. Consequently, if the trader learns by reinforcement, we would expect that their probability of selecting the same strategy (higher than median stake size) will increase on the next trade.

4.3.3.2 Comparison among scenarios

An important feature of DSM is its ability to enable us to assess which scenario best fits the real trading data. Since the strategy space is limited in a scenario under DSM, we can estimate learning models using our empirical trading data and examine the explanatory power of the models. In particular, we can report model fitness measures, such log-likelihood, to compare how well different models describe the data. Consequently, if the EWA model estimated for the 'stake size' scenario (i.e., stake size is the only relevant decision considered by traders) is shown to fit the trading data better than the model developed for 'holding time' scenario (i.e. holding time is considered as the relevant decision to make for profit), then this will suggest that it is more reasonable to assume the 'stake size' scenario than the 'holding time' scenario. In other words, it would suggest that when seeking to increase profits, the results suggest that traders pay most attention to the stake size of their trade rather than holding time.

4.3.4 Applying Learning Models to Real Markets

Figure 5 illustrates how we estimate the learning models using the empirical trading data. We examine five scenarios in this study: stake size, buy/sell, holding time, profit time and loss time. In each scenario, only one decision is assumed to be considered as relevant when considering which outcome will be achieved (i.e., the strategy space is limited). For example, in the 'stake size' scenario, we assume that traders believe that the only thing affecting the size and sign of their payoff is whether they select a stake size above or below their median stake size. Whilst, this may appear a restrictive assumption, we are able to test how well such a model fits the actual trading data.

For each scenario, we produce 'strategy data' from trading data by converting the relevant decisions into strategies. We can estimate learning models directly from the strategy data (consisting of pairs of strategies and payoffs). For example, in the 'stake size' scenario: if the stake size of a trade is lower than his/her median stake size, we convert that stake size in the 'lower stake size' strategy; otherwise we convert it to 'higher stake size' category. The conversion is unnecessary in the 'buy/sell' scenario as the number of options is naturally two. For our study, we assume that the profit point (similar to return rate) obtained as a result of a chosen strategy is the payoff received by the trader.

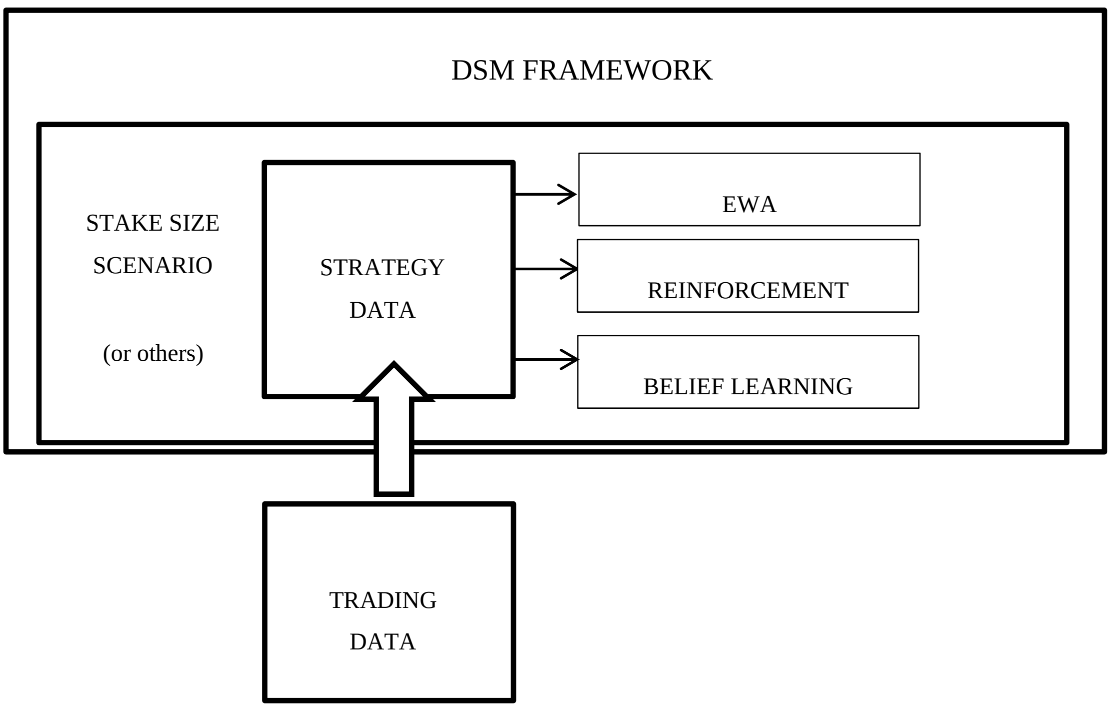


Figure 5: DSM Framework Illustration Chart

This figure illustrates the relationship among the components of the DSM framework.

We estimate the parameters of each model (EWA, reinforcement, and belief based learning model) using the maximum-likelihood estimation method for each trader and for each scenario. We compare the degree to which the various models fit the data across the five scenarios. If one model fits a particular scenario better than for all other scenarios, we can conclude that the assumptions of the scenario are more likely to be legitimate than those relating to the other scenarios. We measure model fitness using log-likelihood, Akaike's information criterion (AIC)(Akaike, 1974) and Bayesian information criterion (BIC), and we present the results of log-likelihood ratio tests.

We aim to examine the relationship between the behavioral characteristics and the trading behavior/performance. The behavioral characteristics which we aim to examine are essentially the parameters of the EWA model. Since we have estimated those parameters for each trader, we have the estimated values of the behavioral characteristics for each trader. We test our hypotheses by regressing the trading behavior/performance on the behavioral characteristics.

4.3.5 Variables

4.3.5.1 Dependent variables

4.3.5.1.1 Trading behavior

Due to the highly volatile nature of spread-trading markets(Chordia et al., 2001), it is generally agreed that traders who choose longer holding times are exposing themselves to greater risk (as they are exposed to greater market uncertainty). The average holding time of the trader i, H_i, is defined as the average number of seconds between the opening and closing of the trades.

Stake size is directly associated with the potential profit/loss of a trade, and larger stake sizes are generally regarded as higher risk options (Y.-J. Liu et al., 2010);(Fehr-Duda et al., 2010). The average stake size of the trader i is denoted by ST_i.

We measure trading frequency with the average number of trades per day for trader i during the period trader i participates the market (TF_i).

4.3.5.1.2 Trading performance

We measure the success of a trader by the total profit/loss earned by trader i (TP_i: the sum of the profit/loss of all trades placed by trader i) and to account for the risk taken by a trader we measure the volatility or variance of returns. . Specifically, the overall trading risk for trader i, V_i, is defined as the variance of returns from their all trades:

$$V_i = variance(R_{ij}), j \in [1, TN_i],$$

, where R_{ik} is the number of points won or lost on the j^{th} trade of trader i.

Trader i's risk-adjusted performance is measured with the Sharpe ratio (Sharpe, 1998), S_i across their all trades, as follows:

$$S_i = \frac{mean(R_{ij})}{\sqrt{variance(R_{ij})}}, j \in [1, TN_i].$$

4.3.5.1.3 Disposition effect

The disposition effect is a behavioral bias influencing investment behavior and performance (Seru et al., 2010). We follow Odean (1998b) and Dhar and Zhu (2006) in estimating disposition effect, i.e., the tendency for a trader to realize gains more quickly than losses.

Odean (1998b) and Dhar and Zhu (2006) determine 'realized gains' and 'realized losses' and 'paper gains' and 'paper losses' for an investor at the time a stock is sold. One is added to the realized gain for a given investor if a sale is profitable. Otherwise, one is added to 'realized loss.' The remaining open trades in the investor's portfolio contribute to the paper gain or paper loss in a similar fashion at the time of the sale. Odean (1998b) then determine the proportion of gains realized (PGR) and proportion of losses realized (PLR) as follows:

$$PGR = \frac{Realized\ Gain}{Realized\ Gain + Paper\ Gain}$$

$$PLR = \frac{Realized\ Loss}{Realized\ Loss + Paper\ Loss}.$$

The disposition effect is defined as the difference between PGR and PLR:

$$Disposition = PGR - PLR.$$

In particular, if this difference is positive, it suggests a greater tendency to realize gains than losses.

We adapt the method of Odean (1998b) and Dhar and Zhu (2006) to make it suitable for examining spread trades. Specifically, since spread traders usually have few positions opened, we employ a method first proposed by Fraser-Mackenzie et al. (2013) to measure disposition effect for spread traders. This effectively counts the paper gains/losses associated with a particular trade. Due to the short-term nature of spread trading, we count the number of minutes a trade is in profit/loss as the paper gains/losses associated with that trade. In other words, it is assumed that, after a trader opens a position, a series of hold/close decisions must be made, every minute. One is added to paper gain/loss if a position in profit/loss is held for one minute. We estimate the disposition effect for each trader i (D_i).

4.3.5.2 Control variables

We control for a number of factors which might be expected to affect the manner in which an individual trades, including the demographic variables age A_k and gender G_k (1: male; 0: female), two trading intensity variables: total number of trades placed by trader i (TN_i), and the total amount invested by trader i (TM_i : the sum of all the amounts invested by trader).

4.3.6 Testing Hypotheses

The hypotheses were constructed to examine the relationship between the behavioral characteristics associated with the learning models and the trading behavior/performance and disposition effect. To test these, we regress the dependent variables on the behavioral characteristics (i.e., the parameters of the EWA model).

We test the hypothesis H1a with the following regression model:

$$ST = \alpha + \beta_A A + \beta_N N + \beta_\phi \phi + \beta_\rho \rho + \beta_\delta \delta + \beta_\lambda \lambda + \beta_k k \tag{31}$$
$$+\beta_{Age} Age + \beta_{Male} Male + \beta_{TN} TN + \beta_{TM} TM + \varepsilon,$$

where ST is the average stake size of the trader, A is the initial attraction of the first strategy, N is the initial experience, ϕ is the depreciation rate of previous attraction, ρ is the depreciation rate of previous experience, δ is the weight of foregone payoffs, λ is the sensitivity of players to attractions, k is the extent to which attractions average or cumulate, Age is the trader's age, $Male$ is 1 if the trader is male and 0 otherwise, TN is the trader's total number of trades, and TM is the trader's total amount of stake size. We test the hypothesis H1b and H1c by changing the dependent variable of this model to the average holding time of the trader (HT) and the trading frequency (average number of traders per day) of the trader (TF).

Similarly, we use the regression model (6) with the dependent variable changed to TP (the total profit point of the trader), V (the volatility of returns of the trader) and S (the Sharpe ratio of the trader) to test the hypothesis H2a, H2b and H2c, respectively. Hypothesis H3 is tested with the the regression model (6) with DE (the disposition effect of the trader) as the dependent variable.

4.4 Results

4.4.1 Scenario/learning model comparison

We estimate the EWA, reinforcement and belief learning models for each of the five scenarios (i.e., stake size, holding time, total profit, time in profit and time in loss) and measure the goodness of fit of these models (i.e., to what extent the models represent the data). These results are presented in .

Table 25: Model Fitness in Five Scenarios

. We measure model fitness in terms of log-likelihood, AIC, and BIC. The results of likelihood ratio tests to measure whether one model is a significantly better fit than another are also presented. Models with smaller AIC/BIC and higher log-likelihoods better fit the trading data.

These results demonstrate three important findings. First, the results of likelihood ratio tests associated with all five scenarios show that all the behavioral models outperform a model based on a random guess. This provides support for the applicability of behavioral models for understanding trading behavior in real financial markets.

Second, the EWA model outperforms the reinforcement and belief learning models in all scenarios. This finding is consistent with the results from experiments (Camerer and Ho, 1999). The results imply that in each of the scenarios, the EWA model describes the real trading behavior of individuals better than the reinforcement and belief models. In other words, the EWA model best describes the manner in which traders learn to change their behavior when seeking to secure trading profit, if we assume that traders these decisions focus on just one relevant factor (i.e. stake size, buy/sell direction, holding time, time in profit, time in loss).

Third, in the 'stake size' scenario, the model fit differences between a model assuming traders act in a random fashion and the three learning models is much higher than that in all other scenarios. This indicates that the behavior of the traders is better explained by the behavioral models in the 'stake size' scenario than in other scenarios. This suggests that it is most reasonable to assume the 'stake size' scenario. In other words, when seeking to increase profits, the results suggest that traders pay most attention to the stake size of their trade rather than holding time, buy/sell direction, time in profit and time in loss.

Table 25: Model Fitness in Five Scenarios

Stake size scenario	AIC	BIC	LL	LR	df	p
EWA	25948.4	26002.4	-12968.2	56697.8	6	$<10^{-8}$

Reinforcement	28290.5	28317.4	-14142.2	54349.7	3	$<10^{-8}$
Belief	26903.2	26939.1	-13447.5	55739.0	4	$<10^{-8}$
Random guess	82634.2	82634.2	-41317.1			

Buy/sell scenario

	AIC	BIC	LL	LR	df	p
EWA	76958.7	77012.7	-38473.3	5687.4	6	$<10^{-8}$
Reinforcement	81735.7	81762.7	-40864.8	904.49	3	$<10^{-8}$
Belief	80753.1	80789.1	-40372.5	1889.1	4	$<10^{-8}$
Random guess	82634.2	82634.2	-41317.1			

Holding time scenario

	AIC	BIC	LL	LR	df	p
EWA	81484.3	81538.3	-40736.1	1161.9	6	$<10^{-8}$
Reinforcement	82552.6	82579.6	-41273.3	87.631	3	$<10^{-8}$
Belief	82279.4	82315.4	-41135.7	362.70	4	$<10^{-8}$
Random guess	82634.2	82634.2	-41317.1			

Profit time scenario

	AIC	BIC	LL	LR	df	p
EWA	81593.3	81647.3	-40790.6	1052.8	6	$<10^{-8}$
Reinforcement	82432.6	82459.6	-41213.3	207.55	3	$<10^{-8}$
Belief	82330.6	82366.6	-41161.3	311.56	4	$<10^{-8}$
Random guess	82634.2	82634.2	-41317.1			

Loss time scenario

	AIC	BIC	LL	LR	df	p
EWA	81288.1	81342.1	-40638.0	1358.0	6	$<10^{-8}$
Reinforcement	82003.9	82030.9	-40998.9	636.30	3	$<10^{-8}$
Belief	81905.8	81941.8	-40948.9	736.34	4	$<10^{-8}$
Random guess	82634.2	82634.2	-41317.1			

This table presents the results of model fitness across scenarios. Akaike's information criterion (AIC) is calculated with the following formula:

$$AIC = -2LL + 2d + \frac{2d(d+1)}{n-d-1}$$

, where LL is log-likelihood, d is the number of parameters and n is the number of samples. Bayesian information criterion (BIC) is calculated with the following formula:

$$BIC = -2LL + d * log(n)$$

LR is likelihood ratio, df is the degree of freedom and the p-value derived from likelihood ratio test which assumes chi-square distribution.

4.4.2 Hypothesis test

As indicated above, the results suggest that traders pay most attention to stake size when seeking to increase profits. This suggests that the 'stake size' scenario is the one most likely to reveal the behavioral factors related to a given trader. In addition, we found that the EWA model outperforms both the reinforcement and belief learning models. Hence, to test the hypotheses, we apply the EWA model in the 'stake size' scenario, and we examine the relationship between the behavioral characteristics, and the trading behavior, trading performance and disposition effect.

4.4.2.1 Testing hypotheses concerning the relationship between behavioral characteristics of individuals and their trading behavior

The results of estimating equation (31) to examine the relation between average stake size and behavioral characteristics of traders (H1a) are presented in Table 26. The results show that the initial attraction of the first strategy (A) is positively related with the average stake size (0.1411, $p < 0.001$). As the first strategy in the 'stake size' scenario is the lower stake size strategy, a high value of the initial attraction of the first strategy (A) indicates that a trader is likely to choose a low stake size for the first trade. That is, cautious traders, who are more likely to use a lower stake size strategy in the beginning, end up displaying higher average stake sizes than other traders. The weight of foregone payoffs (δ) is positively related with the average stake size (7.0571, $p < 0.001$).

Table 26: Regression Result – Estimated Parameters of EWA Models on Trading Behavior

	Stake Size Scenario					
	ST		HT		TF	
	Coef.		Coef.		Coef.	
(Intercept)	5.4003	***	1637.6	**	12.687	
	(5.5456)		(548.55)		(7. 8156)	
A	0.1411	*	-9.8555	***	-0.0128	
	(0. 0276)		(2.7688)		(0. 0444)	
N	-0.00005		-0.0088	***	0.00005	
	(0. 00002)		(0.00212)		(0. 00003)	
ϕ	1.4864		-590.91	*	2.4248	

	(2. 3292)	(230.45)	(3. 688)
ρ	-2.7125 ***	631.28	-8.2548
	(5. 732)	(567.01)	(9. 0857)
δ	7.0571	-67.213	8.3847 **
	(1. 6632)	(166.32)	(2. 6617)
λ	-2.1201	425.52 *	-2.0699
	(1. 8638)	(184.47)	(2. 9286)
k	-4.1638	-1015.4 *	-4.5648
	(4. 9456)	(489.39)	(7. 8438)
Age	-0.0294	1.0155	-0.0699
	(0.0297)	(2.9436)	(0. 0472)
$Male$	0.6974	93.203	-5.9271 **
	(1. 2496)	(123.62)	(1. 9814)
TN	0.0094	-6.273	-
	(0. 0433)	(4.2861)	
TM	-	-0.0125	-0.0001
		(0.0566)	(0. 0009)
R^2	0.1173	0.0773	0.0379
Observations	1005	1005	1615

The regression model is:

$$ST = \alpha + \beta_A A + \beta_N N + \beta_T T + \beta_\phi \phi + \beta_\rho \rho + \beta_\delta \delta + \beta_\lambda \lambda + \beta_k k$$
$$+ \beta_{Age} Age + \beta_{Male} Male + \beta_{TN} TN + \varepsilon,$$

where ST is the average stake size, A is the initial attraction of the first strategy, N is the initial experience, ϕ is the depreciation rate of previous attraction, ρ is the depreciation rate of previous experience, δ is the weight of foregone payoffs, λ is the sensitivity of players to attractions, k is the extent to which attractions average or cumulate, Age is the trader's age, $Male$ is 1 if the trader is male and 0 otherwise and TN is the trader's total number of trades . ***, ** and * denote significance at 0.1%, 1% and 5% levels, respectively.

The results of estimating equation (6) to examine the relation between average holding time and behavioral characteristics (H1b) are presented in Table 26. The results show that the initial attraction of the first strategy (A) is negatively related with the average holding time (-9.8555, $p < 0.001$). This implies that cautious traders hold their positions longer. We also find that the depreciation rate of previous attraction (ϕ) is negatively related with the average holding time (-590.91, $p < 0.05$).

The results of estimating equation (6) to examine the relation between trading frequency and behavioral characteristics (H1c) are presented in Table 26. The result shows that the weight of foregone payoffs (δ) is positively related with the trading frequency (8.3847, $p < 0.01$).

In summary, we find evidence supporting H1 that the behavioral characteristics are related with trading behavior. Although it is usually assumed that risk-averse traders tend to be cautious and avoid taking the trading risk, we find mixed results regarding various trading behaviors. In terms of holding time and trading frequency, our results show that the cautious traders lower their risk by holding positions in a short time and trading less frequently. On the other hand, the cautious traders place higher stake size on average than others. Therefore, our results indicate the complexity of risk behavior, particularly when individuals can learn from experience.

4.4.2.2 Testing hypotheses concerning the relationship between behavioral characteristics of individuals and their trading performance

The results of examining the relation between an individual's behavioral characteristics and the total profit they earn (H2a) are presented in Table 5. The results show that the initial attraction of the first strategy, A, is positively related with total profit (57.091, $p < 0.001$). This implies that those traders who are more likely to use lower stake size strategy when they commence trading tend to make greater total profits throughout their trading history. The results show that k is negatively related with total profit (-8030.9, $p < 0.01$). Since k represents the extent to which attractions either average or cumulate, the result implies that traders who average the payoffs rather than cumulate make higher total profit. I also find that the depreciation rate of previous experience (ρ) is negatively related with the total profit (-11626, $p < 0.001$).

Table 5: Regression Result – Estimated Parameters of EWA Models on Performance

	Stake Size Scenario				
	TP		V		S
	Coef.		Coef.		Coef.
(Intercept)	6370.3	*	9899.4		-0.0906
	(3055.6)		(7103.1)		(0.0786)
A	57.091	***	-77.956	*	-0.00003
	(15.423)		(35.852)		(0.0004)
N	0.0365	**	-0.0803	**	0.00000008
	(0.0118)		(0.0275)		(0. 00000003)

ϕ	733.68 (1283.7)	-6937.2 * (2984.1)	-0.0463 (0. 033)
ρ	-11626 *** (3158.5)	21361 ** (7342.1)	0.1042 (0. 0812)
δ	-1141.8 (926.44)	-804.4 (2153.6)	0.0392 (0. 0238)
λ	-352.73 (1027.6)	9212.3 *** (2388.7)	0.0084 (0. 0264)
k	-8030.9 ** (2726.1)	1280.1 (6337)	0.0662 (0. 0701)
Age	-2.6697 (16.397)	1.1281 (38.117)	-0.0003 (0. 0004)
$Male$	119.1 (688.63)	733.68 (1600.8)	-0.0228 (0. 0177)
TN	33.476 (23.875)	-167.54 ** (55.5)	-0.0005 (0. 0006)
TM	-13.18 *** (0.3151)	-0.1304 (0.7326)	-0.00003 *** (0. 000008)
R^2	0.6934	0.0783	0.0328
Observations	1005	1005	1005

The regression model is:

$$TP = \alpha + \beta_A A + \beta_N N + \beta_T T + \beta_\phi \phi + \beta_\rho \rho + \beta_\delta \delta + \beta_\lambda \lambda + \beta_k k$$
$$+\beta_{Age} Age + \beta_{Male} Male + \beta_{TN} TN + \beta_{TM} TM + \varepsilon,$$

where TP is the total profit of the trader, A is the initial attraction of the first strategy, N is the initial experience, ϕ is the depreciation rate of previous attraction, ρ is the depreciation rate of previous experience, δ is the weight of foregone payoffs, λ is the sensitivity of players to attractions, k is the extent to which attractions average or cumulate, Age is the trader's age, $Male$ is 1 if the trader is male and 0 otherwise, TN is the trader's total number of trades, and TM is the trader's total amount of stake size. ***, ** and * denote significance at 0.1%, 1% and 5% levels, respectively.

The results of examining the relation between behavioral characteristics and the volatility of returns (H2b) are presented in Table 5. We find that the volatility of returns is negatively related with the initial attraction of the lower stake size strategy, A (-77.956, $p < 0.05$) and the prior experience, N (-0.0803, $p < 0.01$). The depreciation rate of previous attraction, ϕ, is negatively related with the volatility of returns (-6397.2, $p < 0.05$). This suggests that traders who weight previous payoffs higher achieve lower volatility of returns. In addition, the depreciation rate of previous experience, ρ, is positively related with the volatility of returns (21361, $p < 0.01$). This suggests that the traders who are more affected by their previous experience have higher volatility of returns. It is important to notice that ϕ and ρ have opposite effects on the volatility of returns, yet both ϕ and ρ are closely related with traders' experience. This suggests that the influence of experience on trading performance is complicated.

We also find that the sensitivity to attractions (λ) is positively related with the return volatility (9212.3, $p < 0.001$).

The results of examining the relation between traders' behavioral characteristics and Sharpe ratios (H2c) are presented in Table 5. The results show that none of the coefficients of the characteristics are significant. Consequently, we fail to find evidence to support H2c that there is a significant relationship between an individual's behavioral characteristics and the Sharpe ratios they achieve throughout their trading history.

In summary, the learning characteristics regarding stake size are significantly related to the total profit and the volatility of returns achieved by traders throughout their trading history. This is not surprising since the stake size is related with the trading risk which will affect the trading performance including profit and volatility of returns. However, our results do not support the view that relation between the learning characteristics regarding stake size and the Sharpe ratios achieved by traders throughout their trading history.

Testing the relationship between behavioral characteristics of individuals and the disposition effect they display

The results of examining the relationship between traders' behavioral characteristics and disposition effect (H3) are presented in Table 6. The results show that the depreciation rate of previous attraction (ϕ) and the depreciation rate of previous experience (ρ) are, respectively, negatively and positively related with the degree of disposition effect displayed by a trader (-0.0883, $p < 0.001$; 0.1474, $p < 0.05$).

The results demonstrate that the degree to which individuals weight their experience and payoffs in the past are influential on the degree of disposition effect they display, thus supporting H3. Although there is evidence from the literature suggesting that traders can learn from their experience to reduce their disposition bias, we find that different elements of experience, i.e., the depreciation rate of previous attraction and that of previous experience, can have a different impact on the disposition effect.

Table 6: Regression Result – Estimated Parameters of EWA Models on disposition effect

	Stake Size Scenario	
	Estimated Coefficient	Std. Error
(Intercept)	-0.0893	0.0581
A	-0.0002	0.0003
N	-0.00000007	0.0000002

ϕ	-0.0883	***	0.0244
ρ	0.1474	*	0.0601
δ	-0.0214		0.0176
λ	0.0007		0.0196
k	0.0907		0.0519
Age	0.0004		0.0003
$Male$	0.0057		0.0131
TN	0.00005		0.0005
TM	0.000006		0.000006
R^2	0.0288		
Observations	1005		

The regression model is:

$$DE = \alpha + \beta_A A + \beta_N N + \beta_T T + \beta_\phi \phi + \beta_\rho \rho + \beta_\delta \delta + \beta_\lambda \lambda + \beta_k k$$
$$+\beta_{Age} Age + \beta_{Male} Male + \beta_{TN} TN + \beta_{TM} TM + \varepsilon,$$

where DE is the disposition effect of the trader, A is the initial attraction of the first strategy, N is the initial experience, ϕ is the depreciation rate of previous attraction, ρ is the depreciation rate of previous experience, δ is the weight of foregone payoffs, λ is the sensitivity of players to attractions, k is the extent to which attractions average or cumulate, Age is the trader's age, $Male$ is 1 if the trader is male and 0 otherwise, TN is the trader's total number of trades, and TM is the trader's total amount of stake size. ***, ** and * denote significance at 0.1%, 1% and 5% levels, respectively.

4.5 Discussion

The DSM framework enables the comparison across models and scenarios based on real trading data. Our results show that, across all the scenarios we examined (i.e., stake size, buy or sell, holding time, time in profit and time in loss), the trading data supports the view that belief learning models better represent the manner in which the past experiences of traders influence their future behavior and performance than reinforcement learning models. However, the EWA model better represents traders' behavior than both these learning models. The superior performance of the EWA model is not surprising since the EWA model integrates the features of both belief and reinforcement learning. The finding that the belief learning model better represents the behavior of traders than the reinforcement learning model contrasts with the majority of the results of psychological experiments. This suggests that individuals may adopt different approaches in different circumstances. In particular, it may suggest that behavior in the laboratory and the real world may differ or that the behavior of different types of individuals in different decision making domains may differ. Our results clearly show, that in real markets with real money involved, individuals act in a more rational manner (i.e., in using opportunity losses as a basis of learning from previous experiences). This demonstrates that traders are more subject to belief learning than the more animal instinctive, reinforcement learning.

The parameter δ in the EWA model, which is the weight of foregone payoffs, represents the difference between belief and reinforcement learning. That is, the greater weight attached to foregone payoffs (opportunity cost) provides a measure of greater rational choice on the part of traders. Our results from the stake size scenario show that traders who have higher δ, and who therefore are considered as more rational, do not produce better trading performance (measured by the total profit and Sharpe ratio they achieve over their trading history). A reasonable explanation is that the decisions concerning stake size are not directly related to the total profit they are likely to achieve. Clearly higher stake sizes can result in higher profits on individual trades, but equally they can result in larger losses. In other words, a trader's decisions concerning stake sizes are highly related with trading risk that could harm performance if not managed carefully. In sum, our results question the assumption that rational traders achieve better trading performance.

We find that individuals who trade with higher average stake sizes and who trade with greater frequency tend to put weight on foregone payoffs for unchosen strategies. A rational economic agent is generally considered to be one who considers all payoffs, including foregone ones, in making decisions. Consequently, the weight of foregone payoffs is regarded as representing the degree to which a trader makes rational decisions. Our results show that those traders with a higher degree of rationality trade more actively than other traders. Furthermore, traders with better trading performance, i.e., higher total profit and lower volatility of returns, tend to have a higher initial attraction to the lower stake size and lower depreciation rate of previous experience. That is to say, the traders who are more cautious, more likely to place small stake sizes in their early trading period and value their experience more (discount the previous experience slower) achieve higher returns and lower volatility of returns.

The implication of our results on the individual investor behavior is that the impact of individual behavior on market efficiency is complex due to the heterogeneity among individuals. The existence of market anomaly indicates that the level of the market efficiency is not high. One of the well-documented market anomalies is return persistence with which excess returns can be realized by buying past winners and selling past losers. Return persistence is considered as the result of under-reaction which is irrational behavior. Jegadeesh and Titman (1993) find that traders exploiting the return persistence can realize excess returns in the short term. However, the returns turn negative in the long run. They claim that those traders believing in return persistence move prices away from their long-run values temporarily and thereby cause prices to overreact and hence a lower level of market efficiency. Their prediction is consistent with our results that rational traders do not achieve superior trading performance.

4.6 Conclusion

We use Camerer and Ho's (1999) approach to estimate the parameters of EWA model. The major difference is the way in which we define the strategies. Camerer and Ho (1999) use n-person formal-form games in which an individual faces a finite set of strategies in order to develop data to estimate the models, i.e., the strategy space is limited. We are concerned to develop a procedure to estimate the behavioral characteristics with the EWA model and to examine the influence of the behavioral characteristics on trading behavior and trading performance in real financial markets. In a real market, there are an infinite variety and combination of factors to be considered when making trades (e.g., the level of stakes to invest, the holding time for the trade, the profit time, etc.). It would be impossible to estimate a model which included the full variety of these and all their combinations, i.e., the strategy space is unlimited. Consequently, we employ the DSM framework in order to make the problem tractable and to throw some light on the behavioral factors influencing traders' decisions. Clearly, the restrictive assumptions we make will prevent the model capturing the full complexity of the trading decision, but we expect that this approach will provide a step forward in understanding the behavioral factors that influence real-world trading decisions.

Our results show that traders believe that their decisions on stake size are influential in terms of the percentage profit they achieve. We find that traders who behave like rational economic agents (i.e., put greater weight on foregone payoffs for unchosen strategies) tend to place higher stakes on average and trade with higher frequency. However, those traders who are more rational do not achieve superior trading performance. Rather, the traders who achieve superior trading performance (higher total profit and lower volatility of returns) tend to place small stake sizes at the start of their trading history and discount their previous experience slower than others.

An empirical direction of future research is to estimate further characteristics under the DSM framework. The behavioral characteristics which we estimate in this study are the parameters of the EWA model. Hence, it is plausible to employ other behavioral models to estimate other characteristics depending on the research topics.

Chapter 5: Conclusion

This chapter summarizes the major findings, implications, and contributions of the three papers of the thesis and the overall research objective. The aim of this thesis is to investigate, via empirical evidence, the underlying rationality assumption and efficiency prediction of the EMH by examining the nature of learning by traders and the extent to which machine learning techniques may be used to predict markets. I also examine the relationship between the rationality assumption and the extent to which individuals learn from their experience.

In the first paper, I examine the potential origins of the apparent contradiction between the ML and EMH literature regarding market efficiency: many ML-based financial time series forecasting studies seem to find ways to anticipate market developments with surprisingly high accuracy, while the EMH predicts that excess returns cannot be earned in a systematic way. I perform an extensive forecasting benchmark in which two established ML methods are used to predict price movements in most major stock markets. This is the first study, to my best knowledge, to compare intraday and daily machine learning and econometric prediction models across most major stock markets.

My results show that methodological factors, such as the maturity of a financial market, the prediction method, the horizon for which it generates forecasts, and the methodology to simulate model-based trading, have a significant effect on market predictability and the feasibility of model-based trading. Consequently, decisions that forecasting studies have taken with respect to these factors can help explain the results observed. This is not true for our last experimental factor, covariate composition since we find that popular technical indicators are no more predictive than basic reference prices. The implication of the results to the machine learning literature is that the performance measurement should be performed under various settings in order to avoid the bias in estimating model prediction performance.

On the other hand, I do not find overwhelming evidence which contradicts the prediction of EMH: stock prices are partly predictable, but excess returns cannot be obtained in a systematic way. Although most of the accuracies in my simulations are below 60 percent which is substantially lower than that commonly reported in the ML literature, the positive relation between accuracy and trading profit, which is usually assumed to be true in the ML literature, is not observed. That is, even if the prediction accuracy is increased with more sophisticated covariates and predictive models adopted, the trading profit is not necessarily increased. It is important to notice that transaction cost is not included in my simulation experiments, and the trading profit will decrease further if transaction cost is applied to each trade. The implication is important to the ML literature as many studies only focus on the prediction accuracy. However, the increase of prediction accuracy cannot guarantee a higher trading profit. The ML studies in the future should take more factors, such as transaction cost and trading strategies, into account.

The major contribution of the first paper is to provide evidence for the conjecture that the methodological issues might be the source of the difference between the prevailing view in support of the EMH in the financial economics literature and the high prediction accuracy of stock prices reported in ML studies. Certainly, my results suggest that the methodological factors need to be decided carefully to avoid over-/under-estimation of prediction performance in the ML studies in the future.

The implications of my results for the financial economics literature is the empirical evidence for the prediction ability of ML techniques, such as SVM and ANN, to recognize the patterns of market anomalies across major financial markets. ML techniques are not commonly used in the economic and financial field. My results show that ML techniques, even in the basic forms, can capture complicated patterns in a stock market data set. I also find that ML techniques outperform econometric methods in predicting stock prices. Therefore, my results suggest that ML techniques can provide new insights for financial economists studying informational efficiency.

In the second paper, I investigate the relation between experience, risk taking behavior and trading performance. Specifically, I examine the degree to which individuals change risk taking behavior and learn to improve their trading performance through experience by analyzing decisions of 27,868 individual UK spread-traders over a 10 year period. As other empirical studies focus on the effect of experience on returns and behavioral biases, this study discusses the change of risk taking behavior and the risk-adjusted performance when traders gain experience. The results show that traders increase the level of risk taking behavior while gaining more experience. I also find that traders *do not* improve risk-adjusted performance when they have more experience. Instead, whilst experienced traders make higher profits, they suffer higher volatility of returns and decreases in risk-adjusted performance.

The major contribution of the second paper is to examine the effects of learning through experience on traders in real world financial markets, taking account of their risk taking behavior, volatility of returns and risk-adjusted performance. Risk preference of individuals is reflected by the risk taking behavior, and the literature shows that the change of risk taking behavior is motivated by the returns in the past and is affected by total wealth. My results show that risk taking behavior is also affected by accumulated experience, and the reason may be accumulated knowledge, such as the trading rule of markets and the common practices to manage potential loss. Furthermore, the literature generally agrees that individuals learn to improve profit when they have more experience, while my results tell a different story. I find that traders increase returns by taking higher risk and this leads to lower risk adjusted performance. This indicates that traders are likely to be reinforced by returns, rather than improving their investment skills when they have more experience. Also, it is often assumed that the level of risk taking behavior is positively related with the trading risk which is usually measured as the volatility of returns. I find that this assumption does not always hold. My results show that longer holding time leads to higher volatility of returns, but higher investment size and higher trading frequency does not necessarily increase the volatility of returns. This indicates the complexity nature of risk taking behavior, and the conclusion for one type risk taking behavior should not be applied to another without empirical evidence.

The implication of my results for the financial economics literature is to question one of the rational assumptions underlying the neoclassical approach. It is claimed that individuals learn by following the Bayesian rule, i.e., they appropriately update prior probability estimations conditional on new information or events. Instead, my results suggest that traders learn by reinforcement, choosing strategies which led to 'better' outcomes in the past. It appears that in adopting this approach, traders may over-weight the value of returns in assessing what the 'better' outcomes are and this leads them to under-assess the underlying risk. It is important to notice that risk is not as easy to understand as returns. My results suggest that individuals are likely to ignore the impact of taking the high risk until they suffer real loss. In particular, it is unlikely that individual traders will learn the lessons of excessive risk-taking from previous periods of excessive exuberance in financial markets which led to bubbles and eventual crashes. Consequently, my findings indicate the need for intervention and regulation by financial authorities to adopt measures which make clearer the risk involved in particular assets or investment strategies and/or to provide incentives and guidelines for traders to be aware of the potential trading risk.

In the third paper, I propose the decision-based strategy mapping framework (DSM) to solve the problem of unlimited strategy space, which allows me to examine the manner and extent to which important behavioral characteristics of individual traders associated with the experience-weighted attraction (EWA) model influence the individual's trading behavior and performance. The behavioral characteristics are estimated with the EWA model which is a behavioral model developed from lab experiments.

My results show that traders believe that their decisions on stake size are influential in terms of the percentage profit they achieve. I find that traders who behave like rational economic agents (i.e., put greater weight on foregone payoffs for unchosen strategies) tend to place higher stakes on average and trade with higher frequency. On the other hand, those traders who are more rational do not achieve superior trading performance. Instead, the traders who achieve superior trading performance (higher total profit and lower volatility of returns) place small stake sizes at the start of their trading history.

The major contribution of the third paper is the decision-based strategy mapping framework which is, to my best knowledge, the first attempt to estimate unobservable individual characteristics with trading data. The implications of my results for the financial economics literature is that the boundary between empirical and laboratory experiments is weakened. The models developed based on experimental results can be tested with empirical data under DSM framework with some additional assumptions.

In conclusion, I examine the challenges to the efficient market hypothesis with empirical market data. First, I run a comprehensive benchmark to examine the challenge from machine learning techniques to the prediction of an efficient market, and the results are not against the prediction of the efficient market hypothesis: excess returns cannot be earned in a systematic way. Second, I examine the influence of experience on trading performance and trading behavior. I find that traders are reinforced by high returns resulted from high risk behavior through experience, which strongly questioning the rationality assumption of the efficient market hypothesis. Third, I estimate the behavioral characteristics with the experience-weighted attraction model under the decision-based strategy mapping framework. The results show that the behavioral characteristics, such as the weight on foregone payoffs, are influential on trading performance and vary across individuals significantly. That is, the rational agent assumption of the efficient market hypothesis is likely to over-simplify individual behavior in the real world. Overall, the findings of this thesis agree with the efficient market hypothesis on that the markets are efficient, at least to the extent to which state of the art machine learning techniques cannot be utilized to make excess returns. However, the results from individual-level analysis suggest that rationality assumption of the efficient market hypothesis is unlikely to be true.

List of References

Ackert, L.F., Deaves, R., 2009. Behavioral finance: Psychology, decision-making, and markets. Cengage Learning, 2009.

Adrian, T., Rosenberg, J., 2008. Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk. The Journal of Finance 63, 2997–3030. doi:10.1111/j.1540-6261.2008.01419.x

Akaike, H., 1974. A new look at the statistical model identification. IEEE transactions on automatic control 19, 716–723.

Allen, F., Gale, D., 1994. Limited Market Participation and Volatility of Asset Prices. The American Economic Review 84, 933–955. doi:10.2307/2118039

Armano, G., Marchesi, M., Murru, A., 2005. A hybrid genetic-neural architecture for stock indexes forecasting. Information Sciences 170, 3–33. doi:10.1016/j.ins.2003.03.023

Atsalakis, G.S., Valavanis, K.P., 2009. Surveying stock market forecasting techniques – Part II: Soft computing methods. Expert Systems with Applications 36, 5932–5941. doi:10.1016/j.eswa.2008.07.006

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

Awartani, B.M.A., Corradi, V., 2005. Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. International Journal of Forecasting 21, 167–183. doi:10.1016/j.ijforecast.2004.08.003

Bajtelsmit, V.L., Bernasek, A., Jianakoplos, N.A., 1999. Gender differences in defined contribution pension decisions. Financial Services Review 8, 1–10.

Ballings, M., Van den Poel, D., Hespeels, N., Gryp, R., 2015. Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications 42, 7046–7056.

Barber, B.M., Odean, T., 2002. Online investors: do the slow die first? Review of Financial Studies 15, 455–488.

Barber, B.M., Odean, T., 2001. Boys will be boys: Gender, overconfidence, and common stock investment. Quarterly journal of Economics 261–292.

Barberis, N., Huang, M., Santos, T., 2001. Prospect theory and asset prices. The quarterly journal of economics 116, 1–53.

Beechey, M., Gruen, D.W., Vickery, J., 2000. The efficient market hypothesis: a survey. Reserve Bank of Australia, Economic Research Department.

Benartzi, S., Thaler, R.H., 1995. Myopic loss aversion and the equity premium puzzle. The quarterly journal of Economics 110, 73–92.

Bernasek, A., Shwiff, S., 2001. Gender, risk, and retirement. Journal of Economic Issues 345–356.

Berry, M.J., Linoff, G., 1997. Data mining techniques: for marketing, sales, and customer support. John Wiley & Sons, Inc., USA.

Bitvai, Z., Cohn, T., 2014. Day trading profit maximization with multi-task learning and technical analysis. Machine Learning 1–23.

Bodyanskiy, Y., Popov, S., 2006. Neural network approach to forecasting of quasiperiodic financial time series. European Journal of Operational Research 175, 1357–1366.

Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. Journal of Computational Science 2, 1–8.

Bondt, W.F., Thaler, R., 1985. Does the stock market overreact? The Journal of finance 40, 793–805.

Brady, C., Ramyar, R., 2006. White Paper on spread betting. London: Cass Business School.

Brock, W., Lakonishok, J., LeBaron, B., 1992. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. The Journal of Finance 47, 1731–1764. doi:10.1111/j.1540-6261.1992.tb04681.x

Brooks, C., 2014. Introductory econometrics for finance. Cambridge university press.

Bruhin, A., Fehr-Duda, H., Epper, T., 2010. Risk and rationality: Uncovering heterogeneity in probability distortion. Econometrica 78, 1375–1412.

Camerer, C., Ho, T.-H., 1999. Experience-weighted Attraction Learning in Normal Form Games. Econometrica 67, 827–874.

Camerer, C.F., Ho, T.-H., Chong, J.-K., 2002. Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. Journal of Economic Theory 104, 137–188.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Chang, P.-C., Liu, C.-H., Lin, J.-L., Fan, C.-Y., Ng, C.S., 2009. A neural network with a case based dynamic window for stock trading prediction. Expert Systems with Applications 36, 6889–6898.

Charles, A., Darné, O., Kim, J.H., 2011. Small sample properties of alternative tests for martingale difference hypothesis. Economics Letters 110, 151–154.

Charness, G., Levin, D., 2003. When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. Complexity, and Affect (November 17, 2004).

Chen, A.-S., Leung, M.T., Daouk, H., 2003. Application of neural networks to an emerging financial market: forecasting and trading the Taiwan stock index. Computers and Operations Research 30, 901–923. doi:10.1016/S0305-0548(02)00037-0

Chen, K.-Y., Ho, C.-H., 2005. An improved support vector regression modeling for Taiwan Stock Exchange market weighted index forecasting. Presented at the Neural Networks and Brain, 2005. ICNN&B'05. International Conference on, IEEE, pp. nil15-1638.

Chen, W.-H., Shih, J.-Y., Wu, S., 2006. Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets. International Journal of Electronic Finance 1, 49–67.

Chen, W.-S., Du, Y.-K., 2009. Using neural networks and data mining techniques for the financial distress prediction model. Expert Systems with Applications 36, 4075–4086.

Cherkassky, V., Ma, Y., 2004. Practical selection of SVM parameters and noise estimation for SVM regression. Neural Networks 17, 113–126. doi:10.1016/S0893-6080(03)00169-2

Cheung, Y.-W., Lai, K.S., 1995. A search for long memory in international stock market returns. Journal of International Money and Finance 14, 597–615.

Chiang, Y.-M., Hirshleifer, D., Qian, Y., Sherman, A.E., 2011. Do investors learn from experience? Evidence from frequent IPO investors. Review of Financial Studies 24, 1560–1589.

Choi, J.J., Laibson, D., Madrian, B.C., Metrick, A., 2009. Reinforcement learning and savings behavior. The Journal of finance 64, 2515–2534. doi:10.1111/j.1540-6261.2009.01509.x

Choong, C.-K., Baharumshah, A.Z., Yusop, Z., Habibullah, M.S., 2010a. Private capital flows, stock market and economic growth in developed and developing countries: A comparative analysis. Japan and the World Economy 22, 107–117.

Choong, C.-K., Baharumshah, A.Z., Yusop, Z., Habibullah, M.S., 2010b. Private capital flows, stock market and economic growth in developed and developing countries: A comparative analysis. Japan and the World Economy 22, 107–117.

Chordia, T., Goyal, A., Lehmann, B.N., Saar, G., 2013. High-frequency trading. Journal of Financial Markets 16, 637–645. doi:10.1016/j.finmar.2013.06.004

Chordia, T., Roll, R., Subrahmanyam, A., 2001. Market liquidity and trading activity. The Journal of Finance 56, 501–530.

Claessens, S., Klingebiel, D., Schmukler, S.L., 2006. Stock market development and internationalization: Do economic fundamentals spur both similarly? Journal of Empirical Finance 13, 316–350.

Cnaan, A., Laird, N., Slasor, P., 1997. Tutorial in biostatistics: using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. Statistics in Medicin 16, 2349–2380.

Cohen, J., 1969. Statistical power analysis for the behavioral sciences. Academic press, San Diego, CA.

Cristianini, N., Shawe-Taylor, J., 2000. An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.

Dai, W., Shao, Y.E., Lu, C.-J., 2012. Incorporating feature selection method into support vector regression for stock index forecasting. Neural Computing and Applications 23, 1551–1561.

de Oliveira, F.A., Nobre, C.N., Zárate, L.E., 2013. Applying Artificial Neural Networks to Prediction of Stock Price and Improvement of the Directional Prediction Index-Case Study of PETR4, Petrobras, Brazil. Expert Systems with Applications 40, 7596–7606.

Dhar, R., Zhu, N., 2006. Up close and personal: Investor sophistication and the disposition effect. Management Science 52, 726–740.

Doeksen, B., Abraham, A., Thomas, J., Paprzycki, M., 2005. Real stock trading using soft computing models. Presented at the Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on, IEEE, pp. 162–167.

Donaldson, R.G., Kamstra, M., 1999. Neural network forecast combining with interaction effects. Journal of the Franklin Institute 336, 227–236.

Fama, E.F., 1998. Market efficiency, long-term returns, and behavioral finance. Journal of financial economics 49, 283–306.

Fama, E.F., 1991. Efficient capital markets: II. The journal of finance 46, 1575–1617.

Fama, E.F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance 25, 383–417. doi:10.1111/j.1540-6261.1970.tb00518.x

Fama, E.F., French, K.R., 2012. Size, value, and momentum in international stock returns. Journal of Financial Economics 105, 457–472.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of financial economics 33, 3–56.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. The Journal of Political Economy 607–636.

Fehr-Duda, H., Bruhin, A., Epper, T., Schubert, R., 2010. Rationality on the rise: Why relative risk aversion increases with stake size. Journal of Risk and Uncertainty 40, 147–180.

Feng, L., Seasholes, M.S., 2005. Do investor sophistication and trading experience eliminate behavioral biases in financial markets? Review of Finance 9, 305–351.

Fraser-Mackenzie, P., Ma, T., Sung, M., Johnson, J.E.V., Dolia, A.N., 2013. When the going gets tough: insights into individuals' trading behaviour during the financial crisis. At International Conference on the Global Financial Crisis: European Financial Markets and Institutions, United Kingdom. 25 - 26 Apr 2013.

French, K.R., 1980. Stock returns and the weekend effect. Journal of financial economics 8, 55–69.

Gerlein, E.A., McGinnity, M., Belatreche, A., Coleman, S., 2016. Evaluating machine learning classification for financial trading: An empirical approach. Expert Systems with Applications 54, 193–207.

Gervais, S., Odean, T., 2001. Learning to be overconfident. Rev Financ Stud 14, 1–27. doi:10.1093/Rfs/14.1.1

Glaser, M., Weber, M., 2009. Which past returns affect trading volume? Journal of Financial Markets 12, 1–31.

Gloede, O., Menkhoff, L., 2011. Financial professionals' overconfidence: Is it experience, function, or attitude? European Financial Management.

Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. The Journal of Finance 48, 1779–1801. doi:10.1111/j.1540-6261.1993.tb05128.x

Goetzmann, W.N., Kumar, A., 2008. Equity portfolio diversification. Review of Financ 12, 433–463. doi:10.1093/Rof/Rfn005

Greene, M.T., Fielitz, B.D., 1977. Long-term dependence in common stock returns. Journal of Financial Economics 4, 339–349. doi:10.1016/0304-405X(77)90006-X

Greenwood, R., Nagel, S., 2009. Inexperienced investors and bubbles. Journal of Financial Economics 93, 239–258.

Griffin, D., Tversky, A., 1992. The weighing of evidence and the determinants of confidence. Cognitive Psychology 24, 411–435. doi:10.1016/0010-0285(92)90013-R

Griffin, J.M., Kelly, P.J., Nardari, F., 2010. Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. Review of Financial Studies 23, 3225–3277.

Grinblatt, M., Keloharju, M., 2001a. What makes investors trade? The Journal of Finance 56, 589–616.

Grinblatt, M., Keloharju, M., 2001b. What makes investors trade? The Journal of Finance 56, 589–616.

Grossman, S., 1976. On the efficiency of competitive stock markets where trader have diverse information. The Journal of Finance 31, 573–585. doi:10.1111/j.1540-6261.1976.tb01907.x

Grossman, S.J., Stiglitz, J.E., 1980. On the Impossibility of Informationally Efficient Markets. The American Economic Review 70, 393–408. doi:10.2307/1805228

Gulthawatvichai, S., Sung, M., Johnson, J.E.V., Cheah, E.-T., 2013. Insights into herding behavior in financial spread trading markets. Presented at the 11th INFINITI Conference on International Finance.

Hadavandi, E., Shavandi, H., Ghanbari, A., 2010. Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. Knowledge-Based Systems 23, 800–808.

Hand, D.J., 2006. Classifier technology and the illusion of progress. Statistical science 21, 1–14.

Hartog, J., Ferrer-i-Carbonell, A., Jonker, N., 2002. Linking measured risk aversion to individual characteristics. Kyklos 55, 3–26.

Hassan, M.R., Nath, B., Kirley, M., 2007. A fusion model of HMM, ANN and GA for stock market forecasting. Expert System with Applications 33, 171–180. doi:10.1016/j.eswa.2006.04.007

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning. Springer, New York.

Heckman, J.J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, in: Annals of Economic and Social Measurement, Volume 5, Number 4. NBER, pp. 475–492.

Huang, C.-J., Yang, D.-X., Chuang, Y.-T., 2008. Application of wrapper approach and composite classifier to the stock trend prediction. Expert Systems with Applications 34, 2870–2878.

Huang, G.-B., Chen, L., Siew, C.-K., 2006. Universal approximation using incremental constructive feedforward networks with random hidden nodes. Neural Networks, IEEE Transactions on 17, 879–892.

Huang, W., Nakamori, Y., Wang, S.-Y., 2005. Forecasting stock market movement direction with support vector machine. Comput Oper Res 32, 2513–2522. doi:10.1016/j.cor.2004.03.016

Huck, N., 2010. Pairs trading and outranking: The multi-step-ahead forecasting case. European Journal of Operational Research 207, 1702–1716. doi:10.1016/j.ejor.2010.06.043

Hyndman, R.J., Athanasopoulos, G., 2014. Forecasting: principles and practice. OTexts, Australia.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of Finance 48, 65–91.

Jensen, M.C., 1978. Some anomalous evidence regarding market efficiency. Journal of financial economics 6, 95–101.

Jianakoplos, N.A., Bernasek, A., 1998. Are women more risk averse? Economic inquiry 36, 620.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica: Journal of the Econometric Society 263–291.

Kara, Y., Acar Boyacioglu, M., Baykan, Ö.K., 2011. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert systems with Applications 38, 5311–5319.

Kaustia, M., Knüpfer, S., 2008. Do investors overweight personal experience? Evidence from IPO subscriptions. The Journal of Finance 63, 2679–2702.

Keerthi, S.S., Lin, C.-J., 2003. Asymptotic behaviors of support vector machines with Gaussian kernel. Neural computation 15, 1667–1689.

Keim, D.B., 1983. Size-related anomalies and stock return seasonality: Further empirical evidence. Journal of Financial Economics 12, 13–32.

Kim, J.H., Shamsuddin, A., 2008. Are Asian stock markets efficient? Evidence from new multiple variance ratio tests. Journal of Empirical Finance 15, 518–532.

Kim, K., 2003. Financial time series forecasting using support vector machines. Neurocomputing 55, 307–319.

Kuo, R.J., Chen, C., Hwang, Y., 2001. An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. Fuzzy Set Syst 118, 21–45. doi:10.1016/S0165-0114(98)00399-6

Lee, M.-C., 2009. Using support vector machine with a hybrid feature selection method to the stock trend prediction. Expert Systems with Applications 36, 10896–10904. doi:10.1016/j.eswa.2009.02.038

Lesmond, D.A., Schill, M.J., Zhou, C., 2004. The illusory nature of momentum profits. Journal of financial economics 71, 349–380.

Lessmann, S., Baesens, B., Seow, H.-V., Thomas, L.C., 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research 247, 124–136. doi:10.1016/j.ejor.2015.05.030

Lessmann, S., Sung, M., Johnson, J.E., 2011. Towards a methodology for measuring the true degree of efficiency in a speculative market. Journal of the Operational Research Society 62, 2120–2132.

Linnainmaa, J.T., 2011. Why do (some) households trade so much? Review of Financial Studies 24, 1630–1666.

List, J.A., 2004. Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace. Econometrica 72, 615–625. doi:10.1111/j.1468-0262.2004.00502.x

Liu, Y., Wang, M., Zhao, L., 2010. Narrow framing: Professions, sophistication, and experience. Journal of Futures Markets 30, 203–229.

Liu, Y.-J., Tsai, C.-L., Wang, M.-C., Zhu, N., 2010. Prior Consequences and Subsequent Risk Taking: New Field Evidence from the Taiwan Futures Exchange. Management Science 56, 606–620. doi:10.1287/mnsc.1090.1131

Lo, A.W., 1991. Long-Term Memory in Stock Market Prices. Econometrica 59, 1279–1313.

Lo, A.W., Mamaysky, H., Wang, J., 2000. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. The Journal of Finance 55, 1705–1770.

Malkiel, B.G., 2003. The Efficient Market Hypothesis and Its Critics. The Journal of Economic Perspectives 17, 59–82. doi:10.2307/3216840

Malmendier, U., Nagel, S., 2009. Learning from inflation experiences. Unpublished manuscript, UC Berkley.

Mandelbrot, B.B., 1971. When Can Price be Arbitraged Efficiently? A Limit to the Validity of the Random Walk and Martingale Models. The Review of Economics and Statistics 53, 225–236. doi:10.2307/1937966

Markowitz, H., 1952. Portfolio selection*. The journal of finance 7, 77–91.

Mehra, R., Prescott, E.C., 1985. The equity premium: A puzzle. Journal of monetary Economics 15, 145–161.

Mehra, R., Sah, R., 2002. Mood fluctuations, projection bias, and volatility of equity prices. Journal of Economic Dynamics and Control 26, 869–887. doi:10.1016/S0165-1889(01)00035-5

Menkveld, A.J., 2013. High frequency trading and the new market makers. Journal of Financial Markets 16, 712–740. doi:10.1016/j.finmar.2013.06.006

Mittermayer, M.-A., 2004. Forecasting Intraday stock price trends with text mining techniques. System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on 10 pp. doi:10.1109/HICSS.2004.1265201

Nicolosi, G., Peng, L., Zhu, N., 2009. Do individual investors learn from their trading experience? Journal of Financial Markets 12, 317–336. doi:10.1016/j.finmar.2008.07.001

Odean, T., 1998a. Do investors trade too much? Available at SSRN 94143.

Odean, T., 1998b. Are investors reluctant to realize their losses? The Journal of finance 53, 1775–1798.

Ojah, K., Karemera, D., 1999. Random walks and market efficiency tests of Latin American emerging equity markets: a revisit. The Financial Review 34, 57.

Ormoneit, D., Neuneier, R., 1996. Experiments in predicting the German stock index DAX with density estimating neural networks. Computational Intelligence for Financial Engineering, 1996., Proceedings of the IEEE/IAFE 1996 Conference on 66–71. doi:10.1109/CIFER.1996.501825

Ou, P., Wang, H., 2009. Prediction of stock market index movement by ten data mining techniques. Modern Applied Science 3, 28.

Pai, P.-F., Lin, C.-S., 2005. A hybrid ARIMA and support vector machines model in stock price forecasting. Omega 33, 497–505.

Pålsson, A.-M., 1996. Does the degree of relative risk aversion vary with household characteristics? Journal of Economic Psychology 17, 771–787. doi:10.1016/S0167-4870(96)00039-6

Pan, H., Tilakaratne, C., Yearwood, J., 2005. Predicting Australian stock market index using neural networks exploiting dynamical swings and intermarket influences. J Res Pract Inf Tech 37, 43–56.

Patel, J., Shah, S., Thakkar, P., Kotecha, K., 2015. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications 42, 259–268. doi:10.1016/j.eswa.2014.07.040

Paton, D., Williams, L.V., 2005. Forecasting outcomes in spread betting markets: can bettors use ‘quarbs’ to beat the book? Journal of Forecasting 24, 139–154.

Payzan-LeNestour, E., Bossaerts, P., 2014. Learning About Unstable, Publicly Unobservable Payoffs. Review of Financial Studies.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., 2007. Linear and nonlinear mixed effects models. R package version 3, 57.

Qian, B., Rasheed, K., 2007. Stock market prediction with multiple classifiers. Appl Intell 26, 25–33. doi:10.1007/s10489-006-0001-7

Roth, A.E., Erev, I., 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. Game Econ Behav 8, 164–212. doi:10.1016/S0899-8256(05)80020-X

Samuelson, P.A., 1965. Proof that properly anticipated prices fluctuate randomly. Industrial management review 6, 41–49.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Networks 61, 85–117.

Schumaker, R.P., Chen, H., 2009. Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Transactions on Information Systems (TOIS) 27, 12.

Seru, A., Shumway, T., Stoffman, N., 2010. Learning by trading. The Review of Financial Studies 23, 705–739. doi:10.1093/Rfs/Hhp060

Sharpe, W.F., 1998. The sharpe ratio. Streetwise-The best of the Journal of Portfolio Management, Princeton University Press Princeton, New Jersy 169–185.

Sharpe, W.F., 1966. Mutual Fund Performance. The Journal of Business 39, 119. doi:10.1086/294846

Sharpe, W.F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance 19, 425–442. doi:10.1111/j.1540-6261.1964.tb02865.x

Shefrin, H., Statman, M., 1985. The disposition to sell winners too early and ride losers too long: Theory and evidence. The Journal of finance 40, 777–790.

Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. The Journal of Finance 52, 35–55.

Skinner, B.F., 2014. Contingencies of reinforcement: A theoretical analysis. BF Skinner Foundation.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and computing 14, 199–222.

Steiner, M., Wittkemper, H.-G., 1997. Portfolio optimization with a neural network implementation of the coherent market hypothesis. European journal of operational research 100, 27–40.

Tay, F.E., Cao, L., 2001. Application of support vector machines in financial time series forecasting. Omega 29, 309–317.

Thaler, R.H., 2010. The end of behavioral finance.

Thaler, R.H., Johnson, E.J., 1990. Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. Management Science 36, 643–660.

Toomet, O., Henningsen, A., 2008. Sample Selection Models in R: PackagesampleSelection. Journal of Statistical Software 27, 1–23.

Tsai, C.-F., Hsiao, Y.-C., 2010. Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. Decision Support Systems 50, 258–269.

Tseng, C.-H., Cheng, S.-T., Wang, Y.-H., Peng, J.-T., 2008. Artificial neural network model of the hybrid EGARCH volatility of the Taiwan stock index option prices. Physica A: Statistical Mechanics and its Applications 387, 3192–3200. doi:10.1016/j.physa.2008.01.074

Vanstone, B., Finnie, G., Tan, C., 2005. Evaluating the application of neural networks and fundamental analysis in the Australian stockmarket, in: Information Technology Papers. Presented at the IASTED International Conference on Computational Intelligence (CI 2005), Calgary, AB, Canada.

Wilde, G.J.S., 1982. The Theory of Risk Homeostasis: Implications for Safety and Health. Risk Analysis 2, 209–225. doi:10.1111/j.1539-6924.1982.tb01384.x

Xu, S., Chen, L., 2008. A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining. Presented at the 5th International Conference on Information Technology and Applications (ICITA 2008), Cairns, Queensland, Australia.

Żbikowski, K., 2015. Using Volume Weighted Support Vector Machines with walk forward testing and feature selection for the purpose of creating stock trading strategy. Expert Systems with Applications 42, 1797–1805. doi:10.1016/j.eswa.2014.10.001

Zhang, D., Jiang, Q., Li, X., 2004. Application of neural networks in financial data mining. International Journal of Computational Intelligence 1, 116–119.

Zhang, Y., Wu, L., 2009. Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. Expert systems with applications 36, 8849–8854.

Zunino, L., Zanin, M., Tabak, B.M., Pérez, D.G., Rosso, O.A., 2009. Forbidden patterns, permutation entropy and stock market inefficiency. Physica A: Statistical Mechanics and its Applications 388, 2854–2864.