Centre for Computational Finance and Economic Agents

WP068-13

Working Paper Series

Vince Vella and Wing Lon Ng

Improving risk-adjusted performance profile of intraday trading models with Neuro-Fuzzy techniques and moving average high-frequency price signals

2013

University of Essex

CCFEA

www.ccfea.net

Improving risk-adjusted performance profile of intraday trading models with Neuro-Fuzzy techniques and moving average high-frequency price signals

Vince Vella[a], Wing Lon Ng[a,*]

[a]*Centre for Computational Finance and Economic Agents (CCFEA), University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK*

Abstract

We present a performance comparison of risk-adjusted intraday trading strategies based on dynamic non-linear models using the more traditional Artificial Neural Network, as well as Adaptive Neuro-Fuzzy Systems (ANFIS) and Dynamic Evolving Neuro Fuzzy Systems (DENFIS). The model selection process takes into account the risk-return measures together with flexible position holding periods and a return band filter, employing a dynamic combination of moving average signals. Our results show that these models can be successfully applied to support intraday trading strategies, especially when considering constraints such as transaction costs and trading hours, which existing approaches in the literature do not account for.

Keywords: High-frequency trading, ANFIS, DENFIS, Feed-Forward Network, Dynamic Moving Average

1. Introduction

The profitability of trading rules is an incessant debate. Recent literature however claims that profitability of technical trading rules has possibly moved to higher frequency prices as a result of more efficient markets and faster algorithmic trading (Schulmeister, 2009). Notwithstanding this, other findings show that aggressive high-frequency trading (HFT) does not lead to the expected high excessive returns (Kearns et al., 2010). Whilst many former studies focuses on the application of models solely to predict market movements (e.g. Son et al. (2012)), traders in financial markets are typically interested in risk-adjusted performance rather than

**Corresponding author.* Tel.:0044-1206-874684

Email addresses: vvella@essex.ac.uk (Vince Vella), wlng@essex.ac.uk (Wing Lon Ng)

just price predictions themselves (Choey and Weigend, 1997; Xufre Casqueiro and Rodrigues, 2006).

This paper provides new insights into 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. Our first contribution is to present an analysis of the time-varying risk-adjusted performance profile of the applied models by focusing on their dynamics over a the entire out-of-sample testing period and their daily cumulative performance. In contrast to common approaches in the literature which evaluates models using risk-return measures at an arbitrary single point in time (e.g. the end of the sample period), our goal is to provide a deeper understanding of the time-varying performance profile of the applied models.

Our second major contribution is the simple but yet effective extension of common technical trading strategies by considering a '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?*" We investigate the profitability of models on less aggressive HFT, with holding periods between 10 minutes and 1 hour, using 5 minute prices of a set of stocks listed on the London Stock Exchange during the period 2007-2008. 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, as that 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.

As a third contribution in this paper, we extend our trading systems with decision rules accounting for transaction costs and trading hours, and compare the time series of risk adjusted performance measures obtained from different model optimisation functions such as risk-return functions, Root Mean Squared Error (RMSE) and models not considering transaction costs. When training and evaluating a trading system, most former studies only have very limited view of what constitutes successful investment decisions, defining on 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, a smaller forecast error does not necessarily translate into increased trading profits (Brabazon and O'Neill, 2006). Recently, Krollner et al. (2010) find that over 67% of the investigated studies use the forecast error as an evaluation metric and identify a lack of literature examining if machine learning techniques can improve an investor's risk-return trade-off. They also find that over 80% of the papers report that their model outperformed the benchmark model, but most of them do not consider real world constraints at all (see also Álvarez Díaz, 2010).

An interesting finding by Schulmeister (2009) was that by examining 30 minute prices it was identified that beyond the 1990s the profitability of technical trading rules has moved to higher frequency. He questioned that due to more efficient markets this might have moved to even higher frequencies for the more recent periods.

Another finding in Tsai and Wang (2009) and Krollner et al. (2010) is that Artificial Neural Networks (ANNs) are identified to be the dominant machine learning technique in this area. 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). Successful application of ANFIS in trading applications by predicting stock price was demonstrated in Lin et al. (2002); Gradojevic (2007); Kablan and Ng (2011) and many others, with the latter study the only one identified that is focused on high frequency trading.

Equally important is the fact that to keep a successful trading edge these 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 uniformed traders. Implementing evolution requires an ability to balance learning and changing while still respecting the past accumulated knowledge (Marsland, 2009). With a focus on dynamic learning of rules from data Kasabov and Song (2002) introduced pioneering work on evolving neuro fuzzy systems with the introduction of Evolving Neural-Fuzzy Inference System (DENFIS) and its application for time-series prediction. In contrast to ANFIS which optimises the structure by batch learning, DENFIS evolve through incremental, hybrid (supervised/unsupervised), learning, and accommodate new input data, including new features, new classes, etc., through local element tuning. To our best knowledge DENFIS was not previously applied in a high frequency setting.

With these advances in AI and soft computing techniques this paper presents and compares the performance of buy-sell signals generated from a combination of moving average trading rules with the application of ANNs, ANFIS and DENFIS.

The remainder of the paper is structured as follows. In Section 2 we first introduce the moving average signals and explain on how these can be combined to model stock returns. We then provide details about our experiment approach describing model components and underlying prediction and trading algorithms. Section 3 presents the data, our findings and a discussion in the light of existing literature. Section 4 concludes.

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 Artificial Neural Networks. 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.

Based on the findings in the current literature, our experiment approach focuses on a number of objectives:

1. We explore the debated profitability of moving average rules, particularly focusing on high frequency data. In our experiments we use a set of stocks listed on the London Stock Exchange.
2. In contrast to common trading system designs that focuses on a fixed target returns, we apply of a return band in the region between 0.1% and 0.5% which acts as a filter for unprofitable small trades.
3. Evaluate the profitability of less aggressive HFT strategies, with holding trading positions time (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. Consider real world intraday constraints like trading costs, realistic trading hours and no overnight positions.
5. Compare the risk-adjusted daily cumulative performance attained from the more traditional Artificial Neural Networks (ANNs) with the more recent ANFIS and DENFIS models.

Our experiment setup consisted of two core modules (see Figure 1), the Return Prediction Module, which later feeds trading signals to the Trading System Module. Sections 2.1 to 2.4 describe our Return Prediction Models. Section 2.5 explains our Trading algorithm. In Section 2.6 we explain how we measure and evaluate model performance.

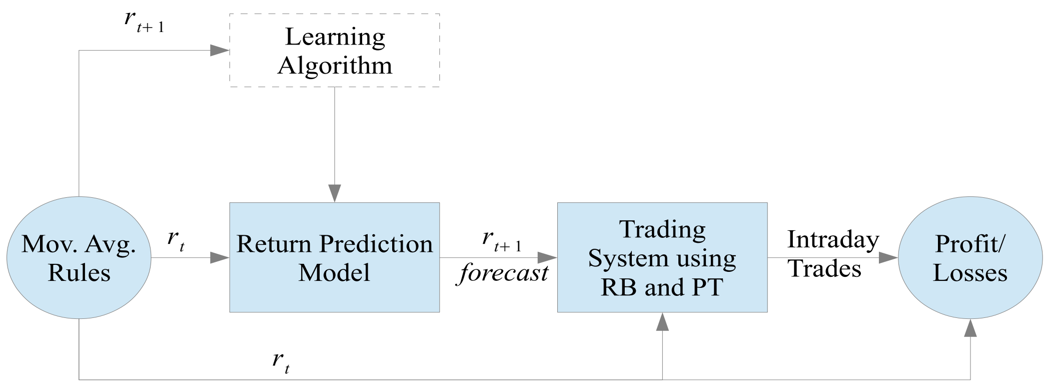

Figure 1: Experiment Setup

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 truncated history of past prices $\{p_t, p_{t-1}, p_{t-N+1}\} \in \mathbb{R}_+^N$. A function $d : I_t \to \Omega$ maps the information set I_t at time t to a space of investment decisions $\Omega = \{short, 0, long\}$, indicating short, neutral or long positions, respectively.

For our time series we use 5 minute continuously compounded returns since returns have much better statistical properties than price levels. These intraday returns are defined as:

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

where $log(.)$ denotes the natural logarithm.

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

$$m_t^n = \frac{1}{n} \sum_{i=0}^{n-1} p_{t-i}, \tag{2}$$

where $i = 0, 1, 2, ..., n-1$ is the "memory span" of the rule. Consider the signal at time t defined as

$$s_t = \begin{cases} long & if \quad p_t \geq (1+\phi)m_t^n \\ 0 & if \quad (1-\phi)m_t^n \leq p_t < (1+\phi)m_t^n \\ short & if \quad p_t < (1-\phi)m_t^n \end{cases} \tag{3}$$

where ϕ is the bandwidth of the rule for whiplash reduction.

Another popular variation of the rule is:

$$s_t^{n_1,n_2} = \begin{cases} (m_t^{n_1} - m_t^{n_2}) & if \quad |m_t^{n_1} - m_t^{n_2}| > \phi \\ 0 & else \end{cases} \tag{4}$$

where n_1 and n_2 are the short and long moving averages, respectively.

We investigate whether intraday high frequency returns can be successfully predicted by making use of buy and sell signals as inputs and use this information to build a profitable trading algorithm. In this setting the linear regression model is

$$r_t = \alpha + \sum_{i=1}^{p} \beta_i s_{t-i}^{n_1,n_2} + \epsilon_t, \tag{5}$$

where the error term ϵ_t is an independent variable with mean 0 and variance σ_t^2.

Kearns et al. (2010) recently 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 in our experiment we investigate the effect of applying less aggressive HFT strategies. We slow down our trading by investigating (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 than compared to the common approaches in the literature that calibrate their trading systems only based on an arbitrary target return. Brabazon and O'Neill (2006) showed that similar use of extended close in intraday trading scenarios can perform better than standard Stop-Loss, Take-Profit and Buy-and-Hold strategies.

We first identify a model that predicts the next five-minute stock return by taking a series of moving average signals as input variables. The three moving average rules utilised are (n_1, n_2) = [(1,5), (5,10), (10,15)], where n_1 and n_2 are in 5 minute time bars. In line with eq. (5), by combining these rules the linear specification of the return prediction model is

$$r_t = \alpha + \beta_i s_{t-i}^{1,5} + \beta_i s_{t-i}^{5,10} + \beta_i s_{t-i}^{10,15} + \epsilon_t, \tag{6}$$

where

$$s_t^{n_1,n_2} = m_t^{n_1} - m_t^{n_2}. \tag{7}$$

In the following, we describe how the ANFIS, DENFIS and feed-forward network (FFN) models are adapted for our experiments.

2.2. 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 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 and predict the stock return.

A popular technique is the Adaptive Neuro-Fuzzy inference system (ANFIS) suggested by Jang (1993). ANFIS presents a Takagi-Sugeno (TS) model in a different architecture that albeit the mathematical underpinnings are similar the structure is formulated to permit ANN learning techniques. Following from the standard ANFIS model and eq. (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 The output of each node $O_{1,i}$ is the membership grade for the moving average signals $\{s_{t-1}^{1,5},\ s_{t-1}^{5,10},\ s_{t-1}^{10,15}\}$. Different types and number of input membership functions, and corresponding premise parameters, are tested in our model calibration process (see Table 1).

Layer 2 Every node in this layer is fixed. In this layer the t-norm is used to "AND" the membership grades, for example the product:

$$O_{2,i} = w_i = \mu_{A_i}(s_{t-1}^{1,5})\mu_{B_i}(s_{t-1}^{5,10})\mu_{C_i}(s_{t-1}^{10,15}). \tag{8}$$

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

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2 + w_3}. \tag{9}$$

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

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i(\alpha_i s_{t-1}^{1,5} + \beta_i s_{t-1}^{5,10} + \gamma_i s_{t-1}^{10,15} + \delta_i). \tag{10}$$

where $\bar{w}_i$ is the normalised firing strength from the previous layer and f_i is a linear function of the moving average signals with consequent parameters $\{\alpha_i, \beta_i, \gamma_i, \delta_i\}$.

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:

$$O_{5,i} = r_t = \sum_i \bar{w}_i f_i = \frac{\sum w_i f_i}{\sum_i w_i}. \tag{11}$$

Table 1: Parameters tested for ANFIS

Parameter	Parameter Value Set
Training Data Size (days)	{5, 10}
Input Membership Functions Shape	{Gaussian, Generalised Bell}
Number of Input Membership Functions	{2, 3}
Training epochs	{10, 20, 50}

Jang (1993) proposes premise and consequent parameters learning using a combination of Steepest Descent and Least Squares Estimation (LSE). The total parameter set is split into two sets, a Set_1 of premise (nonlinear) parameters and a Set_2 of consequent (linear) parameters. ANFIS learning uses a two pass algorithm. In a forward pass Set_1 is unmodified and Set_2 is computed using a LSE algorithm. This is followed by a backward pass where Set_2 is unmodified and Set_1 is computed using a gradient descent algorithm such as back-propagation.

Although the application of ANFIS in finance has been widely studied, most studies only employ daily data whereas applications to intraday HFT are still scarce. Kablan and Ng (2011) successfully applied ANFIS to predict price movement from intraday tick data sampled at high frequency. Due to the intraday volatility seasonality, 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 whether fuzzy logic models could be further improved for trading purposes by automatically get updated 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 1). We test and compare all $2 \times 2 \times 2 \times 6 = 48$ permutations of the parameter combinations in our sensitivity analysis for ANFIS models.

2.3. 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 learning, and their application for dynamic time series prediction. Kasabov and Filev (2006) analyse rule generation from a data stream perspective. DENFIS is based on the ECOS (Evolving COnnectionist Systems) framework suggested by Kasabov and Song (2002).

In our study, this distance is calculated by using the normalised Euclidean distance of a new sample to the cluster centres:

$$dist_i = ||S_i - Cc_i||, \tag{12}$$

where S_i is the current example consisting of a moving average vector [$s_{t-1}^{1,5}$, $s_{t-1}^{5,10}$, $s_{t-1}^{10,15}$] and Cc_i is the centre of cluster i. A threshold value $Dthr$ is defined that limits the cluster size. When each data sample arrives, the steps below are carried out (see also Kasabov and Song, 2002):

1. At the initial step of ECM, the first input data sample is considered as the first cluster with the data itself as the first cluster centre and cluster centre set to zero.
2. As a new data sample arrives, the distance $dist$ between the samples and all other existing cluster centres are determined.
3. The cluster with minimum distance $dist_{min}$ is selected. If $dist_{min}$ is less than the $radius$ then the sample is associated to the cluster and no updates are required.
4. For every existing cluster the respective $dist$ is added to the $radius$ (let this value be range). The cluster with minimum $range$ is selected. If $range$ is less than $2 \times Dthr$ then the sample belongs to the cluster, the $radius$ of this cluster is updated to $(range/2)$ and the cluster centre is updated by positioning it in the line joining the data sample and the cluster centre so that now the distance between the new centre and the sample is equal to the new $radius$ value.
5. If $range$ is greater than $2 \times Dthr$ then a new cluster is created as in step 1. Each new cluster generates a new rule and evolves the structure of the system. As new data samples arrive new rules can be created and existing rules can be updated incrementally.

The inference engine in DENFIS is composed of m fuzzy rules indicated as follows:

$$\begin{aligned} IF \quad & (s_{t-1}^{1,5}\ is\ R_{i,1})\ AND\ (s_{t-1}^{5,10}\ is\ R_{i,2})\ AND\ (s_{t-1}^{10,15}\ is\ R_{i,3}) \\ THEN \quad & r_t\ is\ f_i(s_{t-1}^{1,5}, s_{t-1}^{5,10}, s_{t-1}^{10,15}), \end{aligned}$$

where i=1,2,...,m and $j = 1, 2, 3$; $(s_{t-1}^{n_1,n_2}\ is\ R_{i,j})$ are $m \times 3$ fuzzy propositions as m antecedents for m fuzzy rules; and $R_{i,j}$ are fuzzy sets defined by their fuzzy membership functions $\mu_{R_{i,j}} : s_{t-1}^{n_1,n_2} \rightarrow [0, 1]$. In the consequent part, linear functions f_i, where i=1,2...,m are employed. In DENFIS, all fuzzy membership functions are triangular type functions defined by three parameters:

$$\mu(s_{t-1}^{n_1,n_2}, a, b, c) = max\left(min\left(\frac{s_{t-1}^{n_1,n_2} - a}{b - a}, \frac{c - s_{t-1}^{n_1,n_2}}{c - b}\right), 0\right), \tag{13}$$

Table 2: Parameters tested for DENFIS

Parameter	Parameter Value Set
Training Data Size (days)	{10, 20, 30}
ECM Clustering Threshold	{0.04, 0.06, 0.08, 0.1, 0.12}
Number of Rules in Dynamic FIS	{3, 5 ,6}

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 $S = [s_{t-1}^{1,5},\ s_{t-1}^{5,10},\ s_{t-1}^{10,15}]$ the result of the inference, which is the predicted return r_t, is calculated as the weighted average of each rule's output:

$$r_t = \frac{\sum_{i=1}^{m} w_i f_i(S)}{\sum_{i=1}^{m} w_i} \tag{14}$$

where $w_i = R_{i,1}(s_{t-1}^{1,5})R_{i,2}(s_{t-1}^{5,10})R_{i,3}(s_{t-1}^{10,15})$. Following a similar online learning approach presented in Takagi and Sugeno (1985) and Jang (1993), the linear functions in the consequent parts of the rules are updated using a recursive weighted LSE, applying also a forgetting factor. 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 2). We test and compare all $3 \times 5 \times 3 = 45$ permutations of the parameter combinations in our sensitivity analysis of DENFIS models.

2.4. Neural Network Model

The application of ANNs for the moving average rules models in Hudson et al. (1996), Gençay (1996) and Fernandez-Rodrıguez et al. (2000) was based on the fact that their research identified that, under general regularity conditions, a sufficiently complex single hidden-layer feed-forward network can approximate any member of a class of functions to any degree of accuracy, where the complexity of a single hidden-layer feed-forward network (FFN) is measured by the number of units in the hidden layer. Following eq. (5) the single-layer feed-forward network regression model with lagged buy and sell signals and with d hidden units can written as

$$r_t = \alpha_0 + \sum_{j=1}^{d} \beta_j G(\alpha_j + \sum_{i=1}^{p} \gamma_{ij} S_{t-1}^{n_1,n_2}) + \epsilon_t, \tag{15}$$

where

$$G(u) = \frac{1}{1 + exp(-u)} \tag{16}$$

Table 3: Parameters tested for FFN

Parameter	Parameter Value Set
Training Data Size (days)	{5, 10}
Number of Hidden Units	{5, 10, 20}
Max Training epochs	{1000}

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

2.5. Trading Algorithm

The second module of our trading system (see Figure 1) consists of a trading and money management algorithm that takes the 5 minute return predictions from the first module and performs trades based on specific rules (see also Tan et al., 2011; Vanstone and Finnie, 2009, 2010). 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 paper 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 predicted return value has to be greater (smaller) than the upper (lower) limits of a specific return band 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 return bands 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:

$$position_t = \begin{cases} long: & r_t > returnband \\ short: & r_t < returnband \\ 0: & otherwise. \end{cases} \tag{17}$$

In our experiments we train the models on a daily rolling window basis, hence adapting the model on the most recent market scenarios of 100 in-sample days, followed by another 100 out-of-sample days, totalling 10,200 5 minute prices for

Algorithm 1 Pseudo code for position time

if $position\_duration > PT$ **then**
  if $position\_direction = position\_direction_{pred}$ **then**
    $position\_state \leftarrow keep\ open$
    $position\_open\_time \leftarrow current\_time$
  else
    $position\_state \leftarrow close$
  end if
else
  $position\_duration \leftarrow current\_time - position\_open\_time$
end if

each sample. 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 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) 50K GBP worth of stock at the current market price. A total of five open positions are allowed at one point in time, limiting 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 overestimate the performance 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 position time (PT) holding periods between 10 minutes to 1 hour are considered in the model selection process for each stock during the in-sample period. If after the holding period 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). 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.

2.6. Evaluation

Although many algorithms minimise errors such as the mean squared error (MSE) or RMSE (Alves Portela Santos et al., 2007; de Faria et al., 2009), models constructed using these criterion may not perform well when used for trading pur-

Table 4: Models applied in the experiments

Experiment	AI Algorithms Tested	MA Model	Optimisation Criteria
1	ANFIS, DENFIS, FFN	Dynamic	Sharpe Ratio
1	ANFIS, DENFIS, FFN	Dynamic	Sortino Ratio
2	ANFIS, DENFIS, FFN	Dynamic	Sharpe Ratio, No Cost
2	ANFIS, DENFIS, FFN	Dynamic	RMSE
3	-	MA(1,5)	-
3	-	MA(5,10)	-
3	-	MA(10,15)	-

poses since the costs of predictive errors are assumed to be symmetric. Moreover, existing approaches in the literature do not account for realistic transaction costs and trading hour, reporting possibly biased results. Based on these findings and following from the literature above, we construct a number of trading models by applying different AI algorithms and optimisation functions (see Table 4).

In our first experiment combine the dynamic moving average model with the different AI methods and choose either the Sharpe ratio or Sortino ratio (both defined below) as optimisation criteria. In our second experiment, we apply the same model combinations, but either (a) do not account for transaction costs in the training period or (b) optimise the system entirely on forecast accuracy, in order to see whether and how the ignorance of these constraints would have an impact on the trading performance. Finally, to assess the effectiveness of the dynamic moving average model, we also compare our models against the trading performance of fixed moving average models in a third experiment.

To evaluate the trading system and compare the performance across different models, we apply five different measures (Kablan and Ng, 2011): Shape ratio, Sortino ratio, cumulative return, profit ratio, and win ratio. The Sharpe ratio indicates to investors whether the returns of an asset or a portfolio come from a smart trading strategy or excess risk. The Sharpe ratio is defined as

$$Sharpe\ Ratio = \frac{R_p - r_f}{\sigma_p}, \tag{18}$$

where R_p denotes the expected return, r_f the risk-free interest rate and σ_p the portfolio volatility. The Sharpe ratio measures the risk premium per each unit of total risk in an investment asset or a portfolio. Investments with higher Sharpe ratios are often preferred because the higher the Sharpe ratio translates into better risk-adjusted performance. Similarly, the Sortino ratio is defined as

$$Sortino\ Ratio = \frac{R_p - r_f}{\sigma_{neg}}, \tag{19}$$

where σ_{neg} denotes the standard deviation of only negative asset returns. Thus, the Sortino ratio measures the risk premium per each unit of downside risk in an investment asset or a portfolio. The cumulative return indicates the overall probability of the strategy since the first trade, similar to a buy-and-hold scenario.

It should also be noted that the choice of performance function determines how often the system trades and what percentage of its trades are winning trades. The profit ratio indicates a system's ability to generate profits over losses and is defined as

$$Profit\ Ratio = \frac{Total\ Gain/Number\ of\ winning\ trades}{Total\ Loss/Number\ of\ losing\ trades} \tag{20}$$

The win ratio is the ratio between the number of winning trades and losing trades and is defined as

$$Win\ Ratio = \frac{Total\ Number\ of\ winning\ trades}{Total\ Number\ of\ losing\ trades} \tag{21}$$

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

Albeit many researchers claim the results of their algorithmic trading models by analysing a set of performance measures at a single point in time, covering a specified number of days in the out-of-sample period, 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 resulting in $\prod_{i=1}^{n} pset_i$ different models per algorithm, where n is the number of algorithm parameters and $pset_i$ is the number of unique discrete values tested for each parameter i. In order to select a model with good generalisation capabilities each model was trained and tested over a 100 day period using a rolling window approach. Each model was trained on $day_{n-1} - day_{n-1-s}$ days of 5 minute returns and tested on day_n 5 minute returns, where $n = \{1, 2, \ldots, 100\}$ is a day index and s is the training size in days used for the specific model. This requires that the first s days from the data set are reserved for training.

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). By inspecting our 100 day-by-day trading results and analyse 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 Figures 2-4 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.

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 5) during the period 01/06/2007 to 30/06/2008 (excluding weekends, holidays and after hour trading). Data is sampled at 5 minute intervals using the last trade price every five 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 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. This emanates from the 2007-2008 economic crisis situation. In the following, results for experiment 1 are discussed in Section 3.1, and for experiment 2 and 3 in Section 3.2 (see also overview in Table 4).

3.1. Results for Experiment 1

Based on our first 100 day in-sample period, we 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)). As indicated by the heat map plots for ANFIS (Figure 2), DENFIS (Figure 3) and FFN (Figure 4), the plots in general indicated concentrated regions of higher Sharpe ratio in areas of higher holding position times and return bands.

This indicates the effectiveness of applying these two filters in our trading models. These plots also provide an indication that our trading frequency of interest, i.e. taking a less aggressive holding period of between 10 minutes to 1 hour, can show very positive results (unlike the stated difficulty with aggressive high-frequency trading with position holding periods of between 10 milliseconds and 10 seconds, see Kearns et al. (2010)). Of particular interest is the fact that for specific stocks the heat maps identify more than one area of profitable regions,

Table 5: Descriptive Statistics of 5 Minute Returns

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

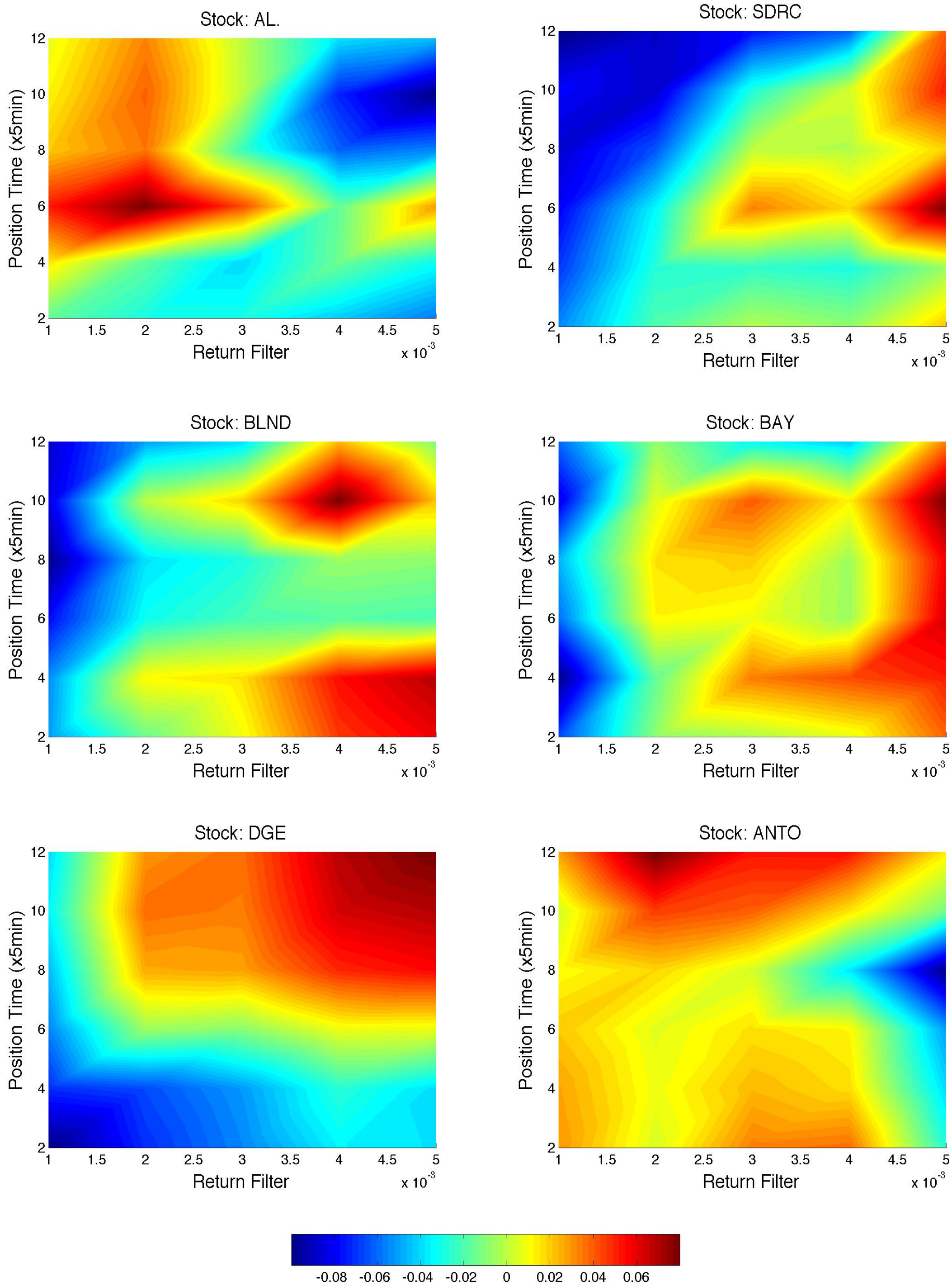

Figure 2: Heat map identifying sensitivity of ANFIS Model and highest Sharpe ratio for different position time and return band regions (in-sample)

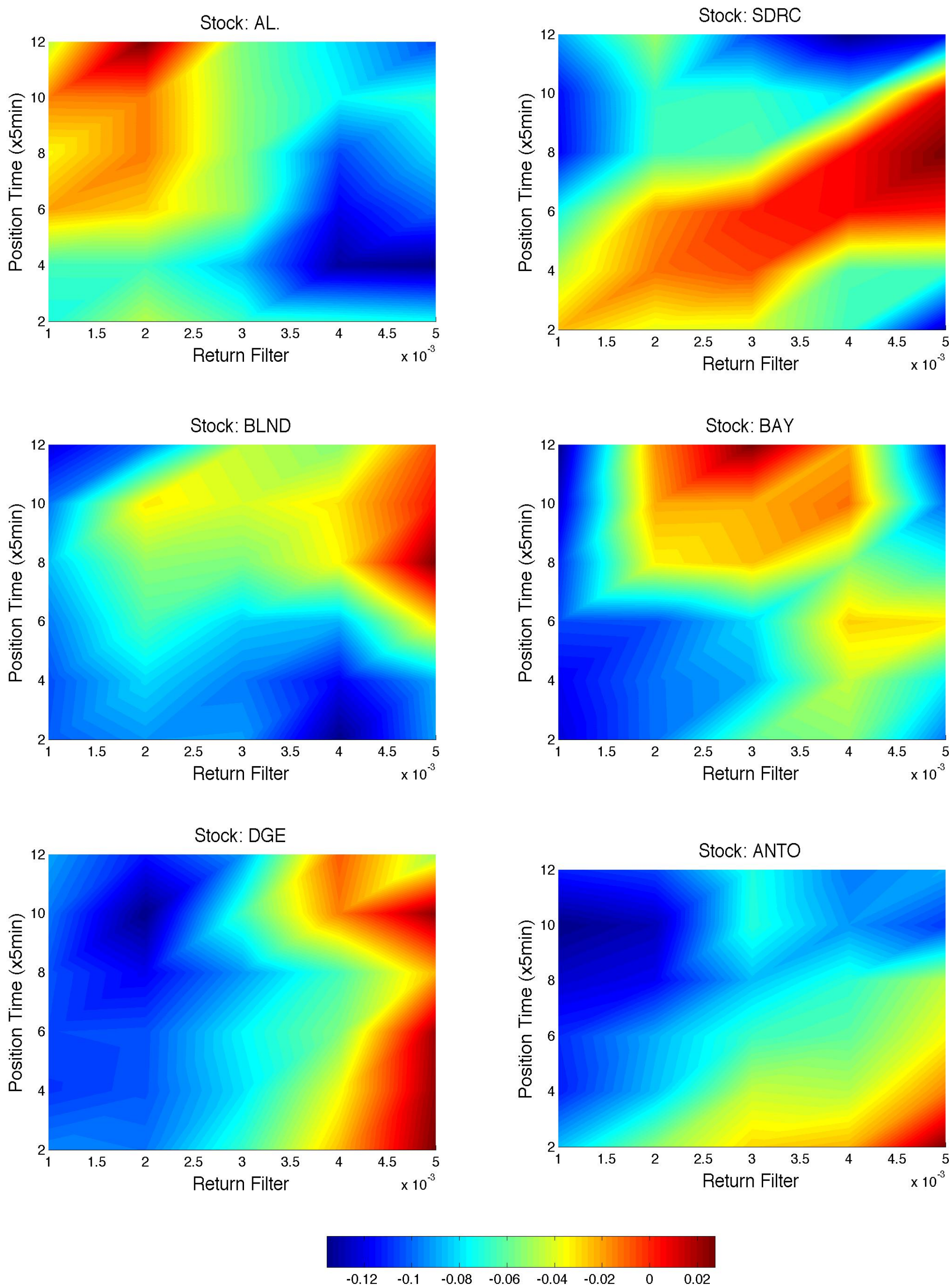

Figure 3: Heat map identifying sensitivity of DENFIS Model and highest Sharpe for different position time and return band regions (in-sample)

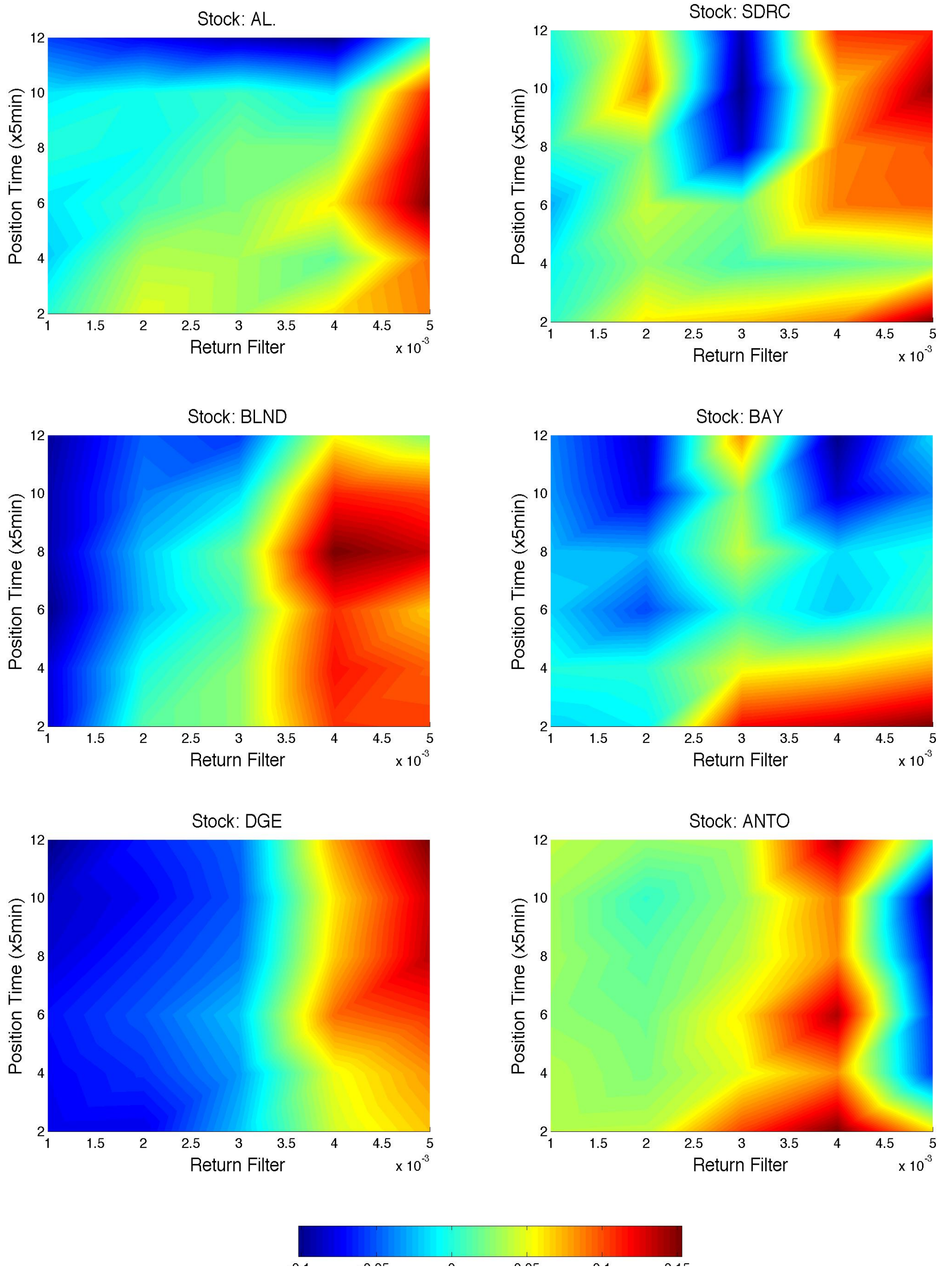

Figure 4: Heat map identifying sensitivity of FFN Model and highest Sharpe ratio for different position time and return band regions (in-sample)

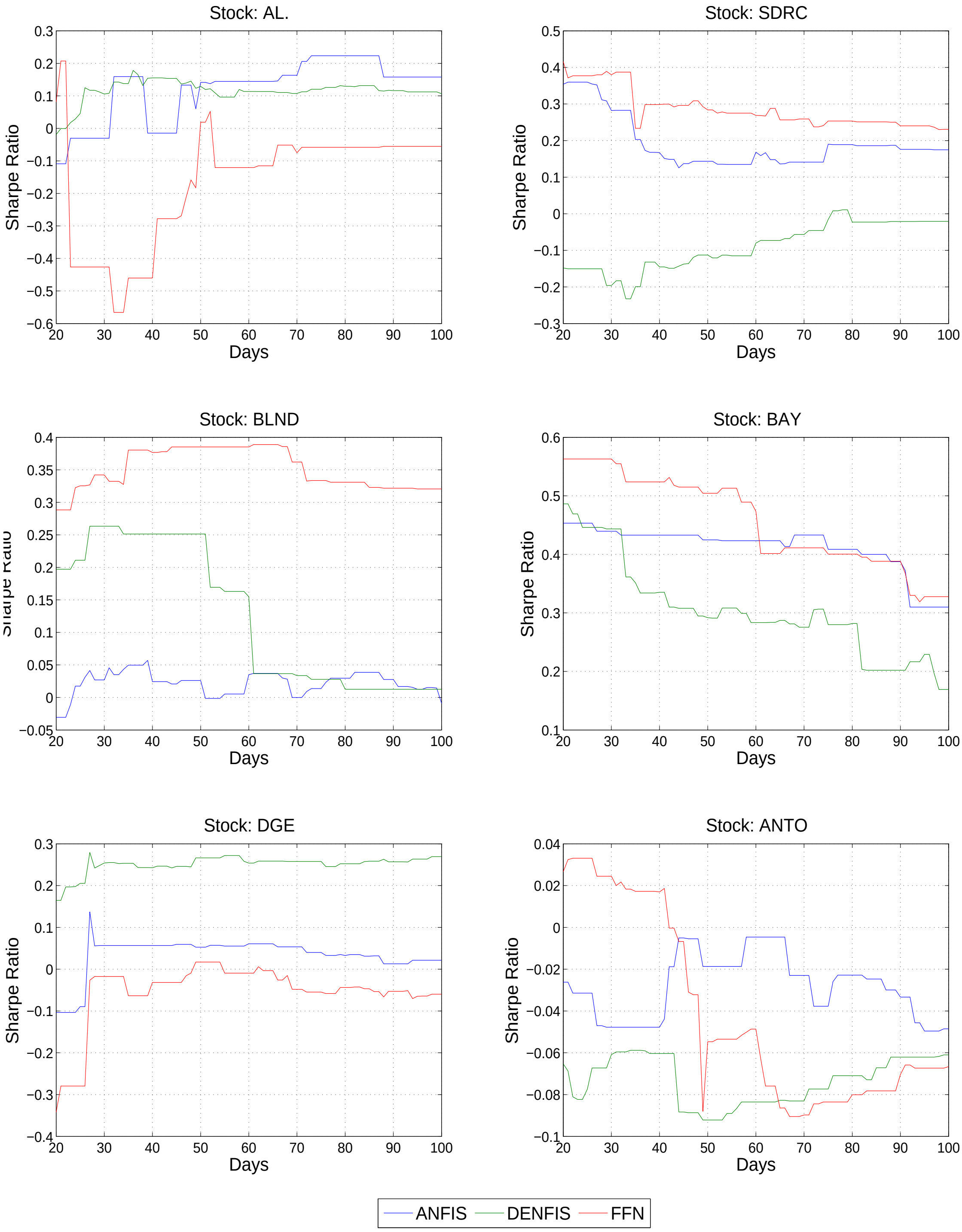

Figure 5: Trading Performance by optimising Sharpe ratio. The plots show for each stock the cumulated Sharpe ratio (y-axis) on the n-th day (x-axis) in the out-of-sample.

Table 6: Model performance using Sharpe ratio optimisation over the 100 day out-of-sample period (bold fond indicates best result among the three AI methods for the specific stock)

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

hence providing a clearer indication to traders on the possible profitable trading strategies.

In the first simulation of the first experiment our model parameter identification was based on applying the Sharpe ratio as our objective function. From the out-of-sample results in Table 6, we see that ANTO was the only stock which has not generated a positive Sharpe ratio across all models. Both ANFIS and DENFIS generated a positive Sharpe ratio in four out of six stocks. In the case of FFN, the model generated positive Sharpe ratio in three out of six stocks, albeit in these three instances it generated the highest Sharpe across the three models.

Although this single point in time performance measurement is the most common approach adopted in literature, this might not be the best way how to investigate the success of a model and our primary interest was to investigate the performance profile of each model over the full 100 day period. Let $measure_{t,t+n}$ represent the aggregated measure from day t to day $t+n$ of the out-of-sample period. Figure 5 shows $Sharpe\ Ratio_{1,20}$ up to $Sharpe\ Ratio_{1,100}$. All stocks had at least one model which generated positive results up to $Sharpe\ Ratio_{1,40}$. After the 40th day only one stock does not generate any positive Sharpe ratios (ANTO). Although a number of studies, all based on daily data (Krollner et al., 2010), show that the profitability of technical analysis has strongly declined or even ceased to exist after early 2000s in the stock market (Schulmeister, 2009), our results

Table 7: Sharpe ratio statistics over the 100 day out-of-sample period following Model Sharpe ratio optimisation

	ANFIS		DENFIS		FFN	
	Mean	Variance	Mean	Variance	Mean	Variance
AL.	0.1208	0.0089	0.1153	0.0012	-0.1661	0.0323
ANTO	-0.0284	0.0002	-0.0732	0.0001	-0.0415	0.0020
BAY	0.4120	0.0016	0.3027	0.0066	0.4566	0.0061
BLND	0.0218	0.0003	0.1254	0.0113	0.3508	0.0010
DGE	0.0335	0.0020	0.2508	0.0004	-0.0555	0.0058
SDRC	0.1920	0.0048	-0.0894	0.0041	0.2859	0.0026
Mean	0.1253	0.0030	0.1053	0.0040	0.1384	0.0083

find 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.

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 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 showed that unlike Kablan and Ng (2011), ANNs still provide a valid benchmark when applying more recent Neuro-Fuzzy models on high frequency price series. ANFIS showed a positive Sharpe ratio on the four (AL., SDRC, BAY, DGE) out of six stocks with a minor loss on one stock (BLND) and lowest loss amongst other models on the 6th stock (ANTO). DENFIS showed clear outperformance on the other models on only one stock (DGE). As also indicated by Table 7 although FFNs show higher mean Sharpe when looking across all stock portfolio, FFNs show more abrupt variations than ANFIS and DENFIS, indicating higher sensitivity to changes in the underlying data features. This can either result in quick increases in Sharpe ratio (AL.) but also drops (ANTO). Hence when considering this aspect, ANFIS resulted in a better model. Table 8 indicates that ANFIS and FFN experience similar performance movements on all stocks except one (ANTO), in which case FFN showed a positive result up to the 40th day.

In the second sets of simulations in experiment 1, we base our model parameter identification process on 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 9). This again confirms the possibility of achieving profitable trading strategies by applying moving average signals with artificial intelligence techniques on high frequency data. ANFIS generates profitable trading

Table 8: Model Sharpe ratio performance correlation over the 100 day out-of-sample period (bold figures indicate significance at 95% level)

	ANFIS-DENFIS	ANFIS-FFN	DENFIS-FFN
AL.	**0.4206**	**0.3489**	**-0.3331**
ANTO	**-0.7748**	**-0.3575**	**0.2973**
BAY	**0.7171**	**0.7966**	**0.8614**
BLND	0.0463	**0.2752**	**0.3227**
DGE	**0.8194**	**0.9360**	**0.8840**
SDRC	**-0.4341**	**0.7858**	**-0.7703**

Table 9: Model performance using Sortino ratio optimisation over the 100 day out-of-sample period (bold fond indicates best result among the three AI methods for the specific stock)

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

results on five out of six stocks with the exception on 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 4 out of 6 stocks with highest results obtained for DGE. When looking at the models performance profile based on Sortino ratio over the 100 days in general (Figure 6), one immediately notices that with the exception of FFN on the AL. stock, the plots exhibit less abrupt variations than those obtained in the Sharpe ratio equivalents.

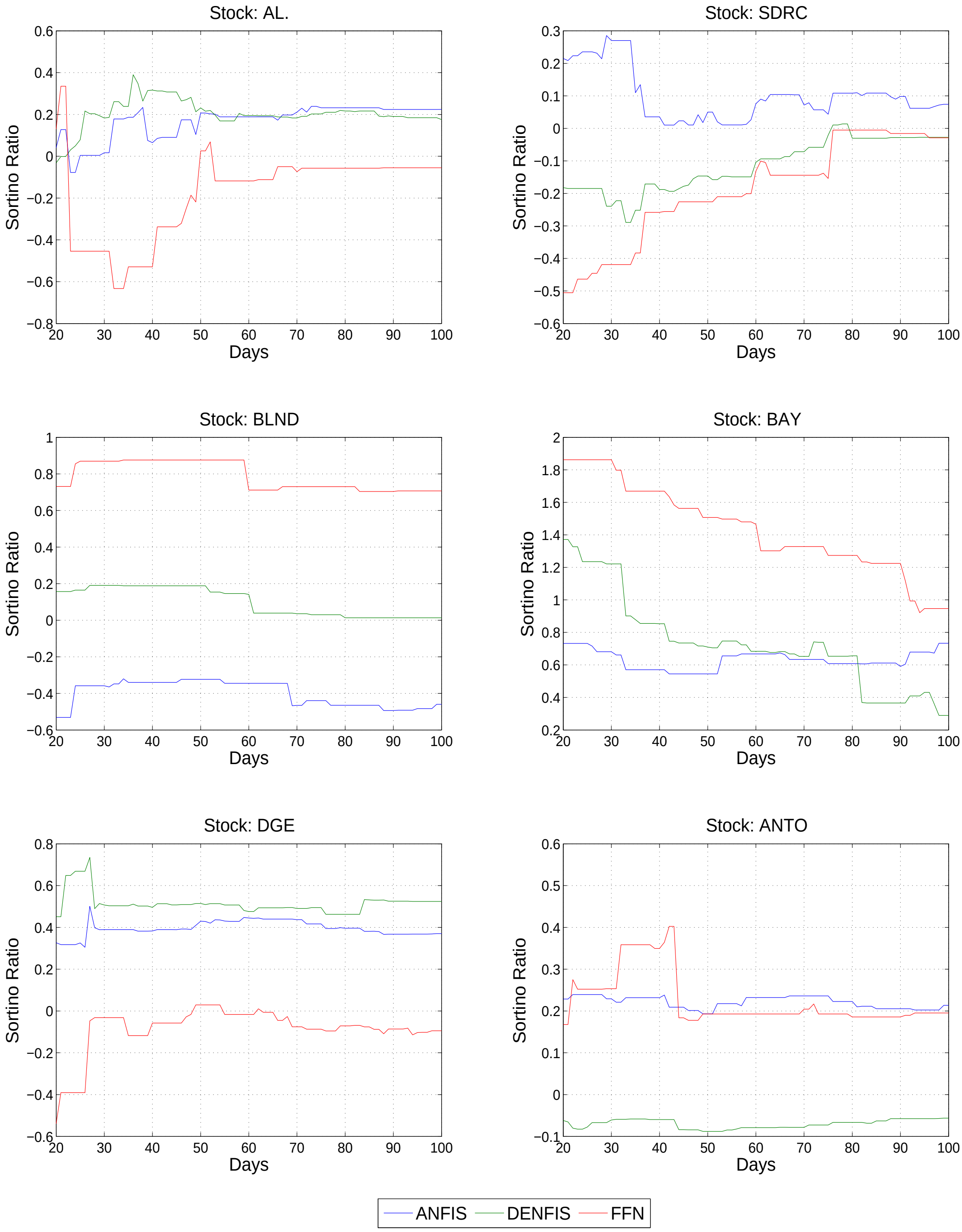

Figure 6: Trading Performance by optimising Sortino ratio. The plots show for each stock the cumulated Sortino ratio (y-axis) on the n-th day (x-axis) in the out-of-sample

Table 10: Sortino ratio statistics over the 100 day out-of-sample period following Model Sortino ratio optimisation

	ANFIS		DENFIS		FFN	
	Mean	Variance	Mean	Variance	Mean	Variance
AL.	0.1686	0.0063	0.2039	0.0048	-0.1775	0.0442
ANTO	0.2210	0.0002	-0.0699	0.0001	0.2244	0.0040
BAY	0.6327	0.0033	0.7309	0.0821	1.4341	0.0737
BLND	-0.4014	0.0049	0.1002	0.0060	0.7870	0.0062
DGE	0.3955	0.0013	0.5143	0.0025	-0.0869	0.0120
SDRC	0.0991	0.0061	-0.1136	0.0064	-0.1891	0.0246
Mean	0.1859	0.0037	0.2276	0.0170	0.3320	0.0275

Table 11: Model Sortino ratio performance correlation over the 100day out of sample period (bold figures indicate significance at 95% level)

	ANFIS-DENFIS	ANFIS-FFN	DENFIS-FFN
AL.	0.2101	**0.5315**	**-0.4892**
ANTO	-0.0972	**0.3273**	**0.4010**
BAY	**0.2369**	-0.0470	**0.9197**
BLND	**0.6732**	**0.7756**	**0.9139**
DGE	**-0.3226**	**0.7714**	**-0.4446**
SDRC	**-0.3424**	**-0.5499**	**0.9030**

From a trading performance perspective this emphasises the importance and effect of the selected risk-return objective function for the applied models. Table 10 summarises the outcome of these plots, a similar conclusion is reached that albeit FFN shows higher mean Sortino ratio when considering the overall portfolio, this however comes at the expense of exhibiting the highest variance over the 100 day out-of-sample period. To the contrary ANFIS shows lowest mean Sortino ratio when considering the whole portfolio but exhibits least variance. This validates the importance of investigating the time varying time-series profile of the cumulative risk-return measures attained when evaluating the underlying models for investment decisions. Table 11 indicates that ANFIS and DENFIS both show similar performance movements with FFN on all stocks except two. In conjunction with the results for the Sharpe ratio discussed earlier, our findings indicate that the selection of specific risk-return measures does not have the same effect on the behaviour of the underlying AI technique and does not guarantee the same time varying performance across all models.

The Sharpe ratio models are compared with the Sortino ratio models by looking at their attained profit ratios (see Figure A.7). In general, the plots show a tendency to group into two performance strata across the 100 day period, a close group competing on the higher side whilst another group giving similar performance on the lower side. With the exception of the last part of the DGE time series, a better performance for the rest of the stocks is attained by ANFIS or FFN models. For SDRC, BAY, DGE and ANTO the plots show that Sortino and Sharpe ratio maximisation has a similar effect on ANFIS and FFN models which result in obtaining a profit ratio time series in a higher or lower band for a specific stock. This distinction is less evident in DENFIS. In contrast to Schulmeister (2009) who demonstrated that aggressive HFT exhibit a surprisingly low profitability, our results show that the cumulative return (Figure A.8) and win ratio (Figure A.9) from a number of models is considerably high. However, in line with Brabazon and O'Neill (2006), a point worth noting is that 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. 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).

3.2. Results for Experiment 2 and 3

In our final part of this paper we present the results attained from benchmark models that are typically found in literature or used in practice (see overview in Table 4). In the first set of simulations in experiment 2, we applied a RMSE minimisation approach for our model selection process. The results in Table 12 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 Figure A.10 we are displaying this for ANFIS over the full out-of-sample period. In the case of FFN, RMSE optimisation clearly outperforms Sharpe optimisation only in AL. These results are in line with Brabazon and O'Neill (2006) and provide clear indication that trading models based on risk-return selection criterion outperform those based on RMSE optimisation.

The recent survey by Krollner et al. (2010) identified that most studies do not consider real world constraints like trading costs (see also Álvarez Díaz, 2010). In our second sets of simulations in experiment 2, we base our model selection criteria on 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

Table 12: Model performance using RMSE optimisation over the 100 day out-of-sample period

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

Table 13: Model performance using Sharpe Ratio optimisation with no transaction costs over the 100 day out-of-sample period

Model	Measure	AL.	ANTO	BAY	BLND	DGE	SDRC
ANFIS	Sharpe Ratio	-0.0762	-0.0273	-0.0647	-0.0220	-0.0464	-0.0263
	Sortino Ratio	-0.0784	-0.0325	-0.0872	-0.0271	-0.0481	-0.0348
	Cum. Return	-3.2531	-0.7929	-1.8076	-0.4576	-0.9125	-0.6583
	Win Ratio	0.5143	0.5214	0.4494	0.5223	0.5375	0.4876
	Profit Ratio	0.7711	0.9205	0.8207	0.9383	0.8595	0.9255
DENFIS	Sharpe Ratio	-0.0523	-0.0395	-0.0878	-0.0425	0.0018	-0.0263
	Sortino Ratio	-0.0576	-0.0473	-0.0987	-0.0473	0.0024	-0.0342
	Cum. Return	-1.7020	-1.1937	-2.4993	-1.4101	0.0232	-0.8667
	Win Ratio	0.5300	0.5178	0.5379	0.5138	0.5144	0.4513
	Profit Ratio	0.8518	0.8898	0.7679	0.8673	1.0046	0.9193
FFN	Sharpe Ratio	-0.0426	-0.0169	-0.0563	-0.0524	-0.0602	-0.0798
	Sortino Ratio	-0.0434	-0.0223	-0.0825	-0.0626	-0.0639	-0.1062
	Cum. Return	-1.6672	-0.4645	-1.5777	-1.4322	-1.1810	-1.6743
	Win Ratio	0.5366	0.4858	0.4630	0.4949	0.5482	0.4905
	Profit Ratio	0.8674	0.9522	0.8446	0.8460	0.8199	0.7976

Table 14: Model performance using Fixed Moving Average (MA) rules over the 100 day out-of-sample period

Model	Measure	AL.	ANTO	BAY	BLND	DGE	SDRC
MA(1,5)	Sharpe Ratio	-0.0792	0.0031	-0.0585	0.0747	-0.1216	-0.0734
	Sortino Ratio	-0.1343	0.0055	-0.0960	-0.1376	-0.2431	-0.1218
	Cum. Return	-1.8815	0.0670	-1.1292	-1.2577	-1.1946	-1.0433
	Win Ratio	0.3169	0.3845	0.3577	0.3448	0.3262	0.3552
	Profit Ratio	0.7673	1.0108	0.8180	0.7721	0.6608	0.8026
MA(5,10)	Sharpe Ratio	-0.0496	-0.0173	-0.0732	-0.1515	-0.1515	-0.0668
	Sortino Ratio	-0.0765	-0.0316	-0.1291	-0.1665	-0.2511	-0.1099
	Cum. Return	-1.3043	-0.4084	-1.6104	-1.3750	-1.3750	-1.1452
	Win Ratio	0.3378	0.3521	0.3058	0.3081	0.3081	0.3256
	Profit Ratio	0.8393	0.9408	0.7724	0.6360	0.6360	0.8002
MA(10,15)	Sharpe Ratio	-0.1562	-0.1351	-0.1393	-0.2099	-0.3059	-0.1422
	Sortino Ratio	-0.2380	-0.2401	-0.3148	-0.3988	-0.5330	-0.2521
	Cum. Return	-4.6426	-3.3269	-2.9649	-3.7596	-3.1388	-2.5714
	Win Ratio	0.2426	0.2495	0.2290	0.2428	0.2167	0.2543
	Profit Ratio	0.5592	0.6229	0.6152	0.4897	0.3952	0.6248

models as in our original Sharpe model in order to simulate realistic trading environments. As indicated in Table 13, negative results are observed for all stocks in all models (see also Figure A.11), except for DENFIS-DGE model which shows a minor positive result. These results show that not considering such costs when training the trading system can lead to biased results in real-world applications.

In our final benchmark experiment 3, 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 (1 and 2). From the results in Table 14 we find that only MA(1,5) had positive results for ANTO and BLND (see also Figure A.12). This provides evidence of the effectiveness of our dynamic moving average approach.

4. Conclusion

In this paper we investigate the trading performance of dynamic moving average rules in conjunction with AI techniques using 5 minute high-frequency intraday prices. In our experiments we consider variable trading position holding periods between 10 minute and 1 hour together with flexible return bands between 0.1% and 0.5%. This span of less aggressive high-frequency trading (HFT) window was chosen to gain more insight on the profitability of intraday trading with respect to the tension created between two literature findings: (i) the view that 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 between 10 milliseconds and 10 seconds does not reap the expected excess returns (Kearns et al., 2010).

We consider the traditional ANN as well as the more recent ANFIS and DENFIS models. Our results, based on applying a trading algorithm to a set of stocks listed on the London Stock Exchange, show that the application of these models can be used to support profitable trading strategies. The sensitivity analysis of our models with respect to holding position time and return band provide clear indication that financial markets are not fully efficient at less aggressive HFT windows and there exist temporary "pockets of predictability" which could be exploited for realising excess returns.

In our out-of-sample evaluation, we show that overall FFN models perform well when compared with the more recent Neuro Fuzzy techniques. However, FFN models also show higher sensitivity to the underlying data features. Looking at the 100 day aggregated trading performance, we find that ANFIS provides the most robust performance measure exhibiting least variance, both in the case of Sharpe and Sortino ratios optimisation. In our experiments, DENFIS did not outperform FFN and ANFIS (see also Tan et al., 2011). Results also show that ANNs still provide a valid benchmark when applying more recent Neuro-Fuzzy models on high frequency price series.

Our results indicate that the selection of specific risk-return measures does not have the same effect on the behaviour of the underlying AI technique and does not guarantee the same time varying performance across all models. Hence the selection of the risk-return optimisation measure has to be based on the investor risk profile, the underlying technique being applied and the dynamic nature of the underlying price distribution.

This validates the importance of investigating the time varying daily time-series profile of the cumulative risk-return measures attained when evaluating the underlying models for investment decisions rather than just looking at performance measures following an arbitrary out of sample period. Simple risk-free metrics used frequently in literature such as percentage of successful trades only convey little information on the actual profitability of the algorithm.

We also compare our models with a set of benchmark models commonly found in literature or used in practice (Krollner et al., 2010). Our results show that trading models based on risk-return selection criterion outperform those based on RMSE optimisation. In the second comparison we base our model selection criteria on Sharpe ratio but exclude transaction costs, a feature that most studies do not consider, possibly reporting overestimated profitability. In our final comparison, we investigate the application of fixed moving average trading signals. Again the results here did not outperform our dynamic moving average approach.

Another area of interesting research is the application of heat maps to identify regions of profitable areas and how these regions change with time. These findings also encourage further research into stacked models which involve the combination of different AI models in an investment decision portfolio with varying risk-return features.

Acknowledgement

The authors would like to thank Hani Hagras and Edward Tsang for their valuable feedback and helpful discussions.

References

Álvarez Díaz, M. (2010). Speculative strategies in the foreign exchange market based on genetic programming predictions. *Applied Financial Economics*, 20(6):465–476.

Alves Portela Santos, A., Carneiro Affonso da Costa, N., and dos Santos Coelho, L. (2007). Computational intelligence approaches and linear models in case studies of forecasting exchange rates. *Expert Systems with Applications*, 33(4):816–823.

Brabazon, A. and O'Neill, M. (2006). *Biologically inspired algorithms for financial modelling*, volume 20. Springer Berlin.

Choey, M. and Weigend, A. S. (1997). Nonlinear trading models through sharpe ratio maximization. *International Journal of Neural Systems*, 8(04):417–431.

de Faria, E., Albuquerque, M. P., Gonzalez, J., Cavalcante, J., and Albuquerque, M. P. (2009). Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. *Expert Systems with Applications*, 36(10):12506–12509.

Enke, D. and Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications*, 29(4):927–940.

Fernandez-Rodrıguez, F., Gonzalez-Martel, C., and Sosvilla-Rivero, S. (2000). On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid stock market. *Economics Letters*, 69(1):89–94.

Gençay, R. (1996). Non-linear prediction of security returns with moving average rules. *Journal of Forecasting*, 15(3):165–174.

Gradojevic, N. (2007). Non-linear, hybrid exchange rate modeling and trading profitability in the foreign exchange market. *Journal of Economic Dynamics and Control*, 31(2):557–574.

Hudson, R., Dempsey, M., and Keasey, K. (1996). A note on the weak form efficiency of capital markets: The application of simple technical trading rules to UK stock prices-1935 to 1994. *Journal of Banking & Finance*, 20(6):1121–1132.

Jang, J. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics*, 23(3):665–685.

Kablan, A. and Ng, W. (2011). Intraday high-frequency FX trading with adaptive neuro-fuzzy inference systems. *International Journal of Financial Markets and Derivatives*, 2(1):68–87.

Kasabov, N. and Filev, D. (2006). Evolving intelligent systems: methods, learning, & applications. In *2006 International Symposium on Evolving Fuzzy Systems*, pages 8–18. IEEE.

Kasabov, N. and Song, Q. (2002). DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *IEEE Transactions on Fuzzy Systems*, 10(2):144–154.

Kearns, M., Kulesza, A., and Nevmyvaka, Y. (2010). Empirical limitations on High-Frequency Trading Profitability. *The Journal of Trading*, 5(4):50–62.

Krollner, B., Vanstone, B., and Finnie, G. (2010). Financial time series forecasting with machine learning techniques: A survey. In *2011 proceedings, European symposium on artificial neural networks: Computational and machine learning, Bruges (Belgium)*, pages 28–30.

Lin, C., Khan, H., and Huang, C. (2002). Can the neuro fuzzy model predict stock indexes better than its rivals? *CIRJE F-Series*.

Marsland, S. (2009). *Machine Learning: an Algorithmic Perspective*. Chapman & Hall/CRC.

Medeiros, M. C., Teräsvirta, T., and Rech, G. (2006). Building neural network models for time series: a statistical approach. *Journal of Forecasting*, 25(1):49–75.

Resta, M. (2009). Seize the (intra)day: Features selection and rules extraction for tradings on high-frequency data. *Neurocomputing*, 72(16-18):3413–3427.

Schulmeister, S. (2009). Profitability of technical stock trading: Has it moved from daily to intraday data? *Review of Financial Economics*, 18(4):190–201.

Son, Y., Noh, D.-j., and Lee, J. (2012). Forecasting trends of high-frequency KOSPI200 index data using learning classifiers. *Expert Systems with Applications*, 39(14):11607–11615.

Takagi, T. and Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *Systems, Man and Cybernetics, IEEE Transactions on*, (1):116–132.

Tan, Z., Quek, C., and Cheng, P. Y. (2011). Stock trading with cycles: A financial application of ANFIS and reinforcement learning. *Expert Systems with Applications*, 38(5):4741–4755.

Tsai, C. and Wang, S. (2009). Stock price forecasting by hybrid machine learning techniques. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 1, pages 755–760.

Tsang, E. (2009). Forecasting where computational intelligence meets the stock market. *Frontiers of Computer Science in China*, 1(3):53–63.

Vanstone, B. and Finnie, G. (2009). An empirical methodology for developing stockmarket trading systems using artificial neural networks. *Expert Systems with Applications*, 36(3):6668–6680.

Vanstone, B. and Finnie, G. (2010). Enhancing stockmarket trading performance with ANNs. *Expert Systems with Applications*, 37(9):6602–6610.

Xufre Casqueiro, P. and Rodrigues, A. J. (2006). Neuro-dynamic trading methods. *European journal of operational research*, 175(3):1400–1412.

Zadeh, L. (1975). The concept of a linguistic variable and its application to approximate reasoning – Part I. *Information Sciences*, 8(3):199–249.

Appendix A. Appendix

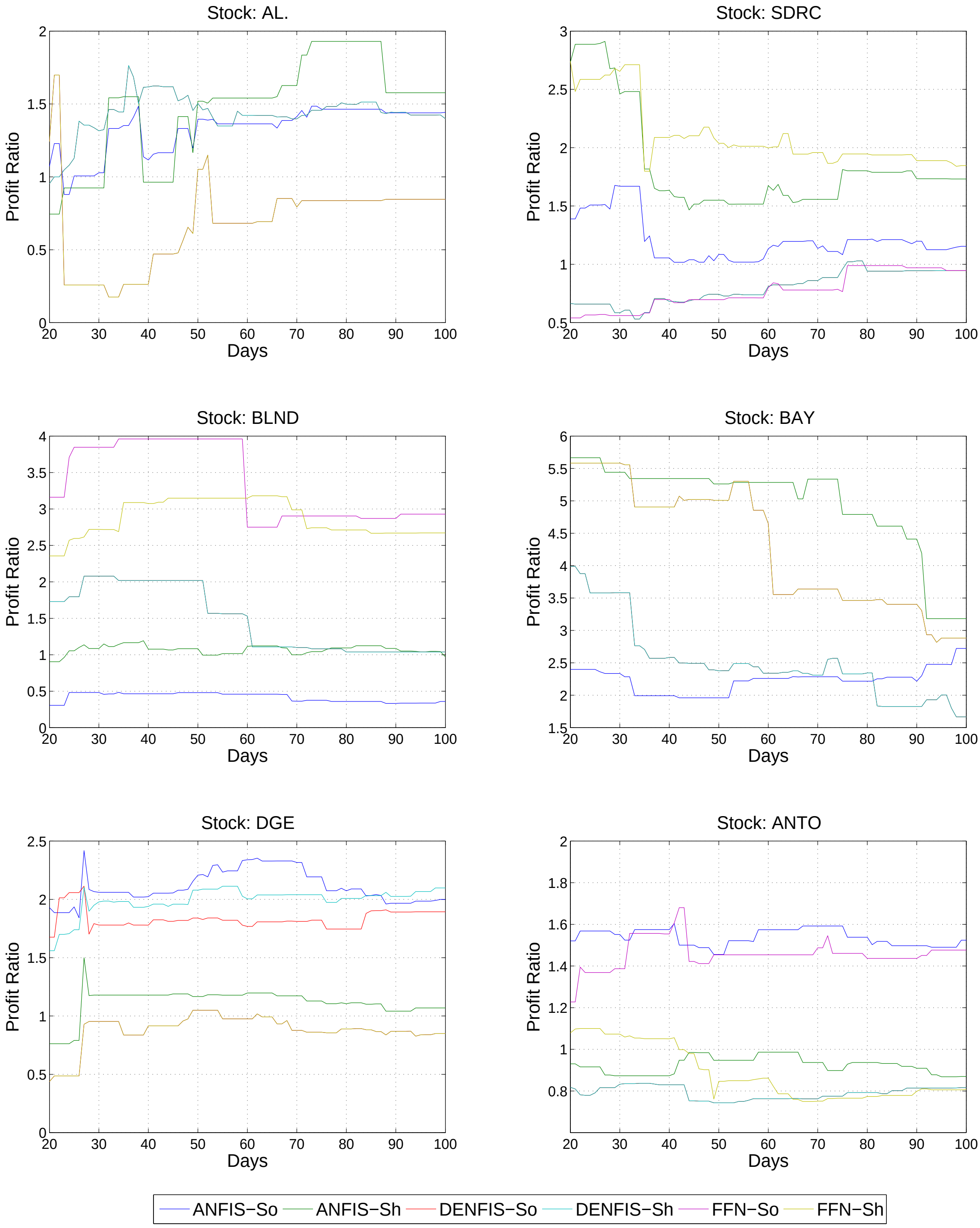

Figure A.7: Profit ratio performance of ANFIS-Sharpe (ANF-Sh), ANFIS-Sortino (ANF-So), DENFIS-Sharpe (DEN-Sh), DENFIS-Sortino (DEN-So), FFN-Sharpe (FFN-Sh) and FFN-Sortino (FFN-So) models. The plots show for each stock the cumulated profit ratio (y-axis) on the n-th day (x-axis) in the out-of-sample.

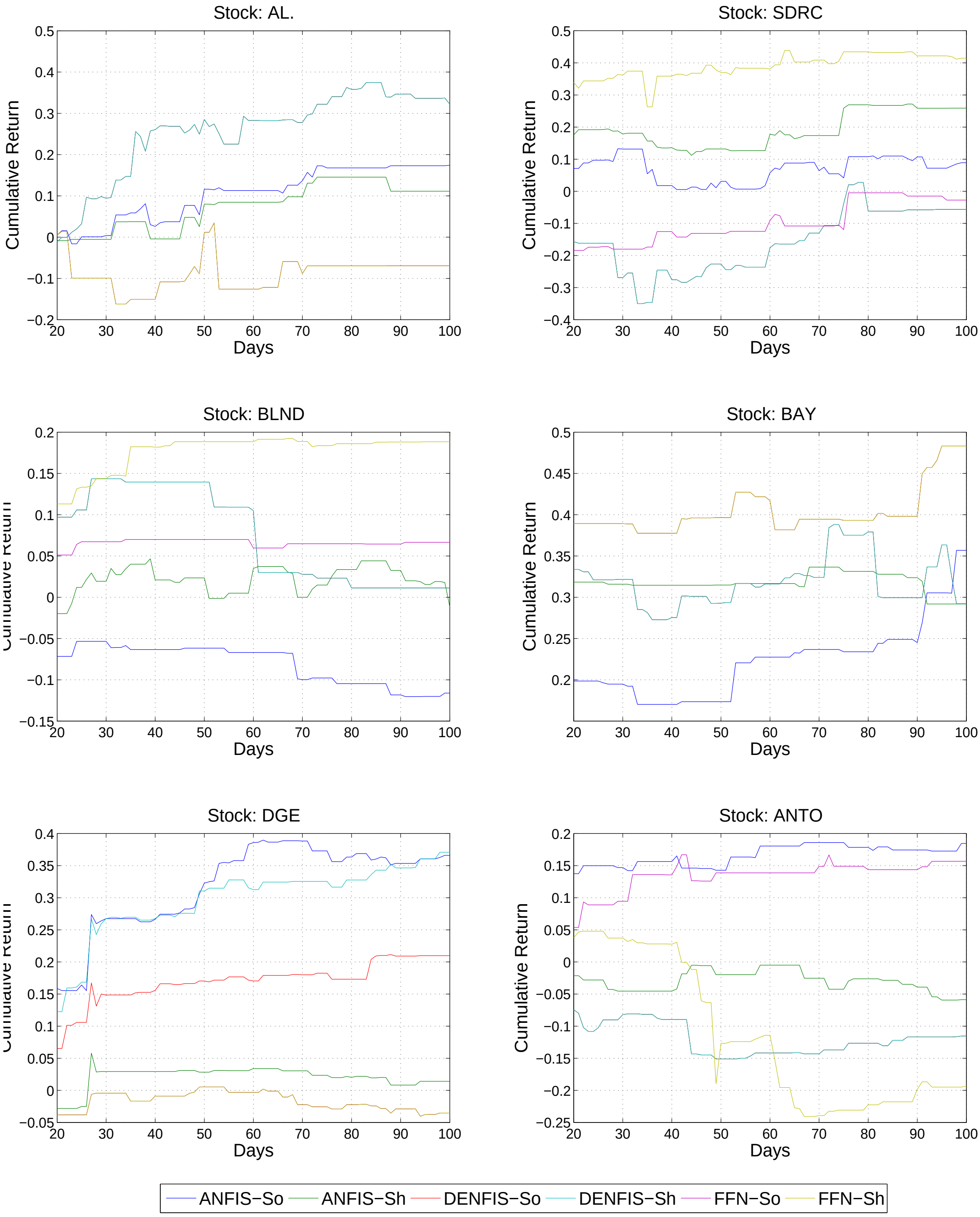

Figure A.8: Cumulative Return performance of ANFIS-Sharpe (ANF-Sh), ANFIS-Sortino (ANF-So), DENFIS-Sharpe (DEN-Sh), DENFIS-Sortino (DEN-So), FFN-Sharpe (FFN-Sh) and FFN-Sortino (FFN-So) models. The plots show for each stock the cumulated return (y-axis) on the n-th day (x-axis) in the out-of-sample.

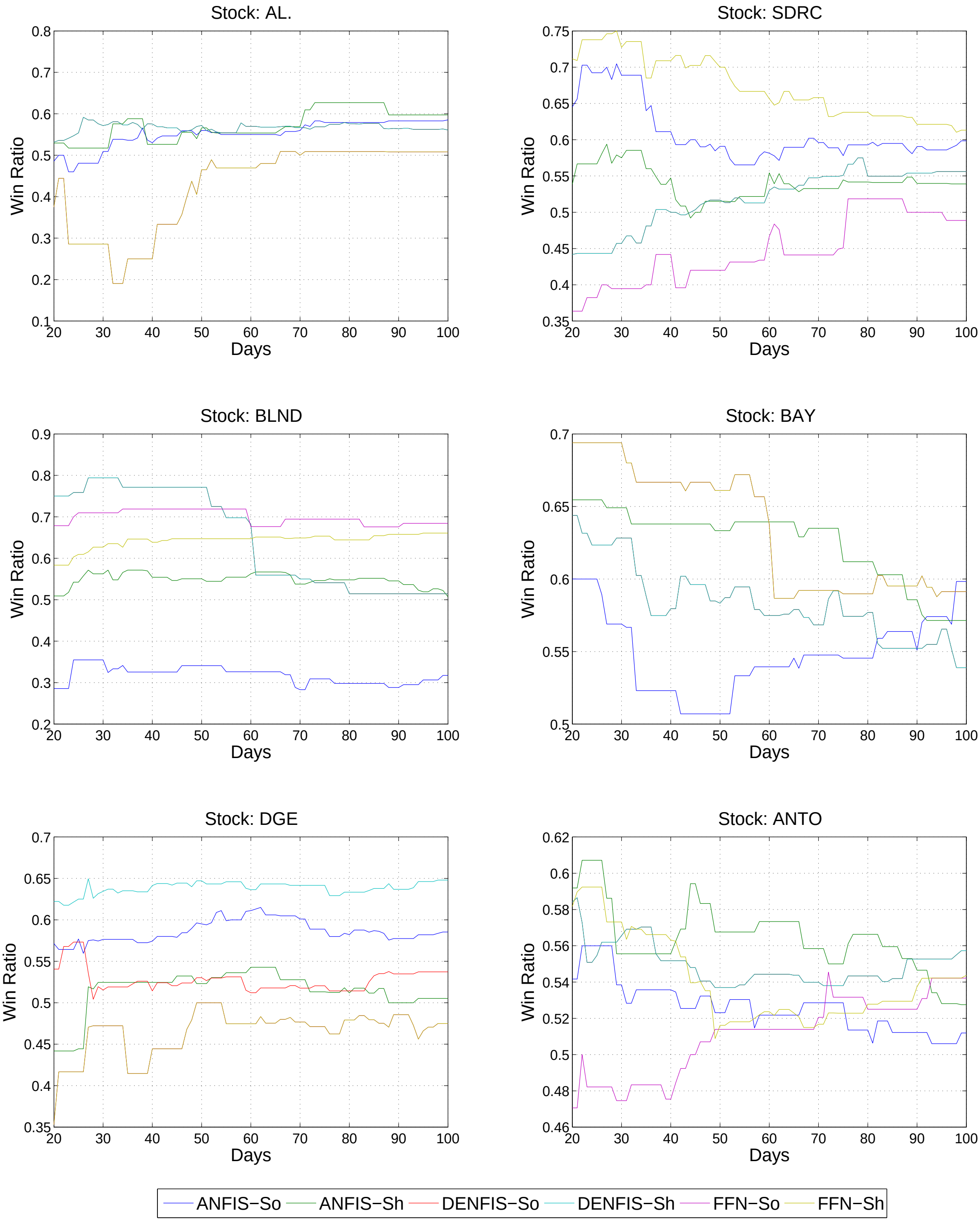

Figure A.9: Win ratio performance of ANFIS-Sharpe (ANF-Sh), ANFIS-Sortino (ANF-So), DENFIS-Sharpe (DEN-Sh), DENFIS-Sortino (DEN-So), FFN-Sharpe (FFN-Sh) and FFN-Sortino (FFN-So) models (out-of-sample). The plots show for each stock the cumulated win ratio (y-axis) on the n-th day (x-axis) in the out-of-sample.

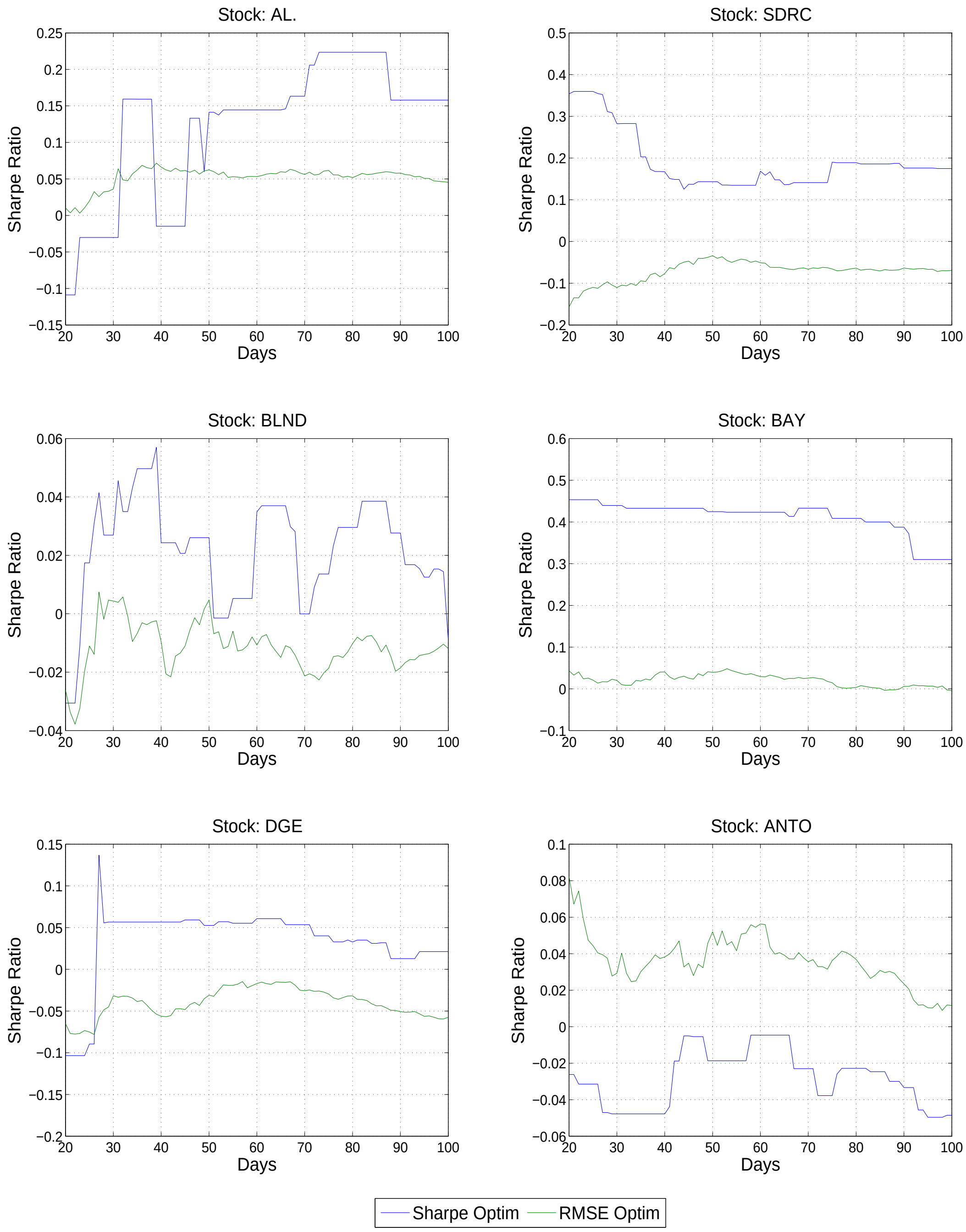

Figure A.10: The plots compare the results obtained from ANFIS Sharpe optimisation (experiment 1) with those from RMSE optimisation (experiment 2), and show for each stock the corresponding cumulated Sharpe ratio (y-axis) on the n-th day (x-axis) in the out-of-sample.

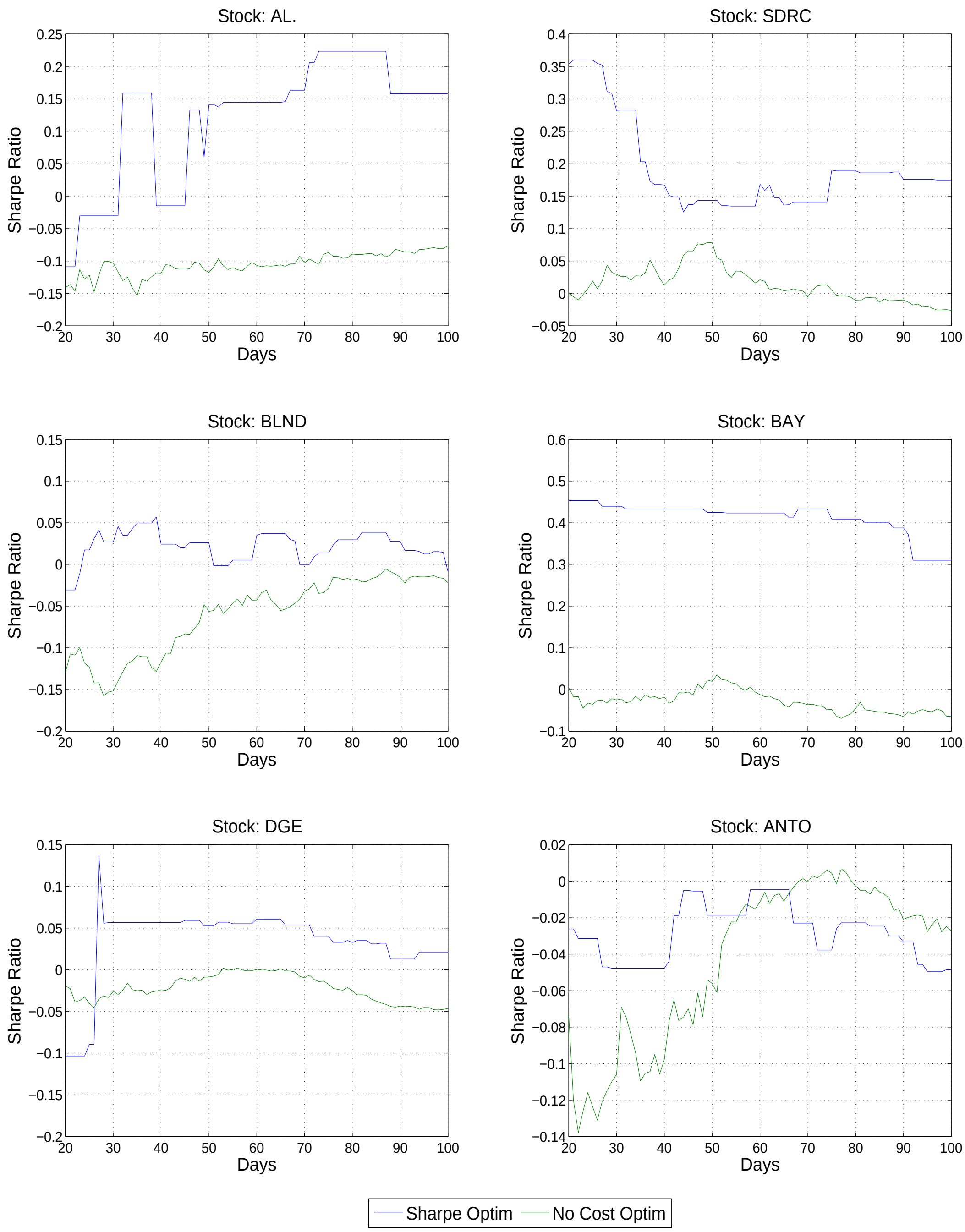

Figure A.11: The plots compare the results obtained from ANFIS Sharpe optimisation (experiment 1) with those from No-Transaction-Costs optimisation (experiment 2), and show for each stock the corresponding cumulated Sharpe ratio (y-axis) on the n-th day (x-axis) in the out-of-sample.

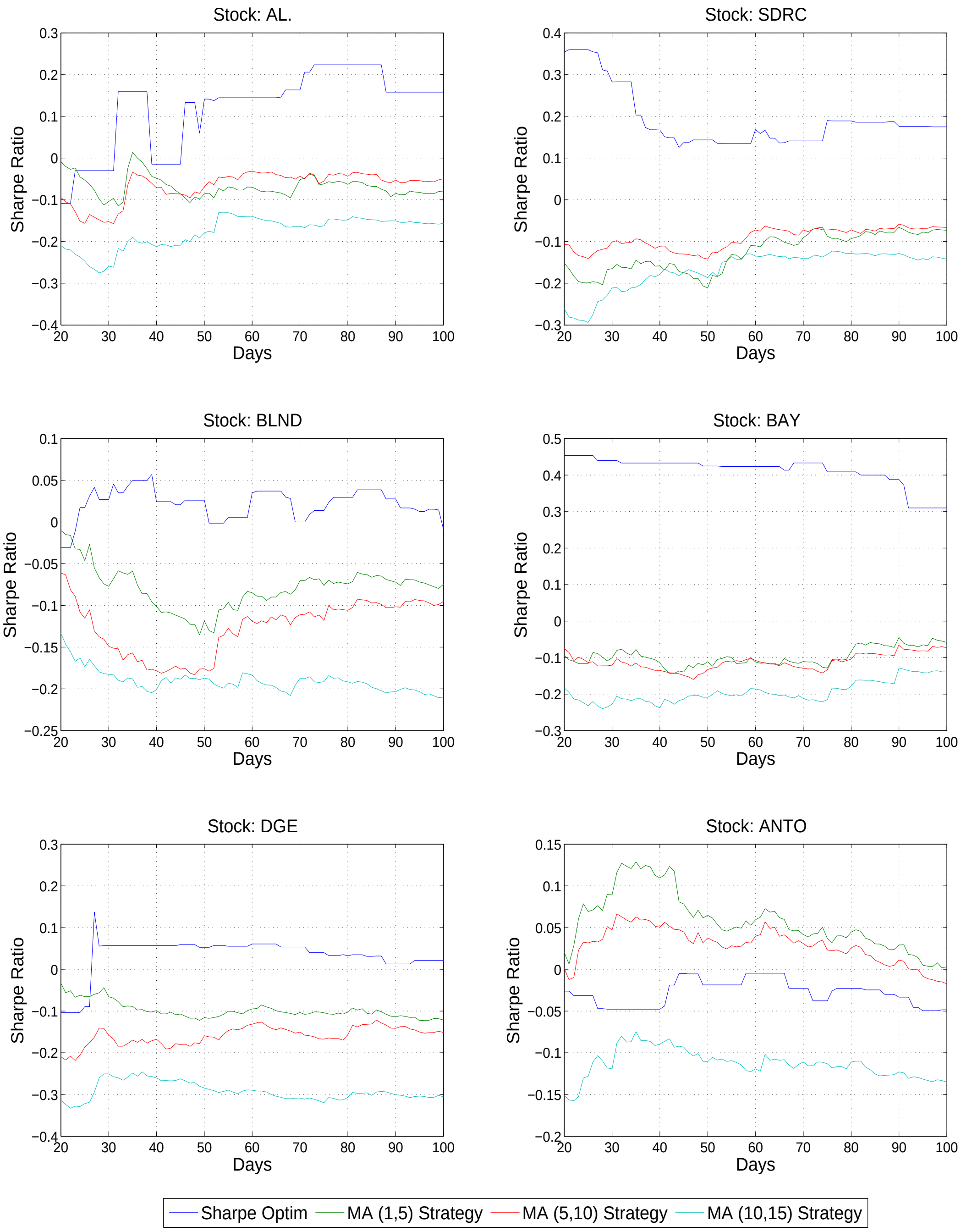

Figure A.12: The plots compare the results obtained from ANFIS Sharpe optimisation (experiment 1) with those from MA optimisation (experiment 3), and show for each stock the corresponding cumulated Sharpe ratio (y-axis) on the n-th day (x-axis) in the out-of-sample.