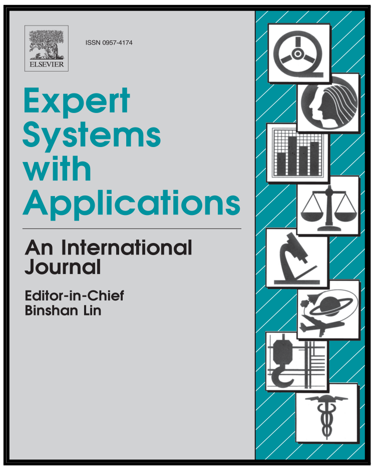
Accepted Manuscript



Improving Risk-adjusted Performance in High Frequency Trading Using Interval Type-2 Fuzzy Logic

Vince Vella, Wing Lon Ng

PII: S0957-4174(16)30020-3
DOI: 10.1016/j.eswa.2016.01.056
Reference: ESWA 10514

To appear in: *Expert Systems With Applications*

Received date: 1 October 2015
Revised date: 30 January 2016
Accepted date: 31 January 2016

Please cite this article as: Vince Vella, Wing Lon Ng, Improving Risk-adjusted Performance in High Frequency Trading Using Interval Type-2 Fuzzy Logic, *Expert Systems With Applications* (2016), doi: 10.1016/j.eswa.2016.01.056

Highlights

- We investigate the viability of Type-2 fuzzy systems in high frequency trading.
- We propose Type-2 models based on a generalisation of the popular ANFIS model.
- Type-2 models score significant risk adjusted performance improvements over Type-1.
- Benefits of Type-2 models increase with higher trading frequencies.

Improving Risk-adjusted Performance in High Frequency Trading Using Interval Type-2 Fuzzy Logic

Vince Vella[a,*], Wing Lon Ng

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

Abstract

In this paper, we investigate the ability of higher order fuzzy systems to handle increased uncertainty, mostly induced by the market microstructure noise inherent in a high frequency trading (HFT) scenario. Whilst many former studies comparing type-1 and type-2 Fuzzy Logic Systems (FLSs) focus on error reduction or market direction accuracy, our interest is predominantly risk-adjusted performance and more in line with both trading practitioners and upcoming regulatory regimes. We propose an innovative approach to design an interval type-2 model which is based on a generalisation of the popular type-1 ANFIS model. The significance of this work stems from the contributions as a result of introducing type-2 fuzzy sets in intelligent trading algorithms, with the objective to improve the risk-adjusted performance with minimal increase in the design and computational complexity. 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 type-2 model and higher levels of microstructure noise. The results resolve 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 HFT.

Keywords: High-Frequency Trading, ANFIS, Type-2 Fuzzy Logic, ANFIS/T2.

*Corresponding author

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

1. Introduction

Most transactions in modern, highly computerised, financial markets are being greatly 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 high frequency trading (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, 1970) that states that current prices incorporate all relevant information with no possibility of predictability or excess returns. A number of authors (e.g. Schulmeister, 2009; Zhang, 2010; Rechenthin and Street, 2013; Holmberg et al., 2013; Brogaard et al., 2014) insist that the presence of efficient pricing becomes more questionable when investigating short-lived (milli-seconds to few minutes) trades. However, Kearns et al. (2010) validate the EMH in their study and argue that generating profits from aggressive high-frequency trading (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 for sub-second financial phenomena during which the human traders loose 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. As an example, in October 2011, the EU Commission proposed a new version of the Markets in Financial Instruments Directive (MiFID2). MiFID2 will apply from January 2017 and will introduce a new regulatory regime for firms which engage in algorithmic or high-frequency trading on European venues or who provide investment management services directly to clients in the EU. The new regulations are intended to ensure that trading systems are adequately designed and tested to mitigate the risks to which they are exposed.

This acts as a reminder for model or algorithm designers that both investors and regulators are more concerned about risk-adjusted performance rather than directional accuracies, error measures or just profitability. Unfortunately, surveys show that the great majority of computational finance research disregards the essential requirement that an investor always measures investment returns in line with risk measures in order to compare relative trading performance (Tsai and Wang, 2009; Krollner et al., 2010). 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 paper we focus on risk-adjusted performance.

Zadeh (1975) proposed that increased systems 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 (Tsang, 2009; Krollner et al., 2010; Sahin et al., 2012) 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.

A frequently cited technique in non-stationary and chaotic time series prediction, which combines ANN and type-1 (T1) fuzzy logic, is the Adaptive Neuro-Fuzzy Inference System (ANFIS) by Jang (1993). Successful application and active continuous research in improving ANFIS based techniques in trading applications is demonstrated by numerous publications (Gradojevic, 2007; Boyacioglu and Avci, 2010; Chang et al., 2011; Tan et al., 2011; Kablan and Ng, 2011; Chen, 2013; Vella and Ng, 2014b; Wei et al., 2014). Moreover Vella and Ng (2014b) showed the increased stability of ANFIS in terms of risk-adjusted performance when compared to ANN alone. Recently, type-2 (T2) fuzzy logic have 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. Our intention in this paper is to investigate the possible improvement that can be obtained by generalising ANFIS to interval T2 (IT2) FLS. 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 applications of such systems. This challenge was a key consideration that inspired our innovative and practical IT2 approach that we present in this paper.

The investigation of possible improvements using T2 in HFT is appealing in view of increased uncertainties which are inherent in high frequency data. Albeit the concepts of risk and uncertainty have often been used interchangeably, economists have long distinguished between the two (e.g. Knight, 1921) and also in recent literature (e.g. Nelson and Katzenstein, 2014; Heal and Millner, 2014). Our view is that overall risk can be divided into measurable risk (e.g. 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 study the objectives can be summarised as follows:

1. To identify practical methods of how the popular ANFIS model can be generalised to an interval T2 Takagi-Sugeno-Kang (IT2 TSK) fuzzy system. We aim to address this with minimal increase in design and computational complexity.
2. To investigate the ability of higher order fuzzy systems to handle increased uncertainties inherent in a HFT scenario.
3. To identify *if* T2 FLSs provide a viable alternative for trading purposes in view of improving risk-adjusted performance.
4. To explore *when* can T2 models offer a more viable approach than T1 alternatives. We analyse this from the perspective of different levels of trading frequencies.

This paper aims to convey 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 overwhelming (Wu and Mendel, 2014), design choices to be made which includes the shape of membership functions, number of membership functions, type of fuzzifier, kind of rules, type of i-norm, method to compute the output, and methods for tuning the parameters. We address this from a number of aspects. Firstly, we make use of a fuzzy clustering algorithm for rule identification in order to reduce the number of rules and hence simplify the 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 popular ANFIS as a solid benchmark model as it is computationally fast and also has been successfully applied in high frequency trading (Kablan and Ng, 2011; Vella and Ng, 2014b). 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 in HFT. Schulmeister (2009) points towards possible market inefficiencies and 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 in 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. Contrary to findings in AI surveys (Tsai and Wang, 2009; Krollner et al., 2010), 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; Xufre Casqueiro and Rodrigues, 2006; Vanstone and Finnie, 2010). 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 models rather than T1. Although previous literature found that T2 models can perform better under increasing uncertainties (Sepulveda et al., 2006; Aladi et al., 2014), it is however not clear at which uncertainty level this would reflect in a reasonable improvement in risk-adjusted trading performance. Birkin and Garibaldi (2009) even showed that if the level of noise is too low, T2 models show no significant improvement on T1. A number of authors (Gençay, 1996; Vanstone and Finnie, 2009, 2010; Holmberg et al., 2013) 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, 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 at decreasing (increasing) levels of return thresholds, which reflect 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 on *out*-of-sample data demonstrate that the proposed ANFIS/T2 model outperforms the standard ANFIS and Buy & Hold methods. Statistically significant improvements in both risk adjusted performance and profitability were registered in higher trading frequency scenarios but disappeared when trading activity was lowered.

The structure of the paper is organised as follows. In Section 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 3. In Section 4, we present our results and analyse model performance. Finally, in Section 5 we draw our conclusions in view of existing literature.

2. Method

Our experiment setup consists of five modules (see Figure 1). Section 2.1 explains our variable selection and data pre-processing. Sections 2.2 and 2.3 explain

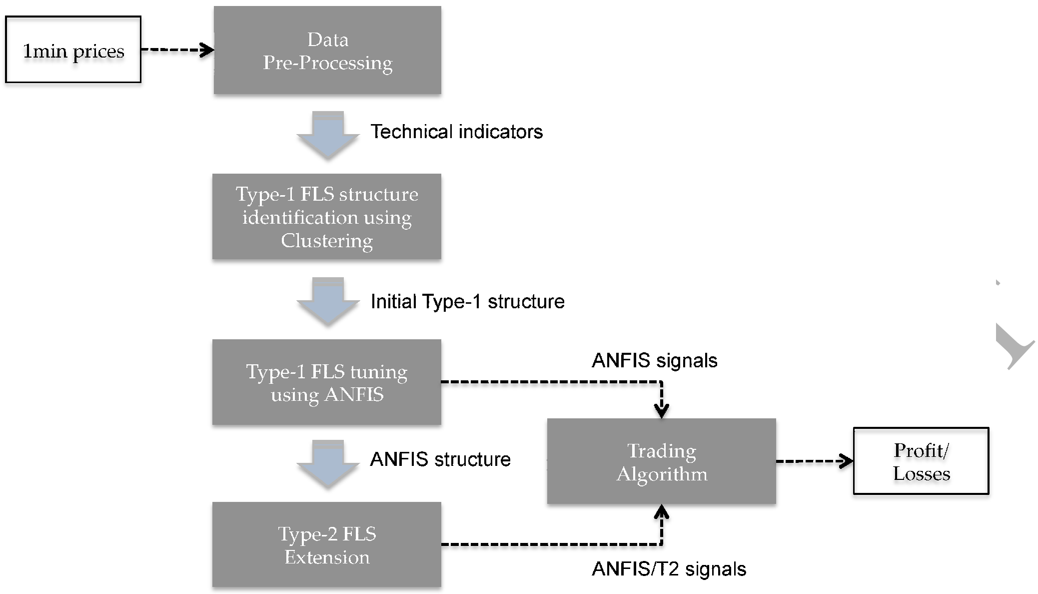


Figure 1: Experiment Setup

our Fuzzy Inference System (FIS) design and the learning approach using ANFIS and ANFIS/T2. In Section 2.4 we explain our trading algorithm.

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 phenomena, but to identify candidate features that can be used by our trading models.

A profitable trading algorithm essentially requires constructing a model which can determine if, under certain conditions and time horizons, prices will be trending or mean-reverting (Pardo, 2011). When designing an HFT model, an important challenge that a model designer is faced with 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 be 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 study is based on findings by Zhang (2010); Brogaard et al. (2014) 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 state that the stock price typically broke the price reversal pattern due to bid-ask bounce after 5 to 10 seconds, and that traces of predictability existed up to 30 minutes, beyond which markets became efficient.

Recent studies (Gradojevic and Gençay, 2013; Vella and Ng, 2014a; Naranjo et al., 2015) show the effectiveness of combining moving average signals with fuzzy logic to capture trend information. In this study we use 1 minute stock prices and define the expected mean return, y, at time t, as

$$y_t = log(m_{t+2}^2) - log(p_t) \ , \tag{1}$$

where

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

p_t is the stock price at time t, and $j = 0, 1, 2, ..., n-1$ is the "memory span". For our trend signals we apply three moving average rules

$$MA_t^{n_1,n_2} = m_t^{n_2} - m_t^{n_1} \tag{3}$$

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 min time bars.

For our mean-reversion indicator we use the popular Relative Strength Index (RSI) (Murphy, 1987). To calculate RSI we consider 1 min prices in the last 15 minutes. The indicator is intended to convey signals about stocks that are likely overvalued or undervalued, which can possibly result in price trend reversals.

A number of authors (e.g. Choudhry et al., 2012; Gradojevic and Gençay, 2013; Holmberg et al., 2013; Schulmeister, 2009; Vella and Ng, 2014a) claim a relationship between breaks in market efficiency and different levels of intraday volatility. For this reason, we consider a fifth indicator using a measure of volatility, which provides an indication of movement strength but not the direction. The ARCH and GARCH frameworks (Engle and Patton, 2001; Hansen and Lunde, 2005; Poon and Granger, 2003, 2005), which address the problems resulting from heteroscedasticity when applying the least squares model, became the standard econometric tools to measure the volatility of returns. ARCH and GARCH models treat heteroskedasticity as a variance to be modelled. However, albeit their popularity, these models are typically applied in the context of "low frequency" daily returns.

In a high-frequency setting, the literature proposes an alternative approach. With the increased availability of intraday data, Merton (1980) proposed the use of intraday daily data to estimate volatility using the sum of intraday squared returns. In particular, Andersen and Bollerslev (1997) show that, under the diffusion assumptions, realised volatility (RV) computed from high-frequency intraday returns, is effectively an error-free volatility measure. Contrary to traditional models in finance, where volatility is considered latent, the realised volatility approach has a key advantage that volatility is considered as an observable variable (Andersen et al., 2001), hence can be more readily applied for modelling purposes. RV, at time t, can be calculated as

$$RV_t = \sum_{j=0}^{n} \left(\log(p_{t-j}) - \log(p_{t-j-1})\right)^2 \quad . \tag{4}$$

To minimise the effect of microstructure noise, we implement a simple RV estimator called Average RV (Christoffersen, 2011), and calculate the last 15 minutes average RV at time t using 5 minute return intervals (as suggested by McAleer and Medeiros, 2008) as

$$RV_t^{avg} = \frac{1}{3} \sum_{j=0}^{2} RV_{t-j} \quad . \tag{5}$$

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

$$y_t = \theta_0 + \sum_{k=1}^{5} \theta_k s_{k,t-1} + \epsilon_t \tag{6}$$

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

$$s_{k,t} = \begin{cases} MA_t^{1,2} & \text{for} \quad k = 1 \\ MA_t^{1,5} & \text{for} \quad k = 2 \\ MA_t^{1,10} & \text{for} \quad k = 3 \\ RSI_t & \text{for} \quad k = 4 \\ RV_t^{avg} & \text{for} \quad k = 5 \end{cases} \quad . \tag{7}$$

2.2. Designing and Tuning 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. In our paper 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.

In the following two sub-sections we describe our approach for (i) the initial identification of a T1 TSK FIS model (Section 2.2.1) and (ii) model tuning (Section 2.2.2).

2.2.1. 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, $(\mathbf{x}^{(1)} : y^{(1)}), (\mathbf{x}^{(2)} : y^{(2)}), ..., (\mathbf{x}^{(q)} : y^{(q)})$ where

$$\begin{aligned}
&\mathbf{x}^{(1)} = [s_{1,t-q}, s_{2,t-q}, ..., s_{5,t-q}], && y^{(1)}_{t-q+1} \\
&\mathbf{x}^{(2)} = [s_{1,t-q+1}, s_{2,t-q+1}, ..., s_{5,t-q+1}], && y^{(2)}_{t-q+2} \\
&... \\
&\mathbf{x}^{(q)} = [s_{1,t-1}, s_{2,t-1}, ..., s_{5,t-1}], && y^{(q)}_{t}.
\end{aligned} \tag{8}$$

In Equation (8), for each data instance at a specific time t, $\mathbf{x}$ is a vector consisting of $\{x_1, x_2, ..., x_5\}$ input elements which represent the $\{s_{1,t-1}, s_{2,t-1}, ..., s_{5,t-1}\}$ technical indicator signals (Equation (7)), and y represents the mean return over the next 2 minutes (Equation (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 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 is, the higher the number of rules, hence improving model accuracy. However, a resulting effect is the increased number of optimisation parameters and hence 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, which is a popular approach to identify fuzzy partitions in data (Bezdek, 1981; Dutta Baruah and Angelov, 2010). By controlling the number of clusters, this gives us the opportunity to identify the best model structure which balances between model accuracy and complexity. This is possible because each cluster center essentially exemplifies a characteristic behaviour of the system in a specific region. Hence, each cluster center can be used as the basis of a membership function for each input variable and are combined in a rule that describes the local system behaviour. Since our input variables have 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 provides equal influence weight of each variable on the clustering algorithm.

Let z denote the fuzziness index. Furthermore, define α as the number of clusters, $d_{uv} = \|\mathbf{x}_u - \mathbf{c}_v\|^2$ as the Euclidean distance between the u-th realisation and the current v-th cluster center $\mathbf{c}_v$, and $d_{uo} = \|\mathbf{x}_u - \mathbf{c}_o\|^2$ as the Euclidean distance from the u-th realisation and the other cluster centres $\mathbf{c_o}$. For each data point $\mathbf{x}_u$, $\forall u \in [1, q]$, and cluster $\mathbf{c}_v$, $\forall v \in [1, \alpha]$, the FCM algorithm iteratively updates the membership grade μ of the u-th data point to the v-th cluster

$$\mu_{uv} = \left(\sum_{l=1}^{\alpha} \left(\frac{d_{uv}}{d_{uo}} \right)^{\frac{2}{z-1}} \right)^{-1} , \tag{9}$$

and the center of the v-th cluster

$$\mathbf{c}_v = \frac{\sum_{u=1}^{q} \mu_{uv}^{z} \mathbf{x}_u}{\sum_{u=1}^{q} \mu_{uv}^{z}} , \tag{10}$$

such that the objective function

$$J_z = \sum_{u=1}^{q} \sum_{v=1}^{\alpha} \mu_{uv}^{z} d_{uv} \tag{11}$$

is minimised. 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, α, we test values between 2 to 5 clusters. This range selection is based on identifying a well distributed set of clusters representing the variable distribution, whilst at the same time avoiding 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 paper we adopt Gaussian membership functions (MFs), where each fuzzy set is represented by

$$Gaussian(x; \overline{x}, \sigma) = e^{-(x-\overline{x})^2 / 2\sigma^2}. \tag{12}$$

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 $\overline{x}$ and standard deviation σ) 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. In the first step, the identified α clusters are projected on each input variable space. This results in α Gaussian MFs for each variable with the mean represented by the cluster centres $\mathbf{c}$. The standard deviation σ is obtained by re-arranging Equation (12) and utilise cluster centres $\mathbf{c}$ and membership grades μ. As a second step, a set of α rules are created in the form

$$\begin{aligned} IF \quad & (x_1 \ is \ A_{i,1}) \ AND \ (x_2 \ is \ A_{i,2}) \ AND \ ... \ AND \ (x_5 \ is \ A_{i,5}) \\ THEN \quad & y_i \ = b_i + \sum_{k=1}^{5} w_{i,k} x_k \end{aligned} \tag{13}$$

where $A_{i,k}$'s are T1 Gaussian MFs, projected from the identified clusters, for the i-th rule and the k-th input ($i = 1, 2, ..., \alpha; k = 1, 2, ..., 5$). In the consequent, y_i is the rule output, defined as the mean return over the next 2 minutes (Equation 1), consisting in a linear function of the input variables $\{x_1, x_2, ..., x_5\}$ with parameters b_i and $w_{i,k}$. Following the identification of the initial FIS using FCM clustering, the structure is fed into ANFIS for model tuning.

2.2.2. FIS Tuning With ANFIS

From literature we identify two major classes of optimisation algorithms for FLSs: gradient-based algorithms and heuristic algorithms, in the latter case most studies focusing on evolutionary computation (EC) algorithms (see discussion in Wu and Mendel, 2014). ANFIS follows the former approach. It is due to the popularity of ANFIS in the finance domain 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.

ANFIS configures a TSK model in a network architecture, and albeit the mathematical underpinnings follow the traditional TSK models, the structure is formulated to permit ANN learning techniques. Our ANFIS structure can be defined by a number of connected layers 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 α nodes with connections from the first input x_1 are in the form

$$O_{1,i} = \mu_{A_{i,1}}(x_1) \text{ for } i = 1, 2, ..., \alpha. \tag{14}$$

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 α represents both the number of rules and also the number of MFs for each input variable. Different values for α are tested as part of our model calibration process (see Table 1).

Layer 2 This layer contains α fixed nodes. In each node, $O_{2,i}$, where $i = 1, ..., \alpha$, the product t-norm ($*$) 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:

$$O_{2,i} = f_i = \mu_{A_{i,1}}(x_1) * \mu_{A_{i,2}}(x_2) * ... * \mu_{A_{i,5}}(x_5). \tag{15}$$

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

$$O_{3,i} = \hat{f}_i = \frac{f_i}{\sum_i f_i}. \tag{16}$$

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

$$O_{4,i} = \hat{f}_i y_i = \hat{f}_i (b_i + \sum_{k=1}^{5} w_{i,k} x_k) , \tag{17}$$

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

$$O_{5,i} = r_t = \sum_i \hat{f}_i y_i = \frac{\sum_i f_i y_i}{\sum_i f_i}. \tag{18}$$

The ANFIS learning process consists of an iterative two-pass algorithm. In a forward pass, the premise parameters (in Layer 1) defining the membership functions are unmodified and the consequent parameters (in Layer 4) are computed using 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 (full detail is provided in Jang, 1993). 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 1). These parameters are tested in combination with an additional set of parameters that are defined for our trading algorithm (see Section 2.4). In the next section we propose how the standard ANFIS model can be extended to a T2 TSK model.

Table 1: Parameters tested for ANFIS

Parameter	Parameter Value Set
Training Data Size (days)	{2, 3}
Number of Input Membership Functions and Rules (α)	{2, 3, 4, 5}
Training epochs	{10, 20, 40}

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.
- Use of non precise terms: "rising steadily", "high volatility", "small loss", "high activity".
- Microstructure noise in the observations.
- Inconsistencies effecting 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 a good contender.

Our primary interest is on identifying model improvements that can result from 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 what 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 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 better management of uncertainty and increased model complexity.

2.3.1. 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 (Mendel et al., 2014; Wu and Mendel, 2014; Aladi et al., 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 FLS from scratch, hence avoiding the use of on 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 (Sepulveda et al., 2006; Aladi et al., 2014), we do not seek to achieve an optimal performance in the error reduction, instead the primary objective is to compare T1 FLS and IT2 FLS 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 partially dependent approach, in Section 3.4 (IT2 Design Considerations) we highlight the strengths and weaknesses of this approach.

2.3.2. 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 (Knight, 1921; Nelson and Katzenstein, 2014; Heal and Millner, 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:

$$\begin{aligned} IF \quad & (x_1 \; is \; \widetilde{A}_{i,1}) \; AND \; (x_2 \; is \; \widetilde{A}_{i,2}) \; AND \; ... \; AND \; (x_5 \; is \; \widetilde{A}_{i,5}) \\ THEN \quad & y_i \; = \; b_i + \sum_{k=1}^{5} w_{i,k} x_k \end{aligned} \tag{19}$$

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

For our proposed ANFIS/T2a, the objective is to convert the T1 rules defined in Equation (13) to A2-C0 rules defined in Equation (19). 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), $\underline{\mu}_{A_{i,k}}$, 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], membership grades of a T2 fuzzy set are themselves T1 fuzzy sets. Therefore, T2 fuzzy sets offer the ability to model higher levels of uncertainty (Mendel et al., 2006; John and Coupland, 2007). 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 ($i = 1, 2, ..., \alpha; k = 1, 2, ..., 5$), whilst keeping the mean, $\overline{x}_{i,k}$, fixed. Hence, for each T1 MF, the LMF and UMF are defined as follows:

$$\underline{\mu}_{A_{i,k}} = Gaussian(x_k; \overline{x}_{i,k}, (1 - \beta)\sigma_{i,k}) \tag{20}$$

$$\overline{\mu}_{A_{i,k}} = Gaussian(x_k; \overline{x}_{i,k}, (1 + \beta)\sigma_{i,k}) \tag{21}$$

where $\overline{x}_{i,k}$ 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 2).

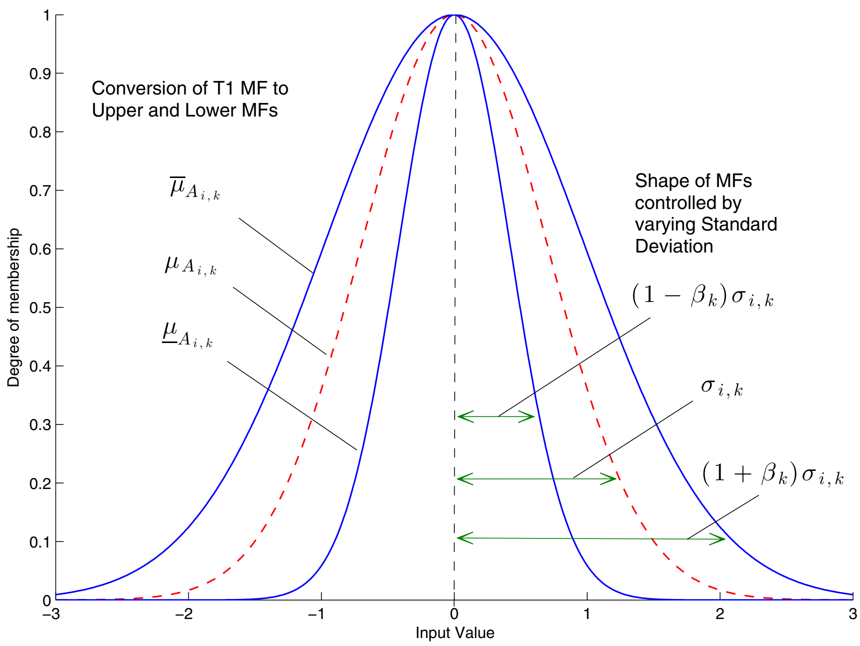


Figure 2: Conversion of T1 fuzzy set to IT2 with fixed mean and uncertain standard deviation. Upper and lower MFs are defined using an additional parameter β_k which represents the width of the IT2 MF as a percentage increase or decrease on the base T1 MF standard deviation respectively. In our first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

We transform the complete T1 to IT2 rule base in this manner. The final output of the model is obtained as follows

$$\mathbf{Y}_{A2-C0} = [y_l, y_r] = \int_{f_1 \in [\underline{f}_1, \overline{f}_1]} \cdots \int_{f_\alpha \in [\underline{f}_\alpha, \overline{f}_\alpha]} 1 / \frac{\sum_i f_i y_i}{\sum_i f_i} \tag{22}$$

where 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, which in the case of IT2 is simplified to 1. The firing strength for each rule i, where $i = 1, 2, ..., \alpha$, is calculated as

$$\underline{f}_i(\mathbf{x}) = \underline{\mu}_{A_{i,1}}(x_1) * \underline{\mu}_{A_{i,2}}(x_2) * ... * \underline{\mu}_{A_{i,5}}(x_5) \tag{23}$$

$$\overline{f}_i(\mathbf{x}) = \overline{\mu}_{A_{i,1}}(x_1) * \overline{\mu}_{A_{i,2}}(x_2) * ... * \overline{\mu}_{A_{i,5}}(x_5) \tag{24}$$

where, like in the case of ANFIS (Equation (15)), $*$ represents the product t-norm. For further details, the interested reader is directed to Mendel et al. (2014).

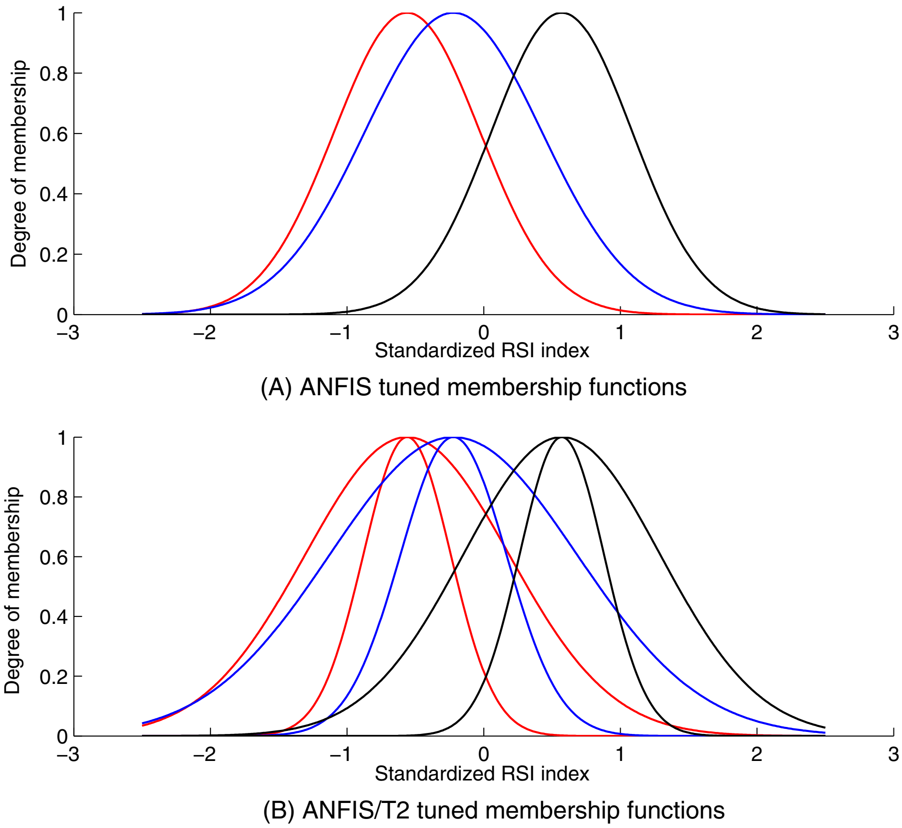


Figure 3: An example showing the MFs for the input variable RSI index following ANFIS training (top), and the corresponding ANFIS/T2 tuned MFs (bottom). Both ANFIS and ANFIS/T2 MFs are dynamically adapted on a daily basis, reflecting the price volatility in that period. During training, the latter IT2 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%.

Following the ANFIS optimisation, as a second training pass we train our model with values of β 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. This 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 example in Figure 3). This also guarantees that during the training process the model achieves at least the same level of performance 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. So in the case of ANFIS/T2b, we convert every rule i, where $i = 1, 2, ..., \alpha$, by transforming the MFs of each input variable to IT2 MFs using β_k, where $k = 1, 2, ..., 5$. In this case the lower and upper MFs are defined as:

$$\underline{\mu}_{A_{i,k}} = Gaussian(x_k; \overline{x}_{i,k}, (1 - \beta_k)\sigma_{i,k}) \tag{25}$$

$$\overline{\mu}_{A_{i,k}} = Gaussian(x_k; \overline{x}_{i,k}, (1 + \beta_k)\sigma_{i,k}) \tag{26}$$

In our training algorithm we apply the same discrete range for possible β_k values. This approach results in a much larger search space of possible β_k combinations, hence rather than performing a parameter sweep we speed up the training process by optimising the β_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.

2.3.3. Computing the Output

The T2 fuzzy logic community has proposed a number of methods for computing the output (the interested reader is directed to Wu, 2013; Mendel et al., 2014, for a review). At a high level, the methods can be divided into two groups. The first group require 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 skip 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:

$$y = \hat{r} = \frac{\sum_{i=1}^{\alpha} y_i(\underline{f}_i + \overline{f}_i)}{\sum_{i=1}^{\alpha}(\underline{f}_i + \overline{f}_i)}. \tag{27}$$

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 (Wu, 2013; Greenfield and Chiclana, 2013).

2.4. Trading Algorithm

Our trading system takes decisions on a minute by minute basis. Every minute, the prediction of the average return over the next two minutes is passed to the trading algorithm which in turn recommends a *buy* ($\Phi = long$) or *sell* ($\Phi = short$) action depending on whether the signal is above or below a specific return threshold parameter RT, or *stand-by* ($\Phi = 0$) if not. 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 with the forecast signal generated in the previous 1 min time bar.

For our simulation we allocate a starting capital of 250,000 GBP for each stock. When our model indicates a buy (sell) signal, our trading algorithm opens a long (short) position worth of $50,000$ GBP. When a trade is closed, the net proceeds is added back to the capital balance, hence maximising the utilisation of the available capital. In line with realistic commercial prices, in our experiments we apply a transaction cost of 10 GBP per trade per direction.

We also use a second parameter, TD, which represents the duration of each trade in minutes. This parameter defines the lifetime of a trade, after which the trade is automatically closed. In our experiments we consider TD values between 2 min to 10 min. Following from 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 end of each daily trading session to avoid having positions overnight.

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

```
Φ ← 0
if signal > RT and signal > prevsignal and balance-tradesize > 0 then
    Φ ← long
end if
if signal < −1 ∗ RT and signal < prevsignal and balance-tradesize > 0 then
    Φ ← short
end if
OPENTRADE(Φ)
for each open trade t do
    tradeDuration(t) ← tradeDuration(t)+1
    if tradeDirection(t) == Φ then
        tradeDuration(t) ← 0
    end if
    if tradeDuration(t) >= TD then
        CLOSETRADE(t)
    end if
end for
```

The pseudo code presented in Algorithm 1 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 1), and are applied in the model selection process to identify the base T1 model for each stock.

3. Experiment Approach

In the following sub-sections we describe the different aspects of our experiment approach and also aim to present a critical analysis in view of our decisions.

3.1. Data

The data we use in this paper is high-frequency trade data for 15 stocks listed at the London Stock Exchange (see Table 2) during a 250 day period between 28/06/2007 to 25/06/2008 (excluding weekends, holidays and after hour trading). Data is sampled at 1 min intervals using the last trade price every 1 min. 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 127500 price points per stock over the entire period. The sample skewness and kurtosis in Table 2 indicate that

Table 2: Descriptive Statistics of 1 Minute Returns

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

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 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 can be seen from the numerous machine learning and artificial intelligence studies surveyed in Krollner et al. (2010) and Tsai and Wang (2009), this is rarely considered. Pardo (2011) warns that including only stocks that follow similar trends can lead to ungeneralised models which 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 2 we include a mix of stocks with both positive and negative mean returns (see also Figure 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.

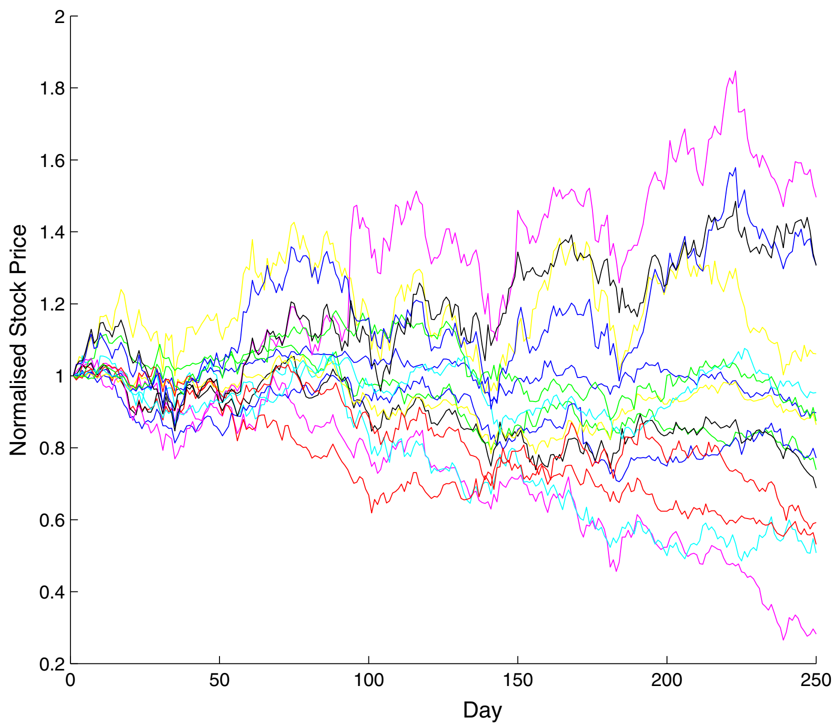


Figure 4: Times series of stock prices (normalised, for better comparison) which shows the mix of trends followed by the selected stocks over the in-sample and out-of-sample period.

3.2. Performance Measures

Surveys (Krollner et al., 2010; Tsai and Wang, 2009) show that the great majority of machine learning studies with application to 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; Pardo, 2011; Vella and Ng, 2014b). 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 misguiding since few large losses can still cancel out a higher number, but smaller in size, wins. Finally, focusing on just 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 lower risk of unexpected huge losses. For this reason, we apply three key measures to assess our models: Shape ratio, percentage profit and profit per trade.

The Sharpe ratio is one of the most popular reward-to-risk ratios and was proposed by the seminal paper of Sharpe (1966) based on the mean-variance theory. The ratio, which represents our risk-adjusted performance measure that we are ultimately interested in, is defined as

$$Sharpe\ Ratio = \frac{R_a - R_b}{\sigma}, \tag{28}$$

where R_a denotes the expected return, R_b the risk-free interest rate and σ the volatility of the return. The Sharpe ratio measures the risk premium per unit of risk in an investment. For a given level of percentage profit, investments with higher Sharpe ratios are preferred due to equivalent returns at lower levels of risk.

The profit per trade measure provides an indication of the efficiency of 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.

The combination of these three measures provides a clear overall picture of model performance both in terms of profitability and risk. However, we note two limitations of Sharpe ratio. Firstly, 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. Secondly, Lo (2002) warns that 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 done by Eling (2008) and Prokop (2012) show that Sharpe ratio measures lead to similar rankings of more sophisticated performance ratios. In balance, we decide in favour of 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.

3.3. Model Training and Testing

In the first experiment (Section 4.1), we compare the performance of T1 models trained using ANFIS and ANFIS/T2 models. The model selection process is based on identifying the best model parameters (defined in Table 1 in conjunction with trading algorithm parameters RT and DT) that result in the highest Sharpe ratio during the 150 day in-sample period. In the case of ANFIS/T2 models, parameter selection is extended to the identification of β parameters (as described in Section 2.3.2).

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 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.

We adopt a day-by-day moving window approach (see Figure 5), whereby at day_d, where $(d = 1, 2, ..., 150)$, the model is trained on 1 minute data points (Equation (7)) 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 (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 produce spurious results during back-testing (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.

3.4. 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. Albeit in Section 2.3.1 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 partial dependent approach, the primary advantage is that it makes it easier to directly compare the T1 and IT2 FLS. A second advantage is that the training of the IT2 FLS could be much faster since a number parameters would already be

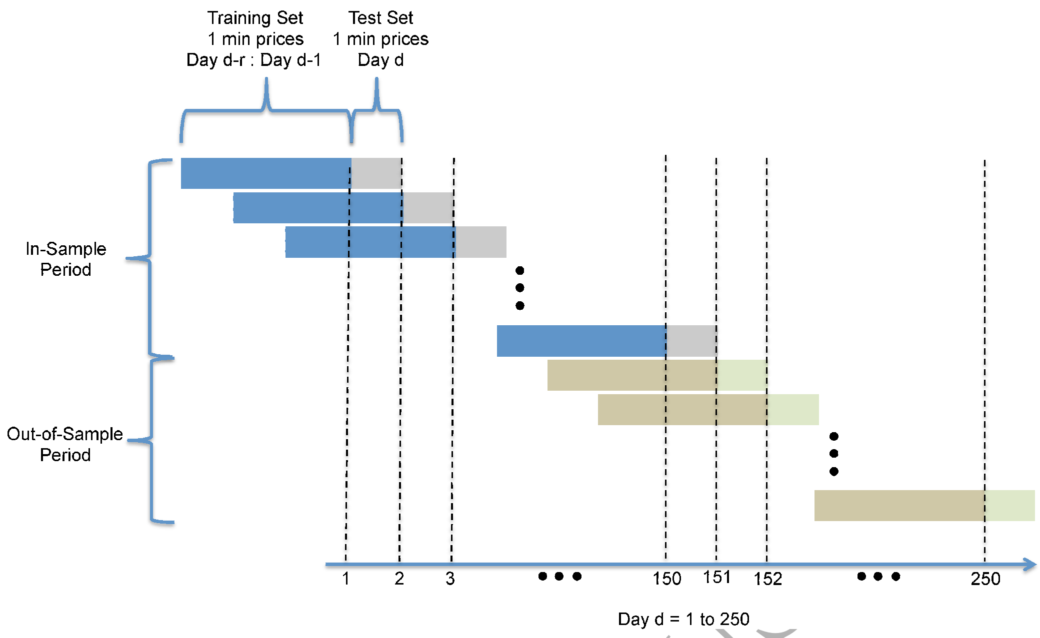


Figure 5: Adopted moving window approach. The first 150 days, each day consisting of 510 1 minute price points, is reserved for the in-sample training and model selection process. The models are trained every day using 1 minute data points from the previous r days and then tested on the next day's prices. 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.

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.

3.5. Assessing Performance Under Different Levels of Noise

An important decision that we 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 Sepulveda et al., 2006; Aladi et al., 2014) 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 that 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; Vanstone and Finnie, 2009, 2010; Holmberg et al., 2013). 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 return threshold. We argue that this proposed approach is more practical and realistic for algorithmic trading scenarios rather than injecting synthetic noise.

Based on this decision, in the second experiment (Section 4.2) 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 on 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.

4. Results and Analysis

In Section 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 4.2, we analyse models performance across increasing levels of uncertainty.

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 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 sell at the end of the 100 day period. The results from Table 3 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,

Table 3: 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.

Company	Symbol	B&H (daily) Profit%	B&H (100 days) Profit%
Antofagasta	ANTO	−8.95%	−8.76%
BHP Billiton	BLT	8.99%	11.37%
British Airways	BA	−3.56%	−40.52%
British Land Company	BLND	−23.43%	−33.67%
Sky	SKY	−8.19%	−16.34%
Cable and Wireless	CW	−19.76%	−10.46%
Aviva	AV	−20.54%	−20.33%
Diageo	DGE	−11.22%	−11.23%
HSBC Holdings	HSBA	5.78%	0.69%
Rio Tinto	RIO	2.20%	3.67%
BP	BP	−22.83%	7.17%
Lloyds Banking Group	LLOY	−4.83%	−30.21%
Tesco	TSCO	1.08%	−13.51%
HBOS	HBOS	−45.73%	−94.18%
Xstrata	XTA	−2.67%	−2.30%

with small to moderate positive period returns (1% to 11%) on 4 stocks (BLT, HSBA, RIO, BP) and heavy losses (>20% losses) in 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 statics in Table 2, 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 effected (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.

For our second benchmark comparison we apply the standard ANFIS. From Table 4 one immediately notices that contrary to the B&H methods, the algorithm is profitable on all 15 stocks. The B&H methods performed better than standard ANFIS only in 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; Tan et al., 2011; Kablan and Ng, 2011; Chen, 2013; Vella and Ng, 2014b; Wei et al., 2014) and the active research in improving the model and application techniques. This also validates our proposal to use ANFIS as our main benchmark model and use it as a basis to seek further improvements.

Table 4: Standard ANFIS performance after 100 day out-of-sample period.

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

From further investigation of the standard ANFIS results (Table 4) 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 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 identify pockets of profitability in shorter time windows, in our case in the 2min to 10min range. Our results also validate the theoretical claims of Zhang (2010); Rechenthin and Street (2013); Brogaard et al.

Table 5: ANFIS/T2a performance comparison against Standard ANFIS after 100 day out-of-sample period. The results depict variations from those produced by ANFIS in Table 4. The number of trades, Sharpe ratio and profit per trade are presented as percentage differences. Profit column shows the percentage point (p.p.) differences.

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

(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 ANFIS/T2a model against standard ANFIS, results in Table 5 show Sharpe ratio improvements in 12 out of 15 stocks. A lower Sharpe ratio was obtained in only 3 stocks (ANTO, HSBA, 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 reduction in profitability by more than 1 percentage points (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. We note that when considering the portfolio of 15 stocks, on average, 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 conjunction with higher Sharpe ratio (lower risk). Another advantage of ANFIS/T2a is the fact that higher risk adjusted performance and profitability is obtained with 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 increased noise filtering.

As a next step we compare the performance of ANFIS/T2b against standard ANFIS. Results in Table 6 indicate that a lower Sharpe ratio was obtained in 5 stocks (BLT, CW, HSBA, BP, and HBOS), 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 6 which show a lower average improvement in terms of Sharpe ratio however at a slightly improved average return per stock.

Our statistics show that at low to moderate intraday trading frequencies, which was 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 resulted in identifying the best balance between complexity and risk-adjusted performance.

We note that both ANFIS/T2a and ANFIS/T2b achieve 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.

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

In our second experiment we investigate the performance of standard ANFIS and the proposed IT2 TSK models under increasing levels of uncertainty. As

Table 6: ANFIS/T2b performance comparison against Standard ANFIS after 100 day out-of-sample period. The results depict variations from those produced by ANFIS in Table 4. The number of trades, Sharpe ratio and profit per trade are presented as percentage differences. Profit column shows the percentage point (p.p.) differences.

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

described in Section 2.3.1, we propose an innovative method to control the level of uncertainty using the signal threshold.

We test fixed thresholds starting from 0.08% down to 0.02% in steps of 0.02%. From the standard ANFIS results (see Table 7) 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 increase 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 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 0.04% level (XTA and RIO) and 4 at the 0.02% level (XTA, RIO, BLT and BLND).

The results presented in Table 7 were used as the basis for comparing the improvements attained from our IT2 models (presented in Figure 6 and Table 8). From Figure 6, we immediately note a steep drop in the average profit per trade and average profitability at the 0.02% level. When investigating further, we see that the ANFIS summary statistics in Table 8 show increasing average

Table 7: ANFIS performance after 100 day out-of-sample across different levels of return threshold (uncertainty).

	Return threshold level					
	0.08%			0.06%		
Symbol	No. of Trades	Sharpe Ratio	Profit /Trade (£)	No. of Trades	Sharpe Ratio	Profit /Trade (£)
ANTO	2036	0.3388	18.90	3687	0.2644	11.03
BLT	475	0.3224	47.56	1016	0.2822	17.50
BA	6248	1.5420	54.92	7830	1.4161	49.34
BLND	921	0.4664	30.12	1956	0.2577	10.48
SKY	1197	0.6189	29.18	2606	0.5404	16.05
CW	4308	1.4777	104.04	4709	1.2952	103.87
AV	1074	0.4136	19.09	2429	0.4510	12.08
DGE	480	0.3605	35.52	1203	0.3807	16.11
HSBA	426	0.4560	37.44	1129	0.2912	16.24
RIO	535	0.2040	23.31	1071	0.1253	9.24
BP	1577	0.8068	33.85	3220	0.8132	20.43
LLOY	3140	0.5685	20.61	5849	0.6705	18.56
TSCO	3211	0.9950	30.97	5505	1.2153	30.90
HBOS	4708	0.4693	24.90	7322	0.4543	20.72
XTA	911	0.1222	16.61	1715	-0.0218	-1.55
	Return threshold level					
	0.04%			0.02%		
Symbol	No. of Trades	Sharpe Ratio	Profit /Trade (£)	No. of Trades	Sharpe Ratio	Profit /Trade (£)
ANTO	6795	0.3356	13.21	14270	0.1378	3.39
BLT	2652	0.1725	5.15	2200	-0.1748	-13.62
BA	9738	1.3835	45.94	12520	1.1199	37.72
BLND	4462	0.2408	7.52	10466	-0.1042	-2.05
SKY	5481	0.6820	12.83	9385	0.4024	7.50
CW	5260	1.4373	97.96	6204	1.3938	85.68
AV	6104	0.2137	3.38	12930	0.0402	0.67
DGE	3348	0.2761	6.05	10495	0.1898	2.37
HSBA	3836	0.3572	6.91	10153	0.0430	0.53
RIO	2643	-0.0750	-2.78	7534	-0.3773	-8.48
BP	6978	0.9543	15.39	11295	0.7763	10.04
LLOY	9621	0.7212	17.63	13693	0.6223	13.12
TSCO	7928	1.5421	26.08	10865	1.2703	20.43
HBOS	11467	0.3684	15.50	16007	0.2740	9.61
XTA	3771	-0.1765	-6.81	2992	-0.2403	-20.21

profitability per stock, from 25.43% at the 0.08% level, up to 33.42% at the 0.04% level. This is attained at reducing average profit per trade however 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 Sharpe ratio (see Table 8 and Figure 6). 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 shows increasing improvements in Sharpe ratio, average profit per stock and average profit per trade. In the case of 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 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 6). The increase in all three measures, especially the increase in profitability at 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 8).

As a final step in our experiment, we validate our results using statistical tests. The tests are carried out on the average Sharpe ratio, average profit per stock and average profit per trade obtained using a paired t-test on the 15 stocks. Table 8 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 improved performance of our proposed ANFIS/T2 models against standard ANFIS at levels of higher uncertainty due to more exposure 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 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 on 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.

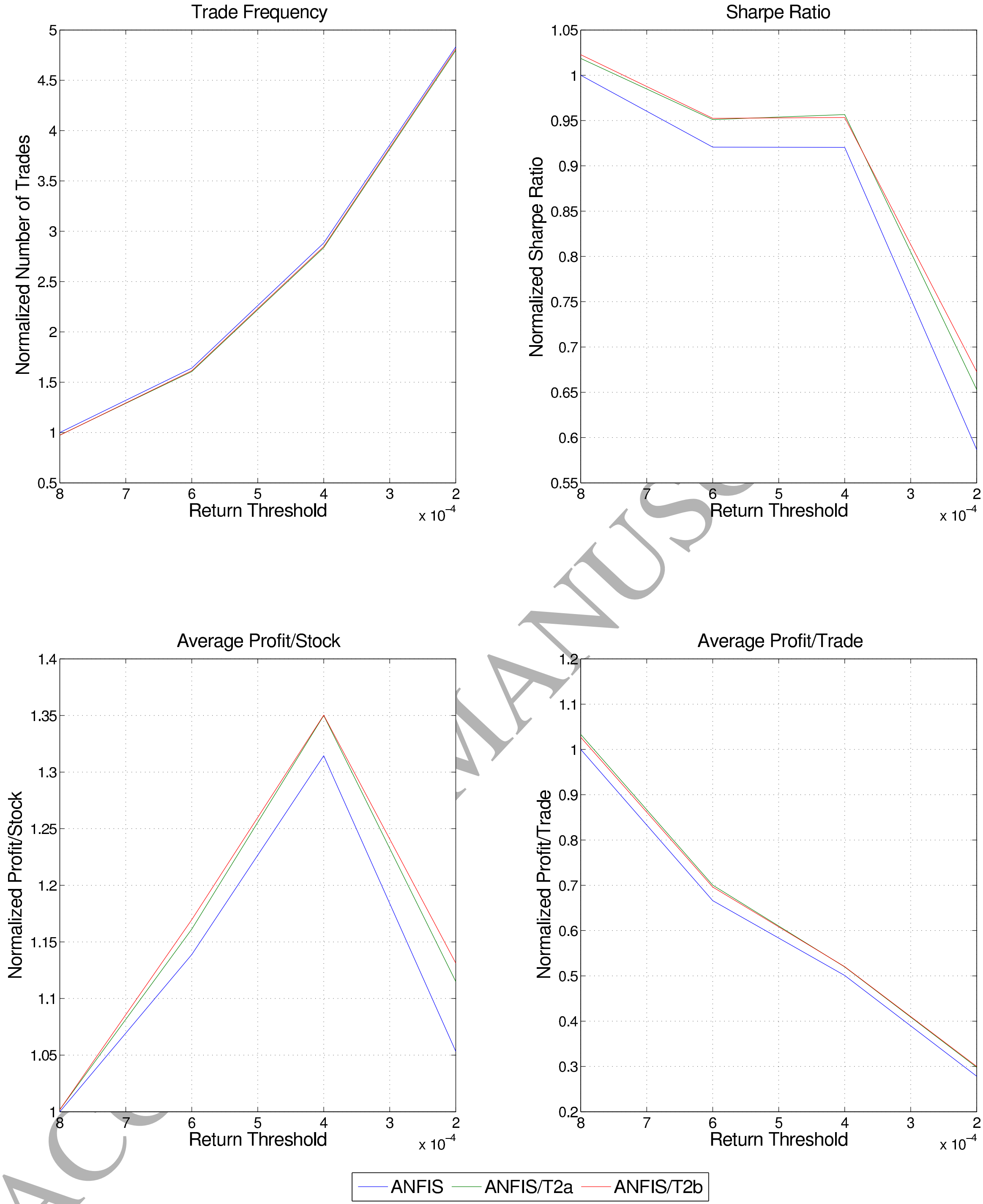


Figure 6: Trends on various measures after 100 days out-of-sample trading and at different degrees of trading frequency (uncertainty).

Table 8: 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.

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

5. Conclusion

Albeit the Efficient Market Hypothesis still resonates with mainstream finance literature, this work extends the thoughts of Johnson et al. (2013) and Brogaard et al. (2014) who highlight the need for new theories in support for high frequency financial phenomena during which the human traders loose the ability to react in real time. Unlike the great majority of computational finance research which focuses on model error reduction, directional accuracy or just profitability, this work is motivated by 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). In line with the latter, and also investors' preferences, our focus is on improving risk-adjusted performance of algorithmic and high frequency trading.

In this paper we convey three contributions. Firstly we propose two innovative and practical methods of how the ANFIS model, a popular AI technique with very active research applications in finance, can be improved by introducing T2 fuzzy sets. The main benefit is to minimise the uncertainty caused by microstructure noise, hence reducing the overall risk. Both of the proposed methods show 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 in HFT. Our results extend the findings from a number of authors (Schulmeister, 2009; Holmberg et al., 2013; Rechenthin and Street, 2013) who claim possible breaks in market efficiency at short time frames. 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. Sepulveda et al., 2006; Aladi et al., 2014) 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 exhibit the highest benefits in trading scenarios reflecting higher exposure to microstructure noise, making our models ideal for HFT environments.

Our contributions also convey a number of management insights. We present an approach of how existing algorithmic and HFT models can be improved by increasing risk-adjusted performance but without compromising overall profitability. Our approach shows a stepwise incremental approach by starting from the popular ANFIS model, and show how by introducing IT2 components on the base model (rather than a whole overhaul to existing investment) this can be enhanced to meet these objectives. The results 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 around risk controls.

In conclusion, this paper opens up a number of avenues for further research. As a start, we propose future research paths that can followed from a fuzzy logic perspective. In this paper we identify risk-adjusted improvements by introducing T2 fuzzy sets and adaptive FOU sizes. This was purposely done to open a research path whilst at the time keeping models simple, practical to use and easier to compare by keeping additional parameters to a minimum. Our first research proposition is that we do not exclude the possibility of further improvements by exploring additional incremental steps in T2 configuration complexity such as different rule extraction methods, membership functions, more complex T2 rules or defuzzification methods. Secondly, in line with additional model complexity, which can mitigate wider adoption, interested researchers can seek to mitigate this risk by investigating the performance of similar Mamdani models which tend to increase model interpretability. We also identify a number of research avenues from a finance perspective. Our first research proposition is to investigate the effect of alternative FOU tuning frequencies, which in this paper we limit to a daily basis. In particular, any possible beneficial relationship between FOU tuning frequency and the diurnal patterns of market intraday activity, a stylised face in finance, remains an open question. Secondly, albeit our study identifies trading improvements by using a set of stocks listed on the London stock exchange, 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. Thirdly, in this paper we adopt the Sharpe ratio as our main risk-adjusted performance measure due to its popularity in both finance literature and practice. In line with Lo (2002), we suggest further research which investigates the underlying model returns and Sharpe ratio statistics, especially in view of further stress testing of the underlying trading models during specific extreme events.

Acknowledgement

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

References

Aladi, J. H., Wagner, C., and Garibaldi, J. M. (2014). Type-1 or interval type-2 fuzzy logic systems on the relationship of the amount of uncertainty and FOU size. In *Proceedings of FUZZ-IEEE 2014, Beiging*, pages 2360–2367.

Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2):115–158.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55.

Anderson, R. M. (2011). Time-varying risk premia. *Journal of Mathematical Economics*, 47(3):253–259.

Anderson, R. M., Eom, K. S., Hahn, S. B., and Park, J.-H. (2013). Autocorrelation and partial price adjustment. *Journal of Empirical Finance*, 24:78–93.

Bailey, D. H., Borwein, J. M., de Prado, M. L., and Zhu, Q. J. (2014). Pseudo-mathematics and financial charlatanism: The effects of backtest over fitting on out-of-sample performance. *Notices of the AMS*, 61(5):458–471.

Benatar, N., Aickelin, U., and Garibaldi, J. M. (2012). An investigation into the relationship between type-2 FOU size and environmental uncertainty in robotic control. In *Proceedings of FUZZ-IEEE 2012, Brisbane, Australia*, pages 1–8.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers.

Birkin, P. A. and Garibaldi, J. M. (2009). A comparison of type-1 and type-2 fuzzy controllers in a micro-robot context. In *Proceedings of FUZZ-IEEE 2009, Jeju Island, Korea*, pages 1857–1862.

Boyacioglu, M. A. and Avci, D. (2010). An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange. *Expert Systems with Applications*, 37(12):7908–7912.

Brabazon, A. and O'Neill, M. (2006). *Biologically Inspired Algorithms for Financial Modelling*. Berlin, Heidelberg: Springer.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-frequency trading and price discovery. *Review of Financial Studies*, 27(8):2267–2306.

Chang, J.-R., Wei, L.-Y., and Cheng, C.-H. (2011). A hybrid anfis model based on ar and volatility for TAIEX forecasting. *Applied Soft Computing*, 11(1):1388–1395.

Chen, M.-Y. (2013). A hybrid anfis model for business failure prediction utilizing particle swarm optimization and subtractive clustering. *Information Sciences*, 220:180–195.

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

Choudhry, T., McGroarty, F., Peng, K., and Wang, S. (2012). High-frequency exchange-rate prediction with an artificial neural network. *Intelligent Systems in Accounting, Finance and Management*, 19(3):170–178.

Christoffersen, P. (2011). *Elements of Financial Risk Management*. Waltham, MA, USA: Academic Press.

Dutta Baruah, R. and Angelov, P. (2010). Clustering as a tool for self-generation of intelligent systems: a survey. *Evolving Intelligent Systems, Leicester, UK*.

Eling, M. (2008). Does the measure matter in the mutual fund industry? *Financial Analysts Journal*, 64(3):54–66.

Engle, R. and Patton, A. J. (2001). What good is a volatility model? *Quantitative Finance*, 1(2):237–245.

Fama, E. (1965). Random Walks in Stock Market Prices. *Financial Analysts Journal*, 21(5):55–59.

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

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.

Gradojevic, N. and Gençay, R. (2013). Fuzzy logic, trading uncertainty and technical trading. *Journal of Banking & Finance*, 37(2):578–586.

Greenfield, S. and Chiclana, F. (2013). Accuracy and complexity evaluation of defuzzification strategies for the discretised interval type-2 fuzzy set. *International Journal of Approximate Reasoning*, 54(8):1013–1033.

Hansen, P. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, 20(7):873–889.

Heal, G. and Millner, A. (2014). Reflections uncertainty and decision making in climate change economics. *Review of Environmental Economics and Policy*, 8(1):120–137.

Holmberg, U., Lönnbark, C., and Lundström, C. (2013). Assessing the profitability of intraday opening range breakout strategies. *Finance Research Letters*, 10(1):27–33.

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

Jang, J.-S. and Sun, C.-T. (1995). Neuro-fuzzy modeling and control. In *Proceedings of the IEEE*, volume 83, pages 378–406.

John, R. and Coupland, S. (2007). Type-2 fuzzy logic: A historical view. *Computational Intelligence Magazine, IEEE*, 2(1):57–62.

Johnson, N., Zhao, G., Hunsader, E., Qi, H., Johnson, N., Meng, J., and Tivnan, B. (2013). Abrupt rise of new machine ecology beyond human response time. *Scientific Reports*, 3.

Kaastra, I. and Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3):215–236.

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.

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

Knight, F. H. (1921). Risk, uncertainty and profit. *Boston: Houghton-Mifflin.*

Krollner, B., Vanstone, B., and Finnie, G. (2010). Financial time series forecasting with machine learning techniques: A survey. In *Proceedings of the European Symposium on Artificial Neural Networks: Computational and Machine Learning, Bruges Belgium*, pages 28–30.

Lo, A. W. (2002). The statistics of sharpe ratios. *Financial Analysts Journal*, 58(4):36–52.

Mamdani, E. H. (1974). Application of fuzzy algorithms for control of simple dynamic plant. In *Proceedings of the Institution of Electrical Engineers*, volume 121, pages 1585–1588. IET.

McAleer, M. and Medeiros, M. C. (2008). Realized volatility: A review. *Econometric Reviews*, 27(1):10–45.

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.

Melin, P. and Castillo, O. (2014). A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition. *Applied Soft Computing*, 21:568–577.

Mendel, J., Hagras, H., Tan, W.-W., Melek, W. W., and Ying, H. (2014). *Introduction to Type-2 Fuzzy Logic Control: Theory and Applications.* John Wiley & Sons, Hoboken, NJ, USA.

Mendel, J. M., John, R. I., and Liu, F. (2006). Interval type-2 fuzzy logic systems made simple. *IEEE Transactions on Fuzzy Systems*, 14(6):808–821.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

Murphy, J. (1987). *Study Guide for Technical Analysis of the Futures Markets: A Self-training Manual*. New York Institute of Finance.

Naranjo, R., Meco, A., Arroyo, J., and Santos, M. (2015). An intelligent trading system with fuzzy rules and fuzzy capital management. *International Journal of Intelligent Systems*, 30(8):963–983.

Nelson, S. C. and Katzenstein, P. J. (2014). Uncertainty, risk, and the financial crisis of 2008. *International Organization*, 68(02):361–392.

Nie, M. and Tan, W. W. (2008). Towards an efficient type-reduction method for interval type-2 fuzzy logic systems. In *Proceedings of FUZZ-IEEE 2008, Hong Kong*, pages 1425–1432.

Pal, N. R. and Bezdek, J. C. (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems*, 3(3):370–379.

Pardo, R. (2011). *The evaluation and optimization of trading strategies*, volume 314. John Wiley & Sons, New Jersey.

Poon, S.-h. and Granger, C. (2005). Practical Issues in Forecasting Volatility. *Financial Analysts Journal*, 61(1):45–56.

Poon, S.-H. and Granger, C. W. J. (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature*, 41(2):478–539.

Prokop, J. (2012). Further evidence on the role of ratio choice in hedge fund performance evaluation. *Journal of Finance and Investment Analysis*, 1(3):181–195.

Rechenthin, M. and Street, W. N. (2013). Using conditional probability to identify trends in intra-day high-frequency equity pricing. *Physica A: Statistical Mechanics and its Applications*, 392(24):6169–6188.

Sahin, S., Tolun, M. R., and Hassanpour, R. (2012). Hybrid expert systems: A survey of current approaches and applications. *Expert Systems with Applications*, 39(4):4609–4617.

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

Sepulveda, R., Melin, P., Díaz, A. R., Mancilla, A., and Montiel, O. (2006). Analyzing the effects of the footprint of uncertainty in type-2 fuzzy logic controllers. *Engineering Letters*, 13(2):138–147.

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, 39(1):119–138.

Sugeno, M. and Kang, G. (1988). Structure identification of fuzzy model. *Fuzzy Sets and Systems*, 28(1):15–33.

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.

Vella, V. and Ng, W. L. (2014a). A dynamic fuzzy money management approach for controlling the intraday risk-adjusted performance of AI trading algorithms. *Intelligent Systems in Accounting, Finance and Management*, 22(2):153–178.

Vella, V. and Ng, W. L. (2014b). Enhancing risk-adjusted performance of stock market intraday trading with neuro-fuzzy systems. *Neurocomputing*, 141:170–187.

Wagner, C. and Hagras, H. (2010). Toward general type-2 fuzzy logic systems based on zslices. *IEEE Transactions on Fuzzy Systems*, 18(4):637–660.

Wei, L.-Y., Cheng, C.-H., and Wu, H.-H. (2014). A hybrid anfis based on n-period moving average model to forecast TAIEX stock. *Applied Soft Computing*, 19:86–92.

Wu, D. (2013). Approaches for reducing the computational cost of interval type-2 fuzzy logic systems: overview and comparisons. *IEEE Transactions on Fuzzy Systems*, 21(1):80–99.

Wu, D. and Mendel, J. M. (2014). Designing practical interval type-2 fuzzy logic systems made simple. In *Proceedings of FUZZ-IEEE 2014, Beijing*, pages 800–807.

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.

Zhang, F. (2010). The effect of high-frequency trading on stock volatility and price discovery. *Working paper, Yale University.*