Improving risk-adjusted performance in high-frequency trading: the role of fuzzy logic systems

Vincent Vella
Centre for Computational Finance and Economic Agents (CCFEA)
University of Essex

This dissertation is submitted for the degree of

Doctor of Philosophy

CCFEA
June 2016
I would like to dedicate this thesis to my loving parents . . .
Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Vincent Vella

June 2016
Related Publications


Acknowledgements

My most heartfelt thanks must go to my supervisor, Dr. Wing Lon Ng. His deep insight, admirable dedication, disciplined and structured approach, and keen eye for detail were not just key attributes that assisted me in the completion of this thesis, but have also inspired me with invaluable life lessons. For this I am forever in his debt.

I would also like to thank Prof. Edward Tsang for his assistance and direction, especially towards the end of my journey. I will remain grateful for his guidance.

My appreciation also goes towards Prof. Hani Hagras for his valuable feedback and helpful discussions.

Last but not least, this thesis would not have been possible without the person who was always behind me - my wife Mary Ann. Her unconditional support was always a source of encouragement that kept me going both in the highs and lows that go with research.
Abstract

In recent years, high-frequency and algorithmic trading have been the subject of increasing risk concerns. A general theme that we adopt in this thesis is that trading practitioners are predominantly interested in risk-adjusted performance. Likewise, regulators are demanding stricter risk controls.

First, we scrutinise different AI model design approaches with the aim to increase the time-varying risk-adjusted performance of trading systems. We show that accounting for transaction costs and using risk-return objective functions improve out-of-sample results. Our experiments identify that neuro-fuzzy logic models exhibit superior performance stability across multiple risk regimes when compared to other common models in AI literature. Moreover, we propose an innovative ensemble model approach which combines multiple risk-adjusted objective functions and dynamically adapts risk-tolerance according to time-varying intraday risk.

Next, we extend our initial findings to the money management aspects of trading algorithms. We propose an innovative fuzzy logic approach which dynamically discriminates across different regions in the trend and volatility space. The model prioritises higher performing regions at an intraday level and adapts capital allocation policies with the objective to maximise global risk-adjusted performance without penalising profitability, rather increase it.

Finally, we explore viable trading improvements that can be attained by advancing our type-1 fuzzy logic ideas to higher order fuzzy systems in view of the increased noise
(uncertainty) that is inherent in high-frequency data. We propose an innovative approach to
design type-2 models with minimal increase in design and computational complexity. As a
further step, we identify a relationship between the increased trading performance benefits of
the proposed type-2 model and higher levels of microstructure noise.

In conclusion, this thesis addresses a desirable need for practitioners, researchers and
regulators in the design of expert and intelligent systems for better management of risk in the
field of high-frequency and algorithmic trading.
# Table of Contents

List of Figures  
List of Tables  

## 1 Introduction  
1.1 Background and motivation  
1.1.1 Trading in financial markets  
1.1.2 The architecture of a trading algorithm  
1.1.3 Aligning algorithmic and investors’ risk preferences  
1.2 Research objectives and thesis structure  
1.3 Contributions  

## 2 Fuzzy logic background  
2.1 Introduction  
2.2 Type-1 fuzzy sets and fuzzy membership functions  
2.3 Type-1 TSK fuzzy logic systems  
2.4 Designing fuzzy logic systems  
2.4.1 Structure identification  
2.4.2 Parameter tuning using least squares and back-propagation  
2.5 Type-2 fuzzy sets and fuzzy membership functions  
2.6 Extending type-1 TSK fuzzy logic systems to type-2
# 3 Improving the risk-adjusted performance of trading algorithms

## 3.1 Introduction

## 3.2 Method

- **3.2.1** Technical trading and moving averages
- **3.2.2** Neural network model
- **3.2.3** ANFIS model
- **3.2.4** DENFIS model
- **3.2.5** ANFIS ensemble
- **3.2.6** Trading algorithm
- **3.2.7** Evaluation

## 3.3 Empirical data and analysis

- **3.3.1** Results for experiment 1
- **3.3.2** Results for experiment 2

## 3.4 Conclusion

# 4 Time-varying intraday volatility and risk-adjusted performance

## 4.1 Introduction

## 4.2 Part 1 - Enhancing the intraday trading performance of neural networks using a dynamic volatility clustering fuzzy filter

- **4.2.1** Method
- **4.2.2** Technical indicators and standard neural network model
- **4.2.3** Standard trading algorithm
- **4.2.4** Intraday volatility and its challenges
- **4.2.5** Enhancing trading model with a dynamic volatility filter
- **4.2.6** Model selection and experiment approach
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.7 Analysis and results</td>
<td>108</td>
</tr>
<tr>
<td>4.2.8 Conclusion</td>
<td>112</td>
</tr>
<tr>
<td>4.3 Part 2 – A dynamic fuzzy money management approach for controlling</td>
<td></td>
</tr>
<tr>
<td>the intraday risk-adjusted performance of AI-trading algorithms</td>
<td>114</td>
</tr>
<tr>
<td>4.3.1 Method</td>
<td>116</td>
</tr>
<tr>
<td>4.3.2 Enhancing money management decisions by considering time-varying</td>
<td></td>
</tr>
<tr>
<td>intraday risk</td>
<td>119</td>
</tr>
<tr>
<td>4.3.3 Experiment approach and considerations</td>
<td>123</td>
</tr>
<tr>
<td>4.3.4 Results and evaluation</td>
<td>127</td>
</tr>
<tr>
<td>4.3.5 Conclusion</td>
<td>137</td>
</tr>
<tr>
<td>5 Market microstructure noise and risk-adjusted performance</td>
<td>141</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>142</td>
</tr>
<tr>
<td>5.2 Method</td>
<td>147</td>
</tr>
<tr>
<td>5.2.1 High-frequency data and technical indicators</td>
<td>148</td>
</tr>
<tr>
<td>5.2.2 Designing and tuning a TSK type-1 fuzzy model</td>
<td>151</td>
</tr>
<tr>
<td>5.2.3 Generalizing ANFIS model to T2 FLS</td>
<td>156</td>
</tr>
<tr>
<td>5.2.4 Trading algorithm</td>
<td>164</td>
</tr>
<tr>
<td>5.3 Experiment approach</td>
<td>165</td>
</tr>
<tr>
<td>5.3.1 Data</td>
<td>165</td>
</tr>
<tr>
<td>5.3.2 Performance measures</td>
<td>167</td>
</tr>
<tr>
<td>5.3.3 IT2 design considerations</td>
<td>169</td>
</tr>
<tr>
<td>5.3.4 Model training and testing</td>
<td>170</td>
</tr>
<tr>
<td>5.3.5 Controlling different levels of noise</td>
<td>171</td>
</tr>
<tr>
<td>5.4 Results and analysis</td>
<td>172</td>
</tr>
<tr>
<td>5.4.1 Experiment 1: Comparison against benchmark models</td>
<td>173</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Experiment 2: Comparison of T1 and T2 models under different noise levels</td>
</tr>
<tr>
<td>5.5</td>
<td>Conclusion</td>
</tr>
<tr>
<td>6</td>
<td>Conclusion</td>
</tr>
<tr>
<td>6.1</td>
<td>Summary of presented work</td>
</tr>
<tr>
<td>6.2</td>
<td>Contributions</td>
</tr>
<tr>
<td>6.3</td>
<td>Limitations and future work</td>
</tr>
<tr>
<td>References</td>
<td></td>
</tr>
<tr>
<td>Appendix A</td>
<td>Further results to Chapter 3</td>
</tr>
</tbody>
</table>
List of Figures

1.1 General components of a trading algorithm ...................................... 7
2.1 Components of a fuzzy rule-based system ...................................... 24
2.2 T1 Gaussian MF ................................................................. 26
2.3 T1 TSK fuzzy model example .................................................. 28
2.4 Projection of identified clusters on the input space for MF identification ...... 30
2.5 T2 primary MF ................................................................. 35
2.6 General T2 secondary MF ..................................................... 37
2.7 Interval T2 secondary MF ..................................................... 38
2.8 T2 fuzzy rule-based system ..................................................... 39
3.1 Experiment setup ................................................................. 49
3.2 Our proposed eANFIS architecture ............................................. 61
3.3 Moving window approach adopted for model training and testing .......... 64
3.4 Heat map identifying the sensitivity of the ANFIS model ..................... 70
3.5 Trading performance by optimising Sharpe ratio ................................ 74
3.6 Plots comparing the results obtained from ANFIS Sharpe optimisation, eAN- FIS, no-transaction-costs optimisation, RMSE optimisation and period return optimisation. ..................................................... 77
4.1 Experiment setup showing our proposed locally enhanced extension ........ 90
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2</td>
<td>NN Setup</td>
<td>96</td>
</tr>
<tr>
<td>4.3</td>
<td>Effect of applying return threshold on the number of trades</td>
<td>98</td>
</tr>
<tr>
<td>4.4</td>
<td>Decision surface example as identified by our volatility clustering fuzzy filter</td>
<td>105</td>
</tr>
<tr>
<td>4.5</td>
<td>Various trends of the normalised stock prices used in the experiments</td>
<td>107</td>
</tr>
<tr>
<td>4.6</td>
<td>Effect of applying a return threshold on the number of trades</td>
<td>118</td>
</tr>
<tr>
<td>4.7</td>
<td>Example of clusters found in Hammerson (HMSO) data as identified by FCM</td>
<td>121</td>
</tr>
<tr>
<td>4.8</td>
<td>Decision surface example for Hammerson (HMSO) as identified by our fuzzy controller showing different degrees of Sharpe ratio regions</td>
<td>122</td>
</tr>
<tr>
<td>4.9</td>
<td>Time series of the stock prices (normalised, for better comparison)</td>
<td>125</td>
</tr>
<tr>
<td>4.10</td>
<td>Performance of the AI models over the 100-day out-of-sample period for Rexam (REX)</td>
<td>134</td>
</tr>
<tr>
<td>4.11</td>
<td>Performance of AI models over the 100-day out-of-sample period for Sage Group (SGE)</td>
<td>134</td>
</tr>
<tr>
<td>4.12</td>
<td>Performance of AI models over the 100 day out-of-sample period for ICAP (IAP)</td>
<td>135</td>
</tr>
<tr>
<td>4.13</td>
<td>Performance of AI models over the 100-day out-of-sample period for ITV (ITV)</td>
<td>135</td>
</tr>
<tr>
<td>4.14</td>
<td>Performance of AI models over the 100-day out-of-sample period for Hammerson (HMSO)</td>
<td>135</td>
</tr>
<tr>
<td>4.15</td>
<td>Performance of AI models over the 100-day out-of-sample period for Associated British Foods (ABS)</td>
<td>136</td>
</tr>
<tr>
<td>4.16</td>
<td>Performance of AI models over the 100-day out-of-sample period for TESCO (TSCO)</td>
<td>136</td>
</tr>
<tr>
<td>5.1</td>
<td>Experiment setup showing our proposed T2 extensions</td>
<td>148</td>
</tr>
<tr>
<td>5.2</td>
<td>Conversion of T1 fuzzy set to IT2 with fixed mean and uncertain standard deviation</td>
<td>161</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>5.3</td>
<td>An example showing the conversion of T1 MFs for one input variable to IT2</td>
<td>162</td>
</tr>
<tr>
<td>5.4</td>
<td>Times series of stock prices showing the mix of trends followed over the in-sample and out-of-sample period</td>
<td>168</td>
</tr>
<tr>
<td>5.5</td>
<td>Trends on various measures after 100-day out-of-sample period and at different degrees of return threshold (uncertainty)</td>
<td>182</td>
</tr>
<tr>
<td>A.1</td>
<td>Trading performance by optimising the Sortino ratio</td>
<td>210</td>
</tr>
<tr>
<td>A.2</td>
<td>Plots comparing the results obtained from ANFIS Sharpe optimisation and eANFIS with those from MA optimisation.</td>
<td>211</td>
</tr>
</tbody>
</table>
**List of Tables**

3.1 Parameters tested for FFN ........................................ 54
3.2 Parameters tested for ANFIS .................................... 56
3.3 Parameters tested for DENFIS .................................. 59
3.4 Models applied in the experiments ............................. 66
3.5 Descriptive statistics of 5-minute returns .................. 69
3.6 Model performance using Sharpe ratio optimisation over the 100-day out-of-sample period ........................................... 72
3.7 Model performance using Sortino ratio optimisation over the 100-day out-of-sample period ........................................... 75
3.8 eANFIS performance using intraday dynamic switching over the 100-day out-of-sample period ........................................... 78
3.9 Model performance using Sharpe ratio optimisation with no transaction costs over the 100-day out-of-sample period .................. 79
3.10 Model performance using RMSE optimisation over the 100-day out-of-sample period ........................................... 80
3.11 Model performance using period return optimisation over the 100-day out-of-sample period ........................................... 80
3.12 Model performance using fixed moving average (MA) rules over the 100-day out-of-sample period ........................................... 81
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Parameters tested for NN model</td>
<td>96</td>
</tr>
<tr>
<td>4.2</td>
<td>Additional parameters tested for the NN and the fuzzy volatility filter</td>
<td>106</td>
</tr>
<tr>
<td>4.3</td>
<td>FTSE 100 companies analysed in Part 1 of this chapter</td>
<td>107</td>
</tr>
<tr>
<td>4.4</td>
<td>Trading performance following the 100-day out-of-sample period</td>
<td>109</td>
</tr>
<tr>
<td>4.5</td>
<td>Performance percentage difference of standard NN, NN-FCM1 and NN-FCM2 when compared with buy-and-hold method following the 100-day out-of-sample period</td>
<td>110</td>
</tr>
<tr>
<td>4.6</td>
<td>Performance percentage difference of NN-FCM1 and NN-FCM2 when compared with the standard NN model following the 100-day out-of-sample period</td>
<td>111</td>
</tr>
<tr>
<td>4.7</td>
<td>Combined parameters tested for the NN model and trading algorithm</td>
<td>118</td>
</tr>
<tr>
<td>4.8</td>
<td>Additional parameters tested for the NN and fuzzy volatility filter</td>
<td>123</td>
</tr>
<tr>
<td>4.9</td>
<td>Descriptive statistics of 5-minute returns for companies analysed in Part 2 of this chapter</td>
<td>124</td>
</tr>
<tr>
<td>4.10</td>
<td>Risk-adjusted performance of tested models following the 100-day out-of-sample period</td>
<td>129</td>
</tr>
<tr>
<td>4.11</td>
<td>Average risk-adjusted performance of tested models for 50 repeated trials over the 100-day out-of-sample period</td>
<td>130</td>
</tr>
<tr>
<td>4.12</td>
<td>Statistical tests on the Sharpe ratio of the AI models over the 100-day out-of-sample period</td>
<td>132</td>
</tr>
<tr>
<td>4.13</td>
<td>Statistical tests on profit (%) of the AI models over the 100-day out-of-sample period</td>
<td>133</td>
</tr>
<tr>
<td>5.1</td>
<td>Parameters tested for ANFIS</td>
<td>156</td>
</tr>
<tr>
<td>5.2</td>
<td>Models applied in the experiments</td>
<td>166</td>
</tr>
<tr>
<td>5.3</td>
<td>Descriptive statistics of 1-minute returns</td>
<td>167</td>
</tr>
<tr>
<td>5.4</td>
<td>Results obtained from the B&amp;H methods in the 100-day out-of-sample period</td>
<td>174</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.5</td>
<td>Standard ANFIS performance after the 100-day out-of-sample period.</td>
<td>175</td>
</tr>
<tr>
<td>5.6</td>
<td>ANFIS/T2a performance comparison against standard ANFIS after 100-day out-of-sample period.</td>
<td>176</td>
</tr>
<tr>
<td>5.7</td>
<td>ANFIS/T2b performance comparison against the standard ANFIS after 100-day out-of-sample period</td>
<td>178</td>
</tr>
<tr>
<td>5.8</td>
<td>ANFIS performance after 100-day out-of-sample period across different levels of return threshold (uncertainty)</td>
<td>179</td>
</tr>
<tr>
<td>5.9</td>
<td>Summary statistics for the standard ANFIS and the corresponding variations in the ANFIS/T2 models over a 100-day out-of-sample period across different levels of return threshold (uncertainty)</td>
<td>183</td>
</tr>
</tbody>
</table>
In this introductory chapter, we describe the adopted rationale in conducting this research. First, we borrow and present reflections from three knowledge domains, namely high frequency trading, financial risk and artificial intelligence, with special focus on fuzzy logic techniques. The intention is to highlight the key concepts from each domain as a reflection of current state of literature. In our review, we take a critical approach by identifying a number of open questions and literature gaps that motivate our research. Secondly, we present the main research objectives and how these are addressed in the subsequent chapters.

1.1 Background and motivation

1.1.1 Trading in financial markets

This section sets the scene of our research problem domain by highlighting the mechanisms of the trading environment as well as the decisions and actions with which traders are faced on a daily basis. The understanding of these elements, together with the interdependence with other key variables, such as price, volume, liquidity and volatility, are key in order to
identify opportunities where computational intelligence can be of beneficial assistance to traders, including the possibility of (intelligently) automating whole processes.

Johnson (2010) divides the trading process into three stages: (i) price formation, (ii) price discovery or trade execution, and (iii) reporting, clearing and settlement. In this thesis, our research overlaps with the first two stages.

Price formation is the process through which the price of an asset is determined. The price can be perceived as the consolidated investors’ view of the future value of an asset. Different investors tend to have different views in terms of asset valuation, and hence, price formation is the result of supply and demand conditions.

An understanding of market structure elements is crucial because they determine what people can know and hence what affects the price formation process. This drives the actions and relations among the different traders, ultimately resulting in determining who will trade profitably. The type of mechanism adopted in a specific market is typically either quote-driven or order-driven. In this thesis, our research is conducted using trade data from the London Stock Exchange’s electronic order book – Stock Exchange electronic Trading Service (SETS). SETS is one of the most liquid electronic order books in Europe, hence benefiting from lower transaction costs. Algorithmic, especially higher frequency, trading is typically characterised by a large number of orders with small order quantities, short holding periods and no overnight positions (Aldridge, 2013; Brogaard et al., 2014). This justifies our selection of SETS (offering high liquidity and volumes) for our experiments. Moreover, as shown by Aitken et al. (2015), HFT studies using SETS data are scarce.

Similar to most exchanges, LSE employs a continuous auction mechanism with orders entered between 8:00 and 16:30. Price discovery happens when supply and demand requirements intersect, hence determining the price of an asset. The constant tension between buyers and sellers dictates the dynamic (non-stationary) movement of market prices. Traders’ actions are driven by the information available and the deductions they extract from it. To
formulate these deductions, two approaches are commonly adopted, namely fundamental and technical analyses. While fundamental analysis involves the study of company fundamentals information, such as financial projections and market positioning, technical analysis is primarily concerned with price and volume patterns. The main aim of both approaches is to assist traders with the analysis of the information they receive and thus be in a position to formulate an opinion on the possible shifts in supply and demand, hence determining the direction in which prices are likely to move.

In this thesis, our interest is in short-term intraday trading, hence it is more speculative in nature. This makes technical analysis a more adequate tool and rests on the underlying assumption that prices move in trends (hence asserting price momentum) (Murphy, 1987). Harris (2002) defines speculators as traders who trade to profit from information and predictions about future prices. This also depends on the accuracy (and interpretation) of the signals generated from the information received. Typically, technical traders apply a set of technical indicators, which normally consist in an arsenal of statistical measures and charts, to assist with the analysis of the noisy price time series and with the prediction of short-term market moves. From the concluded analysis, speculators choose between buy or sell actions based upon which side they expect will be profitable. However, Bao and Yang (2008) show that capturing a true price turning point can be a difficult challenge since most of the technical indicators are price followers and, hence, it is difficult to infer whether a trend will continue or break. The difficulty is intensified due to the prevalence of higher market volatility. This presents the key challenge for trend-following trading strategies, which we investigate in this thesis.

The debatable predictability of financial markets is rooted in the Efficient Market Hypothesis (EMH) (Fama, 1965, 1970). According to EMH, markets are efficient and all attempts at predicting market prices are futile as the prices already incorporate all the information that could affect them. The theory interprets price movements as a random walk by hypothesising
that since only new information moves stock prices significantly, and since new information arrives at random, hence future movements in stock prices follow a random path. EMH (in its three forms) makes two important assumptions, that all public information is immediately available to the market, and that many traders are perfectly rational (given the same information, traders reach the same optimal asset valuation). However, in the literature we find a number of theoretical contributions which attribute departures from EMH. Simon (1972) points out that decision makers act within bounded rationality, hence their rational choice is within the space limited by their knowledge and processing capacity. Tsang (2009) argues that even in the case of two perfect decision makers, the amount of time taken to arrive at the optimal decisions might be different. For example, the time taken for a first agent to pick up an investment opportunity might be quicker than others. Moreover, a number of authors associate decision making with human emotions and psychological factors, such as overconfidence, herding, fear and regret – all sources of irrationality that lead to market inefficiencies (see Lo et al., 2005; Shiller, 1999).

These theoretical conflicts on market efficiency have fueled incessant empirical research in the search for approaches that can glean trading profitability. Gençay (1996) and LeBaron (1999) demonstrated that simple technical rules can predict market returns. An interesting study by Schulmeister (2009) found that beyond the 1990s, using daily data, technical trading rules became unprofitable. The same study, however, showed that by shifting to intraday 30-minute data, technical rules registered profitability again until the year 2000. Beyond the year 2000, another decline in technical rules’ profitability followed till 2006. Schulmeister suggested that beyond the year 2006, either markets became more efficient or stock price trends moved to higher frequency prices. However, Kearns et al. (2010) claim that, after taking account of transaction costs, aggressive high-frequency trading leads to relatively low profitability, far below the expected high excessive returns. Later, Holmberg et al. (2013) found empirical evidence of intraday trending in stock prices, linking the profitability of
technical rules with days which fall during periods exhibiting higher levels of volatility. In another study, Rechenthin and Street (2013) conducted empirical tests that indicated that, in general, a timespan of 30 minutes is required for stock prices to become efficient; however, no indication of possible profitability was investigated. In this thesis, we aim to contribute new findings to this ongoing debate.

There are also conflicting opinions amongst researchers in terms of the positive and negative market effects of algorithmic and high-frequency trading. Brogaard et al. (2014) recognise the wider adoption of algorithmic and high-frequency trading as a natural evolutionary path in market development, driven by advances in technology, and attribute this advancement with improvements in price discovery, increased liquidity, and with no negative effects on volatility. Zhang (2010) argues that high-frequency trading amplifies price reactions and volatility. Following incidents like the 2008 financial crisis and the flash crash of 2010, algorithmic trading and high-frequency trading began to attract more attention from regulatory bodies. Regulators’ concern was heightened due to the development of special order types like hidden orders, icebergs, and the use of dark pools. In 2014, the European Commission introduced MiFID II (MiFID, 2014), a new regulatory regime for firms which engage in algorithmic or high-frequency trading. MiFID II and delegated acts under MiFID II will apply from January 2018 (this follows a recent one-year extension proposed by the Commission). Any firm carrying out algorithmic trading (including high-frequency trading) is subject to numerous organisational requirements. In particular, under Article 17 (1)(2), investment firms practising algorithmic trading are subjected to new rules which ensure resiliency, sufficient capacity, adequate risk controls, effective backtesting, and business continuity arrangements. Moreover, these firms are required to be more transparent regarding the type of algorithmic trading strategies adopted, the range of trading parameters or limits to which their systems are subjected, the risk controls applied and details of backtesting results. These new regulations underscore the importance of improving risk controls and the
risk-adjusted performance of trading algorithms. In this thesis, we aim to present innovative techniques of how these objectives can be addressed.

In the next section, we analyse the main building blocks of trading algorithms and how AI can contribute to improving them.

1.1.2 The architecture of a trading algorithm

The advancement in computing power and general network (internet) bandwidth, coupled with increased direct access trading, have boosted the opportunity for the adoption of sophisticated algorithms to search for quick intraday profit opportunities. In this new reality, Tsang (2009) states that traders’ rationality is measured by the solution (algorithm) optimality, hence from a computational point of view, traders’ rationality is reflected by their computational intelligence, the better algorithms translating into increased trading opportunities. This puts into perspective the importance of our area of research.

In the literature, one finds different definitions of algorithmic and high-frequency trading. We adopt the definitions as stated in MiFID II. Article 4(1)(39) of MiFID II defines algorithmic trading as:

“... trading in financial instruments where a computer algorithm automatically determines individual parameters of orders such as whether to initiate the order, the timing, price or quantity of the order or how to manage the order after its submission, with limited or no human intervention, and does not include any system that is only used for the purpose of routing orders to one or more trading venues or for the processing of orders involving no determination of any trading parameters or for the confirmation of orders or the post-trade processing of executed transactions.”

Similarly, Article 4(1)(40) of MiFID II defines high-frequency algorithmic trading as:
1.1 Background and motivation

Fig. 1.1: General components of a trading algorithm

“... an algorithmic trading technique characterised by: (a) infrastructure intended to minimise network and other types of latencies, including at least one of the following facilities for algorithmic order entry: co-location, proximity hosting or high-speed direct electronic access; (b) system-determination of order initiation, generation, routing or execution without human intervention for individual trades or orders; and (c) high message intraday rates which constitute orders, quotes or cancellations.”

These definitions place high-frequency trading as essentially a subset of algorithmic trading but where the focus is on higher trading speeds. From a computational perspective, an algorithmic trading solution can be viewed as a series of interconnected blocks of functions, each representing a loosely decoupled software module, with defined inputs and outputs, that permit independent development, updates and testing. Apart from being good software practice, loosely decoupling the modules is ideal in view of MiFID II obligations, which state that any substantial update of an algorithmic trading system requires rigorous testing and approvals.

In Figure 1.1 we present a high level architecture of the trading solutions that we adopt in this thesis. Below, we provide a description of each module and the corresponding gaps that we identified from the literature:
Data preparation module

This module receives raw price data, typically in the form of ticks (stock, quantity, price) from data providers, and transforms the data into the format which will later be required by the model. Although it depends on the type of model adopted, a typical transformation converts irregular time interval points into regular ones (daily, hourly, 30-minute, 10-second, etc.) by taking the last price point before the end of each time bar. In spite of the persisting contention about the feasibility of trading in the high-frequency space, surveys (Krollner et al., 2010; Tsai and Wang, 2009) show that the majority of former studies still focus on daily price time series, making studies in the higher frequency space more scarce. In this thesis, we address this gap by focusing on intraday trading using high-frequency data.

Signal generation module

This module takes as an input the time series from the data preparation module and generates another time series of either transformed or smoothed variables. Kaastra and Boyd (1996) suggest that most common data transformations include first differencing and taking natural logs. The same authors advise to use smoothing techniques for both input and output data (e.g. using moving or exponential moving average), since trend prediction using only price changes or no noise filtering can be a difficult challenge for the underlying model. Tsai and Wang (2009) and Krollner et al. (2010) indicate that the majority of former studies make use of lagged index data, including the use of technical indicator signals like moving averages, RSI, Bollinger bands, etc. Typically, the data output from this module is also converted into a format as required by the prediction model. For example, if the prediction model takes, as its input, three technical signals to predict the next return, the data is organised as a matrix consisting of four columns, where three columns represent the inputs (lagged signals) and one column represents the output. The inputs might also include feedback from the model output - this typically helps the prediction model to adapt to changing market conditions. A similar combination of signals is used in this thesis.
A common approach that is adopted when designing a machine learning model (with the objective of identifying a model with good generalisation capabilities) involves splitting the data into three distinct data sets that are used for training, testing and validation (out-of-sample). Bailey et al. (2014) warn that many computational finance studies only use in-sample data or else short out-of-sample periods, hence increasing the possibility of obtaining spurious results. In this thesis, we follow the suggestions from Kaastra and Boyd (1996) and Pardo (2011), who suggest the adoption of a moving window (walk forward) approach. This approach is typically used for testing trading systems and constitutes a more rigorous and realistic methodology. This involves splitting the data into overlapping training-test-validation sets, and on each cycle moving each set forward through the time series. This approach tends to result in more robust models due to more frequent retraining and large out-of-sample data set (increasing training processing requirements but also resulting in models which adapt more quickly to changing market conditions).

Prediction module

Foucault and Roşu (2016) state that a stylised fact of high-frequency traders is that their aggressive (marketable) orders anticipate very short-run price changes. Brogaard et al. (2014) claim that high-frequency traders predict short-term price changes; by using marketable orders and trade in the direction of true price changes and filter out transitory pricing errors, sufficient informational advantage is generated in order to break the underlying costs threshold. Gençay et al. (2001) and Johnson (2010) state that the ability to predict trends, for example using data mining or artificial intelligence techniques, can offer an edge to traders. In view of the stochastic nature of market prices and increased volatility and unexpected market shifts (hence increased uncertainty), this poses a difficult challenge. Depending on the granularity that a time series is analysed, the more trends can be identified, hence increasing trading opportunities; however, being overly sensitive to short-term fluctuations (hence resulting in position times which are too short) can result in losses, especially due to
transaction costs (Kearns et al., 2010). The objective for an optimal algorithm is to identify trends at an early stage until the model signals that the trend has reversed. In this thesis we aim to contribute to existing literature by proposing new robust prediction models and approaches.

In the literature, we find that popular time series models include regression methods and the ARIMA models (Box et al., 2015), but these models exhibit limitations particularly because of their linear nature (see Lin et al., 2002). Surveys by Tsai and Wang (2009) and Krollner et al. (2010) show a clear trend to use established Artificial Neural Networks (ANNs) in stock price forecasting and enhance them with new training algorithms or combine ANNs with evolutionary and optimisation techniques into hybrid systems. Following the emergence of fuzzy logic (Zadeh, 1973), neural networks and fuzzy inference systems were brought together as general structures for approximating non-linear functions and dynamic processes. Fuzzy logic models were designed to better manage the prevalent uncertainties of the underlying complex systems, which is the case of financial markets. The learning capabilities of neural networks and the increased ability of fuzzy logic models to handle uncertainties are the key features which drove our interest to identify innovative hybrid models that can improve the performance of trading algorithms.

Following the surveys conducted by Tsai and Wang (2009) and Krollner et al. (2010), we argue that the criteria that is typically used to evaluate and compare prediction models do not necessarily translate into profitable trading solutions. For example, Brabazon and O’Neill (2006) state that achieving the best root mean squared error does not indicate profitability for trading purposes, since smaller errors, but wrong side predictions, will still result in losses. Similarly, a high directional accuracy model might still suffer from fewer, but larger, losses. We address this gap by proposing better AI model design methods for algorithmic trading purposes, especially in higher frequency intraday trading scenarios.

**Trading algorithm module**
1.1 Background and motivation

The trading algorithm can be considered as a rule engine which, based on the technical signals and predictive signals as inputs, automates trading decisions. Covel (2009) identifies a number of decisions that are typically managed by a trading algorithm: (i) what to buy or sell, (ii) position sizing, (iii) market entries, (iv) trading stops, and (v) trading exits. Each condition is typically conditional on specific thresholds which need to be tuned as part of the model calibration process. For example, Vanstone and Finnie (2009, 2010) suggest a fixed return threshold filter to avoid small unprofitable movements; however, Holmberg et al. (2013) warn that although the use of higher fixed thresholds increase prediction accuracies, this also results in reduced trading opportunities. Similarly, Brabazon and O’Neill (2006) suggest a volatility filter which limits trading during high volatility (risk) periods, hence avoiding possibly strong adverse market movements. On the other hand, a number of authors (Han et al., 2013; Holmberg et al., 2013; LeBaron, 1999; Schulmeister, 2006) indicate that during high volatility periods, markets tend to become less efficient, hence the better chances for technical rules’ profitability. We conclude that more research is required that investigates the automation of adaptive thresholds along intraday time-varying risk. We address this gap in our thesis.

In this thesis, we propose a methodology which jointly calibrates the trading algorithm parameters with the prediction model parameters, and hence the adopted performance measure considers the joint performance of both the prediction model and the trading algorithm. Although Vanstone and Finnie (2009) and Brabazon and O’Neill (2006) indicate approaches which combine the tuning of the prediction model and the trading algorithm parameters jointly, rather than separately, we note that most published studies stop short at the prediction model.

From the literature we also identify two other shortcomings. Pardo (2011) states that many AI studies focus solely on trend prediction, but decisions like position sizing are often not appreciated and poorly understood in trading strategy design. This possibly leads to
sub-optimal solutions. Moreover, firms engaging in algorithmic trading need to incorporate rules in their system that ensure that they abide by MiFID II regulations – for example, based on the capital base, there are limitations on trading style, maximum order value, maximum order volume, and maximum long-short positions. Vanstone and Finnie (2010) argue that many published studies do not consider real-world intraday market characteristics like trading costs, realistic trading hours and no overnight positions, possibly leading to biased results (Brabazon and O’Neill, 2006; Pardo, 2011). These two shortcomings indicate a gap in current literature which requires further research, especially in intraday trading scenarios. This thesis fills this gap by considering real-world constraints.

**Trade execution module**

This module handles the decisions passed on by the trading algorithm. It typically consists in a set of instructions to execute orders in a specific way. Johnson (2010) states that hundreds of trading algorithms exist in this category (for example, time-weighted average price, percentage of volume and implementation shortfall). Although AI algorithms can be used to improve these execution algorithms (for example, Kablan and Ng, 2010), it was our conscious decision to apply a simple execution mechanism and focus on improving aspects related to the prediction and algorithmic trading modules only (specifically, trend-following strategies). This does not exclude the need for possible future research in other areas.

### 1.1.3 Aligning algorithmic and investors’ risk preferences

The new regulations, like the upcoming MiFID II, emphasise the importance that trading systems are adequately designed and tested to mitigate the risks to which they are exposed. This has to be demonstrated by the (resilient) performance obtained through simulations and backtesting of algorithms. MiFID II lists a number of risks of algorithmic trading, like the overloading of systems at trading venues, the generation of duplicative or erroneous orders, and overreactions to market events, hence boosting volatility. We take the perspective of an
1.1 Background and motivation

Investor and revert to a fundamental concept in standard economics literature which states that investors are concerned with risk and return (Markowitz, 1952). Markowitz argues that if an investor knew the future returns with certainty, then choosing an investment option would merely boil down to picking the security which offers the highest return. However, in the presence of uncertainty, which is inherent in financial markets, the Markowitz mean-variance model allows an investor to seek the highest return at an acceptable level of risk – risk being measured by variance. Motivated by the mean-variance model, Sharpe (1966) proposed the Sharpe ratio as a risk-adjusted measure to compare investment performance. To date, it remains one of the most popular risk-adjusted performance measures due to its practical use, applying standard deviation as the measure of risk. Eling (2008) and Prokop (2012) show the proliferation of risk-adjusted performance measures and their tests indicate that in many cases, the Sharpe ratio leads to similar rankings of the more sophisticated performance ratios. In view of these findings, in this thesis the Sharpe ratio is adopted as one of the key risk-adjusted performance measures.

In finance, volatility is the statistical measure of stock price dispersion, hence the movement of a stock price without regard to direction. Large (small) average daily stock price movements mean high (low) volatility. From one perspective, volatility is linked with unstable market scenarios with the possibility of abrupt larger losses. However, from a different perspective, higher volatility can be interpreted as an opportunity for higher profits. This dilemma is what links volatility with market uncertainty (risk) and more challenging forecasting. Andersen et al. (2009) state that measuring and forecasting volatility is a core topic in finance literature. The ARCH model, introduced by Engle (1982), followed by the GARCH model, generalised by Bollerslev (1986), promoted time-varying volatility to a very active area of research. Over the years, numerous improvements were proposed, such as the ability to incorporate the nonlinearity, asymmetry and long memory properties in the volatility process. Popular extensions include the EGARCH model (Nelson, 1991) and the
GJR-GARCH model (Glosten et al., 1993) (for a review on ARCH models, see Hansen and Lunde, 2005; Poon and Granger, 2003).

Merton (1980) noted that the volatility of a Brownian motion can be approximated to an arbitrary precision using the sum of intraday squared returns, a measure which was coined as realised variance. Later, Andersen et al. (2000b) showed that ex-post daily foreign exchange volatility is best measured by aggregating intraday squared five-minute returns. Consequently, volatility literature has steadily progressed towards the use of higher frequency data which has become more accessible. Andersen and Bollerslev (1997) and Andersen et al. (2000a) claim that volatility exhibits strong intraday periodicity and high persistence, and can be predicted with a good degree of accuracy up to an intraday level. Concurrently, research lead to the investigation of alternative volatility estimators (for example, Barndorff-Nielsen and Shephard, 2004, 2006), primarily focusing on reducing the effect of microstructure noise in high-frequency prices.

We identify a strong disconnect between the investors’ risk-return concept and existing computational finance literature. The literature in computational finance (Brabazon and O’Neill, 2006; Krollner et al., 2010; Pardo, 2011; Tsai and Wang, 2009) clearly indicates that risk is rarely considered, especially in intraday trading scenarios. Many former studies focus on the application of models solely to predict market movements (a more recent example can be found in Son et al., 2012) and, for model evaluation, it is common practice that error measures, win ratios or profitability are applied, but with little attention to risk-adjusted performance. Moreover, although new measures of variation make it possible to predict volatility with a good degree of accuracy up to an intraday level, the use of this information for intraday trading purposes is rarely considered.

We extend the thoughts of Lo et al. (2005) and Shiller (1999) with regard to the effect of human emotions in trading to a computational perspective. We argue that algorithms are programmed encapsulations of the thought processes of the traders who design them,
1.1 Background and motivation

hence the same human elements can be incorporated in the algorithms (e.g. applying different objective functions and risk considerations). Consequently, in line with this thesis, researching the application and performance of risk-adjusted objective functions exhibits higher resonance with investors’ risk sensitive preferences.

Gradojevic and Gençay (2013) indicate the benefits of combining type-1 fuzzy logic with standard technical indicators in view of increased trading uncertainties. Other recent applications of fuzzy logic models in finance have been demonstrated by a number of authors (Boyacioglu and Avci, 2010; Chang et al., 2011; Chen, 2013; Kablan and Ng, 2011; Tan et al., 2011; Wei et al., 2014). We argue that although one of the key claimed advantages of fuzzy logic, when compared to other AI techniques, is the better management of uncertainty, the link (and feasibility) between fuzzy logic and risk-adjusted trading performance gains, to our knowledge, was never explored in previous studies. Mendel et al. (2006) and Wagner and Hagras (2010) demonstrate the increased capabilities in managing uncertainty that is possible by using type-2 fuzzy logic when compared to the earlier type-1. However, this comes at the cost of additional complexity, which is typically the main reason why many practitioners still favour the use of type-1 (Wu and Mendel, 2014). Aladi et al. (2014) suggest that it is not clear what level of uncertainty warrants the (feasible) use of type-2 fuzzy logic (in some cases, no benefits were achieved). Moreover, we have not identified any previous studies which explore the relationship between the move from type-1 to type-2 fuzzy logic and any potential additional gains in profitability and risk-adjusted performance in an intraday trading scenario.

In the spirit of the Sharpe ratio, we contend that investors’ risk is related to the performance stability of the underlying trading models across continuously changing market (risk) conditions. From a computational perspective, Marsland (2009) suggests that systems have to adapt, and hence evolve, to address recurring and changing patterns in the intraday environment which are driven by the actions of informed and uninformed traders.
Implementing evolution requires an ability to balance learning about new information while still respecting past accumulated knowledge. We also note that common approaches only compare model performance at a single point in time following a defined out-of-sample period (Kaastra and Boyd, 1996; Vanstone and Finnie, 2009, 2010). We contend that this approach does not necessarily reflect the performance stability of the underlying models across the different market conditions of the out-of-sample period. Pardo (2011) emphasises the importance of backtesting algorithms using a set of stocks with different trends – this avoids any bias towards particular trends and hence examines the stability of the tested models across different market scenarios. From the reviewed studies and surveys, we contend that these conditions are rarely considered, leaving a research gap for further investigation on the profitability and risk-adjusted performance of trading algorithms under stricter (and more realistic) conditions.

1.2 Research objectives and thesis structure

The approach developed in this thesis is rooted in finding better ways of how to manage the intimately intertwined concepts of uncertainty and risk as an integral part in the design of algorithmic and high-frequency trading solutions. The adopted framework cuts across three knowledge domains, namely high-frequency trading, financial risk and artificial intelligence, with special focus on fuzzy logic techniques.

Uncertainty (risk) is an intrinsically important issue in the design of trading systems because much of the underpinning information is imprecise, incomplete or not totally reliable. The source of this uncertainty originates from a number of dynamic and complex microstructure factors. This makes the problem of predicting financial markets, and achieving consistent profitable results, a very difficult one and the subject of ongoing research.

The management of uncertainty is a central theme within the broad framework of fuzzy set theory. (Zadeh, 1997, p. 123) states that “the guiding principle of soft computing is to exploit
the tolerance of imprecision, uncertainty, and partial truth to achieve tractability, robustness, low solution cost, and better rapport with reality”. This precludes using conventional approaches that require a detailed description of the problem being solved. In this thesis, we investigate the use of soft computing as a decision-making support for short-term investment strategies.

We summarise our primary research objectives as follows:

1. Identify elements in the AI modelling process, when applied in order to enhance trading algorithms, that result in a positive impact on risk-adjusted performance.

2. Investigate the additional risk-adjusted performance that can be gained from the application of fuzzy logic models when compared to popular AI models in the finance literature.

From these primary objectives, we identify a number of secondary objectives which need to be addressed in order to achieve the primary objectives:

1. Investigating the risk-adjusted performance of money-losing algorithms is pretty much pointless. Hence, we aim to explore the debated profitability of technical trading rules, in conjunction with AI, focusing particularly on high-frequency data in an intraday trading scenario.

2. Profitability and risk can be considered as two sides of the same coin. In our research, we intend to identify ways of how risk-adjusted performance can be improved without negatively impacting profitability, but rather improving it.

3. Identify the effect of including (or omitting) real-world intraday trading constraints like trading costs, realistic trading hours and no overnight positions, which are typically ignored in existing studies, possibly leading to biased results.
4. Identify a measure of time-varying intraday risk and how AI algorithms can benefit from this information to dynamically adapt the risk profile and improve the risk-adjusted performance of trading algorithms.

5. Identify whether (and when) advancing from type-1 (T1) to type-2 (T2) fuzzy logic systems provides viable benefits for trading purposes.

6. Increasing model complexity can limit the widespread adoption. In our approach, we consciously aim to identify stepwise improvements on existing models, using fuzzy logic, with minimal increase in model complexity. This also helps in clearly identifying the advantages gained from the proposed approaches (rather than obtaining performance improvements through increased model complexity).

To address these objectives, we take the following approach:

In Chapter 2 we cover the core mathematical and theoretical aspects of fuzzy logic as a preliminary background to the more advanced fuzzy models and optimisation techniques that we will make use of in subsequent chapters.

The objective of Chapter 3 is twofold. Firstly, we analyse the out-of-sample performance of trading models based on our exploration of different objective functions. In this chapter we address these issues and present an approach which optimises risk-adjusted performance whilst considering realistic costs and constraints. Secondly, we explore the positive performance and stability contributions that can be attained by adopting fuzzy logic models in comparison to more popular models, such as neural networks.

Next, in Chapter 4 we extend our research by addressing the problem of time-varying intraday volatility and hence exploring risk-return optimisations at a more granular intraday level. We investigate the overall performance gains that can be attained by
applying fuzzy logic controllers that automatically identify intraday risk scenarios and adapt money management decisions to varying degrees of risk.

As a further step, in Chapter 5 we expand our research on T1 fuzzy logic findings by addressing modelling problems resulting from microstructure noise and focusing on improving risk-adjusted trading performance from an even deeper market microstructure perspective. We investigate the effects on risk and return measures by introducing T2 fuzzy logic at various levels of trading frequencies and microstructure noise.

Finally, in Chapter 6 we consolidate our findings and bring together the final reflections emerging from this thesis. We conclude by providing suggestions for further research.

### 1.3 Contributions

We will now briefly highlight the contributions of the three research chapters in this thesis.

- **Chapter 3**

  Our first contribution is the simple yet effective extension of common technical moving average rules by considering a dynamic combination of moving average prediction models. In our approach, we use AI techniques to combine different moving averages and dynamically tune the trend signals according to the changing speeds of the market. Our results demonstrate that the proposed dynamic moving approach outperforms the risk-adjusted performance obtained from standard moving average technical rules. Contrary to the Efficient Market Hypothesis, we also make use of heat maps which identify pockets of profitability in the 10-minute to one-hour range. As our second contribution, we show that risk-adjusted performance functions lead to better out-of-sample performance than the more common, error-based, objective functions (Brabazon and O’Neill, 2006). To our best knowledge this has not been studied in an intraday high-frequency setting before (as indicated by Bahrammirzaee, 2010, and references therein). Moreover, we show that a
large number of publications, which do not consider transaction costs and other trading constraints in the training process (see surveys by Krollner et al., 2010; Tsai and Wang, 2009), can lead to inflated and unrealistic results in the out-of-sample. Thirdly, we propose an innovative approach which combines multiple risk-return objective functions using an ANFIS ensemble dynamic selection method, and show how it can improve the intraday trading performance of AI models. Finally, we compare the time-varying Sharpe ratio results of the underlying models. We identify the performance stability enhancements that can be attained by adopting fuzzy logic models in comparison to more popular models, such as neural networks.

• Chapter 4

Contrary to many studies that limit the application of AI solely to focus on market direction (see Krollner et al., 2010; Tsai and Wang, 2009), in this chapter we extend our model improvements, as identified in Chapter 3, to the money management decisions of trading algorithms. As our first contribution we present an effective approach to dynamically adjust trading frequency and capital allocation depending on the varying degrees of risk at an intraday level with the objective to improve overall risk-adjusted trading performance. Core to our technique is a fuzzy clustering approach which can identify preferable regions across the trend and volatility space. We use intraday realised volatility as a proxy for uncertainty. This approach goes contrary to studies suggesting the use of fixed return and volatility thresholds (Holmberg et al., 2013; Kaastra and Boyd, 1996; Vanstone and Finnie, 2009, 2010). We show how our aim to improve risk-adjusted performance can be achieved without incurring a reduction in profitability (which is the dilemma reported by Holmberg et al., 2013). Our results show significant improvements in both profitability and risk-adjusted performance when compared to standard NN and buy-and-hold methods. As our second contribution, our results suggest a possible martingale property breakdown during specific trend-volatility states. From a theoretical perspective, we further support the
1.3 Contributions

claims of a number of authors (Gradojevic and Gençay, 2013; Han et al., 2013; Holmberg et al., 2013; LeBaron, 1999; Schulmeister, 2006) who support this relationship, and we extend this research to shorter-term intraday horizons.

• Chapter 5

In our last research chapter, we explore further improvements by advancing our type-1 fuzzy logic ideas to type-2. As our first contribution we propose two innovative and practical methods of how the ANFIS model can be improved by introducing IT2 fuzzy sets with a minimal increase in complexity. This presents an improvement when compared to our ANFIS findings presented in Chapter 3. The main benefit is to minimise the effect of uncertainty caused by microstructure noise, hence reducing the overall risk. By extending the clustering data partitioning ideas that we use in Chapter 4, especially in view of different intraday volatility levels, here we apply a similar approach for model structure identification, which is also used as a basis for our IT2 extensions. This approach results in more compact and efficient fuzzy models. The proposed T2 methods show a significant increase in both risk-adjusted trading performance and profitability when compared to standard ANFIS and B&H methods. Secondly, our results extend the findings of a number of authors (Holmberg et al., 2013; Rechenthin and Street, 2013; Schulmeister, 2009) who claim possible breaks in market efficiency at short time frames. By utilising a combination of technical rules on 2-minute returns with holding periods ranging from between 2 to 10 minutes, we extend our market efficiency investigations that we present in Chapters 3 and 4 to a more microscopic perspective and identify possible pockets of profitability in shorter time windows. As a result of this, we manage to identify a positive link between higher order fuzzy systems and risk-adjusted trading performance. Thirdly, although a number of authors (e.g. Aladi et al., 2014; Sepulveda et al., 2006) demonstrate the increased capability of IT2 models to handle increased uncertainty when compared to T1, we provide deeper insight on the benefits of adopting IT2 models from the perspective of different levels of
trading risk (uncertainty) and trading frequency. We conduct our analysis by adopting an innovative technique to compare how type-1 and type-2 models cope at decreasing (increasing) levels of return thresholds, which is reflected in an increase (reduction) in uncertainty but also in increased (reduced) return potential. To our knowledge, this had never been studied previously, and is thus of great value to HFT.
Chapter 2

Fuzzy logic background

This chapter provides a preliminary mathematical background on fuzzy logic with a specific focus on Type-1 Takagi-Sugeno-Kang (T1 TSK) FLSs (Sugeno and Kang, 1988) and Interval Type-2 (IT2) FLSs (Mendel et al., 2006; Zadeh, 1975). The selected review is presented in such a way so as to set out the necessary building blocks for the more sophisticated fuzzy logic techniques which are extensively used in the subsequent chapters.

2.1 Introduction

Fuzzy logic saw its inception in the proposal of Fuzzy Set Theory by Zadeh (1965). The main motivation behind fuzzy logic can be attributed to the Principle of Incompatibility (Zadeh, 1973) which, in the face of the increasing complexity of underlying systems, supports a departure from the traditional quantitative techniques in favour to models which can model and minimise the effect of uncertainty. In essence, fuzzy logic and fuzzy set theory offer an alternative approach to conventional system theory, which relies on crisp mathematical models of systems, such as algebraic or difference equations.

Fuzzy Rule-Based Systems (FRBS) are a popular application of fuzzy logic and fuzzy set theory and present a competitive alternative to other classic models and algorithms in
classification and regression problems. An FRBS (Figure 2.1) typically consists of four components (Mendel et al., 2014):

- a fuzzifier which takes crisp values as inputs and converts them into degrees of membership of the fuzzy term of each variable.

- a knowledge base which consists of a database holding fuzzy set definitions, and a rule-base storing a list of fuzzy IF-THEN rules.

- an inference engine responsible for inference operations by applying the fuzzy IF-THEN rules.

- a defuzzification component to convert fuzzy values into crisp outputs.

The first fuzzy models were focused on linguistic fuzzy modelling and their semantic interpretability capabilities. One of the most popular of these models is the Mamdani Model (Mamdani, 1977). In these systems, words or sentences are used in place of numbers to describe relationships which are too complex to be described in quantitative terms. Another
popular model is the TSK model which replaces the fuzzy sets as defined in the consequent of a Mamdani rule with a function, typically linear, of the input variables. TSK rules are much more popular in practice due to their simplicity and flexibility (Wu and Mendel, 2014). Secondly, by avoiding the defuzzification process of the Mamdani approach makes the TSK model more computationally efficient. The TSK model was successfully applied in numerous financial applications and still the subject of ongoing research (Boyacioglu and Avci, 2010; Chang et al., 2011; Chen, 2013; Kablan and Ng, 2011; Tan et al., 2011; Wei et al., 2014). It is for this reason that in this thesis we adopt the TSK approach to identify innovative solutions to improve risk-adjusted trading performance.

2.2 Type-1 fuzzy sets and fuzzy membership functions

Fuzzy set theory lies at the heart of FRBSs. In classical set theory, the membership of elements in a set is a bivalent condition defined by a characteristic function which can take crisp values of 0 or 1. Boolean operators AND, OR, and NOT are used to perform the intersect, union and complement operations. A fuzzy set extends this idea to multi-valued logic and expresses the degree to which an element belongs to a set, hence allowing the representation of vagueness and uncertainty. If \( X \) is a collection of objects denoted generically by \( x \), then a fuzzy set \( A \) in \( X \) is defined as a set of ordered pairs:

\[
A = \{(x, \mu_A(x)), x \in X\}, \tag{2.1}
\]

where \( \mu_A \) is called a membership function (MF) denoting the degree of membership of an element in a given fuzzy set \( A \). Hence, rather than a crisp 0 or 1, the function is allowed to return values between 0 and 1. When \( X \) is continuous, \( A \) is commonly written as

\[
A = \int_X \mu_A(x)/x, \tag{2.2}
\]
where the integral operator does not denote integration, but it denotes the collection of all points, \( x \in X \), and the slash (/) operator associates the elements in \( X \) with their membership grades using membership function \( \mu_A(x) \). If \( X \) is discrete, \( A \) is commonly written as

\[
A = \sum_x \mu_A(x)/x,
\]  

(2.3)

where the summation sign does not denote arithmetic addition, but the collection of all points \( x \in X \) with associated MF \( \mu_A(x) \). Corresponding to the ordinary set operations of union, intersection, and complement, fuzzy sets have similar operations. Mathematical definitions of the generalised AND and OR operators are called \( t \)-norm (\( \ast \)) and \( t \)-conorm (\( \oplus \)) respectively.

In the following chapters, we apply the algebraic product \( t \)-norm for fuzzy intersection and the maximum \( t \)-conorm for fuzzy union. As suggested by Mendel et al. (2014), these operators are the ones which are most commonly applied in engineering applications of fuzzy logic.

There are several classes of parameterised functions commonly used to define MFs. These parameterised MFs play an important role in adaptive fuzzy inference systems. In our models proposed in the next chapters, we adopt Gaussian MFs, where each fuzzy set is represented by

\[
\mu_A(x; \bar{x}, \sigma) = e^{-\frac{(x-m)^2}{2\sigma^2}}.
\]  

(2.4)

Unlike other MFs, this shape has only two parameters (the mean \( m \) and variance \( \sigma \)) and it always spreads out over the entire input domain (Wu and Mendel, 2014).

### 2.3 Type-1 TSK fuzzy logic systems

TSK fuzzy models are characterised by a rule-base consisting of rules with fuzzy sets in the antecedents and functions, typically linear, in the consequents. A typical T1 TSK model,
2.3 Type-1 TSK fuzzy logic systems

with \( k \) inputs and \( M \) rules, has rules in the following form:

\[
IF \quad (x_1 \text{ is } A_{i,1}) \ AND \ (x_2 \text{ is } A_{i,2}) \ AND \ ... \ AND \ (x_k \text{ is } A_{i,k})
\]

\[
THEN \quad y_i(x) = w_{i,0} + \sum_{j=1}^{k} w_{i,j}x_j
\]

(2.5)

where \( i = 1, ..., M \) is the rule number, \( w_{i,v}(v = 0, 1, ..., k) \) are the consequent parameters, \( y_i(x) \) is the output of the \( i \)th rule, and \( A_{i,l}(l = 1, 2, ..., k) \) are T1 antecedent fuzzy sets. The output is computed as

\[
y_{TSK,1}(x) = \frac{\sum_{i=1}^{M} f_i(x)y_i(x)}{\sum_{i=1}^{M} f_i(x)}
\]

(2.6)
Fuzzy logic background

![Diagram of T1 TSK fuzzy model example](image)

Fig. 2.3: T1 TSK fuzzy model example

which can be expanded to

\[
y_{TSK,1}(\mathbf{x}) = \frac{\sum_{i=1}^{M} f_i(\mathbf{x})(w_{i,0} + w_{i,1}x_1 + w_{i,2}x_2 + \ldots + w_{i,k}x_k)}{\sum_{i=1}^{M} f_i(\mathbf{x})}, \tag{2.7}
\]

where

\[
f_i(\mathbf{x}) = \mu_{A_{i,1}} \times \mu_{A_{i,2}} \times \ldots \times \mu_{A_{i,k}}. \tag{2.8}
\]

Figure 2.3 represents an example of a T1 TSK model consisting of 2 inputs \( \{x_1, x_2\} \), 2 rules \( \{R_1, R_2\} \) and 2 MFs \( \{A_{i,1}, A_{i,2}\} \) describing each input space, where \( i = \{1, 2\} \).

### 2.4 Designing fuzzy logic systems

A popular fuzzy modelling approach, originally proposed by Sugeno and Kang (1988), involves mapping input-output data. The procedure mainly consists of two parts: structure identification and parameter identification. The next step for a model designer is to define
the structure of the fuzzy logic model. This consists in defining the number and position of antecedent membership functions and corresponding consequents. Ruano et al. (2002) recognized the similarity between neural networks and neuro-fuzzy systems, as they share a common structure: they can be envisaged as a two-stage model, the first performing a non-linear mapping from an input space to an intermediate space, usually of greater dimensionality and using basis functions, and a latter stage, consisting of a linear mapping between the intermediate space and the output space.

### 2.4.1 Structure identification

From the literature we identify two common sources of information for building fuzzy models, prior expert knowledge and directly from data. Traditionally, fuzzy rules would have been elicited via discussions with domain experts. The disadvantage of this approach is that typically the elicitation process requires long discussions with the domain experts. Moreover, subjective or conflicting opinions from domain experts can lead to imprecise information (Buchanan and Wilkins, 1993). Gradojevic and Gençay (2013) and Cheung and Kaymak (2007) demonstrated examples of such an approach for the construction of trading algorithms. In their approach, fuzzy logic and technical analysis were combined by incorporating fuzzy rules based on expert knowledge.

Core to the rule induction process is data partitioning, each rule being a local model representation of the identified subspace. Published studies suggest numerous approaches for learning fuzzy rules from data, such as, directly from data (Wang and Mendel, 1992), decision trees (Chen et al., 2001), neural networks (Nayak, 2009) and support vector machines (Chiang and Hao, 2004). The focus of earlier data-driven methods was on achieving rule extraction automation using simple methods. However, the partitioning algorithms applied resulted in a large number of membership fuzzy sets, with the number of rules generated in many cases being proportional to the cartesian product of each membership fuzzy set. As noted
Fuzzy logic background

Fig. 2.4: Projection of identified clusters on the input space for MF identification

by Mohammadian (1995), as the number of parameters of a system increase, the number of fuzzy rules of the system grows exponentially. Apart from the computational penalty, this also reduces the interpretability of the model.

A survey by Dutta and Angelov (2010) indicates that data clustering is one of the most popular approaches that is used to generate fuzzy rules automatically from input-output data. Advancement in clustering techniques lead to various improvements. These can be split into three categories, offline, online and evolving clustering techniques. The distinguishing factors between these categories are mainly the learning approach and the ability of the clusters to adapt (evolve) as opposed to using a fixed number of clusters. In this thesis we make extensive use of clustering methods to generate fuzzy rules. In Chapter 3 we use an evolving clustering method to generate fuzzy rules from data. In Chapter 4 we use a fuzzy clustering method to identify trading rules based on risk-return regions. This is again used in Chapter 5 to generate an initial T1 rule base which is then extended to IT2.

In order to generate fuzzy rules, fuzzy clustering can be applied in the input data space only, the output data space only or jointly together. Here we present an example to describe
the concept of generating fuzzy rules using clustering. We consider a scenario where we have two inputs \( \{x_1, x_2\} \). As indicated in Figure 2.4, Gaussian MFs can be extracted by projecting clusters identified in the input space onto the respective axes. In the above example two data clusters are identified resulting in two MFs \( \{C1, C2\} \) on each input space. In the TSK model, each obtained cluster is represented by one rule. In this example, the generated rule base, consisting of two rules, will be defined as follows:

\[
\text{R1: IF } (x_1 \text{ is } C1) \text{ AND } (x_2 \text{ is } C1) \text{ THEN } y_{1}(x) = w_{1,0} + w_{1,1}x_1 + w_{1,2}x_2 \\
\text{R2: IF } (x_1 \text{ is } C2) \text{ AND } (x_2 \text{ is } C2) \text{ THEN } y_{1}(x) = w_{2,0} + w_{2,1}x_1 + w_{2,2}x_2
\]  

(2.9)

It is evident that each rule generates a sub-model for each approximate position in the relevant input space. When a new input is presented, the rule (model) closest to the input position will be given more weight. Further parameter identification and rule-base tuning follows. This is described in the next section.

### 2.4.2 Parameter tuning using least squares and back-propagation

Fuzzy logic literature promotes both gradient-based and evolutionary algorithms (Wu and Mendel, 2014) for parameter optimisation. In this thesis we follow the former approach. Our decision is based on strong published evidence about the successful implementation of gradient-based optimisation in finance (Boyacioglu and Avci, 2010; Chang et al., 2011; Chen, 2013; Kablan and Ng, 2011; Tan et al., 2011; Wei et al., 2014). This also provided us with a good benchmark approach to explore possible risk-adjusted performance improvements. Two popular gradient-based optimisation approaches in machine learning literature are the method of Least Squares (LS) and the Back-Propagation (steepest descent) (BP) algorithms. Both methods can be used separately, however they are not mutually exclusive. In Chapters 3 and 4 we make extensive use of the Adaptive Neuro-Fuzzy Inference System (ANFIS) (Jang,
which tunes model parameters by iteratively switching between the two algorithms. This results in a more efficient training algorithm.

**Least Squares Method**

In the LS approach, the MF shapes and parameters of the rules antecedent are fixed, however the consequent parameters are optimised using training data. We follow a model-free approach (Jang, 1993; Mendel et al., 2014) with the objective to completely specify the FLS using training data. The process starts from a given collection of \( N \) input-output data training pairs, \((x^{(1)} : y^{(1)}), (x^{(2)} : y^{(2)}), \ldots, (x^{(N)} : y^{(N)})\). Because the FLS is linear in the consequent parameters, this leads to a LS optimisation problem. In this case, the problem consists in optimising \( w_{i,0}, w_{i,1}, w_{i,2}, \ldots, w_{i,k} : k + 1 \) consequent parameters (see Equation (2.5)) for \( M \) rules, for a total of \( (k + 1)M \) parameters.

The algorithm in the LS optimisation approach starts by re-expressing Equation (2.7) as

\[
y_{TSK,1}(x) = g_{1,0}(x)w_{1,0} + g_{1,1}(x)w_{1,1} + g_{1,2}(x)w_{1,2} + \ldots + g_{1,k}(x)w_{1,k} + \ldots + g_{M,0}(x)w_{M,0} + g_{M,1}(x)w_{M,1} + g_{M,2}(x)w_{M,2} + \ldots + g_{M,k}(x)w_{M,k},
\]

(2.10)

where

\[
g_{i,j}(x) = \frac{f_i(x)x_j}{\sum_{i=1}^{M} f_i(x)},
\]

(2.11)

where \( i = 1, \ldots, M, j = 0, 1, \ldots, k \), and \( x_0 = 1 \). Equation (2.10) can be expressed in vector format as

\[
y_{TSK,1}(x) = g^T(x)w
\]

(2.12)
where \( g(x) \) and \( w \) are \((k+1)M \times 1\) vectors. If Equation (2.12) is used to represent all the \( N \) elements in the training set, we get

\[
y_{TSK,1} \left( x^{(t)} \right) = g^T \left( x^{(t)} \right) w \quad t = 1, \ldots, N
\]  

(2.13)

The \( N \) equations in the form of Equation (2.13) can be grouped together in a vector-matrix format as

\[
y_{TSK,1} = Gw
\]  

(2.14)

where \( G \) is an \( N \times (k+1)M \) matrix. The parameter vector \( w \) of the linear system of equations can be solved using the standard LS approach by minimising the cost function

\[
J(w) = \frac{1}{2} [y - Gw]^T [y - Gw]
\]  

(2.15)

where the solution can be expressed as

\[
[G^T G] w = G^T y.
\]  

(2.16)

**Back-Propagation Method**

In the second approach, both the antecedent and consequent parameters are optimised. Since the FLS antecedent parameters are not linear, in this approach optimisation can be solved using back-propagation. In addition to the consequent parameters, in this case the problem consists in optimising 2 parameters per antecedent Gaussian MF \((m_{A_{i,j}} \text{ and } \sigma_{A_{i,j}})\), \( k \) antecedents, and \( M \) rules, for a total of \( 2kM \) antecedent parameters. When combining both antecedent and consequent parameters, this results in a total of \( 3kM + M \) parameters.

To optimise both antecedent and consequent parameters, we use Gaussian MFs (defined in Equation (2.4)) and start from Equation (2.10) and Equation (2.11) but expand \( g_{i,j} \) in more
detail as

\[ g_{i,j}(x^{(t)}) = \frac{\exp \left( -\frac{1}{2} \sum_{c=1}^{k} \left( \frac{x^{(t)}_c - m_{A_{i,c}}}{\sigma_{A_{i,c}}} \right)^2 \right) x^{(t)}_j}{\sum_{i=1}^{M} \exp \left( -\frac{1}{2} \sum_{c=1}^{k} \left( \frac{c^{(t)}_j - m_{A_{i,c}}}{\sigma_{A_{i,c}}} \right)^2 \right)} \]  

(2.17)

where \( j = 0, 1, \ldots, k, i = 1, \ldots, M, t = 1, \ldots, N \) and \( x_0 = 1 \). Using our training set of \( N \) input-output data pairs, \((x^{(1)}: y^{(1)}), (x^{(2)}: y^{(2)}), \ldots, (x^{(N)}: y^{(N)})\), the objective is to minimise the sum of squares function

\[ J^{(t)} = \frac{1}{2} \left( y_{TSK,i} \left( x^{(t)} \right) - y^{(t)} \right)^2 = \frac{1}{2} e^2 \]  

(2.18)

The gradient descent algorithm involves a number of iterations. Using Equation (2.10) and Equation (2.17), on each epoch the objective function \( J \) is minimised in steps that are defined by the learning rate \( \alpha \) and the gradient of \( J \) with respect to each model parameter as follows (Mendel et al., 2014):

- Coefficients of the consequent function update:

\[ w_{i,j}(t+1) = w_{i,j}(t) - \alpha \frac{\partial J^{(t)}}{\partial w_{i,j}} \]  

(2.19)

which yields

\[ w_{i,j}(t+1) = w_{i,j}(t) - \alpha g_{i,j}(x^{(t)}) \]  

(2.20)

- Mean of the Gaussian antecedent MFs update:

\[ m_{i,j}(t+1) = m_{i,j}(t) - \alpha \frac{\partial J^{(t)}}{\partial m_{A_{i,j}}} \]  

(2.21)
which yields

\[ m_{A_{i,j}}(t+1) = m_{A_{i,j}}(t) - \alpha \epsilon \left( w_{i,0} + w_{i,1}x_1 + w_{i,2}x_2 + \cdots + w_{i,k}x_k - y_{TSS,1}\left(x^{(t)}\right) \right) \]
\[ \times \frac{\left(x_{j}^{(t)} - m_{A_{i,j}}(t)\right)}{\sigma_{A_{i,j}}^2(t)} \eta_{i,j}(x^{(t)}) \] (2.22)

- Variance of the Gaussian antecedent MFs update:

\[ \sigma_{A_{i,j}}(t+1) = \sigma_{A_{i,j}}(t) - \alpha \frac{\partial J(t)}{\partial \sigma_{A_{i,j}}} \] (2.23)

which yields

\[ \sigma_{A_{i,j}}(t+1) = \sigma_{A_{i,j}}(t) - \alpha \epsilon \left( w_{i,0} + w_{i,1}x_1 + w_{i,2}x_2 + \cdots + w_{i,k}x_k - y_{TSS,1}\left(x^{(t)}\right) \right) \]
\[ \times \frac{\left(x_{j}^{(t)} - m_{A_{i,j}}(t)\right)^2}{\sigma_{A_{i,j}}^4(t)} \eta_{i,j}(x^{(t)}) \] (2.24)

2.5 Type-2 fuzzy sets and fuzzy membership functions

Type-2 (T2) fuzzy sets were introduced by Zadeh (1975) and are a generalisation of T1 fuzzy sets. This gives the ability to T2 fuzzy sets to handle higher levels of uncertainty. T2 fuzzy sets are characterised by their 3D structure, where the third dimension represents the membership degree of its 2D space which is defined by the footprint of uncertainty (FOU). As shown in Figure 2.5 the FOU is defined by an Upper Membership Function (UMF) and a Lower Membership Function (LMF). The FOU represents a blurring effect on top of a T1 MF, with the size of the FOU representing an additional dimension to represent the degree of uncertainty. The FOU is a key parameter in T2 FLSs, and in a later chapter we specifically focus on tuning this parameter.
When comparing the T1 MF (Figure 2.2) and the T2 MF (Figure 2.5), it can be identified that in the case of T2 MF, an input acts like a vertical slice between the UMF and LMF. Hence unlike the T1 case where the input maps to a single MF value, in the case of T2 it translates into a serious of points which we call the primary membership. For general T2 MFs, each point on the primary MF maps to a secondary membership degree which is represented by another MF (Figure 2.6).

More formally, a T2 fuzzy set, denoted as \( \tilde{A} \), is defined as

\[
\tilde{A} = \{(x,u), \mu_{\tilde{A}}(x,u))|\forall x \in X, \forall u \in J_x \subseteq [0,1] \}, \tag{2.25}
\]

which shows that T2 sets depend on pairs of the two variables \( x \) and \( u \). A T2 MF is typically represented as

\[
\tilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{A}}(x,u)/(x,u) \quad J_x \subseteq [0,1] \tag{2.26}
\]
2.6 Extending type-1 TSK fuzzy logic systems to type-2

In the case of a T2 FLS (Figure 2.8), the high-level structure is very similar to a T1 FLS, except in the output processing, which consists of two components, namely the type-reducer and the defuzzifier. Although T1 rules can be extended to T2 in various forms (see Mendel et al., 2014), in this thesis we make use of rules in which the consequent consists of crisp...
numbers. In this form, the rules in a T2 TSK model take the following form

\[
\text{IF} \quad \left( x_1 \text{ is } \widetilde{A}_{i,1} \right) \text{ AND } \left( x_2 \text{ is } \widetilde{A}_{i,2} \right) \text{ AND } \cdots \text{ AND } \left( x_k \text{ is } \widetilde{A}_{i,k} \right)
\]

\[
\text{THEN} \quad y_i(x) = w_{i,0} + \sum_{j=1}^{k} w_{i,j}x_j
\]

(2.27)

where \( i = 1, \ldots, M \) is the rule number, \( w_{i,v} (v = 0, 1, \ldots, k) \) are the consequent parameters, \( y_i(x) \) is the output of the \( i \)th rule, and \( \widetilde{A}_{i,l} (l = 1, 2, \ldots, k) \) are T2 antecedent fuzzy sets.

The final output of the model is as follows (Mendel et al., 2014):

\[
Y_{TSK,2}(x) = \int_{f_1} \cdots \int_{f_M} \left( \sum_{i=1}^{M} f_i y_i \right) \frac{\sum_{l=1}^{M} \mu_{F_l}(f_i)}{\sum_{l=1}^{M} f_i}
\]

(2.28)
where $M$ represents the number of rules, $f_i \in F_i$ where $F_i$ represents the firing strength, and $\tau$ represents the $t$-norm. The integral sign represents the fuzzy union operation and the slash operator (/) associates the elements of the rules’ output and firing strength with their secondary membership grade.

The computational complexity derives from the enormous number of embedded sets that have to be individually processed in order to effect type-reduction and defuzzification. In this thesis we make use of IT2 MFs, resulting in A2-C0 models Mendel et al. (2014). In the IT2 case, the combined firing interval of the rules and the corresponding rule consequents (Equation (2.28)) are simplified as

$$Y_{A2-C0}(x) = [y_l, y_r] = \left[ \int_{f_1 \in [f_{l1}, f_{r1}]} \ldots \int_{f_a \in [f_{la}, f_{ra}]} \right]^{1/\sum_i f_i} \sum_i f_i y_i \quad (2.29)$$

where in this case the the association between the elements of the rules’ output and firing strength with their secondary membership grade is simplified to 1. The firing strength for
each rule $i$, where $i = 1, 2, ..., M$, is calculated as

$$f_i(x) = \mu_{A_{i,1}}(x_1) \cdot \mu_{A_{i,2}}(x_2) \cdot ... \cdot \mu_{A_{i,k}}(x_k)$$

(2.30)

$$\overline{f}_i(x) = \overline{\mu}_{A_{i,1}}(x_1) \cdot \overline{\mu}_{A_{i,2}}(x_2) \cdot ... \cdot \overline{\mu}_{A_{i,k}}(x_k)$$

(2.31)

where, like in the case of T1, $\cdot$ represents the product $t$-norm. In an IT2 FLS, $Y_{A2-C0}$ is an interval T1 set defined by $y_l$ and $y_r$ where

$$y_l = \min_{L \in [1,M-1]} \frac{\sum_{i=1}^{L} y_l^f + \sum_{i=L+1}^{M} y_i^f}{\sum_{i=1}^{L} f^i + \sum_{i=L+1}^{M} f^i}$$

(2.32)

$$y_r = \min_{R \in [1,M-1]} \frac{\sum_{i=1}^{R} y_l^\overline{f} + \sum_{i=R+1}^{M} y_i^\overline{f}}{\sum_{i=1}^{R} f^i + \sum_{i=R+1}^{M} \overline{f}^i}$$

(2.33)

It is evident from Equation (2.32) and Equation (2.33) that computing $y_l$ and $y_r$ using this method requires an iterative approach. An evaluation of the most common algorithms that follow this approach can be found in Greenfield and Chiclana (2013) and Wu and Mendel (2014). Once $y_l$ and $y_r$ are computed, the final defuzzified output can be calculated as follows:

$$y_{TSK,2}(x) = \frac{y_l + y_r}{2}.$$  

(2.34)

More recently, it was shown that closed-form approaches for calculating the output offer a good compromise between speed and complexity (Greenfield and Chiclana, 2013; Wu and Mendel, 2014). Closed-form approaches bypass the type-reduction step and the defuzzified output is calculated directly. In this thesis we make use of the Nie-Tan method (Nie and Tan, 2008). This method presents an efficient type-reduction method for interval type-2 fuzzy sets, which involves taking the mean of the lower and upper MFs of the interval set, thus
creating a type-1 fuzzy set. In this case the output computation is simplified as

\[ y_{TSK,2}(x) = \frac{\sum_{i=1}^{M} y_i (f_i + \bar{f}_i)}{\sum_{i=1}^{M} (f_i + \bar{f}_i)} \]  (2.35)

2.7 Conclusion

In this chapter we presented the mathematical foundations for the more sophisticated fuzzy logic techniques that we will be presenting in the following chapters. The next chapters are motivated by the essential idea behind fuzzy logic which supports a departure from the traditional quantitative techniques in favour of models which can model and minimise the effects of uncertainty. In our research, we use uncertainty as a proxy for risk and hence present innovative techniques for how to improve the risk-adjusted performance of trading algorithms.
Chapter 3

Improving the risk-adjusted performance of trading algorithms

The research problem that we investigate in this chapter addresses the concerning number of published studies which promote unsuitable approaches in model training and selection when applied for trading purposes. Applying a wrong approach poses a high possibility of poor out-of-sample trading performance or, possibly worse, the risk of falling far short of the extraordinary profit expectations when applied in real trading scenarios.

This chapter presents a coherent framework for the design and tuning of AI-controlled trading algorithms with the aim of improving the level and stability of risk-adjusted performance. We show the effectiveness of our approach using three milestone models from neuro-fuzzy systems literature, namely Artificial Neural Networks (ANN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and Dynamic Evolving Neuro-Fuzzy Inference Systems (DENFIS). Contrary to most former studies which focus on daily predictions, we seek model improvements in an intraday stock trading scenario using high-frequency data.

As our first contribution, we propose an approach to address a typical problem which is faced when using popular moving average signals which tend to be over/under reactive to price movements. This is further enhanced with a model validation methodology using
heat maps to analyse favourable profitability in specific holding time and signal regions. Secondly, we study the effect of realistic constraints (risks) such as transaction costs and intraday trading hours, which many existing approaches in the literature ignore. We show that ignoring trivial constraints introduces a risk of overestimated profitability. Finally, unlike most former studies that only aim to minimise statistical error measures, we seek improvements using financially more relevant risk-adjusted objective functions and extend this to a deeper analysis on model stability and time-varying performance profiles. We also propose an innovative approach which combines a set of risk-return objective functions using an ANFIS ensemble dynamic selection method, and show how it can improve the intraday trading performance of AI models.

3.1 Introduction

The profitability of technical trading rules is an incessant debate. By examining 30-minute prices, Schulmeister (2009) claimed that beyond the 1990s the profitability of technical trading rules has possibly moved to higher frequency prices as a result of faster algorithmic trading and more efficient markets. Kearns et al. (2010) presented opposing views, claiming that aggressive high-frequency trading (HFT) does not lead to the expected high excessive returns. In spite of this persisting contention in the high frequency space, surveys (Krollner et al., 2010; Tsai and Wang, 2009) showed that the majority of former studies still focus on daily price forecasts using lagged index data and technical trading indicators. Additionally, these surveys indicated that many former studies focus on the application of models solely to predict market movements (a more recent example in Son et al., 2012) and seem to disregard the fact that investors are more interested in risk-adjusted performance rather than just price predictions themselves (Choey and Weigend, 1997; Xufre Casqueiro and Rodrigues, 2006).

Tsai and Wang (2009) and Krollner et al. (2010) identified that Artificial Neural Networks (ANNs) are the dominant machine learning technique in AI-based financial applications.
On the downside, ANNs are regarded as black boxes that cannot describe the cause and effect. Moreover, hybrid models were again found to provide better forecasts compared to ANNs used alone or traditional time series models. Following the emergence of fuzzy logic (Zadeh, 1975), neural networks and fuzzy inference systems were brought together as general structures for approximating non-linear functions and dynamic processes. A popular cited technique in non-stationary and chaotic time series prediction is the Adaptive Neuro-Fuzzy Inference System (ANFIS) by Jang (1993). The successful application of ANFIS in trading applications by predicting stock price was demonstrated in Gradojevic (2007) and Kablan and Ng (2011) and many others. With a focus on the dynamic learning of rules from data, Kasabov and Song (2002) introduced pioneering work on evolving neuro-fuzzy systems with the introduction of the Dynamic Evolving Neuro-Fuzzy Inference System (DENFIS) and its application for time-series prediction. The proposition of evolving models is to keep systems continuously adapting, and hence evolving, and address recurring and changing patterns in the underlying environments. To our best knowledge DENFIS was not previously applied in a high-frequency setting.

Recently, Melin et al. (2012) and Lei and Wan (2012) identified that ANFIS ensembles (which we denote as eANFIS in this chapter) provided better generalisation and a reduction in the mean squared error when compared to conventional ANFIS and other chaotic time series models. Their ensemble integration approach was based on applying the average or weighted average methods across the regression predictions of all components of the ensemble (see also Soto et al. (2013) who extended this by applying a fuzzy integrator approach). Despite these claims, Faulina et al. (2012) showed that ensemble models involving more complex ANFIS combinations do not necessarily lead to improved performance. In this chapter, we explore potential trading performance improvements in a high-frequency trading setting with the application of ANFIS ensembles and the identification of integration methods driven by risk-return objectives.
This chapter presents innovative approaches with the intention to improve model tuning processes and stability for trading purposes. The ultimate aim is to improve the risk-adjusted performance of simple technical trading rules in an intraday stock trading scenario using high-frequency data with the application of artificial intelligence and soft computing techniques. In our experiments we use a set of stocks listed on the London Stock Exchange with a focus on a number of objectives:

1. We explore the debated profitability and an augmentation of moving average rules, particularly focusing on high-frequency data in an intraday trading scenario rather the more common day ahead predictions.

2. In contrast to common trading system designs that focus on fixed target returns, we investigate return bands in the region between 0.1% and 0.5% which act as a threshold for unprofitable small trades.

3. We evaluate the profitability of less aggressive HFT strategies, with a holding time of trading positions (PT) in the region between 10 minutes to 1 hour, in view of stated claims of unattainable high excessive returns from more aggressive HFT strategies.

4. We consider real-world intraday trading constraints like trading costs, realistic trading hours and no overnight positions, which are often ignored in existing studies dealing only with daily trading frequencies.

5. We investigate the risk-adjusted performance attained from three representative milestone models in neurocomputing, namely ANNs, ANFIS and DENFIS models, and also explore the effectiveness of the more sophisticated eANFIS architecture.

6. We analyse model stability by comparing the time series of risk-adjusted performance measures obtained using different model optimisation functions such as single risk-return functions, an innovative combination of different risk-return functions via an
ensemble, Root Mean Squared Error (RMSE), period return and models optimised without considering transaction costs.

Our first contribution is the simple, yet effective, extension of common technical trading strategies by considering a dynamic “portfolio” of moving average prediction models controlled by neuro-fuzzy systems. This is further extended by applying dynamic rules for return bands and trade position times. In line with Tsang (2009), our models try to answer questions of the following form: “Will the price go up (or down) by r% in the next t minutes?” An important challenge in this study is the choice of moving average window length. For example, if the price over an interval is, in general, trending up, there are also several short-term downtrends in the price data. Some of them are real trend reversal points and others are just noise. The trend identifying mechanism should not be overly sensitive to short-term fluctuations, hence applying a too short moving average would result in falsely reporting a break in trend. On the other hand, choosing a too long moving average will result in late reaction to price movement. We suggest a combination of multiple moving average rules as input to the prediction models. This is further enhanced by applying a model validation methodology using heat maps to analyse favourable risk-return regions that identify profitability in specific holding time and signal regions.

As our second major contribution we further extend our trading systems with decision rules accounting for transaction costs and trading hours. Krollner et al. (2010) found that although more than 80% of the published studies in their survey stated that their proposed framework surpassed the benchmark model, most of them ignored trivial constraints that are typically incurred in the real world, possibly introducing bias in the results (see also Álvarez Díaz, 2010). This introduces a risk of overestimated profitability when realistic constraints are applied.

When training and evaluating a trading system, most former studies only have a very limited view of what constitutes successful investment decisions, defining performance on
the grounds of forecast accuracy and win ratios, and often choose to minimise the forecast error of the price prediction, setting this as the objective function (Alves Portela Santos et al., 2007; de Faria et al., 2009; Enke and Thawornwong, 2005; Medeiros et al., 2006). However, small forecast errors do not guarantee trading profits (Brabazon and O’Neill, 2006). Krollner et al. (2010) also identify a lack of literature studying whether AI-based algorithms can enhance an investor’s risk-return trade-off. While a few attempts were reported recently, mainly for other financial applications (Dempster and Leemans, 2006; Esfahanipour and Mousavi, 2011; Evans et al., 2013), the implementation of risk-adjusted performance control has to our best knowledge not been studied in an intraday high-frequency setting before (see also a comprehensive survey by Bahrammirzaee (2010) and the references therein).

As our third contribution we explore how the high-frequency trading performance can be improved by analysing the time series of risk-adjusted performance measures. To assess performance, we also compare our findings against results obtained from other model optimisation functions such as Root Mean Squared Error (RMSE) and models not considering transaction costs. Recent lessons, such as the flash crash in 2010, suggest that not all risks are worth taking in intraday trading. Our chapter aims to shed more light on the risk-return profile for selected stocks traded at the London Stock Exchange and explains how this information can be capitalised. We also consider an innovative combination of risk-return functions using an ANFIS ensemble dynamic selection method, and show how it can improve the intraday trading performance of AI models. Finally, in contrast to common approaches in the literature which evaluate models using performance measures at an arbitrary single point in time (e.g. only at the end of the sample period), our goal is to provide a deeper understanding of the time-varying performance profile of the applied models.

The remainder of the chapter is structured as follows. In Section 3.2 we first introduce the moving average signals and explain how these can be combined to model stock returns. We then discuss our experiment and describe model components and underlying prediction
3.2 Method

A central theme in the technical trading approach is the ability to recognise patterns in market prices that supposedly repeat themselves and hence can be used for predictive purposes. A number of authors showed the predictive capabilities of simple trading rules in conjunction with the application of ANNs. For a survey, e.g. see Vanstone and Finnie (2009) and Vanstone and Finnie (2010), and the references therein. This body of research showed the predictive ability of simple trading rules on daily returns with the application of ANNs and contrasted the weaknesses with traditional econometric models which fail to give satisfactory forecasts for some series because of their linear structure and some other inherent limitations such as the underlying distribution assumptions.

Our experiment setup consists of two core modules (see Figure 3.1). Sections 3.2.1 to 3.2.5 describe our return prediction models. Section 3.2.6 explains our trading algorithm. In Section 3.2.7 we explain how we measure and evaluate model performance.
3.2.1 Technical trading and moving averages

Traders typically employ two classes of tools to decide what stocks to buy and sell: fundamental and technical analysis, both of which aim at analysing and predicting shifts in supply and demand and hence determining the direction that prices are likely to move. While fundamental analysis involves the study of company fundamentals, such as revenues and expenses, market position, annual growth rates, and so on, technical analysis is solely concerned with price and volume data, particularly price patterns and volume spikes. Consider a set of \( n \) historical prices \( \{p_t, p_{t-1}, \ldots, p_{t-n+1}\} \in \mathbb{R}^n_+ \). Similarly to technical analysis, we aim to find a function \( d : I_t \rightarrow \Omega \) that maps the information set \( I_t \) at time \( t \) to a set of trading decisions \( \Omega = \{\text{short}, 0, \text{long}\} \), indicating short, neutral or long positions, respectively.

It is known from finance and economics literature that intraday prices can be very volatile over the course of the trading day (McAleer and Medeiros, 2008) due to market microstructure effects and trading behaviour (such as a change of trading sessions, lunch breaks, time zone effects, etc.). Another often reported source of (perceived) volatility is the bid-ask bounce (e.g. Roll, 1984), where the price appears to heavily fluctuate due to the random arrival of buyers and sellers. To minimise the biasing effect of this “technically” introduced noise and its propagation in our modelling approach, we apply a moving average filter to obtain a smoothed signal. This is a popular approach in technical analysis as it represents a simple yet effective means to account for the stochastic nature of the trading process. For a survey on the application of moving averages in trading, see Krollner et al. (2010) and Ahmed et al. (2010). In contrast to the common framework, however, we will augment the moving average model by jointly considering different lag structures to account for market microstructure effects arising from different intraday trading horizons.

Essentially, a moving average represents a low pass filter which removes higher frequency “noise”, allowing the investor to more clearly identify the lower frequency trend. A typical
moving average, $MA$, is calculated as:

$$MA^n_t = \frac{1}{n} \sum_{i=0}^{n-1} p_{t-i},$$

(3.1)

where $i = 0, 1, 2, \ldots, n - 1$ represents the “memory span” of the rule. One popular application of this rule is to trigger a buy signal if the price goes beyond its moving average, and a sell signal if it falls below it. Consider the signal at time $t$ defined as

$$s_t = \begin{cases} 
  \text{long} & \text{if } p_t \geq (1 + \phi)MA^n_t \\
  0 & \text{if } (1 - \phi)MA^n_t \leq p_t < (1 + \phi)MA^n_t \\
  \text{short} & \text{if } p_t < (1 - \phi)MA^n_t
\end{cases}$$

(3.2)

where $\phi$ is the bandwidth of the rule for whiplash reduction. Another popular variation of the rule is to consider moving averages of different lengths instead, i.e. buy if the short moving average is above the long moving average, and sell if otherwise:

$$s_{t,n_1,n_2} = \begin{cases} 
  (MA^{n_1}_t - MA^{n_2}_t) & \text{if } |MA^{n_1}_t - MA^{n_2}_t| > \phi \\
  0 & \text{else}
\end{cases}$$

(3.3)

where $MA^{n_1}$ and $MA^{n_2}$ are the short and long moving averages, respectively.

We investigate whether intraday high-frequency returns can be predicted by making use of buy and sell signals as inputs and use this information to build a profitable trading algorithm. For our time series we use 5-minute continuously compounded returns as they have much better statistical properties than price levels. These intraday returns are defined as:

$$y_t = \log(p_t) - \log(p_{t-1}),$$

(3.4)

where $\log(\cdot)$ denotes the natural logarithm.
Recently, Gradojevic and Gençay (2013) have shown the effectiveness of combining fuzzy logic with moving average signals over conventional moving average filters due to their non-zero phase shift nature. Gençay et al. (2002) suggested that since a short moving average has a smaller phase shift than a long moving average, it would also indicate a turning point earlier than a long moving average. Although the authors argue that both filters would still indicate a turning point after the event due to their inherent nonzero phase shift nature, their combination is convenient since the trend signals from longer moving averages further confirm signals indicated by the shorter moving averages. In our experiments we propose a model that predicts the next 5-minute stock return by taking a combination of moving average rules as input variables. The three moving average signals utilised are \((s^{n_1,n_2}) = \{s^{1.5}, s^{5,10}, s^{10,15}\}\), where \(n_1\) and \(n_2\) are in 5-minute time bars.

By combining these \(k\) input signals, the linear specification of the return \(y_t\) prediction model is defined as:

\[
y_t = \theta_0 + \sum_{k=1}^{3} \theta_k s_{k,t-1} + \epsilon_t \tag{3.5}
\]

with the error term \(\epsilon_t \sim N(0, \rho)\) and

\[
s_{k,t} = \begin{cases} 
MA_t^{1.5} & \text{for } k = 1 \\
MA_t^{5,10} & \text{for } k = 2 \\
MA_t^{10,15} & \text{for } k = 3 
\end{cases} \tag{3.6}
\]

Kearns et al. (2010) noted that when taking transaction costs into account, aggressive HFT strategies considering holding periods between 10 milliseconds and 10 seconds can have surprisingly modest profitability. For this reason, we investigate the effect of (a) longer holding periods ranging from 10 minutes to 1 hour, and (b) the application of a return band ranging from 0.1% to 0.5%. This approach is more versatile when compared to common approaches in the literature that calibrate their trading systems based only on an arbitrary

In the following sections, we describe how the feed-forward network (FFN), ANFIS, DENFIS and eANFIS models are adapted for our experiments. The process starts from a given collection of \( N \) input-output data training pairs, \((x^{(1)} : y^{(1)}), (x^{(2)} : y^{(2)}), \ldots, (x^{(N)} : y^{(N)})\) where

\[
\begin{align*}
\mathbf{x}^{(1)} &= [s_{1,t-N}, s_{2,t-N}, s_{3,t-N}], & y^{(1)}_{t-N+1} \\
\mathbf{x}^{(2)} &= [s_{1,t-N+1}, s_{2,t-N+1}, s_{3,t-N+1}], & y^{(2)}_{t-N+2} \\
& \vdots \\
\mathbf{x}^{(N)} &= [s_{1,t-1}, s_{2,t-1}, s_{3,t-1}], & y^{(N)}_t.
\end{align*}
\tag{3.7}
\]

In Equation (3.7), for each data instance at a specific time \( t \), \( \mathbf{x} \) is a vector consisting of \( \{x_1, x_2, x_3\} \) input elements which represent the \( \{s_{1,t-1}, s_{2,t-1}, s_{3,t-1}\} \) technical indicator signals (Equation (3.6)), and \( y \) represents the mean return over the next 5 minutes (Equation (3.4)).

### 3.2.2 Neural network model

Hudson et al. (1996), Gençay (1996) and Fernandez-Rodrıguez et al. (2000) showed that, under general regularity conditions, a sufficiently complex single hidden-layer feed-forward network can approximate any member of a class of functions. For a more recent survey on the applications of ANNs to model moving averages, see Vanstone and Finnie (2009) and Ahmed et al. (2010), and the references therein. Following Equation (3.5), we design a single-layer feed-forward network (FFN) regression model with 3 lagged buy and sell signals
and with \( d \) hidden units. Each hidden neuron, \( z_j \), can be represented mathematically as

\[
    z_j = f \left( w_{1,0} + \sum_{k=1}^{3} w_{1,k} x_k \right),
\]

(3.8)

where \( w_{1,0} \) represents the bias parameter of a hidden neuron and \( w_{1,k} \) represents the weight parameter between the \( k \)-th input and hidden neuron \( z_j \). The output neuron can then be represented as

\[
    y_t = f \left( w_{2,0} + \sum_{j=1}^{d} w_{2,j} z_j \right),
\]

(3.9)

where \( w_{2,0} \) represents the bias parameter of the output neuron, \( w_{2,j} \) represents the weight parameter between the \( j \)-th hidden neuron and output neuron, and

\[
    f(u) = \frac{1}{1 + \exp(-u)},
\]

(3.10)

where \( f(u) \) is the activation function in our application (see also Gençay, 1996). For our model identification, a number of model configuration parameters are considered during the in-sample training (see Table 3.1). We test and compare all \( 2 \times 3 \times 1 = 6 \) permutations of the parameter combinations in our sensitivity analysis of FFN models.

### 3.2.3 ANFIS model

Neuro-fuzzy techniques synergise ANNs with fuzzy logic techniques by combining the human-like reasoning style of fuzzy systems with the learning and connectionist structure of
neural networks. Algorithms for the acquisition or tuning of fuzzy models from data typically focus on one or all the following aspects: (i) rule consequent parameter optimisation, (ii) membership function parameter optimisation and (iii) rule induction. The tested systems will be taking input from a number of moving average rules, $x$, and predict the stock return, $y$, as defined in Equation (3.7).

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a popular technique suggested by Jang (1993). ANFIS is a T1 TSK model, and although the rules and mathematical underpinnings follow the traditional T1 TSK models (explained in Section 2.3), the model structure is formulated in a connected layered approach to permit ANN learning techniques. Using the standard ANFIS model and Equation (3.5), we apply the ANFIS architecture layers, denoting the output of the $i$-th node in layer $l$ as $O_{l,i}$, as follows:

**Layer 1** Since we have 3 inputs, this layer contains $3 \times \alpha$ adaptive nodes, where $\alpha$ represents the number of MFs used to defined each input space. The output of each node $O_{1,i}$ is the membership grade for the input moving average signals $\{x_1, x_2, x_3\}$. For instance, the nodes with connections from the first input $x_1$ are in the form

$$O_{1,i} = \mu_{A_{1,k}}(x_1) \text{ for } k = 1, 2, ..., \alpha,$$

(3.11)

where $A_{1,k}$ represents the degree of membership of $x_1$ to the $k$-th MF. Different shapes and numbers of input MFs are tested in our model calibration process (see Table 3.2).

**Layer 2** Every node in this layer is fixed and represents the firing strength of each rule. Using grid-based partitioning, this layer enumerates all possible combinations of membership functions of all inputs, in the case of our experiment resulting in $\alpha^3$ nodes. Each node $O_{2,i}$ in this layer calculates the product $t$-norm to “AND” the membership grades,

$$O_{2,i} = f_i = \mu_{A_{i,1}}(x_1) \ast \mu_{A_{i,2}}(x_2) \ast \mu_{A_{i,3}}(x_3),$$

(3.12)
where $\mu_{A_{i,k}}$ represents the degree of membership calculation for the $k$-th input coming from the previous layer.

**Layer 3** This layer contains fixed nodes which calculate the normalised firing strengths of the rules:

$$O_{3,i} = \hat{f}_i = \frac{f_i}{\sum_i f_i}.$$  \hspace{1cm} (3.13)

**Layer 4** The nodes in this layer are adaptive and perform the consequent of the rules:

$$O_{4,i} = \hat{f}_i y_i = \hat{f}_i \left( w_{i,0} + \sum_{k=1}^{3} w_{i,k} x_k \right),$$  \hspace{1cm} (3.14)

where $\hat{f}_i$ is the normalised firing strength from the previous layer and $y_i$ is the rule consequent linear function for the $i$-th rule, $i = 1, \ldots, \alpha^2$, with parameters $w_{i,0}$ and $w_{i,k}$.

**Layer 5** This layer consists of a single node that computes the overall output. The nodes in this layer are adaptive and perform the consequent of the rules:

$$O_{5,i} = r_i = \sum_i \hat{f}_i y_i = \frac{\sum_i f_i y_i}{\sum_i f_i}.$$  \hspace{1cm} (3.15)

Jang (1993) proposed premise and consequent parameters learning using a combination of gradient descent (GD) and least squares estimation (LSE) (see Section 2.4.2 for a mathematical explanation). The total parameter set is split into two sets, a Set$_1$ of premise (nonlinear) parameters (in Layer 1) and a Set$_2$ of consequent (linear) parameters (in Layer 4).

ANFIS learning uses a two-pass algorithm. In a forward pass Set$_1$ is unmodified and Set$_2$ is computed using the LSE algorithm. This is followed by a backward pass where Set$_2$ is unmodified and Set$_1$ is computed with a GD algorithm (e.g. back-propagation).
Although the application of ANFIS in finance has been widely studied (Boyacioglu and Avci, 2010; Chang et al., 2011; Chen, 2013; Kablan and Ng, 2011; Tan et al., 2011; Wei et al., 2014), most studies only employ daily data, whereas applications to intraday HFT are still scarce. Kablan and Ng (2011) successfully applied ANFIS to predict price movements from intraday tick data sampled at high-frequency. Due to the intraday seasonality of volatility, they applied a volatility filter based on a directional changes threshold to filter out training data from the specific time-bins that do not exceed the specific activity threshold. Iteratively choosing the right number of epochs was also identified as an important step to avoid over-fitting. In their experiment, Kablan and Ng (2011) had the actual membership functions pre-defined and, consequently, the number of rules were fixed, hence limiting model adaptation to membership function and consequent parameter tuning. This raises the question of whether fuzzy logic models could be further improved for trading purposes by automatic updates in terms of rule base, membership function parameters and consequent parameters in view of new data. A number of model calibration parameters are explored for the in-sample training (Table 3.2). We test and compare all $2 \times 2 \times 2 \times 3 = 24$ permutations of the parameter combinations in our sensitivity analysis for ANFIS models.

### 3.2.4 DENFIS model

Kasabov and Song (2002) introduced a new type of TS fuzzy inference systems, denoted as DENFIS, for adaptive on-line and off-line dynamic time series prediction. Research on evolving intelligent systems (refer to Kasabov and Filev, 2006) stems from the fact that

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter Value Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Data Size (days)</td>
<td>{5, 10}</td>
</tr>
<tr>
<td>Input Membership Functions Shape</td>
<td>{Gaussian, Generalised Bell}</td>
</tr>
<tr>
<td>Number of Input Membership Functions</td>
<td>{2, 3}</td>
</tr>
<tr>
<td>Training Epochs</td>
<td>{10, 20, 50}</td>
</tr>
</tbody>
</table>
data streams are often non-stationary or influenced by various conditions and modes of the corresponding environment, hence motivating models that are structurally dynamic, that is, they evolve. In relation to the models described earlier in this chapter, evolving systems can also represent an augmentation of the on-line identification of fuzzy rule-based models and on-line learning neural networks with flexible structures.

Core to DENVIS is a so-called evolving, on-line, maximum distance-based clustering method, called the Evolving Clustering Method (ECM) (for a detailed algorithm description refer to Kasabov and Song, 2002). The rule base is continuously updated and new fuzzy rules are created with new cluster identification (see Section 2.4.1 for a description of how clusters are projected on the input space to generate rules). In our study, the distance between the input examples is calculated by using the normalised Euclidean distance of a new sample to the cluster centres

\[ dist_i = ||x_i - C_i||, \tag{3.16} \]

where \( x_i \) is the current example consisting of a moving average vector \( \{x_1, x_2, x_3\} \) as defined in Equation (3.7) and \( C_i \) is the centre of the \( i \)-th cluster. A threshold value \( D_{thr} \) that limits the cluster size is identified during our in-sample training.

The inference in DENVIS considers \( M \) fuzzy rules of the form:

\[
IF \quad (x_1 \text{ is } A_{i,1}) \text{ AND } (x_2 \text{ is } A_{i,2}) \text{ AND } (x_3 \text{ is } A_{i,3})
\]

\[
THEN \quad y_t = f_i(x),
\]

where \( i = 1, 2, \ldots, M \) and \( j = 1, 2, 3; (x_j \text{ is } A_{i,j}) \) are \( M \times 3 \) fuzzy propositions as \( M \) antecedents for \( M \) fuzzy rules respectively; and \( A_{i,j} \) are fuzzy sets defined by their fuzzy membership functions \( \mu_{A_{i,j}} : x_j \rightarrow [0, 1] \). In the consequent part, linear functions \( f_i \) with \( i = 1, 2, \ldots, M \) are employed.
3.2 Method

Using a similar principle to that employed by the clustering rule projection example in Section 2.4.1, here all fuzzy membership functions are triangular type functions with three parameters \( \{a, b, c\} \) which represent the left, top, right triangle edges and the membership degree is calculated as

\[
\mu(x, a, b, c) = \max \left( \min \left( \frac{x - a}{b - a}, \frac{c - x}{c - b} \right), 0 \right),
\]

(3.17)

where \( b \) is the value of the cluster centre on the \( x \) dimension, \( a = b - d \times Dthr \), and \( c = b + d \times Dthr, d \in [1.2; 2] \). For a given input vector \( x \) the result of the inference, which is the predicted return \( y_t \), is calculated as the weighted average of each rule’s output using

\[
y_t = \frac{\sum_{i=1}^{M} w_i f_i(x)}{\sum_{i=1}^{M} w_i},
\]

(3.18)

where the rule firing strength \( w_i = \mu_{A_{1,1}}(x_1) \times \mu_{A_{1,2}}(x_2) \times \mu_{A_{1,3}}(x_3) \). Following a similar online learning approach to that presented in Takagi and Sugeno (1985) and Jang (1993), the linear functions in the consequent parts of the rules are updated using a recursively weighted LSE, applying also a forgetting factor. For each example, the weights are defined as the distance between the example and the corresponding cluster centre. Kasabov and Song (2002) demonstrated that DENFIS can effectively capture complex time-varying sequences in an adaptive manner and outperform some well-known existing neuro-fuzzy models (including ANFIS). To our best knowledge, DENFIS has not been applied to high-frequency trading yet. In our in-sample training we again consider a number of different model calibration parameters (see Table 3.3). We test and compare all \( 3 \times 5 \times 3 = 45 \) permutations of the parameter combinations in our sensitivity analysis of DENFIS models.
3.2.5 ANFIS ensemble

Melin et al. (2012), Lei and Wan (2012) and Soto et al. (2013) have recently shown the superiority of the ANFIS ensemble over singular ANFIS for chaotic time series prediction. Since their research was focused on evaluating different integration methods of the underlying ensemble regressors and on reducing statistical error, this still left an open question on whether such an architecture can be used to improve risk-adjusted performance when applied to an intraday stock trading scenario. This lead to our hypothesis that rather than optimising a single risk-adjusted objective function, better overall performance can be attained by combining multiple objective functions in an ensemble architecture.

Following these recently claimed successful results, in our experiment we adapted this architecture by combining it with the dynamic selection approach originally studied by Puuronen et al. (1999) for classification applications and later adapted for regression as in Rooney et al. (2004). The latter applied the dynamic selection method using a localised selection based on the lowest cumulative error for the nearest neighbours to the test instance. For our ensemble (eANFIS) we decided to combine the 3 ANFIS models (see Figure 3.2) that were obtained for each stock with respect to the maximisation of the Sharpe ratio, Sortino ratio and period return (details about these measures are provided in the next section). Effectively, these measures represent different degrees of risk aversion where the Sharpe ratio penalises high variance on both wins and losses, the Sortino ratio penalises high variance on losses only, whilst period return does not take into consideration any risk. Hence creating an ensemble which dynamically selects between these base models permits the final ensemble

Table 3.3: Parameters tested for DENFIS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter Value Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Data Size (days)</td>
<td>{10, 20, 30}</td>
</tr>
<tr>
<td>ECM Clustering Threshold</td>
<td>{0.04, 0.06, 0.08, 0.1, 0.12}</td>
</tr>
<tr>
<td>Number of Rules in Dynamic FIS</td>
<td>{3, 5 ,6}</td>
</tr>
</tbody>
</table>
3.2 Method

Fig. 3.2: Our proposed eANFIS architecture

model to switch to different degrees of risk on an intraday level based on the results attained from the most recent simulated trades. The Sharpe ratio is used to evaluate the underlying ensemble models. Hence if, for example, during a specific time slot a Sortino ratio-optimised ANFIS model generated higher returns for the same level of returns variance than the one optimising Sharpe ratio, it would be the selected model. At the same time the application of Sharpe ratio as the deciding measure used to shift between base models with various degrees of risk ensures that this is done in a “safe” manner by switching to riskier models when a specific time slot is identified, resulting in less abrupt movements or larger losses.

A memory component is introduced in our proposed ensemble architecture to keep track of the most recent trades for each base model. For our memory capacity, we define a new parameter, \( h \), which is a measure defining the history length (in tick time) of closed trades that the model should store. Only newly suggested trades from the currently selected base model move to the actual trade book. However, newly suggested trades from non-selected
models are still simulated based on the new arriving prices and are then passed to the memory unit. For each stock, a memory capacity value between 2 and 12 past trades was tested in the 100-day in-sample period. The value resulting in the best Sharpe Ratio performance was then applied in our out-of-sample period. Although Tsymbal et al. (2008) warn about the possible issues of using a fixed window approach, in our case this was addressed by the fact that the underlying base ANFIS models with different parameters, including the return band and position holding time parameters, suggest trades with a different rate intensity and hence time slots. Thus although the memory capacity parameter $h$ is fixed, since this is reflected in tick time (keeping the last $h$ trades) rather than calendar time (keeping the trades closed in a defined number of minutes), the actual time window covered by $h$ trades used in our dynamic selection method represents different overlapping time periods.

### 3.2.6 Trading algorithm

The second module of our trading system (see Figure 3.1) consists of a trading and money management algorithm that takes the return prediction for the next 5 minutes from the first module and executes trades based on specific rules (see also Tan et al., 2011; Vanstone and Finnie, 2009, 2010). In Chapter 1 we explained how an HFT algorithm has to automate a number of decisions: (i) what to buy or sell (markets), (ii) how much to buy or sell (position sizing), (iii) when to buy or sell (entries), (iv) when to go out of a losing position (stops), (v) when to go out of a winning position (exits), and (vi) how to buy or sell (tactics). Our focus in this chapter is particularly on decisions (iii) to (v).

The objective of our trading algorithm is to generate *buy*, *sell* or *do-nothing* signals. For *buy* or *sell* signals the output signal has to be greater (smaller) than the upper (lower) limits of a specific return threshold ($RT$), otherwise the trade signal is set to *do-nothing*. This was introduced in order to filter whiplash effects when the short and long moving averages are close and also limit the number of small trades which even if profitable would result in a loss
due to transaction costs. In contrast to the common approach in the literature focusing on a single target return, we consider different $RT$'s between 0.1% and 0.5% for each stock to search for the optimal return during the in-sample period of 100 trading days. Based on the selected band size, the position taken at time $t$ is:

$$
\text{position}_t = \begin{cases} 
\text{long} & \text{if } y_t > \text{return band} \\
\text{short} & \text{if } y_t < \text{return band} \\
0 & \text{otherwise.}
\end{cases}
$$

In our experiments we train the models on a daily rolling window basis. In this case, at $day_d$, where $(d = 1, 2, \ldots, 100)$, the model is trained on 5-minute data points (Equation (5.3)), from $day_{d-r}$ to $day_{d-1}$, and $r$ represents the training data size in days. The trained model is then used to predict, every 5 minutes, the return over the next 5 minutes (Equation (3.4)) during $day_d$. This is repeated for the whole 100-day in-sample period, for each parameter combination. The best performing model against each measure is finally tested on the following 100-day period out-of-sample on a moving window approach (see the next section on evaluation). In this setup, we apply a constant transaction cost of 10 GBP per trade, per direction, and assume that a trader is willing to invest a fixed 50,000 GBP per position. Every five minutes the trading algorithm takes a decision based on the predicted trading direction, the selected return band and the position holding time. If the signal is to go long (short), the system will buy (sell) 50,000 GBP worth of stock at the current market price. A total of five open positions are allowed at one point in time, limiting the total investment to 250,000 GBP. For this experiment, only positions in the same direction are allowed at the same time. This was done to specifically eliminate the hedging effect of opposing positions which, as a result, can bias the performance measure of the algorithm.

For a trade to be profitable, we defined each position to be held long enough for favourable price movement sufficient to overcome the trading costs. Different trade durations ($TD$)
Fig. 3.3: Adopted moving window approach. The first 100 days, each day consisting of 5-minute price points, is reserved for the in-sample training and model selection process. The models are trained every day using 5-minute data points from the previous \( r \) days and then tested on the following day’s prices. For out-of-sample testing, the same approach is applied and the selected model is moved forward, day-by-day, for the next 100 days.

between 10 minutes to 1 hour are considered in the model selection process for each stock during the in-sample period. If after, a specific \( TD \), the signal is still in the same direction, then the position is kept for another period of the same length. If, on the contrary, the signal has changed, then the position is closed (see Algorithm 1).

It is known that prices tend to be most volatile during the first hour of trading, due to the opening auction and the queuing and backlog of pre-opening orders in the order book. To overcome this period of volatility, we do not trade in the first hour of each day. Furthermore, all open positions are closed at end of day, resulting in the system not holding any positions overnight. Since we are interested in active intraday trading algorithms, models
with parameter combinations that generated less than 100 trades over the 100-day in-sample period are excluded from the experiment.

\textbf{Algorithm 1} Pseudo code of the trading algorithm, where $RT$ is the predicted return threshold and $TD$ is the trade duration.

\begin{verbatim}
\Phi \leftarrow 0

\textbf{if} signal > RT and signal > prevsignal and balance-tradesize > 0 \textbf{then}
    \Phi \leftarrow \text{long}
\textbf{end if}

\textbf{if} signal < \text{\text{-}1} \times RT and signal < prevsignal and balance-tradesize > 0 \textbf{then}
    \Phi \leftarrow \text{short}
\textbf{end if}

OPENTRADE(\Phi)

\textbf{for} each open trade \( t \) \textbf{do}
    tradeDuration(\( t \)) \leftarrow \text{tradeDuration}(\( t \)) + 1
    \textbf{if} tradeDirection(\( t \)) == \Phi \textbf{then}
        tradeDuration(\( t \)) \leftarrow 0
    \textbf{end if}

    \textbf{if} tradeDuration(\( t \)) \geq TD \textbf{then}
        CLOSETRADE(\( t \))
    \textbf{end if}
\textbf{end for}
\end{verbatim}

3.2.7 Evaluation

As identified in Chapter 1, although many algorithms minimise errors such as the mean squared error (MSE) or the root mean squared error (RMSE) (Alves Portela Santos et al., 2007; de Faria et al., 2009), trading systems optimised with respect to these criteria are not guaranteed to excel as the costs of prediction errors are assumed to be symmetric. Moreover,
Table 3.4: Models applied in the experiments

<table>
<thead>
<tr>
<th>Experiment</th>
<th>AI Algorithms Tested</th>
<th>MA Model</th>
<th>Optimisation Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>ANFIS, DENFIS, FFN</td>
<td>Dynamic</td>
<td>Sharpe Ratio</td>
</tr>
<tr>
<td>1b</td>
<td>ANFIS, DENFIS, FFN</td>
<td>Dynamic</td>
<td>Sortino Ratio</td>
</tr>
<tr>
<td>1c</td>
<td>ANFIS Ensemble (eANFIS)</td>
<td>Dynamic</td>
<td>Sharpe Ratio, Sortino Ratio, Period Return</td>
</tr>
<tr>
<td>2a</td>
<td>ANFIS, DENFIS, FFN</td>
<td>Dynamic</td>
<td>Sharpe Ratio, No Cost</td>
</tr>
<tr>
<td>2b</td>
<td>ANFIS, DENFIS, FFN</td>
<td>Dynamic</td>
<td>RMSE</td>
</tr>
<tr>
<td>2c</td>
<td>ANFIS, DENFIS, FFN</td>
<td>Dynamic</td>
<td>Period Return</td>
</tr>
<tr>
<td>2d</td>
<td>-</td>
<td>MA(1,5)</td>
<td>-</td>
</tr>
<tr>
<td>2d</td>
<td>-</td>
<td>MA(5,10)</td>
<td>-</td>
</tr>
<tr>
<td>2d</td>
<td>-</td>
<td>MA(10,15)</td>
<td>-</td>
</tr>
</tbody>
</table>

existing approaches in the literature do not account for realistic transaction costs and trading hours, reporting possibly biased results. Based on these findings, we construct a number of trading models by applying different AI algorithms and optimisation functions (see Table 3.4).

In our first two experiments (1a and 1b) we combine the dynamic moving average model with the different AI methods and choose either the Sharpe ratio or the Sortino ratio (both defined below) as optimisation criteria. We then extend our risk-adjusted performance improvement ideas (experiment 1c) by investigating the effect of combining multiple rather than single objective function optimisation using an ensemble architecture approach. In our second part of the experiments, we apply the same models, but either (2a) do not account for transaction costs in the training period or (2b) optimise the system entirely on forecast accuracy or (2c) optimise the system to maximise return without accounting for risk, in order to see whether and how the ignorance of these constraints would have an impact on the trading performance. Finally (2d), to assess the effectiveness of the dynamic moving average model, we also compare our models against the trading performance of standard fixed moving average models (Krollner et al., 2010).
To evaluate the trading system and compare the performance across different models, we apply five different measures (Kablan and Ng, 2011): Sharpe ratio, Sortino ratio, cumulative return, profit ratio, and win ratio. The Sharpe ratio is defined as

\[
    \text{Sharpe Ratio} = \frac{R - r_f}{\sigma}, \tag{3.20}
\]

where \( R \) denotes the expected return, \( r_f \) the risk-free interest rate and \( \sigma \) the volatility of the return. The Sharpe ratio measures the risk premium per unit of risk in an investment. Investments with higher Sharpe ratios are often preferred due to their higher risk-adjusted performance. The Sortino ratio is defined as

\[
    \text{Sortino Ratio} = \frac{R - r_f}{\sigma_{neg}}, \tag{3.21}
\]

where \( \sigma_{neg} \) refers to the standard deviation of negative asset returns. The Sortino ratio measures the risk premium per unit of downside risk in an investment. The cumulative return indicates the overall probability of the strategy since the first trade, similar to a buy-and-hold scenario. The win ratio is the ratio between the number of winning trades and losing trades and is defined as

\[
    \text{Win Ratio} = \frac{\text{Total Number of winning trades}}{\text{Total Number of losing trades}}. \tag{3.22}
\]

The profit ratio indicates a system’s ability to generate profits over losses and is defined as

\[
    \text{Profit Ratio} = \frac{\text{Total Gain/Number of winning trades}}{\text{Total Loss/Number of losing trades}}. \tag{3.23}
\]

It must be noted that although the profit and win ratios give an indication of the system’s performance, it does not, however, take into consideration the underlying risk (a single loss of $100 cannot compensate 99 winning trades of $1). These ratios are however included in
our evaluation to assess whether the models showing higher risk reflect higher returns or a higher number of wins.

Although many researchers claim positive results for their algorithmic trading models by analysing a set of performance measures at a single point in time, our interest is to validate our models by looking at cumulated risk-return measures on a day-by-day basis. This method provides a clearer analysis of the models’ behaviour and performance pattern over time. A range of model parameter combinations are tested (see Tables 3.1 to 3.3). In order to select a model with good generalisation capabilities, each model is trained and tested using a rolling window approach with a window width of 100 days and a step size of 1 day.

Furthermore, we also conduct a sensitivity analysis of the different models in order to investigate the uncertainty in the predicted output (see also Resta, 2009). The generation and visual inspection of heat maps was conducted in our model validation step. The objective of our sensitivity analysis was to perturb each model with small alterations in its parameters and check the possibility of any resulting large changes in the model’s performance. Any such major changes could be identified as spurious results, hence well-defined performance regions across our parameter space are desirable. A parameter sweep application was implemented in which the same code is run multiple times using unique sets of input parameter values (see Algorithm 2). This included varying one parameter at each step over the range of our parameter multi-dimensional space. Each individual run is executed independently of all other runs over the 100-day in-sample period applying a day-by-day moving window

Algorithm 2 Pseudo code for heat map data generation

\[
P_1 \leftarrow \text{set of vectors storing unique model parameter value combinations} \\
P_2 \leftarrow \text{set of vectors storing unique RB and PT parameter value combinations} \\
\text{for each parameter combination vector } p_1, p_1 \in P_1 \text{ do} \\
\quad \text{for each parameter combination vector } p_2, p_2 \in P_2 \text{ do} \\
\quad \quad \text{trades} \leftarrow \text{prediction/trading algorithm } (p_1, p_2) \\
\quad \quad \text{resultsGrid}_{RB \times PT} \leftarrow \text{sharpe ratio } (trades) \\
\quad \text{end for} \\
\text{end for}
\]
3.3 Empirical data and analysis

The trading systems in this experiment are developed using high-frequency trade data for a set of stocks listed on the London Stock Exchange (see Table 3.5) during the period 01/06/2007 to 30/06/2008 (excluding weekends, holidays and after-hours trading). Data is sampled at 5-minute intervals using the last trade price every 5-minute period. Since the London Stock Exchange operates between 8:00 and 16:30 GMT, this produced 102 price data points per day for each stock. The sample skewness and kurtosis in Table 3.5 indicate that the return distributions are far from being normal. The sample statistics also indicate that only one data set shows an overall positive trend whilst the other five all show an overall negative trend over the selected period during the 2007-2008 economic crisis. In the following, results for experiment 1 are discussed in Section 3.3.1, and for (control) experiment 2 in Section 3.3.2 (see also the overview in Table 3.4).

### Table 3.5: Descriptive statistics of 5-minute returns

<table>
<thead>
<tr>
<th>Company</th>
<th>Symbol</th>
<th>Mean $\times 10^{-5}$</th>
<th>Std. Dev.</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alliance &amp; Leicester</td>
<td>AL.</td>
<td>-0.4982</td>
<td>0.0051</td>
<td>-1.4962</td>
<td>356.4600</td>
</tr>
<tr>
<td>Schroders</td>
<td>SDRC</td>
<td>-0.1363</td>
<td>0.0034</td>
<td>-1.3003</td>
<td>122.5400</td>
</tr>
<tr>
<td>British Land</td>
<td>BLND</td>
<td>-0.2614</td>
<td>0.0031</td>
<td>0.2752</td>
<td>30.1670</td>
</tr>
<tr>
<td>British Airways</td>
<td>BAY</td>
<td>-0.2870</td>
<td>0.0035</td>
<td>-0.1561</td>
<td>59.2020</td>
</tr>
<tr>
<td>Diageo</td>
<td>DGE</td>
<td>-0.0604</td>
<td>0.0021</td>
<td>-0.3675</td>
<td>162.0200</td>
</tr>
<tr>
<td>Antofagasta</td>
<td>ANTO</td>
<td>0.0556</td>
<td>0.0041</td>
<td>1.2677</td>
<td>97.5120</td>
</tr>
</tbody>
</table>

approach. By inspecting our 100 day-by-day trading results and analysing these across the regions in the space of input factors, we can utilise a heat map approach to identify areas which maximised the Sharpe ratio criterion (for illustration, see Figure 3.4 in the next section). In particular, we are interested to see how the models behave across different levels of position time and return band parameters.
Fig. 3.4: Heat map identifying the sensitivity of the ANFIS model and the highest Sharpe ratio for different position time and return band regions (in-sample).
3.3.1 Results for experiment 1

We first perform a sensitivity analysis of our models to identify the robustness of our models and also to investigate the effect of position time and return band on our results (see also Resta, 2009). For illustration, consider the heat map for ANFIS (Figure 3.4), which indicates concentrated regions of a higher Sharpe ratio in areas of higher holding position times and return bands. This demonstrates the effectiveness of applying these two filters in our trading models. The heat map also provides an indication that, although a number of studies are mostly based on daily data (Krollner et al., 2010), technical rules do manage to identify pockets of profitability in the higher frequency range (Schulmeister, 2009). In view of the stated difficulty with aggressive high-frequency trading with position holding periods of between 10 milliseconds and 10 seconds (see Kearns et al., 2010), taking a less aggressive holding period of between 10 minutes to 1 hour can show very positive results. Of particular interest is the fact that for specific stocks the heat maps identify more than one area of profitable regions, hence providing a clearer indication to traders on possible profitable trading strategies.

In experiment 1a, our model parameter identification was based on applying the Sharpe ratio as our objective function. From the out-of-sample results in Table 3.6, we see that ANTO was the only stock which has not generated a positive Sharpe ratio across all models. This indicates that the combination of moving average signals with artificial intelligence techniques can indeed be applied to generate profitable trading strategies in an intraday trading setting. Both ANFIS and DENFIS generated a positive Sharpe ratio in four out of six stocks. In the case of FFN, the model generated a positive Sharpe ratio in three out of six stocks. A point worth noticing is that in these three instances FFN generated the highest Sharpe ratio across the three models; however, in the other three it generated the lowest performance. This indicates the sensitivity of FFN models to changes in the underlying data features.
Table 3.6: Model performance using Sharpe ratio optimisation over the 100-day out-of-sample period (bold font indicates best result among the three AI methods for the specific stock).

<table>
<thead>
<tr>
<th>Model</th>
<th>Measure</th>
<th>AL.</th>
<th>ANTO</th>
<th>BAY</th>
<th>BLND</th>
<th>DGE</th>
<th>SDRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFN</td>
<td>Sharpe Ratio</td>
<td>-0.0553</td>
<td>-0.0666</td>
<td><strong>0.3278</strong></td>
<td><strong>0.3206</strong></td>
<td>-0.0599</td>
<td><strong>0.2307</strong></td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>-0.0549</td>
<td>-0.0636</td>
<td><strong>0.9469</strong></td>
<td><strong>0.6616</strong></td>
<td>-0.0947</td>
<td><strong>0.3676</strong></td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>-0.0691</td>
<td>-0.1931</td>
<td><strong>0.4831</strong></td>
<td><strong>0.1883</strong></td>
<td>-0.0352</td>
<td><strong>0.4144</strong></td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.5082</td>
<td>0.5437</td>
<td><strong>0.5913</strong></td>
<td><strong>0.6607</strong></td>
<td>0.4750</td>
<td><strong>0.6131</strong></td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>0.8459</td>
<td>0.8074</td>
<td><strong>2.8829</strong></td>
<td><strong>2.6709</strong></td>
<td>0.8494</td>
<td><strong>1.8460</strong></td>
</tr>
<tr>
<td>ANFIS</td>
<td>Sharpe Ratio</td>
<td><strong>0.1579</strong></td>
<td><strong>-0.0485</strong></td>
<td>0.3099</td>
<td>-0.0090</td>
<td>0.0213</td>
<td>0.1748</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td><strong>0.2966</strong></td>
<td>-0.0567</td>
<td>0.6471</td>
<td>-0.0103</td>
<td>0.0348</td>
<td>0.3074</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>0.1114</td>
<td><strong>-0.0588</strong></td>
<td>0.2919</td>
<td>-0.0115</td>
<td>0.0142</td>
<td>0.2589</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td><strong>0.5972</strong></td>
<td>0.5275</td>
<td>0.5714</td>
<td>0.5072</td>
<td>0.5053</td>
<td>0.5391</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td><strong>1.5770</strong></td>
<td><strong>0.8698</strong></td>
<td><strong>3.1816</strong></td>
<td>0.9727</td>
<td>1.0691</td>
<td>1.7313</td>
</tr>
<tr>
<td>DENFIS</td>
<td>Sharpe Ratio</td>
<td>0.1067</td>
<td>-0.0610</td>
<td>0.1690</td>
<td>0.0126</td>
<td><strong>0.2692</strong></td>
<td>-0.0207</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>0.1756</td>
<td><strong>-0.0563</strong></td>
<td>0.2895</td>
<td>0.0135</td>
<td><strong>0.3734</strong></td>
<td>-0.0274</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td><strong>0.3222</strong></td>
<td>-0.1154</td>
<td>0.2922</td>
<td>0.0113</td>
<td><strong>0.3708</strong></td>
<td>-0.0565</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.5607</td>
<td><strong>0.5573</strong></td>
<td>0.5389</td>
<td>0.5143</td>
<td><strong>0.6480</strong></td>
<td>0.5561</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>1.3985</td>
<td>0.8164</td>
<td>1.6663</td>
<td>1.0377</td>
<td><strong>2.0977</strong></td>
<td>0.9465</td>
</tr>
</tbody>
</table>

Although this single point in time performance measurement is the most common approach adopted in literature, this value alone might not be sufficient to confirm the success of a model and our primary interest was to assess the performance profile of each model over the full 100-day period. Firstly, the Sharpe ratio for a certain sample on a certain day may be positive just by chance, in which case the reported result is not very conclusive. Secondly, even if the Sharpe ratio on a certain day or for a certain period is statistically significant and positive, different investors might not have the same holding period. The cumulative Sharpe ratio is therefore more informative as it indicates the overall performance of a certain strategy over a longer period. Let \( \text{measure}_{t, t+n} \) represent the aggregated measure from day \( t \) to day \( t+n \) of the out-of-sample period.

Figure 3.5 shows \( \text{Sharpe Ratio}_{1, 20} \) up to \( \text{Sharpe Ratio}_{1, 100} \). All stocks had at least one model which generated positive results up to \( \text{Sharpe Ratio}_{1, 40} \). After the 40th day, only one stock does not generate any positive Sharpe ratios (ANTO). As the cumulative Sharpe ratio
is relatively stable for the majority of our stocks, we can conclude that our proposed models outperform the benchmark strategy not only on a certain day but consistently over a longer period.

In three (BAY, BLND, SDRC) out of six stocks, FFN models showed substantially higher performance measures than ANFIS and DENFIS models, with a fifth one (ANTO) showing the highest obtained positive Sharpe ratio up to the 40th day. This validates the popularity of neural networks in non-linear time series applications as identified in Tsai and Wang (2009) and Krollner et al. (2010). This also shows that unlike Kablan and Ng (2011), ANNs still provide a valid benchmark when applying more recent neuro-fuzzy models on high-frequency price series. The plots however also show that FFN tend to experience the highest variation in performance (for example, in the case of AL and ANTO), which indicates higher model sensitivity. ANFIS shows a positive Sharpe ratio on four (AL, BAY, DGE, SDRC) out of six stocks with a minor loss on one stock (BLND) and lowest loss amongst the other models on the last stock (ANTO). DENFIS only outperforms the other models on one stock (DGE).

In experiment 1b, we base our model parameter identification process on the maximisation of the Sortino ratio. In this case, all six stocks have at least one model which generated positive results over the full 100-day out-of-sample period (Table 3.7). This again validates the possibility of achieving profitable trading strategies by applying a hybrid of moving average prediction models with artificial intelligence techniques on high-frequency data. ANFIS generates profitable trading results on five out of six stocks with the exception of one stock (BLND). In three (AL, ANTO and SDRC) out of six stocks, ANFIS clearly shows a better performance than the other models. FFN performance is positive on three (ANTO, BAY and BLND) out of three stocks, with two best performances across all models on BAY and BLND. DENFIS has positive results in four out of six stocks with the highest results obtained for DGE. When looking at the models’ performance profile based on the Sortino ratio over the 100 days in general (Figure A.1), one immediately notices that with
Improving the risk-adjusted performance of trading algorithms

Fig. 3.5: Trading performance by optimising Sharpe ratio. The plots show the cumulated Sharpe ratio (y-axis) for each stock on the n-th day (x-axis) in the out-of-sample.
3.3 Empirical data and analysis

Table 3.7: Model performance using Sortino ratio optimisation over the 100-day out-of-sample period (bold font indicates best result among the three AI methods for the specific stock).

<table>
<thead>
<tr>
<th>Model</th>
<th>Measure</th>
<th>AL.</th>
<th>ANTO</th>
<th>BAY</th>
<th>BLND</th>
<th>DGE</th>
<th>SDRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFN</td>
<td>Sharpe Ratio</td>
<td>-0.0553</td>
<td>0.1408</td>
<td><strong>0.3278</strong></td>
<td>0.3785</td>
<td>-0.0599</td>
<td>-0.0203</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>-0.0549</td>
<td>0.1954</td>
<td><strong>0.9469</strong></td>
<td><strong>0.7069</strong></td>
<td>-0.0947</td>
<td>-0.0291</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>-0.0691</td>
<td>0.1567</td>
<td><strong>0.4831</strong></td>
<td><strong>0.0665</strong></td>
<td>-0.0352</td>
<td>-0.0273</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.5082</td>
<td><strong>0.5422</strong></td>
<td>0.5913</td>
<td><strong>0.6842</strong></td>
<td>0.4750</td>
<td>0.4886</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>0.8459</td>
<td>1.4763</td>
<td><strong>2.8829</strong></td>
<td><strong>2.9289</strong></td>
<td>0.8494</td>
<td>0.9471</td>
</tr>
<tr>
<td>ANFIS</td>
<td>Sharpe Ratio</td>
<td><strong>0.1282</strong></td>
<td><strong>0.1583</strong></td>
<td>0.3273</td>
<td>-0.3656</td>
<td>0.2164</td>
<td><strong>0.0531</strong></td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td><strong>0.2242</strong></td>
<td><strong>0.2136</strong></td>
<td>0.7331</td>
<td>-0.4597</td>
<td>0.3702</td>
<td><strong>0.0741</strong></td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>0.1742</td>
<td><strong>0.1845</strong></td>
<td>0.3568</td>
<td>-0.1161</td>
<td><strong>0.3659</strong></td>
<td><strong>0.0891</strong></td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td><strong>0.5854</strong></td>
<td>0.5119</td>
<td><strong>0.5983</strong></td>
<td>0.3175</td>
<td><strong>0.5854</strong></td>
<td><strong>0.5985</strong></td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td><strong>1.4423</strong></td>
<td><strong>1.5241</strong></td>
<td>2.7228</td>
<td>0.3596</td>
<td><strong>1.9995</strong></td>
<td><strong>1.1546</strong></td>
</tr>
<tr>
<td>DENFIS</td>
<td>Sharpe Ratio</td>
<td>0.1067</td>
<td>-0.0610</td>
<td>0.1690</td>
<td>0.0126</td>
<td><strong>0.2235</strong></td>
<td>-0.0207</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>0.1756</td>
<td>-0.0563</td>
<td>0.2895</td>
<td>0.0135</td>
<td><strong>0.5239</strong></td>
<td>-0.0274</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td><strong>0.3222</strong></td>
<td>-0.1154</td>
<td>0.2922</td>
<td>0.0113</td>
<td>0.2098</td>
<td>-0.0565</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.5607</td>
<td><strong>0.5573</strong></td>
<td>0.5389</td>
<td>0.5143</td>
<td>0.5372</td>
<td>0.5561</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>1.3985</td>
<td>0.8164</td>
<td>1.6663</td>
<td>1.0377</td>
<td>1.8935</td>
<td>0.9465</td>
</tr>
</tbody>
</table>

the exception of FFN on the AL. stock, the plots exhibit less abrupt variations than those obtained in the Sharpe ratio equivalents.

When reviewing other performance results obtained from the Sharpe ratio and the Sortino ratio models (Tables 3.6 and 3.7), one can identify that in contrast to Kearns et al. (2010) who demonstrated that aggressive HFT exhibits a surprisingly low profitability, our results show that the cumulative return and win ratio from a number of models is considerably high. However, as mentioned above, although the win ratio is a common measure used in literature to measure performance, a higher win ratio does not necessarily result in a profitable model, hence albeit indicative, it cannot be used as a performance measure on its own (see also Brabazon and O’Neill, 2006). This is shown for example in DENFIS-ANTO and DENFIS-SDRC results, where the model is successful in attaining high win ratios but still suffers from larger losses (as indicated by the profit ratio).
Following the results attained in the first set of experiments using the optimisation of a single risk-adjusted objective function, in experiment 1c we extend our ideas by investigating the possible risk-adjusted model performance improvement attained by combining multiple objective functions. Since from experiments 1a and 1b we identified that ANFIS performs well when compared to the other models both in terms of risk-return performance and stability, in our third experiment (1c) we decided to compare the conventional ANFIS models (1a and 1b) with our innovative ANFIS ensemble architecture, described earlier, which uses a novel dynamic selection method to integrate three risk-adjusted base ANFIS models optimised using Sharpe ratio, Sortino ratio and period return respectively. As indicated in the results in Table 3.8, one can immediately notice the effectiveness of our approach by observing that in this case the results are positive for all 6 stocks. By comparing the results against those attained by the singular ANFIS base models (see Tables 3.6 and 3.7) one can also identify that in four out of six stocks (except for BAY and DGE), the ensemble approach outperformed both the Sharpe and Sortino ratio ANFIS models. For the stock BAY, the ensemble approach obtained the highest period return but did not achieve a better risk-adjusted performance. As it can be seen from Figure 3.6, our suggested ensemble architecture showed higher stability in the results attained over the full out-of-sample period.

Following the recent successful claims of reduced statistical error when applying ANFIS ensembles compared to singular ANFIS for time series prediction (Lei and Wan, 2012; Melin et al., 2012; Soto et al., 2013), our results extend these findings by showing that ANFIS ensembles composed of different risk-adjusted performance base models can outperform and show more robustness than the individual ANFIS base models that are optimised using a single risk-adjusted optimisation function. Our proposed ensemble integration method introduces a dynamic short memory component in the ensemble architecture and allows for dynamic selection based on recent risk-adjusted performance at an intraday level. The results also show that following our dynamic selection approach (see also Puuronen et al., 1999) is
Fig. 3.6: The plots compare the results obtained from ANFIS Sharpe optimisation (experiment 1a), eANFIS (experiment 1c), no-transaction-costs optimisation (experiment 2a), RMSE optimisation (experiment 2b) and period return optimisation (experiment 2c) and show the corresponding cumulated Sharpe ratio (y-axis) for each stock on the $n$-th day (x-axis) in the out-of-sample.
an effective method that can be applied to boost the profitability attained from high-frequency trading models without incurring an additional risk penalty.

### 3.3.2 Results for experiment 2

We now present the results attained from benchmark models that are typically found in literature or used in practice (see overview in Table 3.4).

The recent survey by Krollner et al. (2010) identified that most studies do not account for real-world constraints such as trading costs (see also Álvarez Díaz, 2010). In experiment 2a, we base our model selection criteria on the Sharpe ratio but exclude transaction costs in the training period. In our 100-day out-of-sample evaluation, we then apply transaction costs to the selected models as in our original Sharpe model in order to simulate realistic trading environments. As indicated in Table 3.9, negative results are observed for all stocks in all models except for the DENFIS-DGE model, which shows a minor positive result. The performance over the full 100-day out-of-sample period can be seen in the plots comparing the model against the Sharpe optimised ANFIS and eANFIS models in Figure 3.6. The plots further support our initial results and provide a clear indication that not considering such costs when training the trading system can lead to biased results in real-world applications.
In experiment 2b, we apply the RMSE minimisation approach for our model selection process. The results in Table 3.10 show that in the case of ANFIS only two (AL. and ANTO) out of six stocks generate positive results: in the case of DENFIS no stock generates a positive result, and in the case of FFN only three stocks (AL., BLND and SDRC) generate positive results. When comparing these results against the results attained by the risk-return based models discussed earlier (in experiment 1), we find that for both ANFIS and DENFIS models the RMSE optimisation provides better results only on ANTO. In the case of FFN, RMSE optimisation outperforms Sharpe optimisation only in AL. In Figure 3.6 we are displaying the RMSE optimisation performance against Sharpe ratio optimised ANFIS and eANFIS over the full out-of-sample period. Further to the above findings, the plots indicate that RMSE optimisation fails to outperform eANFIS on all six stocks. These results are in line with Brabazon and O’Neill (2006) and provide a clear indication that trading models based on risk-return selection criterion outperform those based on RMSE optimisation.
Table 3.10: Model performance using RMSE optimisation over the 100-day out-of-sample period.

<table>
<thead>
<tr>
<th>Model</th>
<th>Measure</th>
<th>AL.</th>
<th>ANTO</th>
<th>BAY</th>
<th>BLND</th>
<th>DGE</th>
<th>SDRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFN</td>
<td>Sharpe Ratio</td>
<td>0.0178</td>
<td>-0.0109</td>
<td>-0.0065</td>
<td>0.0304</td>
<td>-0.0418</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>0.0172</td>
<td>-0.0103</td>
<td>-0.0076</td>
<td>0.0388</td>
<td>-0.0432</td>
<td>0.0013</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>0.9866</td>
<td>-0.4845</td>
<td>-0.2934</td>
<td>1.0218</td>
<td>-0.9356</td>
<td>0.0384</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.6094</td>
<td>0.6081</td>
<td>0.5606</td>
<td>0.5891</td>
<td>0.5654</td>
<td>0.5087</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>1.0654</td>
<td>0.9623</td>
<td>0.9774</td>
<td>1.1060</td>
<td>0.8694</td>
<td>1.0035</td>
</tr>
<tr>
<td>ANFIS</td>
<td>Sharpe Ratio</td>
<td>0.0454</td>
<td>0.0117</td>
<td>-0.0045</td>
<td>-0.0119</td>
<td>-0.0574</td>
<td>-0.0693</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>0.0652</td>
<td>0.0127</td>
<td>-0.0055</td>
<td>-0.0139</td>
<td>-0.0191</td>
<td>-0.0798</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>2.7156</td>
<td>0.5540</td>
<td>-0.2153</td>
<td>-0.4289</td>
<td>-1.2838</td>
<td>-2.5264</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.5681</td>
<td>0.5544</td>
<td>0.5018</td>
<td>0.5437</td>
<td>0.5272</td>
<td>0.4511</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>1.1804</td>
<td>1.0411</td>
<td>0.9840</td>
<td>0.9621</td>
<td>0.8329</td>
<td>0.8004</td>
</tr>
<tr>
<td>DENFIS</td>
<td>Sharpe Ratio</td>
<td>-0.0135</td>
<td>-0.0515</td>
<td>-0.0232</td>
<td>-0.0676</td>
<td>-0.0755</td>
<td>-0.0523</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>-0.0123</td>
<td>-0.0503</td>
<td>-0.0258</td>
<td>-0.0752</td>
<td>-0.0978</td>
<td>-0.0620</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>-0.7546</td>
<td>-2.5468</td>
<td>-0.9960</td>
<td>-2.3346</td>
<td>-1.6061</td>
<td>-1.9094</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.5802</td>
<td>0.5305</td>
<td>0.9233</td>
<td>0.5266</td>
<td>0.4836</td>
<td>0.4705</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>0.9535</td>
<td>0.8287</td>
<td>0.5206</td>
<td>0.8008</td>
<td>0.7929</td>
<td>0.8457</td>
</tr>
</tbody>
</table>

Table 3.11: Model performance using period return optimisation over the 100-day out-of-sample period.

<table>
<thead>
<tr>
<th>Model</th>
<th>Measure</th>
<th>AL.</th>
<th>ANTO</th>
<th>BAY</th>
<th>BLND</th>
<th>DGE</th>
<th>SDRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFN</td>
<td>Sharpe Ratio</td>
<td>-0.0555</td>
<td>0.0115</td>
<td>-0.1350</td>
<td>0.1017</td>
<td>0.0123</td>
<td>-0.0862</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>-0.0791</td>
<td>0.0137</td>
<td>-0.1550</td>
<td>0.1620</td>
<td>0.0157</td>
<td>-0.1148</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>-0.4596</td>
<td>0.1641</td>
<td>-0.2834</td>
<td>0.9249</td>
<td>0.0358</td>
<td>-0.1860</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.4685</td>
<td>0.5274</td>
<td>0.4380</td>
<td>0.5369</td>
<td>0.5178</td>
<td>0.4852</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>0.8417</td>
<td>1.0356</td>
<td>0.6620</td>
<td>1.3646</td>
<td>1.0341</td>
<td>0.7865</td>
</tr>
<tr>
<td>ANFIS</td>
<td>Sharpe Ratio</td>
<td>0.0128</td>
<td>0.1877</td>
<td>-0.0435</td>
<td>0.0611</td>
<td>-0.0530</td>
<td>0.0414</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>0.0122</td>
<td>0.3054</td>
<td>-0.0579</td>
<td>0.0708</td>
<td>-0.0821</td>
<td>0.0622</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>0.1522</td>
<td>0.5134</td>
<td>-0.4942</td>
<td>0.1667</td>
<td>-0.0427</td>
<td>0.5596</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.5606</td>
<td>0.5449</td>
<td>0.5155</td>
<td>0.5458</td>
<td>0.3913</td>
<td>0.5254</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>1.0485</td>
<td>1.7235</td>
<td>0.8816</td>
<td>1.1947</td>
<td>0.8688</td>
<td>1.1263</td>
</tr>
<tr>
<td>DENFIS</td>
<td>Sharpe Ratio</td>
<td>-0.0753</td>
<td>-0.0368</td>
<td>-0.0786</td>
<td>-0.1159</td>
<td>0.2692</td>
<td>-0.0499</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>-0.0783</td>
<td>-0.0334</td>
<td>-0.0918</td>
<td>-0.1277</td>
<td>0.3734</td>
<td>-0.0709</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>-1.9679</td>
<td>-0.4745</td>
<td>-1.3275</td>
<td>-1.2901</td>
<td>0.3708</td>
<td>-0.9081</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.5090</td>
<td>0.5456</td>
<td>0.5058</td>
<td>0.4793</td>
<td>0.6480</td>
<td>0.4931</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>0.7810</td>
<td>0.8695</td>
<td>0.7891</td>
<td>0.6983</td>
<td>2.0977</td>
<td>0.8670</td>
</tr>
</tbody>
</table>
Table 3.12: Model performance using fixed moving average (MA) rules over the 100-day out-of-sample period.

<table>
<thead>
<tr>
<th>Model</th>
<th>Measure</th>
<th>AL.</th>
<th>ANTO</th>
<th>BAY</th>
<th>BLND</th>
<th>DGE</th>
<th>SDRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA(1,5)</td>
<td>Sharpe Ratio</td>
<td>-0.0792</td>
<td>0.0031</td>
<td>-0.0585</td>
<td>0.0747</td>
<td>-0.1216</td>
<td>-0.0734</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>-0.1343</td>
<td>0.0055</td>
<td>-0.0960</td>
<td>-0.1376</td>
<td>-0.2431</td>
<td>-0.1218</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>-1.8815</td>
<td>0.0670</td>
<td>-1.1292</td>
<td>-1.2577</td>
<td>-1.1946</td>
<td>-1.0433</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.3169</td>
<td>0.3845</td>
<td>0.3577</td>
<td>0.3448</td>
<td>0.3262</td>
<td>0.3552</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>0.7673</td>
<td>1.0108</td>
<td>0.8180</td>
<td>0.7721</td>
<td>0.6608</td>
<td>0.8026</td>
</tr>
<tr>
<td>MA(5,10)</td>
<td>Sharpe Ratio</td>
<td>-0.0496</td>
<td>-0.0173</td>
<td>-0.0732</td>
<td>-0.1515</td>
<td>-0.1515</td>
<td>-0.0668</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>-0.0765</td>
<td>-0.0316</td>
<td>-0.1291</td>
<td>-0.1665</td>
<td>-0.2511</td>
<td>-0.1099</td>
</tr>
<tr>
<td></td>
<td>Cum. Return</td>
<td>-1.3043</td>
<td>-0.4084</td>
<td>-1.6104</td>
<td>-1.3750</td>
<td>-1.3750</td>
<td>-1.1452</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.3378</td>
<td>0.3521</td>
<td>0.3058</td>
<td>0.3081</td>
<td>0.3081</td>
<td>0.3256</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>0.8393</td>
<td>0.9408</td>
<td>0.7724</td>
<td>0.6360</td>
<td>0.6360</td>
<td>0.8002</td>
</tr>
<tr>
<td>MA(10,15)</td>
<td>Sharpe Ratio</td>
<td>-0.1562</td>
<td>-0.1351</td>
<td>-0.1393</td>
<td>-0.2099</td>
<td>-0.3059</td>
<td>-0.1422</td>
</tr>
<tr>
<td></td>
<td>Sortino Ratio</td>
<td>-0.2380</td>
<td>-0.2401</td>
<td>-0.3148</td>
<td>-0.3988</td>
<td>-0.5330</td>
<td>-0.2521</td>
</tr>
<tr>
<td></td>
<td>Win Ratio</td>
<td>0.2426</td>
<td>0.2495</td>
<td>0.2290</td>
<td>0.2428</td>
<td>0.2167</td>
<td>0.2543</td>
</tr>
<tr>
<td></td>
<td>Profit Ratio</td>
<td>0.5592</td>
<td>0.6229</td>
<td>0.6152</td>
<td>0.4897</td>
<td>0.3952</td>
<td>0.6248</td>
</tr>
</tbody>
</table>

In experiment 2c, we compare our risk-adjusted models with those optimised using period return, hence not considering risk in their objective function. When comparing the results in Table 3.11 with those obtained for Sharpe ratio optimised models (Table 3.6), one can identify that not accounting for risk shows outperformance only in two out of six stocks (BLND and ANTO). This can also be verified for ANFIS in the Figure 3.6. The plot also indicates that in this case eANFIS was slightly outperformed in only one stock (SBLND). Similar performance is attained when compared to the Sortino optimised models (Table 3.7). The explanation for this lower performance can be attributed to the fact that the profit ratio is markedly lower, albeit the period return optimised models achieve a win ratio which is fairly in line with those achieved by their risk-adjusted counterparts. This indicates that period return optimised models suffer from larger losses. These results further reinforce our argument with respect to the effectiveness of using risk-adjusted optimisation functions as identified in experiment 2b.
In our final benchmark comparison, experiment 2d, we investigate the application of standard moving average trading signals (e.g. Schulmeister, 2009) over the 100-day out-of-sample period. The applied moving average short and long lags represent those used in our dynamic moving average experiments. From the results in Table 3.12 we find that only MA(1,5) had positive results for ANTO and BLND (see also Figure A.2). This brings to light the issues with moving average filters in trend identification due to their nonzero phase shift nature (as identified by Gencay et al., 2002) and provides evidence of the effectiveness of our combined dynamic moving average with neuro-fuzzy systems approach.

3.4 Conclusion

In this chapter, our research problem stems from the shortcomings of numerous published studies in the AI and computational finance fields recommending model tuning and performance measures that do not fall in line with the risk concerns of investors. We corroborate our arguments by showing that the omission of risk-adjusted performance considerations leads to poor out-of-sample performance or, even worse, infers overestimated trading performance expectations.

We present a number of methods with the aim to improve the risk-adjusted trading performance of AI models using three representative milestone models from neurocomputing literature, namely ANN, ANFIS and DENFIS. In our experiments, we focus on a particular high-frequency trading (HFT) window that was chosen to gain more insight into the profitability of intraday trading with respect to the tension created between two literature findings: (i) the view that the profitability of trading rules has possibly moved to higher frequency prices (Schulmeister, 2009), and on the other hand, (ii) the view that aggressive HFT with position holding periods of between 10 milliseconds and 10 seconds does not reap the expected excess returns (Kearns et al., 2010).
Our first contribution is the simple yet effective extension of common technical moving average rules by considering a dynamic “portfolio” of moving average prediction models. A common challenge in selecting a trend identifying mechanism is that on the one hand, applying a too short moving average would result in falsely reporting a break in trend. On the other hand, choosing a too long moving average would result in a late reaction to price movement. In our approach, we use AI techniques to combine different moving averages and dynamically tune the trend signals according to the changing speeds of the market. This is further enhanced by applying a model validation methodology using heat maps to analyse favourable risk-return regions that identify profitability in specific holding time and signal regions. Our results demonstrate that the proposed dynamic moving approach outperforms the risk-adjusted performance obtained from standard moving average technical rules.

Krollner et al. (2010) find that although more than 80% of the surveyed papers state that their proposed framework surpassed the benchmark model, most of them are actually ignoring trivial constraints in the real world, hence introducing a risk of overestimated profitability (see also Álvarez Díaz, 2010). As our second major contribution, we investigate the effect of omitting transaction costs during model training and selection processes (prior to the testing stage). Our findings show that omitting transaction costs from the model selection process fall far short of the expected trading performance when tested with realistic constraints. This raises an important model design consideration with researchers in computational finance.

Krollner et al. (2010) also identify a lack of literature studying whether AI-based algorithms can enhance an investor’s risk-return trade-off. While a few attempts were reported recently, mainly for other financial applications (Dempster and Leemans, 2006; Esfahanipour and Mousavi, 2011; Evans et al., 2013), the implementation of risk-adjusted performance control has to our best knowledge not been studied in an intraday high-frequency setting before (see also a recent comprehensive survey by Bahrammirzaee (2010) and the references therein). As our third contribution, we present different approaches of how the high-frequency
trading performance of AI-driven algorithms can be improved by analysing the time series of risk-adjusted performance measures. In contrast to common approaches in the literature which evaluate models using performance measures at an arbitrary single point in time (e.g. only at the end of the sample period), we provide a deeper understanding of the time-varying performance profile of the applied models.

Finally, when training and evaluating a trading system, most former studies have a very limited view of what constitutes successful investment decisions, defining success on the grounds of forecast accuracy, win ratios or minimum forecast error (Alves Portela Santos et al., 2007; de Faria et al., 2009; Enke and Thawornwong, 2005; Medeiros et al., 2006). However, as we show in our results, these optimisations do not guarantee trading profits. In our out-of-sample evaluation, we show that overall FFN models perform well when compared to the more recent neuro-fuzzy techniques. However, FFN models also show higher sensitivity to the underlying data features. Looking at a 100-day out-of-sample trading performance, we find that ANFIS provides the most stable performance measure, exhibiting the least variance in the case of both Sharpe and Sortino ratio optimisation. In our experiments, DENFIS did not outperform FFN and ANFIS (see also Tan et al., 2011). We further extend our improvement ideas by proposing a novel method of combining different risk-adjusted objective functions and demonstrate its effectiveness using an innovative ANFIS ensemble architecture and integration method. Results show that our ensemble outperformed conventional ANFIS using single risk-adjusted objective functions.

In the next chapter we extend our research problem to address the time-varying intraday volatility. Our ideas on the utilisation of heat maps are further extended to adapt models at more granular intraday time-windows.
Chapter 4

Time-varying intraday volatility and risk-adjusted performance

In Chapter 3, it was demonstrated that using a global risk-adjusted objective function for AI-controlled trading algorithms is more appropriate than other more common objective functions found in machine learning literature. By extending our ideas to a more granular intraday perspective, the research in this chapter emanates from the stylised fact that activity in markets is not constant throughout the trading day. Changes in the underlying conditions can introduce a substantial increase in error variance and large losses. The research problem addresses risk-adjusted performance maximisation of AI-controlled trading algorithms in view of the challenge posed by continuously changing intraday market conditions as a reflection of time-varying volatility (risk).

A wide body of research claims that volatility exhibits strong intraday periodicity. Whilst sufficient market volatility is required to ensure price movement, on the other hand, higher volatility reflects higher uncertainty, more challenging forecasting, and the possibility of incurring larger losses. Although numerous models are presented in literature that are capable of measuring volatility at a good degree of accuracy up to an intraday level, the literature
of how AI trading algorithms can be optimised to handle intraday time-varying volatility is scarce.

As a first contribution, we use NNs as a motivating example and improve on existing methodologies of how to design and tune AI models when these are applied for intraday trading purposes by considering time-varying volatility. A standard NN architecture is extended with an innovative dynamic fuzzy logic layer which permits adaptive local tuning in line with changing volatility states. Our second contribution is a novel AI-driven money management algorithm that identifies and optimises money allocation according to different intraday volatility levels. Finally, a number of authors claim a relationship between the possible periodic breakdown of market efficiency and volatility. Our third contribution extends these theoretical claims to a more granular intraday level by identifying a link between the profitability of technical rules and different levels of price volatility at short intraday time horizons.

4.1 Introduction

Bearing in mind that essentially a model is a representation of a state, this chapter builds on the risk-adjusted model improvements presented in Chapter 3 by introducing a method to discriminate and adapt to the different market states (risk scenarios) that evolve during a typical trading day. Volatility is used as a statistical measure of the dispersion of returns to discriminate between the different market states.

Volatility for the high-frequency trader poses a tricky scenario. Sufficient market volatility is required to ensure that changes in prices exceed transaction costs. On the other hand, higher volatility reflects higher uncertainty and poses the risk of adverse market movements resulting in losses. Financial time series exhibit so-called stylised facts patterns (see Cont, 2001; Gencay et al., 2002). Applying machine learning techniques to non-stationary time series to infer predictions becomes a more difficult challenge with the possibility of increased
error variance. Although a wide body of research claims that volatility exhibits strong intraday periodicity, high persistence and can be predicted with a good degree of accuracy up to an intraday level (Andersen and Bollerslev, 1997; Andersen et al., 2000a), the literature of how AI algorithms can benefit from this information on volatility for trading purposes is lacking.

This chapter is divided into two main parts:

In Part 1 (Section 4.2), as our main contribution we introduce an approach of how AI algorithms can benefit from information on volatility for trading purposes. We provide new insight on using intraday realised volatility as a proxy for uncertainty. We propose how this information can be used as an additional component for NN tuning to improve risk-return trading performance. Due to the popularity of Neural Networks (NNs) in non-linear times series applications (as identified in surveys by Bahrammirzaee, 2010; Krollner et al., 2010; Tsai and Wang, 2009), as a motivating example we make use of a popular Neural Network (NN) model and use it as our benchmark to measure the effectiveness of our approach. A number of authors have suggested methodologies of how to design and tune NNs when applied for predicting financial time series and trading purposes but with little consideration to the time-varying volatility (see Kaastra and Boyd, 1996; Vanstone and Finnie, 2009, 2010). Other authors intentionally avoid higher periods of uncertainty by keeping out of the market especially during the initial and end intraday trading periods (e.g. Brabazon and O’Neill, 2006), which are well documented to exhibit the highest volatility.

By extending the concept of heat maps presented in Chapter 3 to identify preferable trade regions, we introduce an NN and fuzzy logic hybrid model which consists of a standard NN enhanced with an innovative and efficient volatility filter based on fuzzy c-means (FCM) clustering (Bezdek, 1981). The proposed extension monitors the performance of the underlying NN model across various market volatility levels. Subsequently, it dynamically identifies unique intraday scenarios and creates a dynamic and visually apprehensible risk-
return search space to control algorithmic trading decisions. This feature also mitigates the common black box criticisms that are typically attributed to standard NN techniques. The proposed technique does not limit the possible use of the proposed extension with other machine learning techniques. Our results show that this method can be successfully applied to support high-frequency intraday trading strategies, outperforming both standard NN and buy-and-hold models.

In Part 2 (Section 4.3), the findings from the first part are extended by demonstrating how this approach can be applied to control risk-based money management decisions. As our first contribution, we propose an innovative fuzzy logic method which identifies the approximate regional performance across time-varying intraday risk-return states. Contrary to many studies that limit the application of AI to focus solely on market direction (as indicated by Krollner et al., 2010; Tsai and Wang, 2009), we present an effective money management approach to dynamically adjust trading frequency and position size depending on the varying degrees of risk at an intraday level with the objective to improve the overall risk-adjusted trading performance. This enhances the trading model presented in Chapter 3 which considers equally-sized trade positions and does not discriminate across different intraday volatility states.

Our approach in this chapter also goes contrary to studies suggesting the use of fixed return and volatility thresholds (Holmberg et al., 2013; Kaastra and Boyd, 1996; Vanstone and Finnie, 2009, 2010). We propose an innovative fuzzy logic approach which identifies and categorises model performance across different regions in the trend and volatility space. The model dynamically prioritises higher performing regions at an intraday level and adapts money management policies with the objective to maximise global risk-adjusted performance. By adopting rigorous statistical tests, our results show significant performance improvements compared to both standard NN and buy-and-hold approaches. This provides tenable reason to
infer a relationship between volatility levels and possible breaks in market efficiency during short intraday time horizons.

Our second contribution is that improving the risk-adjusted performance of the underlying trading model does not come at the cost of reduced overall profitability (as identified by Holmberg et al., 2013). This means that our approach should not just act as a filter by just keeping the algorithm out of the market in adverse regions and hence underutilise available capital, but also improve model profitability by allocating more capital to preferable intraday trading scenarios.

One assumption of EMH is that prices evolve as a random walk in time (Fama, 1965; ?), hence returns are martingales. Under this assumption there should be no significant increase in the overall trading performance by discriminating trading priority amongst different intraday trend and volatility states over a sufficiently long trading horizon. In line with this, as a direct consequence of our first contribution, increased trading performance using our proposed method suggests a possible martingale property breakdown during specific trend-volatility states. Although this market phenomena is explored by a number of authors (Gradojevic and Gençay, 2013; Han et al., 2013; Holmberg et al., 2013; LeBaron, 1999; Schulmeister, 2006), as our third contribution we extend the literature by exploring the relationship between volatility and profitability of technical trading rules at shorter term intraday horizons.

4.2 Part 1 - Enhancing the intraday trading performance of neural networks using a dynamic volatility clustering fuzzy filter

The efficient market hypothesis (EMH) (Fama, 1965, 1970), which resulted in Eugene Fama being crowned with a Nobel Prize in Economics in 2013, remains without any doubt one of
the most debated theories in economics. According to EMH and the implied random walk of asset prices, the use of trading rules on historical prices should not result in excess returns after accounting for transaction costs. However, although the EMH is a strong landmark in the finance and economics literature, the application of technical rules is a widespread practice. Research claims from a number of authors (Gradojevic and Gençay, 2013; Han et al., 2013; Holmberg et al., 2013; LeBaron, 1999; Schulmeister, 2006) indicate a relationship between the possible periodic breakdown of market efficiency and volatility. We extend this research at a more granular intraday level and seek a link between the different levels of profitability of technical rules with different levels of price volatility at short intraday time horizons. We explore how this information can improve the risk adjusted performance of high-frequency trading algorithms. To our best knowledge this was not investigated in literature so far.

Part 1 of this chapter is structured as follows. In Section 4.2.1 we introduce our experiment approach and describe the main model components. We then describe the base NN model and the data preparation process (Section 4.2.2), followed by a description of the applied trading algorithm (Section 4.2.3). We then discuss the challenges that arise due to intraday volatility (Section 4.2.4) and present our enhanced trading algorithm (Section 4.2.5). Section 4.2.6 presents our experiment approach, followed by our experiment results and discussions in the light of existing literature (Section 4.2.7). Section 4.2.8 concludes Part 1 of this chapter.

4.2.1 Method

Our experiment setup has two main components (see Figure 4.1). First, a standard NN, which is very popular in trading applications (Choudhry et al., 2012; Krollner et al., 2010; Tsai and Wang, 2009), is employed to forecast the average return over the following 5 minutes, \( y_t \), using standard trading rule signals. This model is similar to the NN model presented in Chapter 3, and on the same lines we also employ a global risk-return objective function for
model calibration purposes. The standard NN model is also used as one of our benchmark models.

However, in this chapter we introduce a second component, a locally enhanced NN model. The output from the standard NN model is passed as input to the locally enhanced NN model extension. We base our approach on fuzzy set theory (Zadeh, 1997). Since our problem deals with intraday uncertainty as a result of time-varying volatility, we present a hybrid model consisting of an additional dynamic fuzzy logic module. The module discriminates between different market states and enhances standard optimisation techniques that are typically applied to NN models by additionally taking into consideration different time-varying volatility states. We adopt a novel algorithm which takes past trade suggestions from standard NN as input, limited to the past $h$ days, and infers the approximate regional (rather than global) risk-adjusted performance. Core to our controller is a fuzzy c-means clustering algorithm (Bezdek, 1981; Dutta and Angelov, 2010) that identifies unique trading
performance regions across two dimensions: (a) the intraday realised volatility (Andersen and Bollerslev, 1997) which is used as a proxy for uncertainty, and (b) the predicted return size which indicates a trend direction. The identified fuzzy clusters allow the extraction of fuzzy rules and their combined result produces a decision surface across the trend and volatility space that is used to identify preferable trade regions.

We denote $k$ as the number of clusters, $g$ as a point in the trend-volatility space, $C_j$ as the $j$-th induced fuzzy cluster in the trend-volatility space and $P_j(i)$ as a performance measure based on the profits $i$ from trades which were executed close to that cluster region in the last $h$ days. We define close as being a point whose degree of membership $\mu_{C_j}(g)$ to a specific cluster exceeds a certain threshold $\theta$. For each cluster, a rule is extracted in the form

$$IF \quad g \in C_j \quad THEN \quad i_{loc} = P_j(i|\mu_{C_j}(g) > \theta)$$

for $j = 1, 2, ..., k$. The simple fuzzy rule (Equation 4.1) can be interpreted as “if trend and volatility levels are close to region $C_j$, then the approximate regional risk-adjusted performance is $i_{loc}$”. Detail on our fuzzy rule extraction and local performance calculation is presented in Section 4.2.5.

The full database of rules is updated on a daily basis and their combined output produces a decision surface that is used to approximate the trading algorithm risk-adjusted performance at a granular intraday level according to different trend and volatility states. The global trading performance is then optimised by dynamically accepting trade signals in regions that are identified to yield higher risk-return performance and reduces (or even stops) the trades in less favourable ones.

The following Sections 4.2.2 to 4.2.5 describe in more detail each model component and their calibration and selection.
4.2 Part 1 - Enhancing the intraday trading performance of neural networks using a dynamic volatility clustering fuzzy filter

4.2.2 Technical indicators and standard neural network model

The concept of a trend is absolutely essential to the technical approach (Murphy, 1987). In essence, one of the premises of technical trading is that prices move in trends. The objective is to identify a trend reversal at an early stage and ride on that trend until the underlying algorithm indicates that the trend has reversed. In our experiment, we intentionally make use of simple trading rules in order to investigate the simplest versions of trading rules used in common practice (Vanstone and Finnie, 2009, 2010). Our trading rules to derive a set of trading decisions \( \Omega = \{ \text{long, short, 0} \} \) are primarily based on \( n \) historical prices \( \{p_t, p_{t-1}, \ldots, p_{t-n+1}\} \in \mathbb{R}^n_+ \). Since the objective of this chapter is to be able to discriminate between different intraday market scenarios, we have to ensure that enough trade signals can be generated by our algorithm to be able to cover an adequate number of trade scenarios during the course of a trading day. For this reason, in this chapter we increase the price frequency adopted in Chapter 3 (5 minute) to 1-minute trade prices.

A moving average in essence represents a low pass filter which removes higher frequency noise, the combination of which provides a convenient signal (explained in Section 3.2.1). Short and long moving averages, \( MA^n_t \), are calculated as defined by Equation (3.1) and the corresponding trade signal, \( s_{t-1}^{n_1,n_2} \), defined by Equation (3.3).

In this experiment we include another popular trading signal which is defined using the Relative Strength Index (RSI) indicator

\[
RSl_t = 100 - 100/(1 + RSF),
\]

where the relative strength factor (RSF) is calculated by dividing the average of the gains by the average of the losses within a specified time period. A common RSI-based signal typically suggests that an asset is considered overbought if \( RSl_t \) reaches a value of 70, indicating a
sell signal due to possible overvaluation. Similarly, if $RSI_t$ drops to 30, the asset is likely to become oversold and undervalued, indicating a buy signal.

We investigate whether the mean return for the next 5 minutes can be successfully predicted by making use of moving average and RSI signals and use this information to build a profitable trading algorithm. In our experiment, the mean return, $y_t$, is defined as

$$y_t = \log(MA_{t+5}^5) - \log(p_t). \quad (4.3)$$

The three moving average rules utilised have the lag structures $(n_1, n_2) \in [(1,5), (5,10), (10,20)]$ (see Equation (3.1)), where $n_1$ and $n_2$ are expressed in 1-minute time bars. For our RSI indicator we calculate this based on prices in the last 30-minute time slot. By combining these input signals, the linear specification of the return $y_t$ prediction model is defined as:

$$y_t = \theta_0 + \sum_{k=1}^{4} \theta_k s_{k,t-1} + \epsilon_t \quad (4.4)$$

with the error term $\epsilon_t \sim N(0, \rho)$ and

$$s_{k,t} = \begin{cases} 
MA_{1.5}^1 & \text{for } k = 1 \\
MA_{5,10}^1 & \text{for } k = 2 \\
MA_{10,20}^1 & \text{for } k = 3 \\
RSI_t & \text{for } k = 4 
\end{cases} \quad (4.5)$$

To generate the trend signals, we apply an NN model. We train our NN model with a set of moving averages and RSIs (relative strength indicators) in order to predict the next average 5-minute return every 1-minute time epoch. The process starts from a given collection of $N$ input-output data training pairs, $(x^{(1)} : y^{(1)})$, $(x^{(2)} : y^{(2)})$, ..., $(x^{(N)} : y^{(N)})$ where
In Equation (4.6), for each data instance at a specific time \( t \), \( x \) is a vector consisting of \( \{x_1, x_2, x_3, x_4\} \) input elements which represent the \( \{s_{1,t-1}, s_{2,t-1}, s_{3,t-1}, s_{4,t-1}\} \) technical indicator signals (Equation (4.3)), and \( y \) represents the mean return over the next 5 minutes (Equation (4.5)).

A number of authors indicate that claims of excessive returns from intraday trading is potentially based on spurious results (e.g. Kearns et al., 2010; Schulmeister, 2009). Although more signal combinations and complex NN models can be explored to gain higher accuracy, we decide to adopt a parsimonious approach to reduce the possibility of overfitting and spurious results (see Bailey et al., 2014). Our NN model consists of a single feed-forward network model (Figure 4.2).

Following Equation (4.4), we design a single-layer feed-forward network (FFN) regression model with 3 lagged buy and sell signals and 1 RSI signal as inputs with \( d \) hidden units. Each hidden neuron, \( z_j \), can be represented mathematically as

\[
z_j = f \left( w_{1,0} + \sum_{k=1}^{4} w_{1,k} x_k \right),
\]

(4.7)

where \( w_{1,0} \) represents the bias parameter of a hidden neuron and \( w_{1,k} \) represents the weight parameter between the \( k \)-th input and the hidden neuron \( z_j \). The output neuron can then be represented as

\[
y_t = f \left( w_{2,0} + \sum_{j=1}^{d} w_{2,j} z_j \right),
\]

(4.8)
where $w_{2,0}$ represents the bias parameter of the output neuron, and $w_{2,j}$ represents the weight parameter between the $j$-th hidden neuron and output neuron. Although in Chapter 3 we used the logistic function as our activation function, during preliminary testing we identified that in this configuration the hyperbolic tangent function presented better results. Hence, in this experiment the activation function $f(u)$ takes the form
\begin{equation}
 f(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}.
\end{equation}

For our model identification and calibration, all $2 \times 3 = 6$ permutations of the parameter constellations combined with a range of up to 1000 training epochs are considered during the in-sample training (see Table 4.1). For efficiency purposes, we use the Levenberg–Marquardt algorithm for our backpropagation learning (Marquardt, 1963).

Model selection is performed by testing and comparing all settings with the different combinations of model parameters, identifying the model with the highest Sharpe ratio (see
4.2 Part 1 - Enhancing the intraday trading performance of neural networks using a dynamic volatility clustering fuzzy filter

Table 4.1: Parameters tested for NN model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter Value Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Data Size (price points)</td>
<td>{510, 1020}</td>
</tr>
<tr>
<td>Number of Hidden Units, $d$</td>
<td>{5, 10, 20}</td>
</tr>
<tr>
<td>Training Epochs</td>
<td>{1,...,1000}</td>
</tr>
</tbody>
</table>

Equation (3.20)). Trading strategies with higher Sharpe ratios are always preferred as they translate into better risk-adjusted performance due to their higher risk premium for a given level of risk. This approach falls in line with our findings in Chapter 3 and is purposely selected to compare the improvements that can be attained by our new fuzzy logic extension which is discussed in Section 4.2.5.

4.2.3 Standard trading algorithm

The trading algorithm takes the next 5-minute average return prediction, every minute, from the NN model, and performs trades based on specific rules. The general structure of the trading algorithm is kept similar to the approach presented in Chapter 3, although here we perform a re-calibration due to a change in price frequency from 5 minutes to 1 minute. The similarity in approach was purposely kept in order to attribute any performance improvements to the proposed fuzzy logic extension.

The objective of our trading algorithm is to generate buy ($\Omega = long$), sell ($\Omega = short$) or stand-by ($\Omega = 0$) signals. For buy or sell signals the predicted return value has to be greater (smaller) than the upper (lower) return threshold $RT$, otherwise the trade signal is set to stand-by. We consider different $RT$s between 0.05% and 0.1% for each stock to cover the fixed transaction costs defined in the algorithm, which amount to 0.05%. In our in-sample period we try to identify the least possible threshold until the algorithm resulted in profitable results at the end of the in-sample period. An increased threshold results in a substantial reduction in intraday trades (Figure 4.6). From our tests the lower limit of
0.07% is selected. In line with Vanstone and Finnie (2010), our trade decision should take account of the individual neural network threshold and also whether the signal is increasing in strength from its previous forecast. By denoting the predicted mean return at time $t$ as $\hat{y}_t$, the position $\Omega$ taken at time $t$ is:

$$
\Omega_t = \begin{cases} 
\text{long} : & \hat{y}_t > RT, \hat{y}_t > \hat{y}_{t-1} \\
\text{short} : & \hat{y}_t < -RT, \hat{y}_t < \hat{y}_{t-1} \\
0 : & \text{otherwise.}
\end{cases}
$$

Similar to Chapter 3, in our experiments we train the models on a daily rolling window basis (see Figure 3.3) for a period covering 100 in-sample days. On a daily basis the model is then tested to predict the price movements of the following day. This totals 51,000 1-minute price points for each sample and is repeated for each parameter combination. The best performing model against each measure is finally tested on the following 100-day period.
out-of-sample using a similar moving window approach (see the next section on evaluation).

In this setup, we apply a constant transaction cost of 10 GBP per trade, per direction, and assume that a trader is willing to invest a fixed 50,000 GBP per position. Every 1 minute the trading algorithm takes a decision based on the predicted trading direction, the selected return band and the elapsed position holding time. If the signal is to go long (short) the system will buy (sell) 50,000 GBP worth of stock at the current market price. The total investment at each point in time is limited to 250,000 GBP (which is also the amount of the start capital for trading one specific asset). For this experiment only positions in the same direction are allowed at the same time. This is done to specifically eliminate the hedging effect of opposing positions which, as a result, can overestimate the performance of the algorithm.

When a trade is placed, if after a trade duration \( (TD) \) of 5 minutes the signal is still in the same direction, then the position is kept for another period of the same length. If, on the contrary, the signal has changed, then the position is closed. This short \( TD \) was selected to simulate real live intraday scenarios in which traders might wish to come out of the market quickly. Brabazon and O’Neill (2006) showed that the similar use of extended close in intraday trading scenarios can perform better than standard stop-loss, take-profit and buy-and-hold strategies. All open positions are closed at end of day, resulting in the system not holding any positions overnight.

### 4.2.4 Intraday volatility and its challenges

An important aspect which goes hand in hand with trend identification, and hence on the outlook of the potential return, is the risk involved, which is statistically represented by volatility.

This, however, brings about a dual contrasting agenda for traders. At face value, volatility brings negative risk associations indicating pull-back signs due to the possibility of cliff-like
drops in prices in the underlying unstable scenarios. However, wide ranging moves can also be considered as profitable opportunities for traders to lock extraordinary gains. Adequate volatility is still required for favourable price movements sufficient to overcome the trading costs. Moreover, as stated by Kearns et al. (2010), in the high-frequency trading domain, the typically shorter holding periods demand more extreme (and thus less frequent) relative price movements to break the costs barrier.

The standard econometric tool to model the volatility clustering of daily (“low-frequency”) returns is the GARCH framework (Engle and Patton, 2001; Hansen and Lunde, 2005; Poon and Granger, 2003, 2005). In a high-frequency setting, however, the literature suggests a different approach. The idea of using intraday daily data to measure volatility was first introduced by Merton (1980) who noted that under the diffusion assumption, volatility can be estimated to an arbitrary precision using the sum of intraday squared returns. Concurrent research lead to the investigation of alternative volatility estimators.

By considering a continuous-time stochastic process for log-prices, \( \log(p_t) \), Andersen et al. (2001) introduced a natural estimator for the quadratic variation of the process which is commonly known as realised variance (RV). Suppose that prices \( p_0, \ldots, p_n \) are observed at \( n + 1 \) times, equally spaced on \([0, t]\). Using these returns, the \( n \)-sample RV, defined as

\[
RV_t = \sum_{t=1}^{n} (\log(p_t) - \log(p_{t-1}))^2
\]

(4.11)

converges in probability to the quadratic variation as the time interval between observations becomes small enough Andersen et al. (2003). Contrary to classical models in mathematical finance where volatility is considered latent, the literature on realised volatility (RV) considers volatility as an observable variable (Andersen et al., 2001) and hence can be easily analysed and applied in our model.

A common problem with using RV is that if the interval is “too small”, the RV estimator can suffer from market microstructure noise present in high-frequency financial data, which
4.2 Part 1 - Enhancing the intraday trading performance of neural networks using a dynamic volatility clustering fuzzy filter

can make the estimator biased. The literature usually suggests a sampling interval of 5 minutes (McAleer and Medeiros, 2008). A recent survey by Andersen et al. (2009) shows a growing literature of volatility estimators which can accommodate market microstructure noise. In the subsequent sub-section we make use of the average RV estimator as it has been found to perform very well in simulation studies and in practice (Christoffersen, 2011). This is explained in the next section.

4.2.5 Enhancing trading model with a dynamic volatility filter

A number of studies do not consider volatility when designing NNs (see Kaastra and Boyd, 1996; Vanstone and Finnie, 2009, 2010). Others (e.g. Brabazon and O’Neill, 2006) simply suggest to avoid highly volatile periods. Although volatility is a proxy of uncertainty, merely avoiding volatility, however, can result in missed opportunities for large returns. The objective of our fuzzy volatility filter is to identify regions of predicted returns and intraday realised volatility where, based on a number of past trades, have achieved good risk-return results, and hence filter out those that result in losses. This is done by creating a simple learning mechanism to approximate the NN sub-regional trading performance which is subsequently used by our trading decision model to automatically control trade signals suggested by the NN and filter out those signals falling in regions with a lower likelihood of success. This improves the overall risk-return performance.

Due to the microstructure noise present in high-frequency data, which can invalidate realised volatility estimators (Andersen et al., 2009), we adopted a simple method called average RV (Christoffersen, 2011), which in simulation studies has been found to perform very well. By adopting this method, the last 30-minute average RV at time \(t\) is calculated as

\[
RV_{t}^{avg} = \frac{1}{5} \sum_{j=1}^{5} RV_{t}^{s,j},
\]  
(4.12)
which represents the average of 5 $RV$ measures (Equation (4.11)) using 5 grids of 5-minute returns overlapping by 1 minute over the 30-minute slot. The advantage of using $RV_{t,avg}$ is that it is a moving average itself, and hence consecutive values will be (by construction) highly autocorrelated, providing a good linear predictor in our NN model. After examining the high correlation between the 30-minute $RV_{t,avg}$ measure at time $t$ and the 30-minute $RV_{t,avg}$ at time $t + 5$ minutes (correlation coefficient of 0.85, significant at 95% confidence level), we decided to adopt a parsimonious approach and use the $RV_{t,avg}$ at time $t$ as a proxy for $RV_{t,avg}$ in the next 5 minutes.

Our objective is to apply the incompatibility principle of Zadeh (1973). In this context, reducing the complexity tends to also reduce the uncertainty, and hence trading risk, of the underlying NN model, which is something we aim to achieve. We apply clustering to detect the possible groupings that exist to reduce the complexity of the model (Delgado et al., 1997).

Most common fuzzy systems define relationships between variables by means of fuzzy if-then rules, which can be viewed as a local description of the system under consideration. A recent survey by Dutta and Angelov (2010) indicates that data clustering is one of the approaches that have been applied extensively to automatically generate rules from data. One of the most popular methods for finding fuzzy partitions is FCM clustering (Bezdek, 1981). It has also been shown that FCM can be extended to support big data sets (see Havens et al., 2012), an important requirement when dealing with high-frequency trading data.

The input data is organised into data pairs from trade events which happened in the past $h$ days. We denote the input to the fuzzy controller as a matrix $\psi = \{G, i\}$. We define $G$ as a set of points $g$ in the trend-volatility space, where as defined in Section 4.2.1 each point consists of our space coordinates $\{y_{t+1}, RV_{t+1}\}$. Each point $g$ represents a past simulated NN trade where $y_{t+1}$ is the expected return signal over the next 5 minutes and $RV_{t+1}$ is the realised volatility at the time the trade was placed. Vector $i$ represents the corresponding profit resulting from each trade. Following our findings in Chapter 3, we define our trade
4.2 Part 1 - Enhancing the intraday trading performance of neural networks using a dynamic
volatility clustering fuzzy filter

profits *inclusive of transaction costs*. Since predicted return and volatility variables have
different bases, these are standardised and rescaled to have a mean of zero and a standard
deviation of one before being fed into the algorithm. This provides equal influence weight of
each variable on the clustering algorithm, which is described next.

As a next step, we identify data clusters across the expected return and volatility dimen-
sions. Depending on the memory size parameter defined in number of days, \( h \), \( G \) will consist
of \( l \) data points, each point represented by row vector \( \mathbf{g} \), which is dependent on the number of
executed trades covered by our memory span. Let \( z \) denote the fuzziness index. Furthermore,
define \( \alpha \) as the number of clusters, \( d_{uv} = \| \mathbf{g}_u - \mathbf{c}_v \|^2 \) as the Euclidean distance between the
\( u \)-th realisation and the current \( v \)-th cluster centre \( \mathbf{c}_v \), and \( d_{uo} = \| \mathbf{g}_u - \mathbf{c}_o \|^2 \) as the Euclidean
distance from the \( u \)-th realisation and the other cluster centres \( \mathbf{c}_o \). For each data point \( \mathbf{g}_u \),
\( \forall u \in [1, l] \), and cluster \( \mathbf{c}_v \), \( \forall v \in [1, \alpha] \), the FCM algorithm iteratively updates the membership
grade \( \mu \) of the \( u \)-th data point to the \( v \)-th cluster

\[ \mu_{uv} = \left( \sum_{j=1}^{\alpha} \left( \frac{d_{uv}}{d_{uo}} \right)^{\frac{2}{z+1}} \right)^{-1}, \]  

and the centre of the \( v \)-th cluster

\[ \mathbf{c}_v = \frac{\sum_{u=1}^{l} \mu_{uv}^{z} \mathbf{g}_u}{\sum_{u=1}^{l} \mu_{uv}^{z}}, \]  

such that the objective function

\[ J_z = \sum_{u=1}^{l} \sum_{v=1}^{\alpha} \mu_{uv}^{z} d_{uv} \]  

is minimised.

The selection of the number of clusters, \( \alpha \), is a typical problem in cluster analysis. The
literature proposes a number of approaches to construct \( \alpha \) compact and distinct clusters
that have small distances (variances) between points within the clusters and large distances between points belonging to different clusters Xu et al. (2005). Following our findings in Chapter 3, we take a different approach in selecting $\alpha$ by identifying the best number of clusters as part of our calibration process with the objective to maximise global risk-return ratios based on past trading performance.

A set of $\alpha$ rules are dynamically created using fuzzy sets derived directly from the components of the $\alpha$ centroids found by clustering on the input space and is updated on a daily basis using local fuzzy cluster risk-return trading information. These fuzzy sets are not projections of the clusters but fuzzy sets induced in the input space by the fuzzy clusters. The rules take the form of Equation (4.1). Since the summation of membership for each data point should be equal to 1 (Equation (4.13)), we decided that for our approximate calculation of the local regional performance function we apply a threshold of $\theta = 1/\alpha$. For our risk-return function $P(i)$ we applied the Sharpe ratio (Equation (3.20)). Hence, our rules take the form of

$$IF \quad g \in C_j \quad THEN \quad i_{loc} = \text{Sharpe}(i \mid \mu_{C_j}(g) > 1/\alpha),$$

(4.16)

where $i_{loc}$ is the approximated local risk-adjusted performance.

It is well known that machine learning algorithms may suffer from spurious inferences and that precise relationships that might have been stable in the past may not hold in the longer term future, hence becoming unreliable (?). For this reason, when considering a new data point, rather than assigning the point to a specific cluster, we adopt a fuzzy approach by obtaining the degree of membership of the new point to the identified clusters. By applying a threshold, rather than a crisp cluster boundary, the approximate local performance calculation in each cluster is influenced by the closest points to each cluster centre. This means that points close to cluster boundaries can be used in the approximate local performance calculation.
Fig. 4.4: Decision surface example as identified by our volatility clustering fuzzy filter showing different levels of Sharpe ratio regions (see colour bar).

of multiple clusters. When projecting the cluster performance on a chart, this produces a smooth, and interpretable, decision surface showing the highest Sharpe ratio across the return prediction strength and realised volatility space (Figure 4.4).

For each new data point, the decision surface is used by our trading algorithm to take trade entry decisions. Every new, minute-by-minute, trend signal generated by the underlying NN module, together with the corresponding realised volatility measure, is passed as a new data point $g$ to the fuzzy logic module to obtain the approximate regional performance as

$$S = \sum_{j=1}^{\alpha} \mu_{C_j}(g) \delta_{i_{loc,j}},$$

(4.17)

giving weight to the regional performance of the closest clusters.
Table 4.2: Additional parameters tested for the NN and the fuzzy volatility filter

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter Value Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trade History Data Size, $h$ (days)</td>
<td>${20, 40, 60}$</td>
</tr>
<tr>
<td>Number of Clusters, $\alpha$</td>
<td>${4, 6, 8, 10}$</td>
</tr>
</tbody>
</table>

Two sets of rules were tested for our trading algorithm. The first test (NN-FCM1) enhances the standard trading rules (Equation (4.10)) by allowing trades only if the regional performance, $S$, at point $g$ is greater than the minimum positive risk-return performance attained amongst all clusters:

$$\Phi_t = \begin{cases} 
\text{long} : & \hat{y}_t > RT, \hat{y}_t > \hat{y}_{t-1}, S > \min(i^+_{loc}) \\
\text{short} : & \hat{y}_t < -RT, \hat{y}_t < \hat{y}_{t-1}, S > \min(i^+_{loc}) \\
0 & \text{otherwise.} 
\end{cases} \quad (4.18)$$

In our second test (NN-FCM2) the algorithm was allowed to trade only if the regional performance, $S$, at point $g$ is greater than the average risk-return performance which is attained across all clusters:

$$\Phi_t = \begin{cases} 
\text{long} : & \hat{y}_t > RT, \hat{y}_t > \hat{y}_{t-1}, S > \max(\bar{i}_{loc}, 0) \\
\text{short} : & \hat{y}_t < -RT, \hat{y}_t < \hat{y}_{t-1}, S > \max(\bar{i}_{loc}, 0) \\
0 & \text{otherwise.} 
\end{cases} \quad (4.19)$$

Two additional parameters were included in these extended models. These are the number of clusters to be identified and the number of trades, based on the number of past trade days (see Table 4.8). Similar to the identification and calibration of our standard NN model, we again test and compare all possible combinations of model parameters for the in-sample training, but we now have to additionally consider the $3 \times 4 = 12$ permutations of the new parameters.
4.2 Part 1 - Enhancing the intraday trading performance of neural networks using a dynamic volatility clustering fuzzy filter

4.2.6 Model selection and experiment approach

The trading systems in this experiment are developed using high-frequency trade data for a set of 7 stocks listed on the London Stock Exchange (Table 4.3) during the period 01/06/2007 to 30/06/2008 (excluding weekends, holidays and after-hours trading). The plot of normalised stock prices (Figure 4.5) indicates a variety of trend patterns across the stock selection which helps to avoid bias and ensures firmer results. For our time series we use 1-minute prices from which we generate a combination of moving average signals. Since the London Stock Exchange operates between 8:00 and 16:30 GMT, this produced 510 price data points per day for each stock.

Model selection is then performed by identifying the configuration with the highest Sharpe ratio obtained in the 100-day in-sample period. For evaluation, all models are then tested on the following 100 days in the out-of-sample period, applying a day-by-day moving window approach. A number of statistics are collected to analyse model performance. Similar to Chapter 3, we adopt the number of trades, Sharpe ratio and Win ratio as our first set of performance measures. However, in Part 1 we opt not to include the Sortino ratio since the intention is to focus on a limited set of performance measures and obtain a first indication of the performance improvements attained by our NN fuzzy extension (expanded later in Part 2 of this chapter). Instead, here we decide to include the profit per trade (expressed in basis

Table 4.3: FTSE 100 companies analysed in Part 1 of this chapter

<table>
<thead>
<tr>
<th>Data Set No.</th>
<th>Symbol</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ANTO</td>
<td>Antofagasta</td>
</tr>
<tr>
<td>2</td>
<td>AL.</td>
<td>Alliance &amp; Leicester</td>
</tr>
<tr>
<td>3</td>
<td>BLND</td>
<td>British Land</td>
</tr>
<tr>
<td>4</td>
<td>BAY</td>
<td>British Airways</td>
</tr>
<tr>
<td>5</td>
<td>HMSO</td>
<td>Hammerson</td>
</tr>
<tr>
<td>6</td>
<td>REX</td>
<td>Rexam</td>
</tr>
<tr>
<td>7</td>
<td>JMAT</td>
<td>Johnson Matthey</td>
</tr>
</tbody>
</table>
We also consider the standard buy-and-hold strategy with 100 daily trades, buying at the opening price, holding it over the course of the trading day, and selling at the closing price. The inclusion of this zero-intelligence benchmark model is to assess the usefulness and potential outperformance of AI-controlled algorithmic trading strategies in general. We also construct an equally weighted portfolio (with a start balance of $7 \times 250,000$ GBP) to gauge the success of a particular strategy if applied to several assets independently.
4.2 Part 1 - Enhancing the intraday trading performance of neural networks using a dynamic volatility clustering fuzzy filter

Table 4.4: Trading performance following the 100-day out-of-sample period. Bold figures represent the highest value for a given asset across all four models.

<table>
<thead>
<tr>
<th>Model Company</th>
<th>Number of Trades</th>
<th>Win Ratio</th>
<th>Sharpe Ratio</th>
<th>Profit per Trade [bp]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Buy &amp; Hold</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL.</td>
<td>100</td>
<td>0.5300</td>
<td>0.0815</td>
<td>5.4700</td>
</tr>
<tr>
<td>ANTO</td>
<td>100</td>
<td>0.4700</td>
<td>-0.0478</td>
<td>-2.4381</td>
</tr>
<tr>
<td>BAY</td>
<td>100</td>
<td>0.5100</td>
<td>-0.0722</td>
<td>-4.5299</td>
</tr>
<tr>
<td>BLND</td>
<td>100</td>
<td><strong>0.5200</strong></td>
<td><strong>0.0398</strong></td>
<td><strong>1.5661</strong></td>
</tr>
<tr>
<td>HEMO</td>
<td>100</td>
<td><strong>0.5500</strong></td>
<td>0.0749</td>
<td><strong>2.8716</strong></td>
</tr>
<tr>
<td>REX</td>
<td>100</td>
<td><strong>0.5400</strong></td>
<td>0.0326</td>
<td><strong>1.0190</strong></td>
</tr>
<tr>
<td>JMAT</td>
<td>100</td>
<td>0.5100</td>
<td>-0.0150</td>
<td>-0.4881</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td></td>
<td>0.5186</td>
<td>0.0134</td>
<td>0.1356</td>
</tr>
<tr>
<td><strong>Standard NN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL.</td>
<td>8804</td>
<td>0.5583</td>
<td>0.1173</td>
<td>0.8394</td>
</tr>
<tr>
<td>ANTO</td>
<td>5696</td>
<td>0.5323</td>
<td>0.0622</td>
<td>0.4849</td>
</tr>
<tr>
<td>BAY</td>
<td>4994</td>
<td>0.4880</td>
<td>-0.0034</td>
<td>0.0225</td>
</tr>
<tr>
<td>BLND</td>
<td>3345</td>
<td>0.4876</td>
<td>0.0253</td>
<td>0.1837</td>
</tr>
<tr>
<td>HEMO</td>
<td>4258</td>
<td>0.5021</td>
<td>0.0640</td>
<td>0.4567</td>
</tr>
<tr>
<td>REX</td>
<td>2947</td>
<td>0.5022</td>
<td>0.0225</td>
<td>0.1484</td>
</tr>
<tr>
<td>JMAT</td>
<td>2034</td>
<td>0.5182</td>
<td>0.0276</td>
<td>0.1871</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td></td>
<td>0.5127</td>
<td>0.0451</td>
<td>0.1136</td>
</tr>
<tr>
<td><strong>NN-FCM1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL.</td>
<td>8040</td>
<td>0.5606</td>
<td>0.1364</td>
<td>0.9051</td>
</tr>
<tr>
<td>ANTO</td>
<td>5367</td>
<td>0.5334</td>
<td>0.0683</td>
<td>0.5358</td>
</tr>
<tr>
<td>BAY</td>
<td>2266</td>
<td><strong>0.5009</strong></td>
<td><strong>0.0295</strong></td>
<td><strong>0.2709</strong></td>
</tr>
<tr>
<td>BLND</td>
<td>1990</td>
<td>0.4925</td>
<td>0.0184</td>
<td>0.1176</td>
</tr>
<tr>
<td>HEMO</td>
<td>2941</td>
<td>0.4974</td>
<td>0.0680</td>
<td>0.4247</td>
</tr>
<tr>
<td>REX</td>
<td>1771</td>
<td>0.5138</td>
<td><strong>0.0377</strong></td>
<td>0.2888</td>
</tr>
<tr>
<td>JMAT</td>
<td>1150</td>
<td><strong>0.5287</strong></td>
<td><strong>0.0475</strong></td>
<td><strong>0.3243</strong></td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td></td>
<td>0.5182</td>
<td>0.0580</td>
<td>0.1516</td>
</tr>
<tr>
<td><strong>NN-FCM2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL.</td>
<td>3289</td>
<td><strong>0.5731</strong></td>
<td><strong>0.1626</strong></td>
<td>1.3763</td>
</tr>
<tr>
<td>ANTO</td>
<td>1835</td>
<td><strong>0.5564</strong></td>
<td><strong>0.1078</strong></td>
<td><strong>1.0218</strong></td>
</tr>
<tr>
<td>BAY</td>
<td>2036</td>
<td>0.4838</td>
<td>0.0050</td>
<td>0.0462</td>
</tr>
<tr>
<td>BLND</td>
<td>575</td>
<td>0.4800</td>
<td>0.0237</td>
<td>0.1877</td>
</tr>
<tr>
<td>HEMO</td>
<td>850</td>
<td>0.5424</td>
<td><strong>0.0898</strong></td>
<td>0.7358</td>
</tr>
<tr>
<td>REX</td>
<td>1680</td>
<td>0.4976</td>
<td>-0.0228</td>
<td>-0.1653</td>
</tr>
<tr>
<td>JMAT</td>
<td>966</td>
<td>0.5197</td>
<td>0.0426</td>
<td>0.3098</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td></td>
<td><strong>0.5219</strong></td>
<td><strong>0.0584</strong></td>
<td><strong>0.1746</strong></td>
</tr>
</tbody>
</table>
### Table 4.5: Performance percentage difference of standard NN, NN-FCM1 and NN-FCM2 when compared with buy-and-hold method following the 100-day out-of-sample period.

<table>
<thead>
<tr>
<th>Company</th>
<th>Win Ratio</th>
<th>Sharpe Ratio</th>
<th>Profit per Trade</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard NN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL.</td>
<td>+2.83</td>
<td>+43.96</td>
<td>-84.65</td>
</tr>
<tr>
<td>ANTO</td>
<td>+6.23</td>
<td>+230.08</td>
<td>+119.89</td>
</tr>
<tr>
<td>BAY</td>
<td>-2.20</td>
<td>+95.23</td>
<td>+99.32</td>
</tr>
<tr>
<td>BLND</td>
<td>-3.24</td>
<td>-36.57</td>
<td>-88.27</td>
</tr>
<tr>
<td>HMSO</td>
<td>-4.79</td>
<td>-14.55</td>
<td>-84.10</td>
</tr>
<tr>
<td>REX</td>
<td>-3.78</td>
<td>-30.98</td>
<td>-85.43</td>
</tr>
<tr>
<td>JMAT</td>
<td>+0.82</td>
<td>+284.00</td>
<td>+138.33</td>
</tr>
<tr>
<td><strong>EW-Portfolio</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.59</td>
<td>+236.26</td>
<td>-16.21</td>
<td></td>
</tr>
<tr>
<td><strong>NN-FCM1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL.</td>
<td>+3.06</td>
<td>+67.36</td>
<td>-83.45</td>
</tr>
<tr>
<td>ANTO</td>
<td>+6.34</td>
<td>+242.97</td>
<td>+121.98</td>
</tr>
<tr>
<td>BAY</td>
<td>-0.91</td>
<td>+140.83</td>
<td>+105.98</td>
</tr>
<tr>
<td>BLND</td>
<td>-2.75</td>
<td>-53.79</td>
<td>-92.49</td>
</tr>
<tr>
<td>HMSO</td>
<td>-5.26</td>
<td>-9.21</td>
<td>-85.21</td>
</tr>
<tr>
<td>REX</td>
<td>-2.62</td>
<td>+15.64</td>
<td>-71.66</td>
</tr>
<tr>
<td>JMAT</td>
<td>+1.87</td>
<td>+416.67</td>
<td>+166.43</td>
</tr>
<tr>
<td><strong>EW-Portfolio</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.04</td>
<td>+322.62</td>
<td>+11.82</td>
<td></td>
</tr>
<tr>
<td><strong>NN-FCM2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL.</td>
<td>+4.31</td>
<td>+99.51</td>
<td>-74.84</td>
</tr>
<tr>
<td>ANTO</td>
<td>+8.64</td>
<td>+325.52</td>
<td>+141.91</td>
</tr>
<tr>
<td>BAY</td>
<td>-2.62</td>
<td>+106.93</td>
<td>+101.02</td>
</tr>
<tr>
<td>BLND</td>
<td>-4.00</td>
<td>-40.45</td>
<td>-88.02</td>
</tr>
<tr>
<td>HMSO</td>
<td>-0.76</td>
<td>+19.89</td>
<td>-74.38</td>
</tr>
<tr>
<td>REX</td>
<td>-4.24</td>
<td>-169.94</td>
<td>-116.22</td>
</tr>
<tr>
<td>JMAT</td>
<td>+0.97</td>
<td>+384.00</td>
<td>+163.46</td>
</tr>
<tr>
<td><strong>EW-Portfolio</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+0.33</td>
<td>+335.71</td>
<td>+28.80</td>
<td></td>
</tr>
</tbody>
</table>

### 4.2.7 Analysis and results

Table 4.4 compares the trading results of our four models applied to the seven companies with respect to a number of indicators. To analyse the effectiveness of our models we also present the performance percentage difference against the standard buy-and-hold strategy (Table 4.5) and also against the standard neural network model (Table 4.6).
4.2 Part 1 - Enhancing the intraday trading performance of neural networks using a dynamic volatility clustering fuzzy filter

Table 4.6: Performance percentage difference of NN-FCM1 and NN-FCM2 when compared with the standard NN model following the 100-day out-of-sample period.

<table>
<thead>
<tr>
<th>Company</th>
<th>Win Ratio</th>
<th>Sharpe Ratio</th>
<th>Profit per Trade</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN-FCM1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL.</td>
<td>+0.23</td>
<td>+16.26</td>
<td>+7.82</td>
</tr>
<tr>
<td>ANTO</td>
<td>+0.11</td>
<td>+9.91</td>
<td>+10.50</td>
</tr>
<tr>
<td>BAY</td>
<td>+1.29</td>
<td>+956.50</td>
<td>+982.96</td>
</tr>
<tr>
<td>BLND</td>
<td>+0.49</td>
<td>-27.15</td>
<td>-35.99</td>
</tr>
<tr>
<td>HMSO</td>
<td>-0.47</td>
<td>+6.25</td>
<td>-7.00</td>
</tr>
<tr>
<td>REX</td>
<td>+1.16</td>
<td>+67.56</td>
<td>+94.58</td>
</tr>
<tr>
<td>JMAT</td>
<td>+1.05</td>
<td>+72.10</td>
<td>+73.33</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td>+0.55</td>
<td>+28.66</td>
<td>+33.46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NN-FCM2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AL.</td>
<td>+1.48</td>
<td>+38.59</td>
<td>+63.96</td>
</tr>
<tr>
<td>ANTO</td>
<td>+2.41</td>
<td>+73.37</td>
<td>+110.72</td>
</tr>
<tr>
<td>BAY</td>
<td>-0.42</td>
<td>+245.33</td>
<td>+250.51</td>
</tr>
<tr>
<td>BLND</td>
<td>-0.76</td>
<td>-6.12</td>
<td>+2.17</td>
</tr>
<tr>
<td>HMSO</td>
<td>+4.03</td>
<td>+40.31</td>
<td>+61.12</td>
</tr>
<tr>
<td>REX</td>
<td>-0.46</td>
<td>-201.33</td>
<td>-211.37</td>
</tr>
<tr>
<td>JMAT</td>
<td>+0.15</td>
<td>+54.35</td>
<td>+65.59</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td>+0.92</td>
<td>+29.58</td>
<td>+53.72</td>
</tr>
</tbody>
</table>

Firstly, all four models have a win ratio above 0.5, with the highest attained in NN-FCM2 (52%), implying that the majority of trades executed were profitable. However, it is to be noted that this measure is not indicative of the actual risk-return performance or even the profitability of the strategy in the long run (as one extreme loss may not be compensated by 99 minor profits).

Secondly, while the B&H method generates the highest profit per trade in 4 out of 7 stocks (AL., BLND, HMSO and REX), it should be noted that it generates the highest losses per trade in 3 out of 7 stocks (ANTO, BAY and JMAT). Although this results in a higher overall profit per trade when compared to Standard NN, both NN-FCM1 and NN-FCM2 outperform the B&H method with an overall profit per trade increase of 11.82% and 28.80% respectively.
As a third point, both the NN-FCM1 and NN-FCM2 models achieve a higher profit per trade at lower risk, excelling in the Sharpe ratio on all stocks, with the only exception in BLND. At an aggregate level, results show over 320% improvement in the Sharpe ratio when compared to B&H method and a 29% improvement over the standard NN model. In general, the results show that NN models making use of our extended fuzzy model volatility filter trade less but yield the highest Sharpe ratio and the highest profit per trade among all AI-optimised strategies.

As identified in our literature review, even though a number of authors have suggested methodologies of how to design and tune NNs when applied for predicting financial time series and trading purposes, little consideration has been given to the time-varying volatility (see Kaastra and Boyd (1996); Vanstone and Finnie (2009, 2010)). Other authors intentionally avoid higher periods of uncertainty by keeping out of the market especially during the initial and end intraday trading periods (Brabazon and O’Neill, 2006), which are well documented to exhibit the highest volatility. A wide body of research claims that volatility exhibits strong intraday periodicity, high persistence and can be predicted with a good degree of accuracy up to an intraday level (see Andersen and Bollerslev (1997); Andersen et al. (2000a)). Our results show that this information can be beneficial for enhancing models such as NNs, an area which so far has been scarcely researched in literature.

4.2.8 Conclusion

Following our findings in Chapter 3 with respect to model tuning and selection when applied to trading purposes, in this chapter we take a more granular perspective and seek to address the challenges presented by time-varying volatility (risk) during the course of a trading day.

A well-known stylised fact in finance literature denotes the strong diurnal pattern exhibited by volatility (Andersen and Bollerslev, 1997; Andersen et al., 2000a). This infers that risk is not constant during the course of a trading day. Although numerous models are
presented in literature that are capable of measuring volatility at a good degree of accuracy up to an intraday level, the literature of how AI-trading algorithms can be optimised to handle \textit{intraday} time-varying volatility is scarce, if considered at all. From a different perspective, this argument contradicts the Efficient Market Hypothesis (Fama, 1965) since, according to the latter, discriminating between different volatility scenarios should in the long term lead to no additional benefits.

In this first part of the chapter, our first contribution is a novel approach based on clustering techniques and fuzzy logic which adapts the underlying signalling model to the time-varying intraday risk. To show the effectiveness of our approach, in our experiment we combine our extension with an NN model that time and again has proven its popularity in a number of surveys. Our approach extracts fuzzy rules dynamically using FCM clustering and makes use of local fuzzy cluster information which is subsequently used by our dynamic fuzzy model to approximate the NN trading performance across different volatility states. This produces a decision surface that is used by our trading algorithm to control intraday trade signals suggested by the NN. Results show that our model obtains better risk-return performance than standard NN and buy-and-hold methods, indicating that intraday time-varying volatility information can improve NN trading models.

Despite the long-standing Efficient Market Hypothesis, a number of authors claim a link between the profitability of technical trading rules and volatility (Gradojevic and Gençay, 2013; Han et al., 2013; Holmberg et al., 2013; LeBaron, 1999; Schulmeister, 2006). As our second contribution, we extend these claims to shorter-term intraday horizons. Our positive results formulate the basis for more exploration and rigorous testing of this finding.

Consequently, in the second part of this chapter we seek further trading improvements by extending our proposed approach to automate money management decisions, supported by a number of statistical tests. The objective is to address the challenge of optimising capital
allocation across different levels of intraday risk scenarios with the ultimate objective of increasing overall risk-adjusted performance.
4.3 Part 2 – A dynamic fuzzy money management approach for controlling the intraday risk-adjusted performance of AI-trading algorithms

Part 1 of this chapter provides evidence that by discriminating across intraday trend signals and volatility levels, and subsequently targeting trades in more profitable regions, helps to improve risk-adjusted performance. The corollary to this approach is to ensure that filtering trades from specific unprofitable trend-volatility regions does not reduce the overall intraday trading activity with the possibility of unused capital. In this second part of this chapter we address the problem of optimising capital allocation for intraday trading using AI-driven money management decisions with the objective to increase both risk-adjusted performance and overall profitability. This is further supported by a rigorous experiment approach and statistical tests which also corroborate our initial findings.

Our approach stems from two gaps in the computational finance literature. Firstly, although most AI literature focuses on identifying market direction, traders in financial markets, on a daily basis, are repetitively presented with a sequence of decisions, with market direction being only one piece of the puzzle. This presents a very limited view of the applicability of AI in trading scenarios. Moreover, many studies in the existing literature do not reflect the rigid constraints that typically govern the trading desk (see also Vanstone and Tan, 2003). In particular, Pardo (2011) states that position sizing is often not appreciated and poorly understood in trading strategy design.

Secondly, as was elaborated upon in Section 4.2, although the proliferation of high-frequency data led to new measures of variation (Andersen et al., 2001; Barndorff-Nielsen and Shephard, 2004, 2006), making it possible to be predicted with a good degree of accuracy up to an intraday level (Andersen et al., 2000a), the use of this information for intraday trading purposes is rarely considered. In our opinion, reverting to AI models solely for
market movement predictions with little consideration to the time-varying market uncertainty (risk) (Son et al., 2012) portrays an incongruent view by financial market practitioners since investors are mostly interested in risk-adjusted performance.

Our goal in Part 2 of this chapter is to address these two, albeit interrelated, literature gaps with the ultimate objective to enhance the risk-adjusted performance of trading algorithms in an intraday stock trading scenario. In particular, we present an innovative method which identifies, in addition to trend direction signals, the optimum capital allocation across an intraday trading period by dynamically adapting the levels of trade frequency and position sizes (two common decisions taken by traders) based on different degrees of expected return and volatility (uncertainty) over short intraday horizons. This also sheds more light on the theoretical claims of Holmberg et al. (2013) that market inefficiency, and hence the profitability of technical rules, can be linked to different volatility periods.

Finding the optimum level of trade frequency and position size along continuously changing intraday market conditions can be a non-trivial task. Although volatility is typically linked to risk, sufficient market volatility is required to ensure that changes in prices exceed transaction costs. Several studies suggest a fixed return threshold filter to avoid small unprofitable movements (Kaastra and Boyd, 1996; Vanstone and Finnie, 2009, 2010). Holmberg et al. (2013) show that increasing the return filter size results in a better success rate and average return. However, the same authors show that this leads to a reduction in the number of trades, hence reducing the investors’ potential profits. Other authors (e.g. Brabazon and O’Neill, 2006) prefer to avoid high volatility (uncertainty) completely by staying out of the market during periods when volatility goes beyond a specified fixed threshold, hence avoiding the risk of possibly strong adverse market movements. This, however, reduces the opportunity of possibly extraordinary gains. Also, Pardo (2011) claims that if a sound position sizing strategy or algorithm is not employed, the effective rate at which trading equity is compounded will remain sub-optimal.
These arguments cast the trade frequency and position size decisions in the context of the better management of uncertainty. We make use of the model presented in Section 4.2 and extend the idea to identify fuzzy clusters which in turn allow for the extraction of fuzzy rules. Their combined result produces a decision surface across the trend and volatility space that is used to adapt trade frequency and position sizing levels based on local (rather than global) regional performance.

With respect to the AI literature, many authors (see Lawrence et al., 1997; Luengo et al., 2009; Prechelt, 1996) question the validity of the experimental and statistical frameworks adopted in many published academic studies. Recently, Bailey et al. (2014) stated that although there are many academic studies that claim to have identified profitable investment strategies, their reported results are almost always based on in-sample statistics. Overfitting a trading strategy on in-sample data can produce (seemingly) impressive results during simulation but often also a devastatingly poor performance during real-time trading (see Pardo, 2011). In this second part of the chapter, we address these criticisms and adopt a thorough experimental framework using a number of independent repeated trials conducted on a set of stocks listed on the London Stock Exchange. Our statistical tests on out-of-sample data reveal that the proposed fuzzy money management approach leads to significant improvements in both risk-adjusted performance and profitability when compared to standard NN and buy & hold methods.

The structure of Part 2 of this chapter is organised as follows. In Section 4.3.1, we describe our experiment approach and explain the main model components and their underlying prediction and trading algorithms. In Section 4.3.2, we provide more detail on our proposed enhanced algorithm. This is followed by our experiment approach and considerations, discussed in Section 4.3.3. In Section 4.3.4, we present and discuss our results. Section 4.3.5 concludes.
4.3.1 Method

In this experiment, we re-use the same setup defined in Part 1 (see Figure 4.1). Again, the process starts from a given collection of $N$ input-output data training pairs,

\[
\left( x^{(1)} : y^{(1)} \right), \left( x^{(2)} : y^{(2)} \right), \ldots, \left( x^{(N)} : y^{(N)} \right)
\]

where

\[
\begin{align*}
x^{(1)} &= [s_{1,t-N}, s_{2,t-N}, s_{3,t-N}, s_{4,t-N}], & y^{(1)}_t &= y_{t-N+1} \\
x^{(2)} &= [s_{1,t-N+1}, s_{2,t-N+1}, s_{3,t-N+1}, s_{4,t-N+1}], & y^{(2)}_t &= y_{t-N+2} \\
&\vdots & & \\
x^{(N)} &= [s_{1,t-1}, s_{2,t-1}, s_{3,t-1}, s_{4,t-1}], & y^{(N)}_t &= y_t
\end{align*}
\]

(4.20)

where $y_t$ is the average return over the next 5 minutes (defined in Equation (4.3)), and the input, $x$, is defined as a set of moving averages and RSI trading rule signals $\{s_1, s_2, s_3, s_4\}$ (defined in Equation (4.5)).

As a motivating example, a standard NN is again employed to forecast, every 1-minute time epoch, the average return over the next 5 minutes. The same calibration approach is used, using the Sharpe ratio as the global risk-adjusted objective function.

Every minute, the NN layer passes a signal to the trading algorithm (further detail in Section 4.2.3). The challenge is to identify a threshold that is high enough that can filter out small price movements, mostly as a result of microstructure effects (see McAleer and Medeiros, 2008), however not being so high that it eliminates trading opportunities. From further investigation on a set of stocks we analysed the drop in the number of trades as the threshold was increased. It was decided to improve on the standard trading algorithm described in Part 1 (Section 4.2.3) and introduce more flexibility by permitting 3 levels of $RT$s. The selection of $RT$ was included as part of the parameter calibration process for each stock. Since models are trained daily on a moving window approach, having flexible return threshold also permits each model to adjust to the general volatility levels that are being experienced by the underlying stock across different time periods. Since calibration was
4.3 Part 2 – A dynamic fuzzy money management approach for controlling the intraday risk-adjusted performance of AI-trading algorithms

Fig. 4.6: Effect on the number of trades as the trading algorithm is tested at various levels of return threshold.

Table 4.7: Combined parameters tested for the NN model and trading algorithm

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter Value Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Data Size (price points)</td>
<td>{510, 1020}</td>
</tr>
<tr>
<td>Number of Hidden Units</td>
<td>{5, 10, 20}</td>
</tr>
<tr>
<td>Return Threshold (RT)</td>
<td>{0.05%, 0.08%, 0.1%}</td>
</tr>
<tr>
<td>Training Epochs</td>
<td>{1, ..., 1000}</td>
</tr>
</tbody>
</table>

based on optimising the Sharpe ratio, in our preliminary testing we identified that this can result in a substantial reduction in trades. Hence an additional constraint was introduced that sets a limit on the lowest number of trades for each stock during the calibration process. Based on the fact that in our experiment scenarios we start with an initial capital of 250,000 GBP and a standard position size of 50,000 GBP, we select a threshold that generate 20 daily trades (minimum) to ensure a wider intraday trading coverage. Hence, in conjunction with the NN parameters, we now consider $2 \times 3 \times 3 = 18$ possible combinations of the parameter settings for our benchmark NN model (Table 4.7).
Algorithm 3 Pseudo code for the extended close algorithm

```
if tradeDuration > 5 minutes then
    if {(signal == trade direction) or (price != trade entry price)} then
        State ← keep open
        tradeOpenTime ← get current time
    else
        State ← close
    end if
else
    tradeDuration ← get current time - tradeOpenTime
end if
```

Another constraint that we added to the standard trading algorithm used in Part 1 (Section 4.2.3) is that to simulate real-world intraday trading environments, where durations can be very short, we consider for each trade a default holding period of five minutes (see Algorithm 3). Closing positions quickly in intraday trading does not necessarily represent stop-losses but can outperform common Buy-and-Hold strategies (Brabazon and O’Neill, 2006). However, we suppress any fast exits when the price does not exhibit any movement from the opening price in the 5-minute time window.

Similar to the approach in Part 1 (Section 4.2.5), we transform our set of signals (Equation (4.20)) into a second data set $\psi = \{G, i\}$. Matrix $G$ consists of data points $g = \{y_{t+1}, RV_{t+1}\}$ from simulated NN trades performed in the last $h$ days, where $y_{t+1}$ is the expected return over the next 5 minutes and $RV_{t+1}$ is the realised volatility at the time the trade was placed. Vector $i$ consists of the corresponding trade profits, inclusive of transaction costs. This data set is used by the money management algorithm that is described in the next section.

4.3.2 Enhancing money management decisions by considering time-varying intraday risk

In a recent study, Holmberg et al. (2013) link the profitability of trading rules and periods of market inefficiency with different levels of volatility at a daily level. When considering higher
frequencies at an intraday level, we know from finance literature that market microstructure effects, trading behaviour (such as changes of trading sessions, lunch breaks, time zone effects, etc.) and the bid-ask bounce can drive prices to be very volatile during the course of a trading day (e.g. McAleer and Medeiros, 2008; Roll, 1984). As our main contribution, we extend the ideas of Holmberg et al. (2013) to a more granular intraday level and investigate the effect of volatility on technical rules’ profitability over short-term intraday horizons. We address this by proposing an automated trading algorithm that identifies preferable pockets of intraday risk-adjusted profitability at different trend and volatility levels with the goal to increase our position frequency and size in successful regions and reduce these in regions that are likely to result in losses.

Similar to Part 1, we employ FCM clustering on the data points in \(G\) (see Section 4.2.5) in order to identify approximate (rather than hard bound) trend and volatility regional spaces which more closely reflect the common terms used by traders in practice, such as “strong positive trend” or “low volatility”, which are not precise in nature. This is contrary to a single fixed threshold approach commonly adopted for return or volatility (Brabazon and O’Neill, 2006; Holmberg et al., 2013; Kaastra and Boyd, 1996; Vanstone and Finnie, 2009, 2010). For example, a return filter of 0.1% should not automatically disqualify a trade signal of 0.0999%. We argue that the underlying price process is too complex to model using hard bound filters and hence we believe that, in line with the incompatibility principle of Zadeh (1973), reducing the complexity by using a fuzzy approach also reduces uncertainty.

In addition to the approach presented in Part 1, as suggested by Pal and Bezdek (1995), further exploration is conducted with regard to the fuzziness index that is used by the FCM algorithm. Values between 1.5 and 2.5 are tested. After examining cluster plots (see for example Figure 4.7) on different stocks and time periods, a fuzzy index value of 1.7 is selected. The criteria are based on identifying a well-distributed set of clusters on the return and volatility space, whilst at the same time avoiding the heavy influence of possible outliers.
Fig. 4.7: Example of clusters found in Hammerson (HMSO) data as identified by FCM.

On a daily basis, a set of $\alpha$ rules are dynamically created using fuzzy sets derived directly from the components of the $\alpha$ centroids (see Figure 4.7) found by using clustering on the input space. The approximate local performance calculation in each cluster is influenced by the trade profits (from vector $i$) of the closest points (past trades defined by vector $g$) to each cluster centre rather than by a single crisp cluster boundary. This means that data points can be members of different clusters. The consolidated effect of the rules results in a decision surface which highlights regions of different Sharpe ratio performance across the trend-volatility space. The decision surface (e.g. Figure 4.8) shows smooth transitions between different performance regions and also the possibility of identifying multiple profitable regions. We believe that this representation presents a more sensible and interpretable trading profile rather than by using a set of hard bound rules that would have resulted through the application of fixed return and volatility thresholds. This also alleviates common criticisms on the use of black box approaches for trading purposes.
Our proposed Neural Network – Fuzzy Money Management (NN-FMM) model enhances the standard trading rules by taking into consideration the approximate local regional performance on the space defined by our trend-volatility space. Holmberg et al. (2013) show that increasing the return filter size results in a better success rate and average return, but possibly it also leads to a reduction in the number of trades, hence reducing the investors’ potential profits. On these lines, our aim is to ensure that improving the risk-adjusted performance of the underlying trading model does not come at the cost of reduced profitability. This means that our money management approach should not simply act as a filter by only keeping the algorithm out of the market in adverse regions and hence underutilise available capital, but also maximise model profitability by allocating more capital to preferable intraday trading scenarios (based on Sharpe ratio performance). Once we obtain the localised performance...
measure $S$, the position size $\omega_t$ is categorised as follows:

$$\omega_t = \begin{cases} 
\text{large position} : & S \geq \lambda_1 \\
\text{small position} : & S \geq \lambda_2, S < \lambda_1 \\
\text{no position} : & S < \lambda_2.
\end{cases}$$ (4.21)

and

$$\lambda = [0, 1],$$

$$\lambda_1 > \lambda_2.$$ (4.22)

To limit the number of possible combinations for $\lambda$ values, in our experiment we opt for values of $\lambda_2 \in \{0, 0.1, 0.2, 0.3, 0.4\}$ and set $\lambda_1 = \lambda_2 + 0.3$. This effectively divides the stock risk-adjusted performance space, and our trade positions, into 3 categories. Contrary to our standard NN trading algorithm, which used equally sized positions of 50,000 GBP per trade, in our enhanced trading algorithm we open positions of 70,000 GBP for high-performance trend-volatility regions ($S > \lambda_1$), 30,000 GBP for medium ones ($S \geq \lambda_2, S < \lambda_1$), and filter out trades in low (or negative) performing regions ($S < \lambda_2$). The remaining entry and exit conditions are applied as in the standard trading algorithm, including the same amount of initial capital of 250,000 GBP.

Three additional model parameters are included in this extended NN approach, which are $\alpha$, the trade history data size, and $\lambda_2$ (see Table 4.8). Given this, additional $4 \times 4 \times 5 = 80$ permutations of the new parameters have to be compared in the model calibration. In the next section, we describe the experiment design process.

### 4.3.3 Experiment approach and considerations

When designing our experiment approach, we paid particular attention to the harsh criticisms put forward by a number of authors (Bailey et al., 2014; Kearns et al., 2010; Lawrence et al.,...
4.3 Part 2 – A dynamic fuzzy money management approach for controlling the intraday risk-adjusted performance of AI-trading algorithms

Table 4.8: Additional parameters tested for the NN and fuzzy volatility filter

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter Value Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trade History Data Size (days)</td>
<td>{5, 10, 20, 40}</td>
</tr>
<tr>
<td>Number of Clusters (α)</td>
<td>{3, 6, 8, 10}</td>
</tr>
<tr>
<td>λ² threshold</td>
<td>{0, 0.1, 0.2, 0.3, 0.4}</td>
</tr>
</tbody>
</table>

Table 4.9: Descriptive statistics of 5-minute returns for companies analysed in Part 2 of this chapter.

<table>
<thead>
<tr>
<th>Company</th>
<th>Symbol</th>
<th>Mean × 10⁻⁶</th>
<th>Std. Dev.</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rexam</td>
<td>REX</td>
<td>-2.5397</td>
<td>0.0013</td>
<td>-6.3234</td>
<td>637.7068</td>
</tr>
<tr>
<td>Sage Group</td>
<td>SGE</td>
<td>-1.0049</td>
<td>0.0013</td>
<td>-0.4793</td>
<td>84.7605</td>
</tr>
<tr>
<td>ICAP</td>
<td>IAP</td>
<td>0.8259</td>
<td>0.0016</td>
<td>-0.2355</td>
<td>139.0775</td>
</tr>
<tr>
<td>ITV</td>
<td>ITV</td>
<td>-7.0749</td>
<td>0.0015</td>
<td>0.4318</td>
<td>144.2745</td>
</tr>
<tr>
<td>Hammerson</td>
<td>HMSO</td>
<td>-2.5716</td>
<td>0.0017</td>
<td>0.7493</td>
<td>99.3494</td>
</tr>
<tr>
<td>Associated British Foods</td>
<td>ABF</td>
<td>-0.5896</td>
<td>0.0011</td>
<td>3.2157</td>
<td>298.4418</td>
</tr>
<tr>
<td>Tesco</td>
<td>TSCO</td>
<td>-2.5689</td>
<td>0.0011</td>
<td>-1.3964</td>
<td>136.6789</td>
</tr>
</tbody>
</table>

1997; Pardo, 2011; Schulmeister, 2009; White, 2000) with respect to serious experimental flaws present in several published studies. In the context of AI-driven trading experiments, we have grouped and addressed these criticisms under five areas:

**Experiment data** Research has shown that trend changes, which by nature occur swiftly, and large shifts in both volatility and liquidity, can have a large and often negative impact on trading performance. Of course, a good, robust model will be more capable of toughing out and trading profitably during such changes (Pardo, 2011). Our data set comprises 1-minute trade records for a set of 7 stocks listed on the London Stock Exchange during a 200-day period between 25/08/2007 and 13/06/2008 (excluding weekends, holidays and after-hours trading). Descriptive statistics of 5-minute returns are shown in Table 4.9. The sample skewness and kurtosis imply that the return distributions are far from being normal. As suggested by Pardo (2011), we ensured to avoid a bias in our experiment results by only picking stocks with a positive trend during the selected period (Figure 4.9).
Realistic constraints  Experiment findings in Chapter 3 show that overestimated profitability can be attained when non-realistic constraints are applied. In our experiments we account for realistic trading hours and no overnight positions. Since the London Stock Exchange operates between 8:00 and 16:30 GMT, this produced 510 1-minute prices per day, hence totalling a time series of 102,000 data points per stock over the whole 200-day period. In line with typical broker costs, in our experiments we apply a transaction cost of 10 GBP per trade per direction.

Model optimisation and performance measures  Many studies ignore the fact that the main interest of investors is risk-adjusted performance. Defining success solely on the grounds of forecast accuracy and win ratios has little practical value (Alves Portela Santos
4.3 Part 2 – A dynamic fuzzy money management approach for controlling the intraday risk-adjusted performance of AI-trading algorithms

et al., 2007; de Faria et al., 2009; Enke and Thawornwong, 2005; Krollner et al., 2010; Medeiros et al., 2006). In our experiments, our primary interest is to optimise and then assess models according to their risk-adjusted performance. Following the Sharpe ratio improvements that we obtained using the NN fuzzy logic extension in Part 1 of this chapter, in Part 2 we strengthen our measure of risk-adjusted performance by additionally including the Sortino ratio (described in Chapter 3, Section 3.2.7) and Calmar ratio. Our decision to introduce Calmar ratio as an additional risk-adjusted performance measure is based on our interest to investigate whether discriminating between different trend-volatility states can also reduce drawdown risk. Unlike the Sortino ratio, which uses downside deviation as a proxy for risk (see Equation (3.21)), the Calmar ratio employs the maximum drawdown to penalise risk. In our set of performance measures we also include win ratio and profit (%).

Model training, selection and validation process  Bailey et al. (2014) show that one can achieve almost any desired Sharpe ratio if one explores large parameter combinations or variations of a strategy, especially if backtesting is performed against an insufficiently large historical dataset. They further show that overfitted strategies are not only likely to disappoint, but, in the presence of memory (as real markets possess), they are actually prone to lose money. As an example, they show that if only five years of daily data are available, no more than forty-five independent model configurations should be tried or we are almost guaranteed to produce strategies with an annualised in-sample Sharpe ratio of 1 but an expected out-of-sample Sharpe ratio of zero.

Similar to our previous experiments, we divide the 200-day data of each stock into 100 days in-sample and 100 days out-of-sample, yielding 51,000 price points for each sample (price series equivalent in size to approximately 204 years of daily prices). In the in-sample training and model selection process, for each parameter setting (see Tables 4.7 and 4.8), we roll forward the window for 100 days on a day-by-day basis, training the model on the previous days’ data (defined as one of the parameters) and forecasting the current (unseen) day
data (depicted in Figure 3.3). For example, in the case of the NN model, we repeat the rolling window approach 18 times for each stock since we have 18 possible model configurations. Only positions in the same direction are allowed in order to avoid the spurious hedging effect of opposing positions that would overestimate the risk-adjusted performance of the algorithm. We selected the specification with the highest Sharpe ratio obtained in the 100-day in-sample period. The selected model configuration was then tested out-of-sample for the next 100 days (51,000 price points) using the same rolling window approach.

**Analysing model robustness and additional benchmark models** To avoid any positively biased conclusions on high-frequency trading returns based on possibly spurious results (as indicated by Schulmeister (2009) and Kearns et al. (2010)), we also run a second set of experiments to compare the risk-adjusted performance measures of 50 independent repeated trials for each stock for both NN and NN-FMM. In each trial we use a random initialisation for NN training in the case of the standard NN (see Lawrence et al., 1997), and both neural network and fuzzy clustering in the case of NN-FMM. Moreover, in each trial the models are trained on a random sample from the training data set presented for each specific day-by-day moving window (in line with Luengo et al., 2009). A total of 700 models (50 models × 7 stocks × 2 AI models) are trained separately on the in-sample period, and then evaluated on the 100-day out-of-sample trading period, at the end of which we recorded the respective performance measures (obtaining 50 × 7 × 2 sets of out-of-sample results). We also checked the distribution of the model results since Lawrence et al. (1997) stressed that one cannot assume that random NN initialisations always lead to a Gaussian distribution.

As additional benchmark models, we also consider two standard buy-and-hold strategies with 100 (daily) trades. In the first buy-and-hold strategy (B&H Long) we buy at the opening price, hold it over the course of the trading day and sell at the closing price. In the second buy-and-hold strategy (B&H Short) we go short at the opening price and close trade at end of day. The inclusion of these zero-intelligence benchmark models is to assess the
usefulness and potential outperformance of AI-controlled algorithmic trading strategies in general. Similarly to Part 1, we also compute the hypothetical performance of an equally weighted portfolio to measure the performance of the strategy if applied to several assets independently and simultaneously.

**4.3.4 Results and evaluation**

In this section, we first analyse and compare the results in the 100 out-of-sample trading days. To check the robustness of our approach, we then study the results of 50 independent repeated trials for each stock covering the same out-of-sample period. This is supported by a number of statistical tests.

**Risk-Adjusted Performance**

From Table 4.10 we can see that the B&H Short method reports positive results on five out of seven stocks, whilst the B&H Long reports positive results on only one stock (SGE). This was an expected result following the general negative trend across most stocks identified in Table 4.9. Contrary to most studies surveyed by Kaastra and Boyd (1996) and Vanstone and Finnie (2009, 2010), this also highlights the importance of using opposing trend B&H strategies as benchmark models since picking only one B&H model which goes against the general market trend tends to present an easier benchmark to beat and can potentially lead to overestimating the results attained by the AI models.

With the standard NN model (NN only) we manage to obtain positive results on six out of seven stocks and also at a portfolio level except for one stock (TSCO). When comparing the standard NN with the B&H Short method, the NN model outperforms in five (REX, SGE, ITV, HMSO, ABF) out of seven stocks with respect to the Sharpe, Sortino and Calmar ratios (all computed from 100 out-of-sample daily returns and non-annualised).
Table 4.10: Risk-adjusted performance of tested models following the 100-day out-of-sample period (bold font indicates best result among the models for the specific stock).

<table>
<thead>
<tr>
<th>Model / Stock</th>
<th>Trades</th>
<th>Win Ratio</th>
<th>Sharpe Ratio</th>
<th>Sortino Ratio</th>
<th>Calmar Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Buy and Hold (Long)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td>100</td>
<td>0.4400</td>
<td>-0.0246</td>
<td>-0.0425</td>
<td>-0.8354</td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td>100</td>
<td>0.5100</td>
<td>0.0197</td>
<td>0.0291</td>
<td>2.5712</td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td>100</td>
<td>0.4500</td>
<td>-0.1578</td>
<td>-0.2177</td>
<td>-10.2760</td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td>100</td>
<td>0.3900</td>
<td>-0.2202</td>
<td>-0.3662</td>
<td>-13.8379</td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td>100</td>
<td>0.4400</td>
<td>-0.1843</td>
<td>-0.3383</td>
<td>-11.6002</td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td>100</td>
<td>0.4600</td>
<td>-0.0275</td>
<td>-0.0479</td>
<td>-1.6415</td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td>100</td>
<td>0.4300</td>
<td>-0.0619</td>
<td>-0.1225</td>
<td>-3.5147</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td>700</td>
<td>0.4457</td>
<td>-0.0938</td>
<td>-0.1580</td>
<td>-5.5906</td>
</tr>
<tr>
<td><strong>Buy and Hold (Short)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td>100</td>
<td><strong>0.5600</strong></td>
<td>-0.0019</td>
<td>-0.0024</td>
<td>0.3738</td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td>100</td>
<td>0.4700</td>
<td>-0.0506</td>
<td>-0.0711</td>
<td>-3.0664</td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td>100</td>
<td><strong>0.5600</strong></td>
<td>0.1220</td>
<td><strong>0.2006</strong></td>
<td>8.7857</td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td>100</td>
<td>0.6000</td>
<td>0.1862</td>
<td>0.2256</td>
<td>14.2690</td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td>100</td>
<td>0.5700</td>
<td>0.1540</td>
<td>0.2147</td>
<td>11.7873</td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td>100</td>
<td>0.5300</td>
<td>0.0001</td>
<td>0.0010</td>
<td>0.8565</td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td>100</td>
<td><strong>0.5700</strong></td>
<td>0.0339</td>
<td>0.0396</td>
<td>2.9353</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td>700</td>
<td>0.5514</td>
<td>0.0634</td>
<td>0.0869</td>
<td>5.1345</td>
</tr>
<tr>
<td><strong>Standard NN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td>2214</td>
<td>0.5208</td>
<td>0.0867</td>
<td>0.1097</td>
<td>9.1406</td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td>2490</td>
<td>0.5758</td>
<td>0.1086</td>
<td>0.1769</td>
<td>10.0755</td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td>1463</td>
<td>0.5435</td>
<td>0.0339</td>
<td>0.0369</td>
<td>3.4385</td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td>1895</td>
<td>0.7041</td>
<td><strong>0.2648</strong></td>
<td>0.2292</td>
<td>27.3872</td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td>2637</td>
<td>0.6238</td>
<td>0.2201</td>
<td>0.2387</td>
<td>32.4711</td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td>1664</td>
<td><strong>0.5670</strong></td>
<td>0.0288</td>
<td>0.0242</td>
<td>2.6946</td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td>1533</td>
<td>0.5510</td>
<td>-0.0334</td>
<td>-0.0363</td>
<td>-2.3656</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td>13896</td>
<td>0.5837</td>
<td>-0.1014</td>
<td>-0.1113</td>
<td>11.8346</td>
</tr>
<tr>
<td><strong>NN-FMM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td>2124</td>
<td>0.5313</td>
<td><strong>0.1703</strong></td>
<td><strong>0.2384</strong></td>
<td><strong>18.2755</strong></td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td>2171</td>
<td><strong>0.6465</strong></td>
<td><strong>0.2124</strong></td>
<td><strong>0.2349</strong></td>
<td><strong>16.9375</strong></td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td>1045</td>
<td>0.5422</td>
<td><strong>0.1000</strong></td>
<td>0.1302</td>
<td><strong>11.3554</strong></td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td>1841</td>
<td><strong>0.7143</strong></td>
<td>0.2628</td>
<td><strong>0.2675</strong></td>
<td><strong>29.9808</strong></td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td>2562</td>
<td><strong>0.6832</strong></td>
<td>0.2860</td>
<td>0.3271</td>
<td>40.4895</td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td>1613</td>
<td>0.5567</td>
<td>0.0619</td>
<td>0.0542</td>
<td>6.8433</td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td>1237</td>
<td>0.5417</td>
<td>0.0676</td>
<td>0.0890</td>
<td>5.3336</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td>12593</td>
<td><strong>0.6023</strong></td>
<td><strong>0.1659</strong></td>
<td><strong>0.1916</strong></td>
<td><strong>18.4594</strong></td>
</tr>
</tbody>
</table>
Finally, our results show the outperformance of our proposed NN-FMM model. The NN-FMM model was only outperformed in two instances, namely by the B&H Short method in the case of the IAP Sortino ratio and by the standard NN in the case of the ITV Sharpe ratio. This suggests that in general our proposed method is superior to both standard NN and B&H models with respect to the investigated risk-adjusted performance measures. Also, this is a clear indication of the effectiveness of our risk-based money management fuzzy controller as an extension to NN design methods for trading purposes (as originally proposed by Kaastra and Boyd, 1996; Vanstone and Finnie, 2009, 2010).

In Table 4.11, we present the average risk-adjusted results obtained in the 100-day out-of-sample period for each AI model. The outperformance of the B&H Short method for IAP is clearer in these results. Except for IAP, NN-FMM shows clear outperformance on the remaining six out of seven stocks on both B&H and standard NN methods. Our results also show that in four out of seven stocks (SGE, HMSO, ABF and TSCO) our proposed approach obtains better win ratio statistics. It has to be noted, however, that with respect to the win ratio the improvements were only marginal, as expected, since our model objective is to improve the risk-adjusted performance.

**Model Robustness**

In this section we present a number of statistical tests to validate the hypothesis that the proposed model significantly outperforms the standard NN model both in terms of risk-adjusted performance and overall profitability. We limit our analysis to two measures, namely Sharpe ratio and profit (%) attained from the respective model trials in the 100-day out-of-sample period (in line with Bailey et al. (2014) who argue regarding the possibly inflated performance in the in-sample period due to overfitting). We also make use of box plots to illustrate the respective distribution of performance measures at the end of the out-of-sample period, also including the Calmar ratio to investigate drawdown effects.
Table 4.11: Average risk-adjusted performance of tested models for 50 repeated trials over the 100-day out-of-sample period (bold font indicates best result among the models for the specific stock).

<table>
<thead>
<tr>
<th>Model / Stock</th>
<th>Trades</th>
<th>Win Ratio</th>
<th>Sharpe Ratio</th>
<th>Sortino Ratio</th>
<th>Calmar Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Buy and Hold (Long)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td>100</td>
<td>0.4400</td>
<td>-0.0246</td>
<td>-0.0425</td>
<td>-0.8354</td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td>100</td>
<td>0.5100</td>
<td>0.0197</td>
<td>0.0291</td>
<td>2.5712</td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td>100</td>
<td>0.4500</td>
<td>-0.1578</td>
<td>-0.2177</td>
<td>-10.2760</td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td>100</td>
<td>0.3900</td>
<td>-0.2202</td>
<td>-0.3662</td>
<td>-13.8379</td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td>100</td>
<td>0.4400</td>
<td>-0.1843</td>
<td>-0.3383</td>
<td>-11.6002</td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td>100</td>
<td>0.4600</td>
<td>-0.0275</td>
<td>-0.0479</td>
<td>-1.6415</td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td>100</td>
<td>0.4300</td>
<td>-0.0619</td>
<td>-0.1225</td>
<td>-3.5147</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td>700</td>
<td>0.4457</td>
<td>-0.0938</td>
<td>-0.1580</td>
<td>-5.5906</td>
</tr>
<tr>
<td><strong>Buy and Hold (Short)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td>100</td>
<td>0.5600</td>
<td>-0.0019</td>
<td>-0.0024</td>
<td>0.3738</td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td>100</td>
<td>0.4700</td>
<td>-0.0506</td>
<td>-0.0711</td>
<td>-3.0664</td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td>100</td>
<td>0.5600</td>
<td>0.1220</td>
<td>0.2006</td>
<td>8.7857</td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td>100</td>
<td>0.6000</td>
<td>0.1862</td>
<td>0.2256</td>
<td>14.2690</td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td>100</td>
<td>0.5700</td>
<td>0.1540</td>
<td>0.2147</td>
<td>11.7873</td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td>100</td>
<td>0.5300</td>
<td>0.0001</td>
<td>0.0010</td>
<td>0.8565</td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td>100</td>
<td>0.5700</td>
<td>0.0339</td>
<td>0.0396</td>
<td>2.9353</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td>700</td>
<td>0.5514</td>
<td>0.0634</td>
<td>0.0869</td>
<td>5.1345</td>
</tr>
<tr>
<td><strong>Standard NN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td>2164</td>
<td>0.5469</td>
<td>0.1079</td>
<td>0.1615</td>
<td>12.6900</td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td>2457</td>
<td>0.5510</td>
<td>0.0984</td>
<td>0.1449</td>
<td>9.2177</td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td>1570</td>
<td>0.5652</td>
<td>0.0421</td>
<td>0.0690</td>
<td>5.3970</td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td>1878</td>
<td>0.6815</td>
<td>0.2019</td>
<td>0.2011</td>
<td>22.5048</td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td>2820</td>
<td>0.6352</td>
<td>0.2767</td>
<td>0.5655</td>
<td>36.2804</td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td>1740</td>
<td>0.5883</td>
<td>0.0329</td>
<td>0.0346</td>
<td>3.5157</td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td>1510</td>
<td>0.5625</td>
<td>0.0412</td>
<td>0.0556</td>
<td>3.8648</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td>14139</td>
<td>0.5901</td>
<td>0.1144</td>
<td>0.1760</td>
<td>13.3529</td>
</tr>
<tr>
<td><strong>NN-FMM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td>2079</td>
<td>0.5520</td>
<td>0.1750</td>
<td>0.2921</td>
<td>22.1393</td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td>2171</td>
<td>0.5638</td>
<td>0.1385</td>
<td>0.1999</td>
<td>13.8003</td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td>970</td>
<td>0.5534</td>
<td>0.0522</td>
<td>0.0776</td>
<td>6.1970</td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td>1836</td>
<td>0.6781</td>
<td>0.2224</td>
<td>0.2320</td>
<td>24.7125</td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td>2713</td>
<td>0.6674</td>
<td>0.3256</td>
<td>0.7061</td>
<td>46.5072</td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td>1681</td>
<td>0.5940</td>
<td>0.0885</td>
<td>0.0971</td>
<td>9.4725</td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td>1248</td>
<td>0.5768</td>
<td>0.1122</td>
<td>0.1518</td>
<td>10.6374</td>
</tr>
<tr>
<td>EW-Portfolio</td>
<td>12698</td>
<td>0.5979</td>
<td>0.1592</td>
<td>0.2509</td>
<td>19.0666</td>
</tr>
</tbody>
</table>
Table 4.12: Statistical tests on the Sharpe ratio of AI models over the 100-day out-of-sample period (bold font indicates best result among the models for the specific stock). The Anderson-Darling (AD) test tests the null hypothesis that the data is from a population with a normal distribution. The Wilcoxon ranksum (RS) test is a one-sided test where the alternative hypothesis states that the median Sharpe ratio of Standard NN is less than the median of NN-FNN. The one-sided paired \( t \)-test tests the alternative hypothesis that the population Sharpe ratio mean of standard NN is less than the population mean of NN-FMM.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>95% Conf.</th>
<th>p-Values</th>
<th>AD Test</th>
<th>RS Test</th>
<th>( t )-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock</td>
<td>Sharpe R.</td>
<td>Interval</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Standard NN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td>0.1079</td>
<td>[0.0910, 0.1247]</td>
<td>0.4583</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td>0.0984</td>
<td>[0.0794, 0.1174]</td>
<td>0.4712</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td>0.0421</td>
<td>[0.0173, 0.0668]</td>
<td>0.2311</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td>0.2019</td>
<td>[0.1822, 0.2216]</td>
<td>0.1127</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td>0.2767</td>
<td>[0.2560, 0.2973]</td>
<td>0.1004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td>0.0329</td>
<td>[0.0158, 0.0499]</td>
<td>0.2179</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td>0.0412</td>
<td>[0.0179, 0.0645]</td>
<td>0.8442</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NN-FMM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td><strong>0.1750</strong></td>
<td>[0.1578, 0.1922]</td>
<td>0.0437</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td><strong>0.1385</strong></td>
<td>[0.1182, 0.1587]</td>
<td>0.1381</td>
<td>0.0031</td>
<td>0.0023</td>
<td></td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td><strong>0.0522</strong></td>
<td>[0.0319, 0.0726]</td>
<td>0.0463</td>
<td>0.1805</td>
<td>0.1824</td>
<td></td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td><strong>0.2224</strong></td>
<td>[0.2043, 0.2405]</td>
<td>0.2515</td>
<td>0.0433</td>
<td>0.0631</td>
<td></td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td><strong>0.3256</strong></td>
<td>[0.3040, 0.3471]</td>
<td>0.7056</td>
<td>0.0000</td>
<td>0.0007</td>
<td></td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td><strong>0.0885</strong></td>
<td>[0.0698, 0.1073]</td>
<td>0.7757</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td><strong>0.1122</strong></td>
<td>[0.0906, 0.1338]</td>
<td>0.7323</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

As a first step we analyse and compare the distribution of the obtained Sharpe ratio for NN and NN-FMM models. We run an Anderson-Darling (AD) test on the respective distributions for each stock and each model (see Table 4.12). As suggested by Lawrence et al. (1997), this is done to confirm a symmetric conversion to the mean, thus validating the approach of using a mean comparison. For the standard NN model, the AD test has not rejected the null hypothesis that the Sharpe ratio distribution is from a normal distribution at a 5% significance level. For the NN-FMM model, this was rejected for two out of seven stocks (REX and IAP). For this reason we decide to apply both a parametric one-sided paired \( t \)-test and a non-parametric Wilcoxon rank-sum (RS) test to test the superiority of NN-FMM.
Table 4.13: Statistical tests on profit (%) of the AI models over the 100-day out-of-sample period (bold font indicates best result among the models for the specific stock). The Anderson-Darling (AD) test tests the null hypothesis that the data is from a population with a normal distribution. The Wilcoxon ranksum (RS) test is a one-sided test where the alternative hypothesis states that the median profit (%) of Standard NN is less than the median of NN-FNN. The one-sided paired \( t \)-test tests the alternative hypothesis that the population profit [%] mean of standard NN is less than the population mean of NN-FMM.

<table>
<thead>
<tr>
<th>Model</th>
<th>Stock</th>
<th>Mean</th>
<th>95% Conf.</th>
<th>( p )-Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sharpe R.</td>
<td>Interval</td>
<td>AD Test</td>
</tr>
<tr>
<td><strong>Standard NN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td></td>
<td>5.2913</td>
<td>[4.5115, 6.0712]</td>
<td>0.5264</td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td></td>
<td>4.9540</td>
<td>[4.0330, 5.7821]</td>
<td>0.4907</td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td></td>
<td><strong>2.8414</strong></td>
<td>[1.1350, 4.2540]</td>
<td>0.6431</td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td></td>
<td>15.0951</td>
<td>[14.0820, 15.9973]</td>
<td>0.7230</td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td></td>
<td>18.4539</td>
<td>[17.0155, 19.6778]</td>
<td>0.9548</td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td></td>
<td>1.3133</td>
<td>[0.6013, 1.9686]</td>
<td>0.4582</td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td></td>
<td>1.2769</td>
<td>[0.5329, 1.9593]</td>
<td>0.5282</td>
</tr>
<tr>
<td><strong>NN-FMM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rexam (REX)</td>
<td></td>
<td><strong>9.8480</strong></td>
<td>[8.8374, 10.7476]</td>
<td>0.2487</td>
</tr>
<tr>
<td>Sage Group (SGE)</td>
<td></td>
<td><strong>7.2403</strong></td>
<td>[6.2758, 8.1037]</td>
<td>0.5996</td>
</tr>
<tr>
<td>ICAP (IAP)</td>
<td></td>
<td>2.2608</td>
<td>[1.2203, 3.1853]</td>
<td>0.1272</td>
</tr>
<tr>
<td>ITV (ITV)</td>
<td></td>
<td><strong>16.8787</strong></td>
<td>[15.7085, 17.9029]</td>
<td>0.7895</td>
</tr>
<tr>
<td>Hammerson (HMSO)</td>
<td></td>
<td><strong>25.0043</strong></td>
<td>[23.3802, 26.3605]</td>
<td>0.4373</td>
</tr>
<tr>
<td>A. B. Foods (ABF)</td>
<td></td>
<td><strong>3.8631</strong></td>
<td>[3.0329, 4.6170]</td>
<td>0.5844</td>
</tr>
<tr>
<td>Tesco (TSCO)</td>
<td></td>
<td><strong>3.6639</strong></td>
<td>[2.9588, 4.3136]</td>
<td>0.3335</td>
</tr>
</tbody>
</table>

The tests show that the NN-FNN model is statistically significantly superior in five out of seven cases (REX, SGE, HMSO, ABF, TSCO). In the case of IAP, both tests do not indicate superiority, although further tests indicate that the performance is not statistically worse, hence performance is considered to be at par. In the case of ITV the result was inconclusive since the tests indicated different outcomes. Our conclusion drawn from all the tests and after evaluating the 95% confidence intervals (also presented in Table 4.12) is that NN-FMM generally outperforms the NN model from a risk-adjusted performance perspective.
4.3 Part 2 – A dynamic fuzzy money management approach for controlling the intraday risk-adjusted performance of AI-trading algorithms

As a second step we analyse and compare the distribution of the obtained profit (%) for both models (see Table 4.13). The AD test in this case has not rejected the null hypothesis for all stocks in both models, hence there is no sufficient statistical evidence to reject a normal distribution. The one-sided RS and $t$-tests applied to the NN-FMM model reveal that it outperforms the standard NN model on six out of seven stocks at the 5% significance level, with the only exception being IAP. In conjunction with the above Sharpe ratio results, we conclude that NN-FMM outperforms the standard NN model in terms of both Sharpe ratio and profitability, except in only one out of seven stocks (IAP).

Finally, we visually summarise our results for the Calmar ratio and the profit measure using box plots (see Figures 4.10 to 4.16). In all instances, the median Calmar ratio and the median profit of the NN-FMM is higher than the NN equivalent. The Calmar ratio plots indicate that NN-FMM models in general have lower drawdowns. In the case of IAP, which was the only stock where NN-FMM has not shown significant out-performance in both Sharpe ratio and profit measures, we see that NN-FMM still shows a substantially higher Calmar ratio (see Figure 4.12).

The above results lead us to conclude in favour of our three main contributions that we aim to validate in this paper. Firstly, we find a significant link between intraday trading profitability and two interrelated dimensions, which are NN return prediction strength (profit) and realised volatility (risk), extending the findings in the literature (Gradojevic and Gençay, 2013; Han et al., 2013; Holmberg et al., 2013; LeBaron, 1999; Schulmeister, 2006) at a more granular intraday level. Our results show that our proposed fuzzy logic method to dynamically adjust trade position size to different degrees of risk is an effective money management approach which improves the risk-adjusted performance of intraday trading models. Secondly, improving the risk-adjusted performance of the underlying trading model does not come at the cost of reduced overall profitability (as reported by Holmberg et al., 2013). This means that our approach does not only act as a filter by merely keeping the
Fig. 4.10: Performance of the AI models over the 100-day out-of-sample period for Rexam (REX).

Fig. 4.11: Performance of AI models over the 100-day out-of-sample period for Sage Group (SGE).
4.3 Part 2 – A dynamic fuzzy money management approach for controlling the intraday risk-adjusted performance of AI-trading algorithms

Fig. 4.12: Performance of AI models over the 100-day out-of-sample period for ICAP (IAP).

Fig. 4.13: Performance of AI models over the 100-day out-of-sample period for ITV (ITV).

Fig. 4.14: Performance of AI models over the 100-day out-of-sample period for Hammerson (HMSO).
Fig. 4.15: Performance of AI models over the 100-day out-of-sample period for Associated British Foods (ABS).

Fig. 4.16: Performance of AI models over the 100-day out-of-sample period for TESCO (TSCO).
4.3 Part 2 – A dynamic fuzzy money management approach for controlling the intraday risk-adjusted performance of AI-trading algorithms

Algorithm out of the market in adverse regions, and hence under-utilising available capital, but also improves model profitability by allocating more capital to preferable intraday states. This also validates the claims of Pardo (2011) in view of the disregarded effect of risk-based position sizing methods and should raise interest in the use of AI methods to address a wider set of trading decisions rather than limiting research primarily (if not solely) to market movement predictions (Krollner et al., 2010; Son et al., 2012; Tsai and Wang, 2009; Vanstone and Tan, 2003).

4.3.5 Conclusion

In Part 2 of this chapter, we seek further trading improvements from the identified relationship between the intraday profitability of technical trading rules and volatility that was initially explored in Part 1. The research problem focuses on optimising capital allocation for intraday trading using AI-driven money management decisions with the objective to increase both risk-adjusted performance and overall profitability. In line with Zadeh’s principle of compatibility, we resort to the fact that preciseness remains much less applicable in complex trading environments and hence we propose a novel fuzzy logic approach to control risk-based trading decisions.

This work conveys a number of contributions. Existing computational finance literature shows a great predominance of studies that limit the research on AI primarily (if not solely) to market movement predictions (as indicated in Krollner et al., 2010; Son et al., 2012; Tsai and Wang, 2009; Vanstone and Tan, 2003). As our first contribution, we extend these ideas by proposing an innovative fuzzy money management approach which dynamically adapts trading frequency and position sizing decisions across intraday trend (profit) and volatility (risk) states to improve the overall risk-adjusted performance. Contrary to many studies that suggest trading rules based on a fixed position sizing strategy (as reported by Pardo, 2011), fixed return thresholds (Kaastra and Boyd, 1996; Vanstone and Finnie, 2009,
2010) and fixed volatility thresholds (Brabazon and O’Neill, 2006; Holmberg et al., 2013), our approach dynamically evolves a continuous trading decision surface across the whole intraday trend-volatility space. We demonstrate the applicability of our fuzzy logic approach by presenting a hybrid fuzzy model as an extension to a popular neural network trading model (Choudhry et al., 2012; Krollner et al., 2010; Tsai and Wang, 2009).

Secondly, we show how our aim to improve the risk-adjusted performance can be achieved without incurring a reduction in profitability (dilemma reported in Holmberg et al., 2013). Our hybrid model calculates the approximate risk-adjusted performance across different trend-volatility states and automatically balances capital allocation according to preferable trend and volatility scenarios during the course of a trading day. The results show significant improvements in both profitability and risk-adjusted performance when compared to standard NN and buy-and-hold methods. We further validate our results with a robust statistical analysis on a number of independent trials in order to address the numerous criticisms in the literature (see Bailey et al., 2014; Lawrence et al., 1997; Luengo et al., 2009; Prechelt, 1996) that cast doubts on the experimental and statistical frameworks adopted by many published studies in computational finance literature.

Thirdly, contrary to the EMH, our findings extend the support for studies that claim a relationship between the profitability of technical trading rules and volatility (Gradojevic and Gençay, 2013; Han et al., 2013; Holmberg et al., 2013; LeBaron, 1999; Schulmeister, 2006). Our study explores this link at a more granular intraday level and at higher trading frequencies with the objective to improve risk-adjusted performance. To our best knowledge this was not previously addressed in the literature. From a theoretical perspective this sheds more light on the possible breakdown of the martingale property of prices under certain market conditions.

Following the demonstrated advantages of applying global risk-adjusted objective functions at a daily level (Chapter 3), in this chapter we took a more granular perspective and
presented innovative ways of how AI-driven trading algorithms can be further improved by discriminating across time-varying intraday volatility. In the next chapter we endeavour to gain further improvements by addressing challenges presented at a market microstructure level.
In this chapter, our aim is to investigate the profitability and risk-adjusted performance of trading algorithms from a higher frequency perspective than that presented in Chapters 3 and 4. Finance literature draws a set of market microstructure stylised facts, such as the bid-ask bounce and the discreteness of price change, which induce market microstructure frictions and consequently tamper the theoretical price by an error term, or “noise”. This noise, which accumulates with the sampling frequency, is attributed as one of the key modelling challenges and sources of uncertainty in high-frequency trading (HFT). Our primary objective is to identify model improvements that can result from the better handling of microstructure noise in an HFT setting. Whilst many former studies comparing type-1 (T1) and type-2 (T2) Fuzzy Logic Systems (FLSs) focus on error reduction or market direction accuracy, our interest, in line with our general thesis theme, predominantly lies with risk-adjusted performance.

This chapter extends our research findings that are presented in previous chapters from two aspects. Firstly, in Chapter 3 we show the effectiveness of ANFIS and also identify the increased stability of ANFIS in terms of risk-adjusted performance when compared to ANN alone. In this chapter we improve on these findings by investigating the possible
refinements that can be obtained by generalising ANFIS to interval T2 (IT2) FLS. Secondly, in Chapter 4 we enhance model risk-adjusted performance by applying FCM clustering to discriminate between different intraday volatility scenarios. In this chapter we use a similar data partitioning approach but build on these findings by applying FCM clustering to identify the T1 fuzzy logic rules which are then extended to IT2. This not only results in a more compact and efficient model but also in increased profitability and risk-adjusted performance.

We propose an innovative approach to design an IT2 model which is based on a generalisation of the popular T1 ANFIS model. The significance of this work lies in the identification of risk-adjusted performance improvements that are obtained as a result of introducing T2 fuzzy sets in intelligent trading algorithms. This is achieved with a minimal increase in the design and computational complexity of the underlying T1 models. Overall, the proposed ANFIS/T2 model scores significant performance improvements when compared to both standard ANFIS and buy-and-hold methods. As a further step, we identify a relationship between the increased trading performance benefits of the proposed T2 model and higher levels of microstructure noise. The results satisfy a desirable need for practitioners, researchers and regulators in the design of expert and intelligent systems for the better management of risk in the field of HFT.

5.1 Introduction

Most transactions in modern, highly computerised, financial markets are being largely controlled by specialised algorithms which incessantly sift through masses of data and take split-second trading decisions. According to a recent study by Brogaard et al. (2014), between 2008 and 2010 HFT algorithms accounted for 70% of dollar volume on the NASDAQ exchange. This tends to defy the long-standing Efficient Market Hypothesis (EMH) (Fama, 1965; ?) that states that current prices incorporate all relevant information with no possibility of predictability or excess returns. A number of authors (e.g. Brogaard et al., 2014; Holmberg
et al., 2013; Rechenthin and Street, 2013; Schulmeister, 2009; Zhang, 2010) insist that the presence of efficient pricing becomes more questionable when investigating short-lived (milliseconds to a few minutes) trades. However, Kearns et al. (2010) validate the EMH in their study and argue that generating profits from aggressive HFT is next to impossible. These debates keep this domain a very active area of research.

According to Johnson et al. (2013), this new machine-dominated reality highlights the need for new theories in support of sub-second financial phenomena during which the human traders lose the ability to react in real time. Due to the non-stationary characteristics of financial time series (see Fama, 1965), applying machine learning techniques to infer predictions is a challenging task and prone to increased error variance. Complexity is heightened given the level of noise in high-frequency stock price movements. Incidents like the “flash crash” of 6 May 2010 stress the importance of risk management. As a result, in recent years HFT and algorithmic trading have been the subject of increasing global regulatory attention. The new regulations are intended to ensure that trading systems are adequately designed and tested to mitigate the risks to which they are exposed (see reference to MiFID2 regulation in Chapter 1, Section 1.1.3, which will apply from January 2017).

In line with the main theme adopted in previous chapters, we stress again that this acts as a reminder for model or algorithm designers that both investors and regulators are more concerned with risk-adjusted performance rather than directional accuracies, error measures or solely profitability. Unfortunately, in previous chapters we highlight that the great majority of computational finance research disregards this fact (Bahrammirzaee, 2010; Krollner et al., 2010; Tsai and Wang, 2009). Using a risk-adjusted performance measure is essential in order to compare relative trading performance. A higher return trading strategy does not necessarily outperform other strategies if the associated risk is also substantially higher. This is the reason why in this chapter we continue to seek further improvements in risk-adjusted performance.
We again revert to Zadeh (1975), who proposed that increased system complexity calls for approaches that are significantly different from the traditional methods which are highly effective when applied to mechanistic systems. Fuzzy sets and systems are attributed as an excellent method to deal with situations where the element of uncertainty and imprecision is high, typically prevalent in complex environments (see Wagner and Hagras, 2010). A number of surveys (e.g. Krollner et al., 2010; Sahin et al., 2012; Tsang, 2009) place Artificial Neural Networks (ANNs) amongst the most popular learning techniques in AI-based financial applications and hybrid models. ANNs, especially in conjunction with fuzzy logic, were found to provide better forecasts.

As we note in Chapter 3, one of the most popular combinations of ANN and type-1 (T1) fuzzy logic is the Adaptive Neuro-Fuzzy Inference System (ANFIS) (Jang, 1993). The successful application and active continuous research dedicated to improving ANFIS-based techniques in trading applications is demonstrated by numerous publications (Boyacioglu and Avci, 2010; Chang et al., 2011; Chen, 2013; Gradojevic, 2007; Kablan and Ng, 2011; Tan et al., 2011; Wei et al., 2014). In Chapter 3 we extend this literature by presenting an ANFIS approach which can improve risk-adjusted performance and stability. However, type-2 (T2) fuzzy logic has recently gained significant academic attention (see review in Melin and Castillo, 2014) and as of today it remains a primary area of research in the fuzzy logic domain (Mendel et al., 2014). To our best knowledge, the use of higher order fuzzy logic systems (FLSs) in a high-frequency trading environment has not been addressed in the literature before. This presents an opportunity, which we tackle in this chapter, to investigate further improvements for the standard ANFIS model by extending it to T2 fuzzy logic. However, in line with Wu and Mendel (2014), we argue that although T2 FLSs provide the researcher with extensive freedom in design options, the increased computational and design complexity can possibly hinder the wider application of such systems. This challenge is a
5.1 Introduction

key consideration that inspired our innovative and practical T2 approach that we present in this chapter.

The investigation of possible improvements using T2 in HFT is appealing in view of increased uncertainties which are inherent in high-frequency data. Although the concepts of risk and uncertainty have often been used interchangeably, economists have long distinguished between the two (e.g. Knight, 1921), also in recent literature (e.g. Heal and Millner, 2014; Nelson and Katzenstein, 2014). Our view is that overall risk can be divided into measurable risk (e.g. the flip of a fair coin), and uncertainty, which we categorise as the risk of events to which it is difficult to attach a probability distribution. Our aim is to further reduce the trading uncertainty by utilising T2 FLSs. We have not identified any existing literature that investigates the level of noise (uncertainty) indirectly reflecting the trading frequency that would warrant the (feasible) use of T2 over T1 fuzzy logic methods for algorithmic trading purposes.

With respect to the above literature review and identified gaps, in this chapter the objectives can be summarised as follows:

1. To identify practical methods of how the popular ANFIS model, the stability strengths of which we show in Chapter 3, can be generalised to an interval T2 Takagi-Sugeno-Kang (IT2 TSK) fuzzy system. We aim to address this with a minimal increase in design and computational complexity.

2. To investigate the ability of higher order fuzzy systems to handle the increased uncertainties inherent in an HFT scenario. In this chapter our interest is to investigate higher trading frequencies than those we present in Chapters 3 and 4.

3. To identify if T2 FLSs provide a viable alternative for trading purposes in view of improving risk-adjusted performance.
4. To explore when T2 models can offer a more viable approach than T1 alternatives. We analyse this from the perspective of different levels of trading frequencies.

This chapter aims to put forward a number of contributions. As a first contribution we propose an innovative, but at the same time a more accessible, way of how to design a T2 FLS from an optimised T1 neuro-fuzzy FLS (ANFIS/T2). With IT2 there are many, sometimes an overwhelming amount of (Wu and Mendel, 2014), design choices to be made, which include the shape of membership functions, the number of membership functions, the type of fuzzifier, the kind of rules, the type of $i$-norm, the method to compute the output, and the methods for tuning the parameters. We address this from a number of aspects. Firstly, we improve on the ideas in Chapter 4 by applying a fuzzy clustering algorithm to identify different trend and volatility data segments and use the clustering results to construct the fuzzy rule database. This approach reduces the number of rules and hence simplifies the final model. We apply simple rules where antecedents are T2 fuzzy sets and consequents are crisp numbers (A2-C0). Secondly, as our base structure for the T2 model we use the ANFIS model since, as we identified in Chapter 3, it acts as a solid benchmark model, is computationally fast and has also been successfully applied in high-frequency trading (see also Kablan and Ng, 2011). Thirdly, we reduce the training complexity by reducing the number of tuning parameters, limiting this to varying sizes of the Footprint of Uncertainty (FOU). Our parsimonious approach also reduces the possibility of overfitting and spurious results (see Bailey et al., 2014). Finally, we apply an efficient closed-form type reduction method.

As a second contribution we shed more light on the theoretical market efficiency debate as regards HFT. Schulmeister (2009) points towards possible market inefficiencies and the profitability of technical trading rules at higher frequencies, this being driven by faster algorithmic trading. Recently, Rechenthin and Street (2013) claimed that when price shocks break the bid-ask spread, which was identified to happen anywhere between 5 to 10 seconds,
price movements can be predicted for up to one minute. Beyond this point, prediction probabilities remained significant for about the next 5 minutes, dying out completely beyond 30 minutes. In our case, we make use of HFT trade data from a set of stocks listed on the London Stock Exchange and investigate a combination of technical rules on 2-minute returns with holding periods ranging between 2 to 10 minutes. This extends the market efficiency investigations that we present in Chapters 3 and 4 to a more microscopic perspective and hence seeks possible pockets of profitability in shorter time windows. In line with previous chapters, we align ourselves with the priorities of investors and regulators and focus on comparing the proposed models using risk-adjusted performance (Choey and Weigend, 1997; Vanstone and Finnie, 2010; Xufre Casqueiro and Rodrigues, 2006). We are not aware of any previous studies which investigate the link between higher order fuzzy systems and risk-adjusted performance.

Finally, as our third contribution, we try to answer an important question which explores, from a trading performance perspective, when it is viable to apply T2 rather than T1 models. Although previous literature has found that T2 models can perform better under increasing uncertainties (Aladi et al., 2014; Sepulveda et al., 2006), it is however not clear at which uncertainty level this would be reflected in a reasonable improvement in risk-adjusted trading performance. Birkin and Garibaldi (2009) have even shown that if the level of noise is too low, T2 models show no significant improvement on T1. A number of authors (Gençay, 1996; Holmberg et al., 2013; Vanstone and Finnie, 2009, 2010) suggest the use of a threshold on the predicted signals below which a trading action is not taken into consideration. This is done to reduce the effect of the underlying noise; however, this is at the cost of reduced trades and hence possible return. We propose an innovative experiment approach by extending this technique to analyse how T1 and T2 models cope with decreasing (increasing) levels of return thresholds, which are reflected in an increase (reduction) in uncertainty but also
in increased (reduced) return potential. This ability to handle higher frequency noise is fundamental for HFT.

Our evaluation of out-of-sample data demonstrates that the proposed ANFIS/T2 model outperforms the standard ANFIS and buy-and-hold methods. Statistically significant improvements in both risk-adjusted performance and profitability are registered in higher trading frequency scenarios but disappear when trading activity is lowered.

The structure of this chapter is organised as follows. In Section 5.2, we introduce the main model components and design method. This is followed by a description of our experiment approach and model evaluation presented in Section 5.3. In Section 5.4, we present our results and analyse model performance. Finally, in Section 5.5 we draw our conclusions in view of the existing literature.

5.2 Method

Our experiment setup consists of five modules (see Figure 5.1). Section 5.2.1 explains our variable selection and data pre-processing. Sections 5.2.2 and 5.2.3 explain our Fuzzy Inference System (FIS) design approach using clustering which is later fed into ANFIS and ANFIS/T2 for tuning (this section presents our main model enhancements on previous chapters). In Section 5.2.4 we explain our trading algorithm.

5.2.1 High-frequency data and technical indicators

In this section we explore evidence presented by a number of authors who claim that in HFT scenarios there exist short time windows where past prices can convey information which can be used for predictive purposes. Our interest is not to identify the determining factors of this claimed HFT phenomenon, but to identify candidate features that can be used by our
trading models. We do however reduce our window of interest to a more granular (higher frequency) price structure than that we present in Chapters 3 and 4.

A profitable trading algorithm essentially requires constructing a model which can determine whether, under certain conditions and time horizons, prices will be trending or mean-reverting (Pardo, 2011). When designing an HFT model, an important challenge faced by a model designer is the claim that return autocorrelations in HFT can have both genuine and spurious elements (Anderson, 2011; Anderson et al., 2013). The latter is attributed to market microstructure noise (McAleer and Medeiros, 2008), mainly resulting from non-synchronous trading effect and bid-ask bounce. In view of this, a core consideration in designing HFT models is to manage the tension between moving to higher price frequencies, hoping to benefit from possible price correlations, but at the same time being able to manage the increasing noise levels which give rise to perceived price movements and volatility (see Andersen and Bollerslev, 1997; Rechenthin and Street, 2013).

We investigate whether the average return for the next 2 minutes can be successfully estimated using 5 signals which can provide information on price trend, reversion and
movement strength from a time window of previous prices ranging from 1 to 15 minutes. The set of signals that we select for this chapter is based on findings by Brogaard et al. (2014) and Zhang (2010), who identified that the main determinants of current HFT activity are past returns, liquidity, and HFT activity. The time-window selection is based on claims from Rechenthin and Street (2013), who stated that the stock price typically broke the price reversal pattern due to the bid-ask bounce after 5 to 10 seconds, and that traces of predictability existed for up to 30 minutes, beyond which markets became efficient.

Our findings in Chapters 3 and 4 (see also Gradojevic and Gençay, 2013; Naranjo et al., 2015) show the effectiveness of combining moving average signals with fuzzy logic to capture trend information. In this chapter we use 1-minute stock prices, \( p_t \), and define the expected mean return, \( y_t \), at time \( t \), as

\[
y_t = \log(MA_{t+2}^2) - \log(p_t),
\]

where short and long moving averages, \( MA_t^n \), are calculated as defined by Equation (3.1). We apply corresponding trade signals, \( MA_t^{n_1-n_2} \), defined by Equation (3.3) with the lag structures \((n_1, n_2) \in [(1, 2), (1, 5), (1, 10)]\) where \( n_1 \) and \( n_2 \) are expressed in 1-minute time bars. Compared to Chapters 3 and 4, here we apply shorter lag structures in order to generate a higher number of signals and hence increase our potential trade frequency.

Similar to Chapter 4, for our mean-reversion indicator we use the popular Relative Strength Index (RSI) (Murphy, 1987). However, since in this chapter we wish to investigate higher frequency trading, to calculate RSI we consider 1-minute prices in the previous 15 minutes.

In Chapter 4 we identify model improvements by discriminating between different levels of intraday volatility. For this reason, here we utilise Realised Volatility (RV) as our fifth input variable. This presents a more efficient and compact approach to discriminate between intraday volatility levels by including it as part of the main prediction model (rather than at a
subsequent second step). We present the calculation of RV at time $t$ in Chapter 4, Equation (4.11). Similarly, to minimise the effect of microstructure noise, we again implement a simple RV estimator called average RV (Christoffersen, 2011). Average RV is calculated using Equation (4.12). However, in this case we utilise the last 15 minutes average RV (rather than 30 minutes) at time $t$ using 5-minute return intervals (as suggested by McAleer and Medeiros, 2008). Again, this is done primarily due to our interest in generating a higher number of trade signals.

In summary, the identified $k$ variables yield a linear regression model to describe the relationship with $y_t$ as

$$y_t = \theta_0 + \sum_{k=1}^{5} \theta_k s_{k,t-1} + \varepsilon_t \quad (5.2)$$

with the error term $\varepsilon_t \sim N(0, \rho)$ and

$$s_{k,t} = \begin{cases} \text{MA}_t^{1.2} & \text{for} \quad k = 1 \\ \text{MA}_t^{1.5} & \text{for} \quad k = 2 \\ \text{MA}_t^{1.10} & \text{for} \quad k = 3 \\ \text{RSI}_t & \text{for} \quad k = 4 \\ \text{RV}_{t}^{\text{avg}} & \text{for} \quad k = 5 \end{cases} \quad (5.3)$$

### 5.2.2 Designing and tuning a TSK type-1 fuzzy model

A first consideration is to select the type of FLS to employ. The literature identifies two main types, namely Mamdani (Mamdani, 1974), where the rule consequents are fuzzy sets on the output space, and TSK (Sugeno and Kang, 1988), where the rule consequents are crisp functions of the inputs. We adopt the TSK approach due to their popularity in practice resulting from their simplicity and flexibility (Wu and Mendel, 2014). Moreover, for TSK rules, output calculation is less computationally intensive: the output is a weighted average of
the crisp rule consequents, where the weights are the firing levels of the rules (the preliminary mathematical background is presented in Section 2.3).

In the following two sub-sections, we describe our approach for (i) the initial identification of a T1 TSK FIS model (Section 5.2.2) and (ii) model tuning (Section 5.2.2).

**Initial FIS structure identification**

We follow a model-free approach (Mendel et al., 2014) with the objective to completely specify the FLS using training data. The process starts from a given collection of \( q \) minute-by-minute input-output data training pairs, \((x^{(1)} : y^{(1)}), (x^{(2)} : y^{(2)}), \ldots, (x^{(q)} : y^{(q)})\) where

\[
\begin{align*}
  x^{(1)} &= [s_{1,t-q}, s_{2,t-q}, \ldots, s_{5,t-q}], & y^{(1)}_{t+1} \\
  x^{(2)} &= [s_{1,t-q+1}, s_{2,t-q+1}, \ldots, s_{5,t-q+1}], & y^{(2)}_{t+2} \\
  \vdots \\
  x^{(q)} &= [s_{1,t-1}, s_{2,t-1}, \ldots, s_{5,t-1}], & y^{(q)}_{t}.
\end{align*}
\] (5.4)

In Equation (5.4), for each data instance at a specific time \( t \), \( x \) is a vector consisting of \( \{x_1, x_2, \ldots, x_5\} \) input elements which represent the \( \{s_{1,t-1}, s_{2,t-1}, \ldots, s_{5,t-1}\} \) technical indicator signals (Equation (5.3)), and \( y \) represents the mean return over the next 2 minutes (Equation (5.1)).

The idea of fuzzy inference systems can be broken down into a divide-and-conquer (Jang and Sun, 1995) approach. The first objective is to identify fuzzy regions that partition the input space using the antecedents of fuzzy rules. The second objective is to map a local behaviour within a given region using the rule consequents. The selection of the space partitioning scheme has two important effects on the resulting model. The first effect is that the more granular the space partitioning, the higher the number of rules, hence improving model accuracy. However, a resulting effect is the increased number of optimisation parameters and therefore computational complexity. Increased model complexity can also result in overfitting.
The designer has to balance accuracy and model complexity depending on the structure of the underlying data and the specific context.

For this reason, we apply a clustering algorithm, namely fuzzy c-means (FCM) clustering, that we introduced earlier in Chapter 4. Similar to Chapter 4, we use FCM to identify fuzzy partitions in data (Bezdek, 1981; Dutta and Angelov, 2010); however, in this chapter we use it as a precursor (rather than a subsequent step) to our learning algorithm. By controlling the number of clusters, this gives us the opportunity to identify the best model structure which balances model accuracy and complexity (improving on both performance and complexity when compared to the rule database structure presented in Chapter 3). This is possible because each cluster centre essentially exemplifies a characteristic behaviour of the system in a specific region. Hence, each cluster centre can be used as the basis of a membership function for each input variable and these are combined in a rule that describes the local system behaviour.

Since our input variables have a different basis, these variables are standardised and rescaled to have a mean of zero and a standard deviation of one before being fed into the algorithm. This ensures equivalent influence weighting of each variable on the clustering algorithm. Let \( z \) denote the fuzziness index. Furthermore, define \( \alpha \) as the number of clusters, \( d_{uv} = \|x_u - c_v\|^2 \) as the Euclidean distance between the \( u \)-th realisation and the current \( v \)-th cluster centre \( c_v \), and \( d_{uo} = \|x_u - c_o\|^2 \) as the Euclidean distance from the \( u \)-th realisation and the other cluster centres \( c_o \). For each data point \( x_u, \forall u \in [1, q] \), and cluster \( c_v, \forall v \in [1, \alpha] \), the FCM algorithm iteratively updates the membership grade \( \mu \) of the \( u \)-th data point to the \( v \)-th cluster until the algorithm has converged (a detailed description of the FCM algorithm is presented in Section 4.2.5, page 102).

As suggested by Pal and Bezdek (1995), we test values between 1.5 and 2.5 for the fuzzy index \( z \). For the number of clusters, \( \alpha \), we test values between 2 to 5 clusters. This range selection is based on identifying a well-distributed set of clusters representing the variable
Market microstructure noise and risk-adjusted performance

distribution, whilst at the same time avoiding the heavy influence of possible outliers. After examining cluster plots on different stocks and time periods, we select a fuzzy index value of 1.7. The number of clusters is used as a model parameter which is included as part of the model tuning process in combination with the ANFIS parameters (see next section).

In this chapter, we adopt Gaussian membership functions (MFs), where each fuzzy set is represented by

\[ \text{Gaussian}(x; \bar{x}, \sigma) = e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}. \]  

(5.5)

An MF returns the degree of membership, in the range [0,1], of a specific point in a particular variable region. We select this particular fuzzy set shape because, unlike other MFs, it has only two parameters (the mean \( \bar{x} \) and the standard deviation \( \sigma \)) and it always spreads out over the entire input domain (Wu and Mendel, 2014).

Once the clustering process is complete, we follow two steps to create an initial T1 TSK fuzzy model (preliminary background on FIS structure identification using clustering is presented in Section 2.4.1). In the first step, the identified \( \alpha \) clusters are projected on each input variable space. This results in \( \alpha \) Gaussian MFs for each variable with the mean represented by the cluster centres \( c \). The standard deviation \( \sigma \) is obtained by re-arranging Equation (5.5) and utilising cluster centres \( c \) and membership grades \( m \). As a second step, a set of \( \alpha \) rules are created in the form

\[
\text{IF} \quad (x_1 \text{ is } A_{i,1}) \quad \text{AND} \quad (x_2 \text{ is } A_{i,2}) \quad \text{AND} \ldots \quad \text{AND} \quad (x_5 \text{ is } A_{i,5}) \\
\text{THEN} \quad y_i = b_i + \sum_{k=1}^{5} w_{i,k}x_k 
\]  

(5.6)

where \( A_{i,k} \) represents T1 Gaussian MFs, projected from the identified clusters, for the \( i \)-th rule and the \( k \)-th input \( (i = 1, 2, \ldots, \alpha; k = 1, 2, \ldots, 5) \). In the consequent, \( y_i \) is the rule output, defined as the mean return over the next 2 minutes (Equation 5.1), as a linear function of the input variables \( \{x_1, x_2, \ldots, x_5\} \) with parameters \( b_i \) and \( w_{i,k} \). Following the identification of the
initial FIS using FCM clustering, we seek further model tuning which is described in the next section.

**FIS tuning with ANFIS**

From the literature we identify two major classes of optimisation algorithms for FLSs: gradient-based algorithms and heuristic algorithms, where in the latter case most studies focus on evolutionary computation (EC) algorithms (see the discussion in Wu and Mendel, 2014). ANFIS follows the former approach. It is due to the popular use of ANFIS in finance that we decide to use it as our optimisation technique and hence as a benchmark model to explore possible risk-adjusted performance improvements by extending the model to an IT2 TSK FLS. The performance stability of ANFIS is also demonstrated by our findings in Chapter 3. The optimisation algorithm is also similar to the approach that we adopted in previous chapters, hence we eliminate, as much as possible, any performance discrepancies resulting from alternative optimisation techniques and focus more on identifying performance gains as a result of model structure and identification improvements.

Based on our variable selection (Section 5.2.1), next we define the ANFIS configuration that we adopt in this chapter. The main differences when compared to the ANFIS configuration that we present in Chapter 3 (Section 3.2.3) are that here we have 5 input variables (rather than 3) and the number of membership functions and rules are based on the number of identified clusters $\alpha$ (rather than MF space partitioning). The ANFIS network is defined as follows:

**Layer 1** Since we have 5 inputs, this layer contains $5 \times \alpha$ adaptive nodes, one node for every membership function associated with each input. For instance, the $\alpha$ nodes with connections from the first input $x_1$ are in the form

$$O_{1,i} = \mu_{A_{1,i}}(x_1) \text{ for } i = 1, 2, ..., \alpha,$$

(5.7)
where $\mu_{A_{i,k}}$ is the membership degree for the $i$-th T1 MF and the $k$-th input ($i = 1, 2, ..., \alpha; k = 1, 2, ..., 5$). In our setup, $\alpha$ represents both the number of rules and also the number of MFs for each input variable. Different values for $\alpha$ are tested as part of our model calibration process (see Table 5.1).

**Layer 2** This layer contains $\alpha$ fixed nodes. In each node, $O_{2,i}$, where $i = 1, ..., \alpha$, the product $t$-norm ($\ast$) is used to “AND” the membership grades which are passed from the previous layer. The output is the firing strength, $f_i$, of each rule:

$$O_{2,i} = f_i = \mu_{A_{1,i}}(x_1) \ast \mu_{A_{2,i}}(x_2) \ast \cdots \ast \mu_{A_{5,i}}(x_5).$$  \hspace{1cm} (5.8)

**Layer 3** In this layer, which consists of $\alpha$ fixed nodes, the normalised firing strengths, $\hat{f}_i$, are calculated using

$$O_{3,i} = \hat{f}_i = \frac{f_i}{\sum_i f_i}. \hspace{1cm} (5.9)$$

**Layer 4** The nodes in this layer are adaptive and act as a function

$$O_{4,i} = \hat{f}_i y_i = \hat{f}_i (b_i + \sum_{k=1}^{5} w_{i,k} x_k), \hspace{1cm} (5.10)$$

where $\hat{f}_i$ is the normalised firing strength from the previous layer and $y_i$ is the rule consequent linear function for the $i$-th rule, $i = 1, ..., \alpha$, with parameters $b_i$ and $w_{i,k}$.

**Layer 5** This layer consists of a single node and combines the output from all the nodes in the previous layer to calculate the overall output as

$$O_{5,i} = r_i = \sum_i \hat{f}_i y_i = \frac{\sum_i \hat{f}_i y_i}{\sum_i \hat{f}_i}. \hspace{1cm} (5.11)$$

Similar to the process described in Section 3.2.3, for ANFIS learning we follow an iterative two-pass algorithm. In a forward pass, the premise parameters (in Layer 1) defining
Table 5.1: Parameters tested for ANFIS.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter Value Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Data Size (days)</td>
<td>{2, 3}</td>
</tr>
<tr>
<td>Number of Input Membership Functions and Rules ((\alpha))</td>
<td>{2, 3, 4, 5}</td>
</tr>
<tr>
<td>Training Epochs</td>
<td>{10, 20, 40}</td>
</tr>
</tbody>
</table>

the membership functions are unmodified and the consequent parameters (in Layer 4) are computed using the least squares algorithm. On completion of the forward pass, the consequent parameters are unmodified and a backward pass feeds the errors back into the network using back-propagation to adjust the premise parameters (more detail is presented in Section 2.4.2). In our in-sample training and model selection process we test and compare all \(2 \times 4 \times 3 = 24\) permutations of the parameter combinations (Table 5.1). These parameters are tested in combination with an additional set of parameters that are defined for our trading algorithm (see Section 5.2.4). In the next section, we propose how the standard ANFIS model can be extended to a T2 TSK model.

5.2.3 Generalizing ANFIS model to T2 FLS

There is no mathematical proof that by changing a T1 FLS to T2 FLS, a T2 fuzzy logic controller (FLC) will automatically outperform a T1 FLC (Wu and Mendel, 2014). When considering T2, an initial step for an algorithm designer is to understand the underlying uncertainties and the sources thereof. In an HFT environment, one can identify a number of sources of uncertainty:

- Constantly changing market activity and volatility conditions.

- The use of non-precise terms: “rising steadily”, “high volatility”, “small loss”, “high activity”.

- Microstructure noise in the observations.
Market microstructure noise and risk-adjusted performance

- Inconsistencies affecting trade execution (for example, execution time and transaction costs).

By identifying the levels and sources of uncertainty, the designer can formulate an initial opinion, firstly on the fit for T2 FLS, rather than using T1 FLS, and secondly on the numerous T2 design options to consider (see Wu and Mendel, 2014, for a summary of design options). In our scenario, all the identified uncertainties point towards T2 as being a good contender.

Our primary interest is to identify model improvements that can result from the better handling of the microstructure noise. This noise is attributed as one of the key modelling challenges and sources of uncertainty in HFT (Anderson, 2011; Anderson et al., 2013; Rechenthin and Street, 2013). T1 fuzzy sets cannot fully represent the uncertainty associated with the inputs since, as a contradiction, the membership function of a T1 fuzzy set has no uncertainty associated with it. To address this criticism, Zadeh (1975) introduced T2 fuzzy sets, which has been a very active area of research (see Wu and Mendel, 2014). We consider Interval T2 (IT2) fuzzy sets since they are much less computationally intensive and more popular than the generalised T2 (Mendel et al., 2014; Wu and Mendel, 2014). This projects our focus on investigating whether the introduction of T2 fuzzy sets in the antecedents of fuzzy rules can improve the risk-adjusted performance of trading algorithms, in which scenarios and to what extent. We consider a special IT2 TSK FLS when its antecedents are T2 fuzzy sets but its consequents are crisp numbers (referred to as A2-C0 by Mendel et al., 2014). Although other IT2 models exist (see Mendel et al., 2014), we select this model because it provides a good balance between the better management of uncertainty and increased model complexity. Preliminary mathematical background on IT2 models is provided in Sections 2.5 and 2.6.
5.2 Method

**IT2 FLS design approach**

At this point, we decide on an important design consideration for our experiments. The decision is to select the method to adopt when it comes to tuning our IT2 FLS. In this section, we describe our rationale.

We consider two different approaches that are commonly adopted to design IT2 FLSs (Aladi et al., 2014; Mendel et al., 2014; Wu and Mendel, 2014): a partially dependent approach and a totally independent approach. In the former approach, the designer starts with an optimised T1 FLS, which is then used as a basis for the design of the IT2 FLSs. On the other hand, the totally independent approach is used to design IT2 FLSs from scratch, hence avoiding the use of an intermediate T1 FLS.

We adopt the partially dependent approach for a number of reasons. Firstly, although previous literature found that T2 models can perform better under increasing uncertainties (Aladi et al., 2014; Sepulveda et al., 2006), we do not seek to achieve an optimal performance in the error reduction; instead, the primary objective is to compare T1 FLSs and IT2 FLSs and to shed more light on the possible gains in risk-adjusted performance. Secondly, in line with Wu and Mendel (2014), the increased parameters and design options in IT2 FLS can be overwhelming and possibly limit more widespread use. With the adopted approach, our objective is to propose an incremental step from standard ANFIS, which as we highlighted earlier is a popular and already established technique in finance, and to contribute new improvements in this active area of research. Thirdly, our parsimonious approach also reduces the possibility of overfitting and spurious results (see Bailey et al., 2014). Whilst these reasons present our rationale for selecting a partially dependent approach, in Section 5.3.3 (IT2 design considerations) we highlight the strengths and weaknesses of this approach.
ANFIS/T2 models

In this section, we propose two methods, ANFIS/T2a and ANFIS/T2b, of how the T1 FIS structure resulting from the ANFIS training can be extended to an IT2 FLS (A2-C0) model. In line with the economic theories about the separation of overall risk between risk and uncertainty (Heal and Millner, 2014; Knight, 1921; Nelson and Katzenstein, 2014), our intention is to seek trading performance improvements that can result from the introduction of T2 fuzzy sets. This is obtained by minimising the uncertainty caused by microstructure noise present in high-frequency data, hence reducing the overall risk.

The A2-C0 rules are defined as follows:

\[
\text{IF } (x_1 \text{ is } A_{i,1}) \text{ AND } (x_2 \text{ is } A_{i,2}) \text{ AND } \ldots \text{ AND } (x_5 \text{ is } A_{i,5}) \text{ THEN } y_i = b_i + \sum_{k=1}^{5} w_{i,k} x_k \]  

(5.12)

where in the premise part of the rule, \( A_{i,k} \) are IT2 Gaussian MFs projected from the identified clusters for the \( i \)-th rule and the \( k \)-th input \((i = 1, 2, \ldots, \alpha; k = 1, 2, \ldots, 5)\). In the rule consequent, \( y_i \) is a linear function of the input variables \( \{x_1, x_2, \ldots, x_5\} \) with parameters \( b_i \) and \( w_{i,k} \).

For our proposed ANFIS/T2a, the objective is to convert the T1 rules defined in Equation (5.6) to the A2-C0 rules defined in Equation (5.12). To do this, we start from the ANFIS-optimised T1 fuzzy sets. The next step is to introduce the footprint of uncertainty (FOU) for the MFs in the premise part of the rules whilst keeping the consequent part fixed. The FOU represents the blurring effect of a T1 membership function, \( \mu_{A_{i,k}} \), and is completely described by two corresponding bounding functions, a lower membership function (LMF), \( \mu_{A_{i,k}} \), and an upper membership function (UMF), \( \overline{\mu}_{A_{i,k}} \), both of which are T1 fuzzy sets. Unlike their T1 counterparts, whose membership values are precise numbers in the range [0, 1], the membership grades of a T2 fuzzy set are themselves T1 fuzzy sets. Therefore, T2 fuzzy sets
offer the ability of modelling higher levels of uncertainty (John and Coupland, 2007; Mendel et al., 2006). Aladi et al. (2014) show how T2 fuzzy sets can handle increased noise and claim a direct relationship between FOU size and levels of noise. However, Benatar et al. (2012) warn that selecting too small an FOU will result in no improvements over the T1, whilst if too large, the T2 model will perform worse. In our case, complexity is compounded since noise levels are not fixed but time-varying due to time-varying market activity.

To define the size of the FOU, for ANFIS/T2a we adopt a parsimonious approach by introducing one additional parameter. This parameter, \( \beta \in [0, 1] \), determines the increase or decrease in the standard deviation, \( \sigma_{i,k} \), of all the Gaussian T1 MFs across all input variables \( (i = 1, 2, \ldots, \alpha; k = 1, 2, \ldots, 5) \), whilst keeping the mean, \( x_{i,k} \), fixed. Hence, for each T1 MF, the LMF and UMF are defined as follows:

\[
\mu_{A_{i,k}} = \text{Gaussian}(x_k; x_{i,k}, (1 - \beta)\sigma_{i,k}) \quad (5.13)
\]

\[
\bar{\mu}_{A_{i,k}} = \text{Gaussian}(x_k; x_{i,k}, (1 + \beta)\sigma_{i,k}) \quad (5.14)
\]

where \( x_{i,k} \) and \( \sigma_{i,k} \) are the parameters for the \( i \)-th T1 Gaussian MF and \( k \)-th input tuned by ANFIS. When applied, this results in a new set of IT2 MFs (Figure 5.2). We transform the complete T1 to IT2 rule base in this manner. The final output of the model is obtained as follows

\[
Y_{A2-C0} = [y_l, y_r] = \left( \int_{f_1 \in [\mathcal{F}_1, \mathcal{F}_1]} \ldots \int_{f_\alpha \in [\mathcal{F}_\alpha, \mathcal{F}_\alpha]} \right) 1 / \left( \sum_{i} f_i y_i \right)
\]

where the integral sign represents the fuzzy union operation and the slash operator (/) associates the elements of rules output and firing strength with their secondary membership grade, which in the case of IT2 is simplified to 1. The firing strength for each rule \( i \), where
Conversion of T1 MF to Upper and Lower MFs

Shape of MFs controlled by varying Standard Deviation

$\mu_{A_{1,k}}$, $\mu_{A_{2,k}}$, ..., $\mu_{A_{5,k}}$

$(1 + \beta_k)\sigma_{i,k}$, $(1 - \beta_k)\sigma_{i,k}$

Input Value

Fig. 5.2: Conversion of T1 fuzzy set to IT2 with fixed mean and uncertain standard deviation. Upper and lower MFs are defined using an additional parameter $\beta_k$, which represents the width of the IT2 MF as a percentage increase or decrease on the base T1 MF standard deviation respectively. In our first experiment, we train the model to identify and assign the same $\beta$ value across all inputs $k = 1, 2, ..., 5$. In our second experiment, each $\beta_k$ is tuned to different values.

$i = 1, 2, ..., \alpha$, is calculated as

$$f_i(x) = \mu_{A_{1,i}}(x_1) \ast \mu_{A_{2,i}}(x_2) \ast \ldots \ast \mu_{A_{5,i}}(x_5)$$  \hspace{1cm} (5.16)

$$\bar{f}_i(x) = \mu_{A_{1,i}}(x_1) \ast \mu_{A_{2,i}}(x_2) \ast \ldots \ast \mu_{A_{5,i}}(x_5)$$  \hspace{1cm} (5.17)

where, like in the case of ANFIS (Equation (5.8)), $\ast$ represents the product $t$-norm. For further details, the interested reader is directed to Mendel et al. (2014).

Following the ANFIS optimisation, as a second training pass we train our model with values of $\beta$ ranging from 0% to 40% in discrete steps of 5%. This range is selected from our testing on in-sample data. It is to be noted that we intentionally include 0% in our search space, which results in the reduction of the IT2 FLS back to the corresponding T1 FLS.
Fig. 5.3: An example showing the MFs for the input variable RSI index following ANFIS training (top), and the corresponding ANFIS/T2 tuned MFs (bottom). Both ANFIS and ANFIS/T2 MFs are dynamically adapted on a daily basis, reflecting the price volatility in that period. During training, the latter IT2 MFs optimal thickness is identified by initially adopting the same Gaussian MFs parameters tuned by ANFIS and then increase (decrease) the standard deviation by a factor ranging from 0% to 40%.

allows the training algorithm to select between T1 and IT2 FLS and to dynamically adapt the model FOU according to the level of market uncertainty during the specific training period (see the example in Figure 5.3). This also guarantees that during the training process, the model achieves at least the same level of performance as that of the T1 FLS.

For our second proposed method, ANFIS/T2b, we introduce more flexibility in the model by introducing 5 new parameters in the model, \( \{ \beta_1, \beta_2, ..., \beta_5 \} \), where \( \beta_k \in [0, 1) \). These parameters represent the increase or decrease in the FOU for all T1 fuzzy sets defining the space of the individual 5 input variables. Hence, with this approach, the FOU can dynamically
adapt to different levels of uncertainty across the different input variables. Therefore in the case of ANFIS/T2b, we convert every rule \( i \), where \( i = 1, 2, \ldots, \alpha \), by transforming the MFs of each input variable to IT2 MFs using \( \beta_k \), where \( k = 1, 2, \ldots, 5 \). In this case, the lower and upper MFs are defined as:

\[
\mu_{A_{i,k}} = \text{Gaussian}(x_k; \bar{x}_{i,k}, (1 - \beta_k)\sigma_{i,k}) \tag{5.18}
\]
\[
\Pi_{A_{i,k}} = \text{Gaussian}(x_k; \bar{x}_{i,k}, (1 + \beta_k)\sigma_{i,k}). \tag{5.19}
\]

In our training algorithm, we apply the same discrete range for possible \( \beta_k \) values. This approach results in a much larger search space of possible \( \beta_k \) combinations, hence rather than performing a parameter sweep, we speed up the training process by optimising the \( \beta_k \) values using a mixed integer genetic algorithm. This does not limit our approach to other possible optimisation methods. Following the rule base conversion from T1 to A2-CO, the final step is to decide on how to compute the output. This is described in the next section.

**Computing the output**

The T2 fuzzy logic community has proposed a number of methods for computing the output (the interested reader is directed to Mendel et al., 2014; Wu, 2013, for a review). At a high level, the methods can be divided into two groups. The first group requires an interim defuzzification process which reduces the output T2 fuzzy sets to T1 fuzzy sets. These are typically solved via iterative algorithms. The second group skips this step completely by calculating the output directly. Siding with less computational intensive methods, we chose the Nie-Tan (NT) method (Nie and Tan, 2008), which falls under the latter group. The NT method computes the output as follows:

\[
y = \hat{r} = \frac{\sum_{i=1}^{\alpha} y_i (f_i + \bar{f}_i)}{\sum_{i=1}^{\alpha} (f_i + \bar{f}_i)}, \tag{5.20}
\]
which effectively makes use of a vertical-slice representation of a T2 fuzzy set and involves taking the mean of the lower and upper membership functions, creating a type-1 fuzzy set. When compared to other defuzzification methods, it was demonstrated that the NT method provides a good balance between accuracy and complexity (Greenfield and Chiclana, 2013; Wu, 2013).

5.2.4 Trading algorithm

Our trading algorithm is kept similar to the approach presented in Chapters 3 and 4. The similarity is purposely kept in order to attribute any performance improvements to the proposed T2 enhancements. Every minute, the prediction of the average return over the following two minutes is passed to the trading algorithm which in turn recommends a buy ($\Phi = long$), sell ($\Phi = short$) or stand-by ($\Phi = 0$) action. Similar to other chapters, we allocate a starting capital of 250,000 GBP for each stock, buy or sell positions of 50,000 GBP and a transaction cost of 10 GBP per trade per direction. When a trade is closed, the net proceeds are added back to the capital balance, hence maximising the utilisation of the available capital.

However, in this chapter, the algorithm is calibrated to simulate an increased number of intraday trades. This is done primarily through the re-calibration of the return threshold (RT) and trade duration (TD) parameters. In our experiments, we explore four levels of the $RT$ parameter for each stock, \{0.08\%, 0.06\%, 0.04\%, 0.02\%\}, to avoid small price movements mostly originating from microstructure effects. As suggested by Vanstone and Finnie (2009), in our algorithm we also take account of whether the signal is increasing in strength, or decreasing in strength from its previous forecast. Hence, before opening a position, the algorithm confirms the current signal by comparing this with the forecast signal generated in the previous 1-minute time bar.
For our second parameter, $TD$, which represents the duration of each trade in minutes, we consider values between 2 minutes to 10 minutes. Following evidence by Rechenthin and Street (2013), this range is selected to reduce the possible perceived movements resulting from the bid-ask bounce. The time window is, however, short enough to capture any instances of market inefficiencies due to HFT. If a signal in the same direction is generated prior to closure, $TD$ is reset back to zero. This form of extended close proved to be successful in previous studies (Brabazon and O’Neill, 2006). We close all open positions at the end of each daily trading session to avoid having positions overnight.

The pseudo code presented in Algorithm 1 (presented in Chapter 3, page 65) describes the structure of the trading algorithm that we use in our experiments. Both $RT$ and $TD$ parameters are used in conjunction with the standard ANFIS parameters referred to earlier (see Table 5.1), and are applied in the model selection process to identify the base T1 model for each stock.

### 5.3 Experiment approach

We conduct two experiments (see Table 5.2). In the first experiment (1a-1e), we test the trading performance of our proposed ANFIS/T2a and ANFIS/T2b models against Buy-and-Hold (B&H) and standard ANFIS benchmark models. In this setup, each model is tuned at the optimum level of noise (and hence, trading frequency) that can result in the highest Sharpe ratio by the respective model. In the second experiment (2a-2c), the trading performance of our proposed T2 models is compared with standard ANFIS at different levels of noise with the objective to identify the best performing model at different degrees of uncertainty.

In the following sub-sections, we describe the different aspects of the approach we adopt in our experiments and also aim to present a critical analysis in view of our decisions. Section 5.3.1 describes our data and underlying statistics. Section 5.3.2 explains the performance measures that we adopt to perform model selection and evaluation. A discussion about
Table 5.2: Models applied in the experiments.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>AI Algorithm</th>
<th>Adaptive Tuning</th>
<th>Noise Level (Uncertainty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>B&amp;H (Daily)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1b</td>
<td>B&amp;H (100 days)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1c</td>
<td>ANFIS</td>
<td>T1 MFs &amp; Rules</td>
<td>Optimum</td>
</tr>
<tr>
<td>1d</td>
<td>ANFIS/T2a</td>
<td>IT2 MFs (variable width) &amp; Rules</td>
<td>Optimum</td>
</tr>
<tr>
<td>1e</td>
<td>ANFIS/T2b</td>
<td>IT2 MFs (variable width/MF) &amp; Rules</td>
<td>Optimum</td>
</tr>
<tr>
<td>2a</td>
<td>ANFIS</td>
<td>T1 MFs &amp; Rules</td>
<td>Fixed Levels</td>
</tr>
<tr>
<td>2b</td>
<td>ANFIS/T2a</td>
<td>IT2 MFs (variable width) &amp; Rules</td>
<td>Fixed Levels</td>
</tr>
<tr>
<td>2c</td>
<td>ANFIS/T2b</td>
<td>IT2 MFs (variable width/MF) &amp; Rules</td>
<td>Fixed Levels</td>
</tr>
</tbody>
</table>

our IT2 model design considerations is presented in Section 5.3.3. This is followed by a description of our model training and testing process (Section 5.3.4) and our approach for controlling different levels of noise (Section 5.3.5).

5.3.1 Data

The data we use in this chapter is high-frequency trade data for 15 stocks listed at the London Stock Exchange (see Table 5.3) during a 250-day period between 28/06/2007 and 25/06/2008 (excluding weekends, holidays and after-hours trading). Data is sampled at 1-minute intervals using the last trade price every 1 minute. Since the London Stock Exchange operates between 8:00 and 16:30 GMT, this produces 510 price points per day, resulting in a time series of 127,500 price points per stock over the entire period. The sample skewness and kurtosis in Table 5.3 indicate that the return distributions are far from being normal. In the selection of our data set size, we note the harsh criticism brought forward by Bailey et al. (2014) in view of the number of publications which base their study on a small backtest length given the number of model configurations tested. Bailey et al. (2014) prove that this easily gives rise to possible overfitting with the chance of spurious results (especially in in-sample tests). To mitigate this risk, in our case, each model is tested
Table 5.3: Descriptive statistics of 1-minute returns

<table>
<thead>
<tr>
<th>Company</th>
<th>Symbol</th>
<th>Mean $\times 10^{-6}$</th>
<th>Std. Dev.</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antofagasta</td>
<td>ANTO</td>
<td>0.6509</td>
<td>0.0024</td>
<td>1.2194</td>
<td>330.6081</td>
</tr>
<tr>
<td>BHP Billiton</td>
<td>BLT</td>
<td>2.1716</td>
<td>0.0020</td>
<td>-0.1043</td>
<td>214.8807</td>
</tr>
<tr>
<td>British Airways</td>
<td>BAY</td>
<td>-5.4056</td>
<td>0.0020</td>
<td>0.2329</td>
<td>150.9052</td>
</tr>
<tr>
<td>British Land Company</td>
<td>BLND</td>
<td>-5.0941</td>
<td>0.0017</td>
<td>0.0749</td>
<td>86.1668</td>
</tr>
<tr>
<td>Sky</td>
<td>SKY</td>
<td>-2.2680</td>
<td>0.0013</td>
<td>-0.1850</td>
<td>173.2209</td>
</tr>
<tr>
<td>Cable and Wireless</td>
<td>CW</td>
<td>-2.1766</td>
<td>0.0014</td>
<td>0.0824</td>
<td>311.7595</td>
</tr>
<tr>
<td>Aviva</td>
<td>AV</td>
<td>-2.9633</td>
<td>0.0016</td>
<td>0.2320</td>
<td>280.4653</td>
</tr>
<tr>
<td>Diageo</td>
<td>DGE</td>
<td>-1.0693</td>
<td>0.0013</td>
<td>0.2544</td>
<td>299.0263</td>
</tr>
<tr>
<td>HSBC Holdings</td>
<td>HSBA</td>
<td>-1.1900</td>
<td>0.0014</td>
<td>0.1183</td>
<td>4082.5000</td>
</tr>
<tr>
<td>Rio Tinto</td>
<td>RIO</td>
<td>3.2828</td>
<td>0.0021</td>
<td>0.9144</td>
<td>361.6329</td>
</tr>
<tr>
<td>BP</td>
<td>BP</td>
<td>-0.2378</td>
<td>0.0012</td>
<td>-0.0179</td>
<td>119.9209</td>
</tr>
<tr>
<td>Lloyds Banking Group</td>
<td>LLOY</td>
<td>-4.0709</td>
<td>0.0016</td>
<td>0.0704</td>
<td>313.4039</td>
</tr>
<tr>
<td>Tesco</td>
<td>TSCO</td>
<td>-1.1754</td>
<td>0.0012</td>
<td>-1.1077</td>
<td>133.3191</td>
</tr>
<tr>
<td>HBOS</td>
<td>HBOS</td>
<td>-9.9067</td>
<td>0.0022</td>
<td>0.7628</td>
<td>174.7204</td>
</tr>
<tr>
<td>Xstrata</td>
<td>XTA</td>
<td>2.0661</td>
<td>2.0661</td>
<td>-2.0788</td>
<td>298.5333</td>
</tr>
</tbody>
</table>

on a number of price points which would be equivalent to over 505 years of daily data per stock.

Another important consideration when selecting stocks for back-testing purposes is the importance of picking a mix of stocks which exhibit different trends. As it can be seen from the numerous machine learning and artificial intelligence studies surveyed in Krollner et al. (2010), Tsai and Wang (2009) and Bahrammirzaee (2010), this is rarely considered. Pardo (2011) warns that only including stocks that follow similar trends can lead to ungeneralised models which only work in specific scenarios only, hence introducing a bias in the experiment results. We noted this risk when picking the stocks, and as shown in the descriptive statistics in Table 5.3, we include a mix of stocks with both positive and negative mean returns (see also Figure 5.4) over the selected training and test period. Our selection of stocks is also representative of a number of industry sectors. Further testing on a wider stock selection, instruments and markets is left as future work.
Fig. 5.4: Times series of stock prices (normalised, for better comparison) which shows the mix of trends followed by the selected stocks over the in-sample and out-of-sample period.

5.3.2 Performance measures

As noted in Chapters 3 and 4, surveys (Krollner et al., 2010; Tsai and Wang, 2009) show that the great majority of machine learning studies with trading applications focus on the minimisation of error functions, directional accuracy or else profitability (issues with these measures are discussed in Brabazon and O’Neill (2006) and Pardo (2011), and we explore them further in Chapter 3). The danger with error functions is that a small error does not necessarily translate into profitability since it does not reflect the direction. On the other hand, directional accuracy measures are not enough to ensure overall profitability since they do not incorporate the magnitude of the correct or incorrect predictions. Moreover, a high directional accuracy might be completely misleading since few large losses can still cancel out a higher number of wins, but which are smaller in size. Finally, focusing only on
profitability does not incorporate the possible drawdowns that can be experienced during specific periods. This can be disastrous for an investor. This also reflects the rules that are being proposed by directives like MiFID2 that are intended to ensure that trading algorithms show robustness with a lower risk of unexpected huge losses. For this reason, we apply three key measures to assess our models: the Shape ratio, percentage profit and profit per trade. Next, we define our rationale for choosing these measures.

In this chapter we remain consistent with previous chapters and use the Sharpe ratio (Equation 3.20) as our key measure of risk premium per unit of risk in an investment. In light of the research focus in this chapter, two limitations of the Sharpe ratio are noted. Firstly, the Sharpe ratio does not separate between variability in gains and losses, hence it attributes penalisation to both upside and downside variability. This might not represent the interest of investors who would rather welcome positive variability in gains. In this chapter, however, we favour model stability and hence our interest lies more in identifying algorithmic trading models that can offer steady returns. Secondly, Lo (2002) warns that the Sharpe ratio highly depends on the distribution of the underlying returns. In the case of non-normal distributions which exhibit “fat tails” this might lead to misleading results. However, investigations conducted by Eling (2008) and Prokop (2012) show that Sharpe ratio measures lead to similar rankings of more sophisticated performance ratios. This also transpires from our previous chapters were in addition to the Sharpe ratio we use other risk-adjusted measures, such as the Sortino ratio and the Calmar ratio. In balance, we decide in favour of the Sharpe ratio as our main risk-adjusted performance measure due to its simplicity and thus easy application, and also due to its widespread acceptance both in literature and in practice.

As we show in Chapter 4, the profit per trade measure provides an indication of the efficiency of the underlying algorithm in terms of capital allocation. It also provides an indication of the existing spread between the average return per trade and the underling transaction costs. A higher spread would indicate the possibility of increasing the number of
trades with the chance to increase overall profitability. In view of our interest to investigate model performance at higher trading frequencies, we opt to also include this measure in this chapter. We conclude that the combination of the above two measures, in conjunction with overall profitability, provide a clear overall picture of model performance both in terms of profitability and risk.

5.3.3 IT2 design considerations

For our IT2 design approach, we consider two different options that are commonly adopted to design IT2 FLSs: a partially dependent approach and a totally independent approach. Although in Section 5.2.3 we present both options and the reasons why we opted for a partially dependent approach, it is important for model designers to understand the advantages and disadvantages of both options. In the partially dependent approach, the primary advantage is that it makes it easier to directly compare the T1 and IT2 FLSs. A second advantage is that the training of the IT2 FLS can be much faster since a number parameters would already have been optimised by the T1 model. On the other hand, the advantage of the totally independent approach is that it avoids the assumption that the optimised parameters of the T1 model; for example, the type and number of membership functions, are the best parameters to be inherited by the IT2 FLS, hence possibly leading to a sub-optimal IT2 FLS model. This is a conscious risk that we undertake in this study, the primary reason being that our main objective is the comparison of T1 and IT2 models.

5.3.4 Model training and testing

The AI model selection process is based on identifying the best model parameters (defined in Table 5.1 in conjunction with trading algorithm parameters \( RT \) and \( DT \)) that result in the highest performance during the 150-day in-sample period. Indirectly, this means that the optimum noise levels (and hence trading frequency) is selected for each model. In the case
of ANFIS/T2 models, parameter selection is extended to the identification of $\beta$ parameters (as described in Section 5.2.3) which define the optimum size of FOU.

Common practice in time series and machine learning literature is to divide the time-series into training, testing and validation sets. However, Kaastra and Boyd (1996) and Pardo (2011) argue that in the case of trading scenarios, a more rigorous approach is to adopt moving window (also known as walk-forward) testing which consists in a series of overlapping training-testing-validation sets. Although the moving window approach requires more frequent model re-training, it tries to simulate real-life trading and also permits quicker model adaptation to changing market conditions.

Similar to previous chapters, we adopt a day-by-day moving window approach (see Figure 3.3). In this case, at $day_d$, where $(d = 1, 2, ..., 150)$, the model is trained on 1-minute data points (Equation (5.3)) from $day_d - r$ to $day_d - 1$, and $r$ represents the training data size in days. The trained model is then used to predict minute-by-minute mean returns (Equation (5.1)) during $day_d$. This is repeated for the whole 150-day in-sample period, for each parameter combination. The final selected model is then tested, using the same day-by-day moving window approach, over the next 100-day out-of-sample period. The size of the time series provides a sufficiently large historical dataset which reduces the possibility of over fitting or the production of spurious results during backtesting (Bailey et al., 2014).

Apart from the T1 FLS model trained using ANFIS, we also consider two buy-and-hold (B&H) strategies as additional benchmark models. In the first strategy (B&H daily), for every trading day we buy at the daily opening price, hold it over the course of the trading day and sell at the daily closing price. In the second strategy (B&H 100 days), we buy at the beginning of the out-of-sample period, hold for the duration of the 100-day period, and sell at the closing price of the 100th day. Comparisons against these zero-intelligence models help us to validate the contribution that is attained by introducing AI-controlled algorithmic
trading. For indicative purposes, we also present a number of average statistics across the whole portfolio of stocks.

### 5.3.5 Controlling different levels of noise

An important decision that we need to take is to define the approach to use to simulate different levels of noise. This will in turn enable us to compare T1 FLS with IT2 FLS under different degrees of uncertainty.

We propose an innovative approach which can allow us to adjust, in a controlled fashion, the level of noise and hence be able to compare T1 and IT2 FLSs under different degrees of uncertainty. A number of authors (e.g., Aladi et al., 2014; Sepulveda et al., 2006) take the approach of methodically generating and injecting synthetic noise in the data. We decide to take a different approach by making use of a stylised fact in financial time series where microstructure noise, in its nature, is more pronounced in higher frequency data (Medeiros et al., 2006). Typically, this effect is reduced by using a threshold which acts as a filter on the predicted signals, below which a trading action is not taken into consideration (Gençay, 1996; Holmberg et al., 2013; Vanstone and Finnie, 2009, 2010). We follow this approach in both Chapter 3 and Chapter 4. Here, we extend the use of this method by hypothesising that this approach is effectively controlling for uncertainty (indirectly). An increased (reduced) signal strength effectively translates into reduced (increased) uncertainty that the predicted move is due to microstructure noise with the reduced (increased) risk to result in unprofitable trades. Hence, we test the models under different levels of uncertainty by adjusting different levels of the return threshold. We argue that this proposed approach is more practical and realistic for algorithmic trading scenarios rather than by injecting synthetic noise.

Based on this decision, in the second experiment (see 2a-2c in Table 5.2) the objective is to compare the trading performance of T1 FLS and IT2 FLS at different levels of uncertainty by controlling the return threshold, $RT$. For this reason, after testing the models during
the in-sample period, using the same moving window approach that is applied in the first experiment, the model selection process is based on choosing the model which returned the highest Sharpe ratio at specific levels of $RT$. Finally, we conduct statistical tests on the 100-day out-of-sample results to determine whether there is enough evidence of IT2 superiority at different levels of RT. The results of these experiments are presented in Section 5.4.2.

5.4 Results and analysis

In Section 5.4.1, we conduct a performance comparison between the benchmark models, namely B&H methods and standard ANFIS, and the proposed IT2 FLS models during 100 out-of-sample trading days. In Section 5.4.2, we analyse the models’ performance across increasing levels of uncertainty.

5.4.1 Experiment 1: Comparison against benchmark models

Our approach is to first establish the performance obtained from benchmark models and then identify improvements that can be attained by our proposed IT2 models. As our first set of benchmark models, we consider two B&H methods. In the first method, B&H (daily), we simulate a one round-trip trade, every day, for the full 100-day out-of-sample period. This consists in performing a buy trade at the opening of the exchange and selling the asset at close of day. In the second method, B&H (100 days), we simulate buying the stock at the beginning of the out-of-sample period and then selling at the end of the 100-day period.

The results from Table 5.4 indicate that B&H (daily) obtains small to moderate positive period returns (1% to 9%) only on 4 (BLT, HSBA, RIO and TSCO) out of the 15 stocks, and major losses (>20% losses) on 4 other stocks (BLND, AV, BP and HBOS). In general, the same negative results are obtained in the B&H (100 days) method, with small to moderate
positive period returns (1% to 11%) on 4 stocks (BLT, HSBA, RIO, BP) and heavy losses (>20% losses) on a number of others (BA, BLND, AV, LLOY and HBOS). The huge losses (>30% losses) obtained from the B&H (100 days) on BA, BLND, LLOY and HBOS are partially expected due to the substantially large negative mean returns identified in the descriptive statistics in Table 5.3, hence indicating a strong negative trend. We also note the results of BP and TSCO which show opposing results for B&H (daily) and B&H (100 days). This result is possible due to the fact that whilst B&H (100 days) considers only the first and last price of the testing period, B&H (daily) is affected (positively or negatively) by all the daily trends in the 100-day out-of-sample period. For example, a strong negative trend in the last 20 days can be disastrous for B&H (100 days); however, B&H (daily) can still carry over some profits from the previous 80 days. This confirms our approach in adopting the two methods as our first set of benchmark models.

Table 5.4: Results obtained from the B&H methods in the 100-day out-of-sample period. B&H (daily) performs 100 trades, each position covering one full trading day. B&H (100 days) performs 1 trade, covering the full 100 day period.

<table>
<thead>
<tr>
<th>Company</th>
<th>Symbol</th>
<th>B&amp;H (daily) Profit%</th>
<th>B&amp;H (100 days) Profit%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antofagasta ANTO</td>
<td>-8.95%</td>
<td>-8.76%</td>
<td></td>
</tr>
<tr>
<td>BHP Billiton BLT</td>
<td>8.99%</td>
<td>11.37%</td>
<td></td>
</tr>
<tr>
<td>British Airways BA</td>
<td>-3.56%</td>
<td>-40.52%</td>
<td></td>
</tr>
<tr>
<td>British Land Company BLND</td>
<td>-23.43%</td>
<td>-33.67%</td>
<td></td>
</tr>
<tr>
<td>Sky SKY</td>
<td>-8.19%</td>
<td>-16.34%</td>
<td></td>
</tr>
<tr>
<td>Cable and Wireless CW</td>
<td>-19.76%</td>
<td>-10.46%</td>
<td></td>
</tr>
<tr>
<td>Aviva AV</td>
<td>-20.54%</td>
<td>-20.33%</td>
<td></td>
</tr>
<tr>
<td>Diageo DGE</td>
<td>-11.22%</td>
<td>-11.23%</td>
<td></td>
</tr>
<tr>
<td>HSBC Holdings HSBA</td>
<td>5.78%</td>
<td>0.69%</td>
<td></td>
</tr>
<tr>
<td>Rio Tinto RIO</td>
<td>2.20%</td>
<td>3.67%</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>-22.83%</td>
<td>7.17%</td>
<td></td>
</tr>
<tr>
<td>Lloyds Banking Group LLOY</td>
<td>-4.83%</td>
<td>-30.21%</td>
<td></td>
</tr>
<tr>
<td>Tesco TSCO</td>
<td>1.08%</td>
<td>-13.51%</td>
<td></td>
</tr>
<tr>
<td>HBOS</td>
<td>-45.73%</td>
<td>-94.18%</td>
<td></td>
</tr>
<tr>
<td>Xstrata XTA</td>
<td>-2.67%</td>
<td>-2.30%</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.5: Standard ANFIS performance after 100 the day out-of-sample period.

<table>
<thead>
<tr>
<th>Company</th>
<th>Symbol</th>
<th>No. of Trades</th>
<th>Sharpe Ratio</th>
<th>Profit%</th>
<th>Profit/Trade (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antofagasta</td>
<td>ANTO</td>
<td>2036</td>
<td>0.3388</td>
<td>14.32%</td>
<td>£18.90</td>
</tr>
<tr>
<td>BHP Billiton</td>
<td>BLT</td>
<td>475</td>
<td>0.3224</td>
<td>8.65%</td>
<td>£47.56</td>
</tr>
<tr>
<td>British Airways</td>
<td>BA</td>
<td>6248</td>
<td>1.5420</td>
<td>86.40%</td>
<td>£54.92</td>
</tr>
<tr>
<td>British Land Company</td>
<td>BLND</td>
<td>921</td>
<td>0.4664</td>
<td>10.52%</td>
<td>£30.12</td>
</tr>
<tr>
<td>Sky</td>
<td>SKY</td>
<td>2606</td>
<td>0.5404</td>
<td>15.47%</td>
<td>£16.05</td>
</tr>
<tr>
<td>Cable and Wireless</td>
<td>CW</td>
<td>4709</td>
<td>1.2952</td>
<td>108.40%</td>
<td>£103.87</td>
</tr>
<tr>
<td>Aviva</td>
<td>AV</td>
<td>1074</td>
<td>0.4136</td>
<td>7.88%</td>
<td>£19.09</td>
</tr>
<tr>
<td>Diageo</td>
<td>DGE</td>
<td>1203</td>
<td>0.3807</td>
<td>7.47%</td>
<td>£16.11</td>
</tr>
<tr>
<td>HSBC Holdings</td>
<td>HSBA</td>
<td>1129</td>
<td>0.2912</td>
<td>7.08%</td>
<td>£16.24</td>
</tr>
<tr>
<td>Rio Tinto</td>
<td>RIO</td>
<td>535</td>
<td>0.2040</td>
<td>4.87%</td>
<td>£23.31</td>
</tr>
<tr>
<td>BP</td>
<td>BP</td>
<td>6978</td>
<td>0.9543</td>
<td>35.75%</td>
<td>£15.39</td>
</tr>
<tr>
<td>Lloyds Banking Group</td>
<td>LLOY</td>
<td>9621</td>
<td>0.7212</td>
<td>51.79%</td>
<td>£17.63</td>
</tr>
<tr>
<td>Tesco</td>
<td>TSCO</td>
<td>7928</td>
<td>1.5421</td>
<td>60.27%</td>
<td>£26.08</td>
</tr>
<tr>
<td>HBOS</td>
<td>HBOS</td>
<td>4708</td>
<td>0.4693</td>
<td>38.45%</td>
<td>£24.90</td>
</tr>
<tr>
<td>Xstrata</td>
<td>XTA</td>
<td>911</td>
<td>0.1222</td>
<td>5.88%</td>
<td>£16.61</td>
</tr>
</tbody>
</table>

For our second benchmark comparison, we apply the standard ANFIS. From Table 5.5 one immediately notices that contrary to the B&H methods, the algorithm is profitable on all 15 stocks. The B&H methods perform better than standard ANFIS in only one stock (BLT). Considering that the results are based on a 100-day out-of-sample period, the model shows moderate returns (5% to 10%) as in the case of BLT, AV, DGE, HSBA, RIO and XTA, and excellent returns on all other stocks. These results validate the popularity of ANFIS in finance (Boyacioglu and Avci, 2010; Chang et al., 2011; Chen, 2013; Kablan and Ng, 2011; Tan et al., 2011; Wei et al., 2014) and the active research in improving the model and application techniques. This also validates our findings in Chapter 3 and our proposal to use ANFIS as our main benchmark model and use it as a basis to seek further improvements.

From further investigation of the standard ANFIS results (Table 5.5) we notice that the number of trades performed over the 100-day out-of-sample period varies substantially across the different stocks, ranging from an average of 5 trades a day (BLT) up to 96 trades a day (LLOY). When we examine the number of trades in relation to Profit%, we can identify...
that the highest Profit% was achieved by those stocks with the highest trading frequency (BA, CW, BP, LLOY, TSCO and HBOS) in spite of lower profit per trade. This is a typical outcome of HFT, whereby higher overall profits are obtained from lower profits per trade but increased trading frequency. More importantly, these higher returns are also obtained in conjunction with higher risk-adjusted performance (Sharpe ratio). Contrary to Kearns et al. (2010) who claim the absence of profitability in HFT, our findings support the claims of Schulmeister (2009) who identified pockets of profitability in shorter time windows, in our case in the 2-minute to 10-minute range. Our results also validate the theoretical claims of Zhang (2010), Rechenthin and Street (2013) and Brogaard et al. (2014) regarding the possible market efficiency breakdowns in the high-frequency range.

Our next challenge is to explore any additional performance gains that can be obtained from our proposed IT2 TSK models, namely ANFIS/T2a and ANFIS/T2b. The two models are tested on the same 100-day out-of-sample period used in the benchmark models. From our performance comparison of the ANFIS/T2a model against standard ANFIS, the results in Table 5.6 show Sharpe ratio improvements in 12 out of 15 stocks. A lower Sharpe ratio was obtained in only 3 stocks (ANTO, HSBA and BP). In terms of profitability, it can be noted that 10 out of 15 stocks show higher profitability. This was only marginally lower in the case of HSBA and TSCO, with only 3 stocks (ANTO, BA and CW) showing a reduction in profitability by more than 1 percentage point (p.p.). These results provide a clear indication of the superiority of our proposed ANFIS/T2a over standard ANFIS.

The improved performance of ANFIS/T2a is further assessed by investigating the summary statistics in Table 5.6. We note that when considering the portfolio of 15 stocks, on average, the Sharpe ratio increases by 2.36% and the average return per stock increases by 0.69 p.p. Considering that portfolios of financial institutions typically hold thousands of equities and millions in investments, minor increases in profitability can translate into significant monetary value. More importantly, this increase in profitability is achieved in
Table 5.6: ANFIS/T2a performance comparison against standard ANFIS after the 100-day out-of-sample period. The results depict variations from those produced by ANFIS in Table 5.5. The number of trades, Sharpe ratio and profit per trade are presented as percentage differences. Profit column shows the percentage point (p.p.) differences.

<table>
<thead>
<tr>
<th>Company</th>
<th>Symbol</th>
<th>No. of Trades</th>
<th>Sharpe Ratio</th>
<th>Profit (p.p.)</th>
<th>Profit /Trade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antofagasta</td>
<td>ANTO</td>
<td>−6.53%</td>
<td>−0.89%</td>
<td>−1.91%</td>
<td>−8.19%</td>
</tr>
<tr>
<td>BHP Billiton</td>
<td>BLT</td>
<td>−5.68%</td>
<td>+0.36%</td>
<td>+0.01%</td>
<td>+6.21%</td>
</tr>
<tr>
<td>British Airways</td>
<td>BA</td>
<td>−0.82%</td>
<td>+0.41%</td>
<td>−1.36%</td>
<td>−1.53%</td>
</tr>
<tr>
<td>British Land Company</td>
<td>BLND</td>
<td>−1.52%</td>
<td>+1.53%</td>
<td>+0.0%</td>
<td>+1.57%</td>
</tr>
<tr>
<td>Sky</td>
<td>SKY</td>
<td>−0.42%</td>
<td>+0.39%</td>
<td>+0.16%</td>
<td>+1.52%</td>
</tr>
<tr>
<td>Cable and Wireless</td>
<td>CW</td>
<td>−1.19%</td>
<td>+0.15%</td>
<td>−1.18%</td>
<td>−0.59%</td>
</tr>
<tr>
<td>Aviva</td>
<td>AV</td>
<td>−1.49%</td>
<td>+8.82%</td>
<td>+0.56%</td>
<td>+9.01%</td>
</tr>
<tr>
<td>Diageo</td>
<td>DGE</td>
<td>−0.67%</td>
<td>+1.88%</td>
<td>+0.11%</td>
<td>+2.26%</td>
</tr>
<tr>
<td>HSBC Holdings</td>
<td>HSBA</td>
<td>−1.86%</td>
<td>−0.81%</td>
<td>−0.09%</td>
<td>+0.52%</td>
</tr>
<tr>
<td>Rio Tinto</td>
<td>RIO</td>
<td>−1.50%</td>
<td>+14.63%</td>
<td>+0.56%</td>
<td>+13.41%</td>
</tr>
<tr>
<td>BP</td>
<td>BP</td>
<td>−0.92%</td>
<td>−3.8%</td>
<td>+1.72%</td>
<td>+6.74%</td>
</tr>
<tr>
<td>Lloyds Banking Group</td>
<td>LLOY</td>
<td>−1.43%</td>
<td>+17.79%</td>
<td>+3.01%</td>
<td>+9.12%</td>
</tr>
<tr>
<td>Tesco</td>
<td>TSCO</td>
<td>−0.63%</td>
<td>+0.23%</td>
<td>−0.08%</td>
<td>+0.46%</td>
</tr>
<tr>
<td>HBOS</td>
<td>HBOS</td>
<td>−6.18%</td>
<td>+4.69%</td>
<td>+0.67%</td>
<td>+8.85%</td>
</tr>
<tr>
<td>Xstrata</td>
<td>XTA</td>
<td>−0.99%</td>
<td>+18.28%</td>
<td>+1.0%</td>
<td>+18.85%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td><strong>−1.76%</strong></td>
<td><strong>+2.36%</strong></td>
<td><strong>+0.69%</strong></td>
<td><strong>+3.15%</strong></td>
</tr>
</tbody>
</table>

conjunction with higher a Sharpe ratio (lower risk). Another advantage of ANFIS/T2a is the fact that higher risk-adjusted performance and profitability are obtained with a lower number of trades (1.76% less trades). Holmberg et al. (2013) claim that increased risk-adjusted performance can result from increased signal filtering (reduced trades) but at the cost of reduced overall profitability. On the contrary, our results indicate that ANFIS/T2a shows lower overall trading activity but more efficient capital allocation by instigating more trades during preferable market states and by having increased noise filtering.

As a next step, we compare the performance of ANFIS/T2b against standard ANFIS. The results in Table 5.7 indicate that a lower Sharpe ratio was obtained in 5 stocks (BLT, CW, HSBA, BP, and HBOS), with improvements showing in the remaining 10 stocks. In terms of profitability, ANFIS/T2b also obtained lower results in 5 stocks (ANTO, BLT, BA, CW and HSBA) and an increase in the remaining 10 stocks. The initial indications are
that although, in general, ANFIS/T2b performed better than standard ANFIS, the increased overall performance was less than that obtained by ANFIS/T2a. This is also demonstrated from the summary statistics in Table 5.7 which show a lower average improvement in terms of the Sharpe ratio; however, at a slightly improved average return per stock.

Our statistics show that at low to moderate intraday trading frequencies, which were primarily driven by optimising models to maximise Sharpe ratios, both ANFIS/T2a and ANFIS/T2b performed better than ANFIS, with ANFIS/T2a showing the best risk-adjusted performance at this level of trading frequency. This also conveys an important message for model designers. Increased model complexity, as in the case of ANFIS/T2b when compared to ANFIS/T2a, does not guarantee a better risk-adjusted performance. Moreover, it provides an indication that our approach of adding incremental levels of model complexity results in identifying the best balance between complexity and risk-adjusted performance.

We note that both ANFIS/T2a and ANFIS/T2b achieve a substantial increase in profit per trade when compared to standard ANFIS. In the case of ANFIS/T2a, the model performed better in 12 out of 15 stocks whilst, in the case of ANFIS/T2b, improvements showed in 11 stocks. This is an indication that increasing the number of intraday trades can possibly result in an increase in the overall profitability. However, this comes at a cost of increased uncertainty in trade profitability due to more exposure to microstructure noise. This is investigated in our second experiment, presented in the next section.

5.4.2 Experiment 2: Comparison of T1 and T2 models under different noise levels

In our second set of experiments, we investigate the performance of standard ANFIS and the proposed IT2 TSK models under increasing levels of uncertainty. As described in Section 5.2.3, we propose an innovative method to control the level of uncertainty using the signal threshold.
Table 5.7: ANFIS/T2b performance comparison against the standard ANFIS after 100-day out-of-sample period. The results depict variations from those produced by ANFIS in Table 5.5. The number of trades, Sharpe ratio and profit per trade are presented as percentage differences. Profit column shows the percentage point (p.p.) differences.

<table>
<thead>
<tr>
<th>Company</th>
<th>Symbol</th>
<th>No. of Trades</th>
<th>Sharpe Ratio</th>
<th>Profit (p.p.)</th>
<th>Profit /Trade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antofagasta</td>
<td>ANTO</td>
<td>−5.55%</td>
<td>+3.24%</td>
<td>−1.22</td>
<td>−3.77%</td>
</tr>
<tr>
<td>BHP Billiton</td>
<td>BLT</td>
<td>−5.68%</td>
<td>−0.41%</td>
<td>−0.24</td>
<td>+2.9%</td>
</tr>
<tr>
<td>British Airways</td>
<td>BA</td>
<td>−0.03%</td>
<td>+1.34%</td>
<td>−0.08</td>
<td>−0.1%</td>
</tr>
<tr>
<td>British Land Company</td>
<td>BLND</td>
<td>−2.61%</td>
<td>+4.41%</td>
<td>+0.3</td>
<td>+5.72%</td>
</tr>
<tr>
<td>Sky</td>
<td>SKY</td>
<td>+0.27%</td>
<td>+4.02%</td>
<td>+0.53</td>
<td>+3.46%</td>
</tr>
<tr>
<td>Cable and Wireless</td>
<td>CW</td>
<td>−1.02%</td>
<td>−0.69%</td>
<td>−1.02</td>
<td>−0.51%</td>
</tr>
<tr>
<td>Aviva</td>
<td>AV</td>
<td>−0.93%</td>
<td>+12.72%</td>
<td>+0.71</td>
<td>+10.41%</td>
</tr>
<tr>
<td>Diageo</td>
<td>DGE</td>
<td>−0.83%</td>
<td>+0.64%</td>
<td>+0.07</td>
<td>+1.84%</td>
</tr>
<tr>
<td>HSBC Holdings</td>
<td>HSBA</td>
<td>−1.51%</td>
<td>−9.57%</td>
<td>−0.61</td>
<td>−7.45%</td>
</tr>
<tr>
<td>Rio Tinto</td>
<td>RIO</td>
<td>−0.19%</td>
<td>+9.73%</td>
<td>+0.37</td>
<td>+8.04%</td>
</tr>
<tr>
<td>BP</td>
<td>BP</td>
<td>−0.43%</td>
<td>−1.96%</td>
<td>+1.19</td>
<td>+4.42%</td>
</tr>
<tr>
<td>Lloyds Banking Group</td>
<td>LLOY</td>
<td>−1.08%</td>
<td>+11.76%</td>
<td>+1.79</td>
<td>+5.61%</td>
</tr>
<tr>
<td>Tesco</td>
<td>TSCO</td>
<td>−0.01%</td>
<td>+2.72%</td>
<td>+0.62</td>
<td>+1.38%</td>
</tr>
<tr>
<td>HBOS</td>
<td>HBOS</td>
<td>−6.54%</td>
<td>−5.98%</td>
<td>+0.7</td>
<td>+9.36%</td>
</tr>
<tr>
<td>Xstrata</td>
<td>XTA</td>
<td>−1.54%</td>
<td>+20.04%</td>
<td>+0.92</td>
<td>+18.02%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td><strong>−1.37%</strong></td>
<td><strong>+2.24%</strong></td>
<td><strong>+0.87</strong></td>
<td><strong>+2.83%</strong></td>
</tr>
</tbody>
</table>

We test fixed thresholds starting from 0.08% down to 0.02% in steps of 0.02%. From the standard ANFIS results (see Table 5.8) we can identify the increase in the number of trades in line with decreasing thresholds. For example, in the case of ANTO, the number of trades increases from an average of 20 trades per day at a threshold of 0.08% up to 143 trades per day at a threshold of 0.02%. At the higher thresholds (0.08% and 0.06%), standard ANFIS showed no negative return at the 0.08% threshold level and only 1 negative return (XTA) at the 0.04% threshold level. In the case of the lower thresholds (0.04% and 0.02%), standard ANFIS showed 2 negative returns at the 0.04% level (XTA and RIO) and 4 at the 0.02% level (XTA, RIO, BLT and BLND).

The results presented in Table 5.8 were used as the basis for comparing the improvements attained from our IT2 models (presented in Figure 5.5 and Table 5.9). From Figure 5.5, we immediately note a steep drop in the average profit per trade and average profitability at the
Table 5.8: ANFIS performance after 100-day out-of-sample period across different levels of return threshold (uncertainty).

<table>
<thead>
<tr>
<th>Symbol</th>
<th>No. of Trades</th>
<th>Sharpe Ratio</th>
<th>Profit /Trade (£)</th>
<th>No. of Trades</th>
<th>Sharpe Ratio</th>
<th>Profit /Trade (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANTO</td>
<td>2036</td>
<td>0.3388</td>
<td>18.90</td>
<td>3687</td>
<td>0.2644</td>
<td>11.03</td>
</tr>
<tr>
<td>BLT</td>
<td>475</td>
<td>0.3224</td>
<td>47.56</td>
<td>1016</td>
<td>0.2822</td>
<td>17.50</td>
</tr>
<tr>
<td>BA</td>
<td>6248</td>
<td>1.5420</td>
<td>54.92</td>
<td>7830</td>
<td>1.4161</td>
<td>49.34</td>
</tr>
<tr>
<td>BLND</td>
<td>921</td>
<td>0.4664</td>
<td>30.12</td>
<td>1956</td>
<td>0.2577</td>
<td>10.48</td>
</tr>
<tr>
<td>SKY</td>
<td>1197</td>
<td>0.6189</td>
<td>29.18</td>
<td>2606</td>
<td>0.5404</td>
<td>16.05</td>
</tr>
<tr>
<td>CW</td>
<td>4308</td>
<td>1.4777</td>
<td>104.04</td>
<td>4709</td>
<td>1.2952</td>
<td>103.87</td>
</tr>
<tr>
<td>AV</td>
<td>1074</td>
<td>0.4136</td>
<td>19.09</td>
<td>2429</td>
<td>0.4510</td>
<td>12.08</td>
</tr>
<tr>
<td>DGE</td>
<td>480</td>
<td>0.3605</td>
<td>35.52</td>
<td>1203</td>
<td>0.3807</td>
<td>16.11</td>
</tr>
<tr>
<td>HSBA</td>
<td>426</td>
<td>0.4560</td>
<td>37.44</td>
<td>1129</td>
<td>0.2912</td>
<td>16.24</td>
</tr>
<tr>
<td>RIO</td>
<td>535</td>
<td>0.2040</td>
<td>23.31</td>
<td>1071</td>
<td>0.1253</td>
<td>9.24</td>
</tr>
<tr>
<td>BP</td>
<td>1577</td>
<td>0.8068</td>
<td>33.85</td>
<td>3220</td>
<td>0.8132</td>
<td>20.43</td>
</tr>
<tr>
<td>LLOY</td>
<td>3140</td>
<td>0.5685</td>
<td>20.61</td>
<td>5849</td>
<td>0.6705</td>
<td>18.56</td>
</tr>
<tr>
<td>TSCO</td>
<td>3211</td>
<td>0.9950</td>
<td>20.61</td>
<td>5505</td>
<td>1.2153</td>
<td>30.90</td>
</tr>
<tr>
<td>HBOS</td>
<td>4708</td>
<td>0.4693</td>
<td>24.90</td>
<td>7322</td>
<td>0.4543</td>
<td>20.72</td>
</tr>
<tr>
<td>XTA</td>
<td>911</td>
<td>0.1222</td>
<td>16.61</td>
<td>1715</td>
<td>-0.0218</td>
<td>-1.55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbol</th>
<th>No. of Trades</th>
<th>Sharpe Ratio</th>
<th>Profit /Trade (£)</th>
<th>No. of Trades</th>
<th>Sharpe Ratio</th>
<th>Profit /Trade (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANTO</td>
<td>6795</td>
<td>0.3356</td>
<td>13.21</td>
<td>14270</td>
<td>0.1378</td>
<td>3.39</td>
</tr>
<tr>
<td>BLT</td>
<td>2652</td>
<td>0.1725</td>
<td>5.15</td>
<td>2200</td>
<td>-0.1748</td>
<td>-13.62</td>
</tr>
<tr>
<td>BA</td>
<td>9738</td>
<td>1.3835</td>
<td>45.94</td>
<td>12520</td>
<td>1.1199</td>
<td>37.72</td>
</tr>
<tr>
<td>BLND</td>
<td>4462</td>
<td>0.2408</td>
<td>7.52</td>
<td>10466</td>
<td>-0.1042</td>
<td>-2.05</td>
</tr>
<tr>
<td>SKY</td>
<td>5481</td>
<td>0.6820</td>
<td>12.83</td>
<td>9385</td>
<td>0.4024</td>
<td>7.50</td>
</tr>
<tr>
<td>CW</td>
<td>5260</td>
<td>1.4373</td>
<td>97.96</td>
<td>6204</td>
<td>1.3938</td>
<td>85.68</td>
</tr>
<tr>
<td>AV</td>
<td>6104</td>
<td>0.2137</td>
<td>3.38</td>
<td>12930</td>
<td>0.0402</td>
<td>0.67</td>
</tr>
<tr>
<td>DGE</td>
<td>3348</td>
<td>0.2761</td>
<td>6.05</td>
<td>10495</td>
<td>0.1898</td>
<td>2.37</td>
</tr>
<tr>
<td>HSBA</td>
<td>3836</td>
<td>0.3572</td>
<td>6.91</td>
<td>10153</td>
<td>0.0430</td>
<td>0.53</td>
</tr>
<tr>
<td>RIO</td>
<td>2643</td>
<td>-0.0750</td>
<td>-2.78</td>
<td>7534</td>
<td>-0.3773</td>
<td>-8.48</td>
</tr>
<tr>
<td>BP</td>
<td>6978</td>
<td>0.9543</td>
<td>15.39</td>
<td>11295</td>
<td>0.7763</td>
<td>10.04</td>
</tr>
<tr>
<td>LLOY</td>
<td>9621</td>
<td>0.7212</td>
<td>17.63</td>
<td>13693</td>
<td>0.6223</td>
<td>13.12</td>
</tr>
<tr>
<td>TSCO</td>
<td>7928</td>
<td>1.5421</td>
<td>26.08</td>
<td>10865</td>
<td>1.2703</td>
<td>20.43</td>
</tr>
<tr>
<td>HBOS</td>
<td>11467</td>
<td>0.3684</td>
<td>15.50</td>
<td>16007</td>
<td>0.2740</td>
<td>9.61</td>
</tr>
<tr>
<td>XTA</td>
<td>3771</td>
<td>-0.1765</td>
<td>-6.81</td>
<td>2992</td>
<td>-0.2403</td>
<td>-20.21</td>
</tr>
</tbody>
</table>
0.02% level. When investigating further, we see that the ANFIS summary statistics in Table 5.9 show increasing average profitability per stock, from 25.43% at the 0.08% level, up to 33.42% at the 0.04% level. This is attained by reducing average profit per trade but with an increase in trading activity. However, it is important to note that at the 0.04% level, the return threshold is equivalent to our transaction costs, hence making it much more difficult to obtain profitable trades beyond this level.

When observing the Sharpe ratio summary statistics, it shows a trend that as the threshold moves from higher to lower levels, the model experiences decreasing levels of the Sharpe ratio (see Table 5.9 and Figure 5.5). This pattern is attributed to increasing levels of risk (uncertainty) in line with decreasing thresholds. However, in comparison to standard ANFIS, the proposed IT2 models show increasing improvements in the Sharpe ratio, the average profit per stock and the average profit per trade. In the case of the Sharpe ratio, improvements range from an increase of 1.86% at the 0.08% threshold level, up to 11.33% at the 0.02% level. Improvements in the average profit per stock range from 0.05 p.p. at the 0.08% threshold level, up to 1.57 p.p. at the 0.02% level. A similar trend is achieved by the ANFIS/T2b model (Figure 5.5). The increase in all three measures, especially the increase in profitability at a lower risk, indicates the superior performance of the proposed ANFIS/T2 models when compared to standard ANFIS. The pattern indicates that this increase in performance gets more pronounced at lower thresholds which experience higher effects of microstructure noise (refer to Table 5.9).

As a final step, we validate our results using statistical tests. The tests are carried out on the average Sharpe ratio, the average profit per stock and the average profit per trade obtained using a paired $t$-test on the 15 stocks. Table 5.9 shows that the difference in performance results at the return lower thresholds (0.04% and 0.02%) are all significant. The tests strengthen our earlier claims of the improved performance of our proposed ANFIS/T2 models against standard ANFIS models at levels of higher uncertainty due to more exposure
5.4 Results and analysis

Fig. 5.5: Trends on various measures after 100-day out-of-sample period and at different degrees of return threshold (uncertainty).

to microstructure noise. This makes our proposed models more suitable contenders for HFT environments. At the higher return thresholds (0.08% and 0.06%), both models experience
Table 5.9: Summary statistics for standard ANFIS and the corresponding variations in the ANFIS/T2 models. Results are obtained over a 100-day out-of-sample period across different levels of return threshold (uncertainty). Bold figures for performance measures average Sharpe ratio, average profit/stock and average profit/trade indicate a rejected paired-sample t-test. The test applies the null hypothesis that the difference in results (ANFIS vs. ANFIS/T2) comes from a normal distribution with mean equal to zero and unknown variance at 5% sig. level.

<table>
<thead>
<tr>
<th>Measure</th>
<th>0.08%</th>
<th>0.06%</th>
<th>0.04%</th>
<th>0.02%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard ANFIS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of trades</td>
<td>31247</td>
<td>51247</td>
<td>90084</td>
<td>151009</td>
</tr>
<tr>
<td>Average Sharpe Ratio</td>
<td>0.6108</td>
<td>0.5624</td>
<td>0.5622</td>
<td>0.3582</td>
</tr>
<tr>
<td>Average Profit / Stock</td>
<td>25.43%</td>
<td>28.96%</td>
<td>33.42%</td>
<td>26.78%</td>
</tr>
<tr>
<td>Average Profit / Trade (£)</td>
<td>35.13</td>
<td>23.40</td>
<td>17.60</td>
<td>9.78</td>
</tr>
<tr>
<td><strong>ANFIS/T2a Improvement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of trades</td>
<td>−2.55%</td>
<td>−2.13%</td>
<td>−1.62%</td>
<td>−0.73%</td>
</tr>
<tr>
<td>Average Sharpe Ratio</td>
<td>+1.86%</td>
<td>+3.3%</td>
<td>+3.93%</td>
<td>+11.33%</td>
</tr>
<tr>
<td>Average Profit / Stock (p.p.)</td>
<td>+0.05</td>
<td>+0.57</td>
<td>+0.9</td>
<td>+1.57</td>
</tr>
<tr>
<td>Average Profit / Trade</td>
<td>+3.32%</td>
<td>+5.15%</td>
<td>+3.78%</td>
<td>+7.05%</td>
</tr>
<tr>
<td><strong>ANFIS/T2b Improvement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of trades</td>
<td>−2.73%</td>
<td>−1.63%</td>
<td>−1.22%</td>
<td>−0.42%</td>
</tr>
<tr>
<td>Average Sharpe Ratio</td>
<td>+2.27%</td>
<td>+3.45%</td>
<td>+3.58%</td>
<td>+14.69%</td>
</tr>
<tr>
<td>Average Profit / Stock (p.p.)</td>
<td>+0.05</td>
<td>+0.78</td>
<td>+0.91</td>
<td>+1.98</td>
</tr>
<tr>
<td>Average Profit / Trade</td>
<td>+2.68%</td>
<td>+4.53%</td>
<td>+3.89%</td>
<td>+7.71%</td>
</tr>
</tbody>
</table>

...some insignificant measures due to lower improvements against the standard ANFIS results.

This indicates that at reduced uncertainty, the introduction of IT2 fuzzy sets has less effect on trading performance. Another indication from our results is that although ANFIS/T2b show significant increases in both average Sharpe ratio and average profit per trade across all return thresholds, the highest improvements in both standard ANFIS and ANFIS/T2a are demonstrated in the lowest threshold (0.02%). This highlights the importance of identifying, incrementally, the right balance between model complexity and the specific level of uncertainty.
5.5 Conclusion

In this chapter we investigated the risk-adjusted performance of trading algorithms from a higher frequency perspective. This work extends the research of Johnson et al. (2013) and Brogaard et al. (2014) who have highlighted the need for new theories in support of high-frequency financial phenomena during which the human traders lose the ability to react in real time. The challenge is two-fold. Firstly, from the literature we know that at higher sampling frequencies the effect of microstructure noise becomes more pronounced. This noise presents an additional level of complexity in the design of trading models. Secondly, infamous mishaps like the “flash crash” of 6 May 2010 and the subsequent more stringent regulatory regimes that are coming into force (e.g. introduction of MiFiD2 from January 2017) indicate the higher level of risk involved in HFT and also the increasing responsibility on algorithm designers to implement adequate risk controls.

We address this challenge by consolidating our research findings from previous chapters and extending this to new proposed models based on T2 fuzzy logic. Firstly, we propose two innovative and practical methods of how the ANFIS model, a popular AI technique in finance, can be improved by introducing IT2 fuzzy sets with a minimal increase in complexity. This presents a stepping stone from our ANFIS findings presented in Chapter 3. The main benefit is to minimise the effect of uncertainty caused by microstructure noise, hence reducing the overall risk. By extending the clustering data partitioning ideas that we use in Chapter 4, especially in view of different intraday volatility levels, here we apply a similar approach for model structure identification which is also used as a basis for our IT2 extensions. This approach results in more compact and efficient fuzzy models. The proposed T2 methods show a significant increase in both risk-adjusted trading performance and profitability when compared to standard ANFIS and B&H methods.

Secondly, we shed more light on the theoretical market efficiency debate within HFT. Our results extend the findings of a number of authors (Holmberg et al., 2013; Rechenthin
and Street, 2013; Schulmeister, 2009) who claim possible breaks in market efficiency at short
time frames. By utilising a combination of technical rules on 2-minute returns with holding
periods ranging from 2 to 10 minutes, we extend our market efficiency investigations that we
present in Chapters 3 and 4 to a more microscopic perspective and identify possible pockets
of profitability in shorter time windows. As a result of this, we manage to identify a positive
link between higher order fuzzy systems and risk-adjusted trading performance.

Thirdly, although a number of authors (e.g. Aladi et al., 2014; Sepulveda et al., 2006)
demonstrate the increased capability of IT2 models to handle increased uncertainty when
compared to T1, we provide deeper insight on the benefits of adopting IT2 models from
the perspective of different levels of trading risk (uncertainty) and trading frequency. We
conclude that the introduction of T2 fuzzy sets (in contrast to the models presented in
Chapters 3 and 4) exhibit the highest tangible benefits in trading scenarios that are more
exposed to microstructure noise, making our models more ideal (and feasible) for HFT
environments.
Chapter 6

Conclusion

This thesis proposes a number of improvements to existing artificial intelligence approaches applied to algorithmic and high-frequency trading. Our point of departure stems from two stronghold perspectives in finance. Firstly, while the literature in finance and economics stands firm on the Efficient Market Hypothesis (EMH), many traders continue to make buy and sell decisions based on historical data. Whilst the debate continues, this thesis stimulates discussion against a strict adherence to the EMH. Our interest is in challenging the EMH in its weak form, and in particular our focus is on the application of technical trading rules (in conjunction with AI) at short intraday time horizons. Our second perspective is that investors typically look beyond profits when evaluating trading strategies. The reason is that brilliant results could mask the fact that an algorithm might carry huge risks to achieve the claimed performance. However, from our reviewed literature we identify that time and time again, surveys indicate that the majority of AI research publications with an application to finance put risk as a second priority (if at all). These risk exposures are driving stricter regulations as reflected in the upcoming regulatory regimes like MiFID2 that increase the responsibility on algorithm designers to implement adequate risk controls.

These arguments drive our research to identify innovative techniques directed towards intraday trading algorithms with the objective to improve risk-adjusted performance. This
chapter provides a summary of the presented work, lists the main contributions and discusses limitations and future research.

6.1 Summary of presented work

First, in Chapter 1, we bring together the literature review from three knowledge domains, namely high-frequency trading, financial risk and artificial intelligence, with special focus on fuzzy logic techniques. The intention is to highlight key intersecting concepts from each domain and present arguments from which the research objectives that are addressed in this thesis are derived. Furthermore, as a precursor to the fuzzy logic techniques which are extensively used in subsequent chapters, Chapter 2 provides the necessary mathematical ground work on fuzzy logic with a specific focus on Type-1 Takagi-Sugeno-Kang (TSK) FLSs and Interval Type-2 (IT2) FLSs.

Following the initial introductory chapters, in Chapter 3 we address the concerning number of published studies which promote unsuitable approaches to model training and selection when applied for trading purposes. Contrary to the more common day-ahead predictions, the debated profitability of moving average rules is explored using high-frequency data in an intraday trading scenario. We investigate holding times of trading positions in the region between 10 minutes to one hour. Moreover, rather than using fixed target returns, which is quite a common practice, we investigate a range of return bands in the region between 0.1% and 0.5%. These act as a threshold for unprofitable small trades.

A data set consisting of stock trades effected on the London Stock Exchange is used. We compare the performance attained from three representative milestone models in neurocomputing, namely Neural Network (NNs), Adaptive Neuro-Fuzzy Inference System (ANFIS) and Dynamic Evolving Neuro-Fuzzy Inference System (DENFIS) models, and also propose the more sophisticated eANFIS architecture. Different model optimisation functions are compared using single risk-return functions, an innovative combination of different risk-return
functions via an ensemble, Root Mean Squared Error (RMSE), period return and models optimised without considering transaction costs. Unlike most studies which analyse model performance at a single point in time in the out-of-sample, the stability of these models is compared by analysing the time series (performance profile) of daily performance measures over the full out-of-sample period.

By extending the risk-adjusted model enhancements presented in the Chapter 3, in Chapter 4 we seek further improvements by introducing a method on how to effectively adapt trading algorithms to the different market states (risk scenarios) that evolve during a typical trading day. From the outset of this chapter, it is highlighted that the literature of how AI algorithms can benefit from this information on volatility for trading purposes is scarce. This problem is addressed in two parts. Part 1 (Section 4.2) investigates whether the use of intraday realised volatility as a proxy for uncertainty can benefit AI trading algorithms. As a motivating example we make use of a popular NN model and use it as our benchmark to measure the effectiveness of our approach (although the proposed extension can be used in conjunction with other machine learning techniques). In Part 2 (Section 4.3), the research is extended to explore whether intraday uncertainty information can be used to improve risk-based money management decisions. The challenge is to identify a method which dynamically adjusts trading frequency and position size depending on the varying degrees of risk at an intraday level with the objective to improve overall risk-adjusted trading performance. Another challenge that is addressed in this chapter is to investigate whether model profitability can also be improved by allocating more capital to preferable intraday trading scenarios.

Finally, we investigate profitability and the risk-adjusted performance of trading algorithms by addressing a key challenge presented by microstructure stylised facts which cause the theoretical price to be tampered by an error term, or “noise”. To this aim, Chapter 5 extends the T1 techniques explored in Chapters 3 and 4 by seeking better performance
from higher order fuzzy systems to handle the increased uncertainties inherent in an HFT scenario. Practical methods are presented of how the popular ANFIS T1 model can be generalised to an IT2 TSK fuzzy system. To address the criticism of increased complexity that is normally attributed to T2 models, this problem is addressed with a minimal increase in design and computational complexity. The presented experiments aim to investigate whether T2 FLSs provide a viable alternative for trading purposes in view of improving risk-adjusted performance. As a further step, the viability of T2 FLSs is explored at different levels of trading frequencies.

6.2 Contributions

In this thesis, we convey a number of contributions. We consolidate the main contributions under four different perspectives:

1. **The role of fuzzy logic in algorithmic trading**
   
   Managing uncertainty, which in this context we link to risk, is at the heart of fuzzy logic. Any investigation into the role that fuzzy logic can contribute to high-frequency algorithmic trading is scarce. Furthermore, the relationship between fuzzy logic and risk-adjusted performance has, to our best knowledge, never been investigated.

   In Chapter 3, we show the effectiveness of our approach using three milestone models from neuro-fuzzy literature, namely ANNs, ANFIS and DENFIS. It is shown that although ANNs live up to their popularity in terms of performance, the combination of fuzzy logic using ANFIS provides a more stable performance in terms of both profitability and risk-adjusted performance. Out-of-sample performance stability is a highly appealing feature from an investor’s perspective. We also introduce an innovative ANFIS ensemble model (eANFIS) which combines and dynamically selects models which are optimised using different risk-return objective functions. The model
automatically shifts between various degrees of risk tolerance. On the one hand, the model switches to riskier risk-adjusted objective functions during intraday periods of high performance. This avoids the possible adoption of an overly risk-averse model with the possibility of increased penalisation in profitability. On the other hand, the model dynamically reverts back to more risk-sensitive objective functions during unfavourable time windows, hence reducing the possibility of large losses. It is shown that this approach improves the intraday trading performance of AI models.

In Chapter 4, we propose an innovative fuzzy money management approach which dynamically adapts trading frequency and position sizing decisions across intraday trend (profit) and volatility (risk) states to improve the overall risk-adjusted performance. Contrary to many studies that suggest trading rules based on a fixed position sizing strategy, fixed return thresholds and fixed volatility thresholds, our approach dynamically evolves a continuous trading decision surface across the whole intraday trend-volatility space. We demonstrate the applicability of our fuzzy logic approach by presenting a hybrid fuzzy model as an extension to a popular neural network trading model. Core to our approach is a fuzzy clustering method that automatically identifies preferable pockets of intraday risk-adjusted profitability at different trend and volatility levels with the goal of increasing our position frequency and size in successful regions and reducing these in regions that are likely to result in losses. The results show significant improvements in both profitability and risk-adjusted performance when compared to standard NN and Buy-and-Hold (B&H) methods.

Finally, in Chapter 5 we propose two innovative and practical methods of how the ANFIS model can be improved by introducing IT2 fuzzy sets with a minimal increase in complexity. The main benefit is to minimise the effect of uncertainty caused by microstructure noise, hence reducing the overall risk. We propose a clustering data partitioning approach for model structure identification. This results in more
compact and efficient fuzzy models. The proposed T2 methods show a significant increase in both risk-adjusted trading performance and profitability when compared to standard ANFIS and B&H methods. Moreover, we provide deeper insight into the benefits of adopting IT2 models from the perspective of different levels of trading risk (uncertainty) and trading frequency. We conclude that the introduction of T2 fuzzy sets exhibit the highest tangible (investor) benefits in trading scenarios reflecting higher exposure to microstructure noise, making our models more ideal (and feasible) for HFT environments.

2. Market efficiency during short intraday time windows
This thesis fuels the ongoing debate on market efficiency. In particular, this thesis investigates the predictive performance of technical trading rules against the weak form of the efficient market hypothesis, since only past prices are used as predictor variables. Our contribution focuses on time windows of between a few minutes up to one hour. In our experiments we use high-frequency trade data pertaining to a number of stocks listed on the London Stock Exchange. The research chapters (Chapters 3, 4 and 5) demonstrate how technical signals can be used in conjunction with AI techniques, resulting in profitable returns.

In Chapter 3, we have demonstrated a simple yet effective extension of common technical rules by considering a dynamic amalgamation of moving average signals. In our approach, we use AI techniques to dynamically tune the trend signals based on the current intraday market speed. Our results demonstrate that the proposed dynamic moving approach outperforms the risk-adjusted performance obtained from the standard technical rules. By applying heat maps, we identify concentrated regions of a higher Sharpe ratio in areas of higher holding position times and return bands. The heat maps indicate that, contrary to a number of claims in the literature, technical rules do manage to identify pockets of profitability in the high-frequency range, with holding
periods of between 10 minutes to one hour. Of particular interest is the fact that for specific stocks, the heat maps identify more than one region of profitability, hence providing a clearer indication to traders on the possible profitable trading strategies.

In Chapter 4, we cast further doubts on market efficiency. Although a number of authors claim a relationship between the possible periodic breakdown of market efficiency and volatility, we extend these theoretical claims to a more granular intraday level by identifying a link between the profitability of technical rules with different levels of price volatility at short intraday time horizons. We demonstrate this by presenting a number of experiments which show how algorithmic models can benefit from increased profitability and risk-adjusted performance by dynamically discriminating between different intraday volatility states.

Finally, in Chapter 5 we extend our market efficiency investigations to a more microscopic perspective and identify possible pockets of profitability in shorter time windows. By utilising a combination of technical rules on 2-minute returns with holding periods ranging between 2 to 10 minutes, we manage to identify a positive link between higher order fuzzy systems and risk-adjusted trading performance.

In all of the proposed models, we have also demonstrated that the obtained results outperform B&H methods.

3. **Model design and evaluation for trading purposes**

The implementation of risk-adjusted performance control has, to our best knowledge, not been studied in an intraday high-frequency setting before. In Chapter 3, we show that when applying the win ratio as the objective function, a higher win ratio does not necessarily result in a profitable model, due the possibility of larger losses. Similarly, our results provide a clear indication that trading models based on RMSE and period return optimisation provide weak out-of-sample performance when compared to using risk-adjusted functions. Our experiments also show that not considering transaction
costs during model training can lead to biased results in real-world applications. All the above results can offer a new research direction to the numerous published studies (which from our reviewed surveys form the majority) which adopt non risk-adjusted objective functions.

In contrast to common approaches in the literature which evaluate models using performance measures at an arbitrary single point in time (e.g. only at the end of the out-of-sample period), we propose the application of the cumulative Sharpe ratio. This approach provides a deeper understanding of the time-varying performance profile of the applied models. We observe least variance in the case of both Sharpe and Sortino ratio optimisation models. We further extend our improvement ideas by proposing a novel method of combining different risk-adjusted objective functions and demonstrate its effectiveness, outperforming single risk-adjusted objective functions.

Another important consideration for model designers is the model adaptiveness. In Chapter 3 and Chapter 4 we present fuzzy logic methods that show how the underlying trading models can be designed to be adaptive to different levels of intraday risk. In both scenarios we show improvements in profitability and risk-adjusted performance when compared to the more rigid approaches that are common in the literature. Moreover, in Chapter 5 we demonstrate the importance, and the resulting trading effects, of adopting robust models (such as IT2 models) that can perform well under increased noise levels, especially when applying higher frequency data.

4. Management insights

Our contributions also convey a number of financial management insights. The results presented in this thesis should be of utmost interest for decision makers and also encourage further research and investment by firms which will be impacted by new regulatory regimes, such as MiFID2, that will demand that the employed trading systems meet numerous requirements, particularly when it comes to risk controls. In
6.3 Limitations and future work

Chapters 3, 4 and 5 we present a number of innovative techniques of how existing algorithmic and high-frequency trading models can be improved by increasing risk-adjusted performance but without compromising overall profitability; rather, it is increased.

Our proposed models show a stepwise approach by starting from the popular models used in finance literature, such as NNs and ANFIS, and show how by introducing incremental components on the base AI models (rather than a whole overhaul of existing investment), these can be enhanced to meet the set risk objectives.

6.3 Limitations and future work

In conclusion, this thesis opens up a number of avenues for further research. As a start, we propose future research paths that can be followed from a fuzzy logic perspective. In Chapters 3 and 4, we identify increased model stability and risk-adjusted improvements by introducing T1 models. In Chapter 5, further improvements are identified in higher frequency bands with the introduction of IT2 fuzzy sets and adaptive FOU sizes. In our approach, we adopt incremental enhancements on popular models in AI literature with an application to finance. This choice of approach was consciously followed in order for the proposed models to remain practical to use and easier to compare by keeping the additional parameters to a minimum. Our first research proposition is that we do not exclude the possibility of further improvements (in the management of uncertainty and hence trading performance) by exploring additional incremental steps in model configuration complexity, different rule extraction methods, different method functions, and more complex T2 rules or defuzzification methods.

In this thesis, we primarily make use of Sugeno fuzzy models. In comparison to Mamdani fuzzy models, Sugeno models are known to provide higher accuracy but they do this at the
cost of lower interpretability. In our case, the Sugeno approach was selected since it provides
a more natural extension to NNs, the latter acting as a popular and strong benchmark in
finance. However, we appreciate that in line with additional model complexity, this can
mitigate wider adoption. An interesting research path can seek to mitigate this risk by
investigating the performance of similar Mamdani models which tend to increase model
interpretability. To meet this aim, the exploration might also be extended to the combination
of fuzzy logic models with more interpretable machine learning techniques such as decision
trees.

We also identify a number of research avenues from a finance perspective. By taking
into consideration our conclusions from Chapters 4 and 5, our first research proposition is
to investigate the effect of alternative FOU tuning frequencies, which in this thesis we limit
to a daily basis. In particular, any possible beneficial relationship between the FOU tuning
frequency and the diurnal patterns of market intraday activity, a stylised fact in finance,
remains an open question.

Secondly, in our experiments we limit our data set to trade data from a list of stocks
listed on the London Stock Exchange. However, further research can be expanded to include
wider portfolios and markets (including forex) which can experience higher trading activity,
possibly resulting in higher microstructure noise. This can help to identify wider scale
gains in risk-adjusted performance and increase the practical applicability of the proposed
techniques.

Thirdly, in this thesis we adopt the Sharpe ratio as our main risk-adjusted performance
measure. However, the Sharpe ratio does not separate between variability in gains and losses,
hence it attributes penalisation to both upside and downside variability. This might not
represent the interest of investors who would rather welcome positive variability in gains. In
this study, however, we favour model stability and hence our interest is more in identifying
algorithmic trading models that can offer steady returns. Moreover, from our literature review
we identify that in many studies, Sharpe ratio measures lead to similar rankings of more sophisticated performance ratios. Hence, we decide in favour of the Sharpe ratio as our main risk-adjusted performance measure due to its simplicity and thus easy application, and also due to its widespread acceptance both in literature and in practice. This, however, does not exclude further research to identify any additional improvements that can be gained by adopting more sophisticated risk-adjusted measures or their combination (as indicated in the ensemble approach presented in Chapter 3).
References


References


Appendix A

Further results to Chapter 3
Further results to Chapter 3

Fig. A.1: Trading performance by optimising the Sortino ratio. The plots show the cumulated Sortino ratio (y-axis) on the n-th day (x-axis) in the out-of-sample period.
Fig. A.2: The plots compare the results obtained from ANFIS Sharpe optimisation (experiment 1a) and eANFIS (experiment 1c) with those from MA optimisation (experiment 2d), and show the corresponding cumulated Sharpe ratio (y-axis) on the $n$-th day (x-axis) in the out-of-sample period.