BOUNDEDLY RATIONAL LEARNING AND HETEROGENEOUS TRADING STRATEGIES WITH HYBRID NEURO-FUZZY MODELS

Stelios D. Bekiros*

European University Institute, Via Delle Fontanelle 10, 50014
San Domenico di Fiesole (FI), Italy
&
Center for Nonlinear Dynamics in Economics and Finance (CeNDEF)
Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11,
1018 WB Amsterdam, The Netherlands

Abstract

The present study deals with heterogeneous learning rules in speculative markets where heuristic strategies reflect the rules-of-thumb of boundedly rational investors. The major challenge for "chartists" is the development of new models that would enhance forecasting ability particularly for time series with dynamic time- varying, nonlinear features. This paper introduces fuzzy learning rules with the incorporation of beliefs, preferences and idiosyncratic behavioral patterns for decision-making and trading under uncertainty. The efficiency of a technical trading strategy based on a neurofuzzy model is investigated, in order to predict the direction of the market for NASDAQ Composite, NIKKEI255 and FTSE100. Moreover, it is demonstrated that the incorporation of the estimates of the conditional volatility changes strongly enhances predictability, as it provides valid information for a potential turning point on the next trading day. The total return of the proposed volatility-based neurofuzzy model, including transaction costs, is consistently superior to a markov-switching model, a recurrent neural network as well as to the buy & hold strategy for all indices. The findings can be justified by invoking either the "volatility feedback" theory or the existence of portfolio insurance schemes in the equity markets and are also consistent with the view that volatility dependence produces sign dependence. Overall what leads to optimal prediction is the dynamic update of the expectations and preferences of the heuristic learning rules combined with the adaptive calibration of the "degrees-of-belief" that match agent's "fads".

Keywords: Market heterogeneity; Technical trading rules; Heuristic learning; Neurofuzzy models; Conditional volatility

JEL classification: G10; G14; C53

---

* The author would like to thank Cees Diks and Ramazan Gençay for helpful discussions and comments on an earlier version of the paper. A part of the study was presented at the seminar series of the Center for Nonlinear Dynamics in Economics and Finance (CeNDEF) at the University of Amsterdam (UvA). The author acknowledges financial support from the Max Weber Programme at the European University Institute, Florence. Tel.: +39-055-4685-698; *Fax:* +39-055-4685-804; *E-mail address*: Stelios.Bekiros@eui.eu

1. Introduction

Heterogeneous agents approach challenges the conventional representative, rational agent framework. Heterogeneity in expectations can lead to market instability and complicated dynamics of prices, which are driven by endogenous market forces. Simon (1957) argued that, boundedly rational agents using simple rules-of-thumb for their decisions under uncertainty, provides a more accurate and realistic description of human behavior than perfect rationality with optimal decision rules. Ever since the introduction of the Efficient Markets Hypothesis, fully rational agents were considered the driving forces of markets, which in turn operated in a way to aggregate and process the beliefs and demands of traders reflecting all available information (Fama, 1970, 1991). But the empirical evidence from financial markets was not in full accordance with the Efficient Markets Hypothesis. The alternative behavioral model suggested was based on relaxing strict rational agent assumptions and introducing market frictions. The key arguments of behavioral agent based models as reported by Hommes (2001, 2005, and 2006) are closely related to Keynes view that “expectations matter”, to Simon’s result that humans are boundedly rational and to Kahneman-Tversky analysis in psychology that individual behavior under uncertainty can be described by simple heuristics and biases.

In view of empirical studies that stock prices can be predicted with a fair degree of reliability advocates of Efficient Markets Hypothesis (e.g. Fama and French, 1995) claim that such results are based on time-varying-equilibrium expected returns generated by rational pricing in an efficient market which compensates for the level of risk undertaken. On the contrary, opponents (e.g. La Porta, *et. al.*, 1997; Shiller, 2002) argue that predictability reflects the psychological factors and fashions or "fads" of irrational investors in a speculative market. This irrational behavior has been emphasized by Shleifer and Summers (1990) and Black (1986) in their exposition of noise traders who act on the basis of imperfect information and consequently cause prices to deviate from their equilibrium values. Arbitrageurs dilute a minor part of these shifts in prices, yet the major component of deviation is tradable. Moreover, Black claimed that noise traders play a useful role in promoting market liquidity. Overall, there are two types of agents in heterogeneous agent models: “fundamentalists”, who base their expectations upon dividends, earnings, growth or even macroeconomic factors, and “chartists” (noise traders and technical analysts) who instead base their trading strategies upon historical patterns and heuristics and try to extrapolate trends in future asset prices (Brock and Hommes, 1998). The present study focuses on the latter.

2. Asset returns and volatility dynamics

Fluctuations in actual prices, many times even greater than those implied by changes in the market fundamentals, are inferred by Shiller (1987) as being the result of waves of optimistic or pessimistic market psychology. The sharp stock US market decline of 22% on October 19, 1987, in the complete absence of news about fundamentals, appears to contradict conventional theory as suggested by the Efficient Markets Hypothesis. Moreover, the empirical investigation of the relation between stock return volatility and stock returns has a long tradition in finance literature. According to the "time-varying risk premium theory" (Bekaert and Wu, 2000) the return shocks are caused by changes in conditional volatility. When 'bad' news arrives in the market the current volatility increases and this causes upward revisions of the conditional volatility. This increased conditional volatility has to be compensated by a higher expected return, leading to an immediate decline in the current value of the market. An asymmetric nature of the volatility response to return shocks emerges from the above theory. While bad news generates an increase in conditional volatility, the net impact of good news in not clear. Another explanation to the asymmetric reaction of the conditional volatility may be offered through the "leverage effects" (Christie, 1982). A negative (positive) return increases (reduces) financial leverage, which makes the stock riskier (less risky) and increases (reduces) volatility. The causality however here is different: the return shocks lead to changes in conditional volatility, whereas the time-varying premium theory contends the opposite. An alternative rationalization for the relation of conditional volatility revisions and stock returns may be offered by invoking trigger strategies in the equity markets (Krugman, 1987). Institutional participants in equity markets react whenever the maximum expected loss of portfolios, as measured for example by the Value-at-Risk (VaR), reaches a predetermined level and therefore share price dynamics are being driven, partly, by revisions in the measured conditional volatility[1]. Each time the conditional volatility rises, a number of those portfolios will deviate from their pre-

[1] VaR depends entirely on a multiple of the estimated conditional volatility under the assumption of normally distributed returns.

determined level of VaRs, hitting their risk limits and this will generate a re-allocation of assets towards safer ones. When portfolio insurers leave the market the stock prices must fall in order for the other investors to be given an incentive to hold a larger quantity of stock. If a rational expectations world is further assumed then investors take into account the effects of portfolio insurance schemes and no drop in stock prices is being observed. Furthermore, many researchers study the contemporaneous relationship between the one-day stock index returns and the associated changes in the level of implied volatility indices (Whaley 2000, Simon 2003, Giot 2005). The results indicate the existence of a negative and statistically significant relationship between the returns of the S&P100 (NASDAQ 100) and the implied volatility VIX (VXN) index. Specifically, for the S&P100 index this relationship has been also found to be asymmetric in the sense that negative stock index returns are associated with greater proportional changes in implied volatility measures than are positive returns. The explanation offered for this opposite response is that option traders react to negative returns by bidding up the implied volatility. Nonetheless, empirically there is a growing debate whether the implied volatility can be used as a forward indicator of the underlying equity index. This issue has not been treated properly in the literature with the exception of a paper by Giot (2005) who regressed the forward looking S&P100 index returns, over various time intervals, on 21 dummy variables representing equally spaced percentiles of a rolling two-year history of VIX. Giot (2005) concludes that positive forward-looking returns are to be expected for long positions at high levels of the implied volatility indices.

In one part of this study the trading implications of conditional volatility are examined within a broader framework as concerns the nonlinear functional form of the forecast generating mechanism as well as the presence of past returns that might have forecasting power. Christoffersen and Diebold (2003) in their paper show that volatility dependence produces sign dependence, and therefore forecastability, as long as expected returns are nonzero. The intuition behind this relationship is that volatility changes will alter the probability of observing negative or positive returns. More specifically, the higher the volatility, the higher the probability of a negative return, as long as the expected returns are positive. Moreover, they show that this result is entirely consistent with the existence of no conditional mean dependence, or the absence of conditionally Gaussian distributions. In that context, the predictive return sign ability of trading rules that rely on a simple switching strategy is investigated: positive predicted returns are executed as long positions and negative returns as short positions. A similar strategy has been employed, with considerable success, by a number of other researchers (Gençay 1998b, Gençay 1998a, Fernández-Rodriguez *et. al.*, 2000) etc. In general terms they find that the returns from the switching strategy are higher than those from the passive one for annual returns, even when transaction costs are high. They also find that the asset return predictability is increased during volatile periods. The buy and sell signals are produced from technical trading strategies that incorporate various linear or non-linear econometric models.

3. Nonlinear forecasting and boundedly rational decision-making with Fuzzy inference systems

In heterogeneous markets the major challenge for "chartists" is the development of new models, or the modification of existing methods, that would enhance forecasting ability particularly for time series with dynamic time variant patterns. Conventional time series analysis, based on stationary stochastic processes does not always perform satisfactorily on economic and financial time series (Harvey, 1989). The reason is that economic data are not generally described by simple linear structural models, white noise or even random walks. The most commonly used techniques for financial forecasting are Regression methods and Autoregressive Moving Average (ARMA) models (Box and Jenkins, 1970). These methods have been used extensively in the past, but they often fail to give an accurate forecast for some series because of their nonlinear structures and some other inherent limitations. Even though ARCH/GARCH models (Bollerslev 1986, Engle 1982) deal with non-constant variance, still some series cannot be explained or predicted satisfactorily, due to inherent chaotic or noise patterns, fat tails, or other nonlinear components.

Extensive research in the area of nonlinear modeling has shown that neural networks enhance financial forecasting, mainly because they perform advanced mathematical and statistical processes such as nonlinear interpolation and function approximation. Neural Networks are parallel computational models comprising input and output vectors as well as processing units (neurons) interconnected by adaptive connection strengths (weights), trained to store the "knowledge" of the network. Adya et. al. (1998) demonstrated the advanced predictive ability of neural networks for time series forecasting. White (1989) and

Kuan et. al. (1994), suggested that the relationship between neural networks and conventional statistical approaches for time series forecasting is complementary. Refenes et al. (1994) indicated that traditional time series techniques for forecasting have reached their limitation in applications with nonlinearities within the data sets. Additionally, the function approximation properties of neural networks have been thoroughly investigated by many authors. The results in Cybenko (1989), Funahashi (1989), Hornik (1991), Hornik et. al. (1989, 1990), Gallant et. al. (1988, 1992) and Hecht-Nielsen (1989) demonstrated that feedforward networks with sufficiently many hidden units and properly adjusted parameters can approximate any function to any desired degree of accuracy. Poddig (1993) applied a feedforward neural network to predict the exchange rates between American Dollar and Deutsche Mark, and compared results to regression analysis. Other examples using neural networks in stock and currency markets include Gençay (1998), Green et. al. (1994), Manger (1994), Rawani (1993), Weigend (1991), Yao et. al. (1996) and Zhang (1994). However, conventional time series analysis techniques as well as neural networks incorporate in terms of input variables, only quantitative factors, such as stock returns, indices and other financial or economic magnitudes. A number of qualitative factors, e.g., macroeconomical or political effects as well as traders psychology or "fads", may seriously influence the market trend, thus it is important to capture this unstructured expert knowledge.

Fuzzy logic has been implemented initially in the area of control systems and decision theory and recently in economic applications with highly promising results. It provides a means of representing uncertainty with imprecise data. In that sense it is an excellent tool for the boundedly rational agent who uses rules-of-thumb for decision-making and learning under uncertainty. Specifically, in a fuzzy system numeric variables (inputs and outputs) are translated into fuzzy linguistic terms representing beliefs, e.g. "low" and "high". Each term is described by a membership function, which estimates the *"degree"* to which a variable belongs to a fuzzy set. Finally, *fuzzy inference or learning rules* represented in IF-THEN statements are specified to associate the fuzzy input to the output fuzzy set. Thus, the IF-THEN rules could comprise an efficient mechanism of incorporating heterogeneous beliefs of the agent in a technical trading system. In general, Fuzzy systems are widely applied in fields like classification, decision support, process simulation and control systems, exactly because are effective means of modeling human expert knowledge, experience, intuition, etc., (Sugeno 1988, Kosko 1992, Klir et. al.1995, Jamshidi et. al. 1997, Mamdani 1974, 1977). Financial and marketing applications have also been reported (Altrock, 1997). One important advantage of fuzzy inference systems is their linguistic interpretability. When implementing fuzzy systems, the focus is paid on modeling fuzziness and linguistic vagueness using membership functions. The fuzzy system approach has been applied to different forecasting problems whereby the operator's expert knowledge is used for prediction (Kaneko 1996, Al-Shammari, et. al. 1998). Although the fuzzy logic-based forecasting shows promising results, the process to construct a fuzzy logic system is subjective and depends on some ad-hoc assumptions. The learning rules derived in this way may not always yield the best forecast, and the choice of membership functions depends on trial and error. Neural networks' learning ability can be utilized to adjust and fine-tune the fuzzy membership functions. The combination of both techniques results in a hybrid Neurofuzzy model which incorporates the learning ability of the neural network and the functionality of the fuzzy expert system. In a Neurofuzzy system the basic concept is the derivation of various parameters of a fuzzy inference system by means of adaptive training methods obtained from neural networks (Buckley et. al. 1994, Lin et. al. 1996, Nishina et. al., 1997).

The present study advances the literature on heterogeneous learning rules in speculative markets where strategies reflect the "fads" of boundedly rational investors, by introducing a technical trading system with the incorporation of beliefs and idiosyncratic behavioral patterns represented by fuzzy inference rules. In technical terms, it expands the literature that has utilized separately neural networks or fuzzy logic systems in forecasting applications, by presenting a hybrid Neurofuzzy approach that leads to superior predictions upon the direction-of-change of the market. Moreover, beyond the existing practice that has utilized as inputs return lags, moving averages etc, it is demonstrated that the Neurofuzzy model leads to superior predictions via the incorporation of volatility changes in addition to endogenous return lags. The purpose of this paper is to illustrate this concretely through an investigation of the relative direction-of-change predictability of the proposed volatility-based Neurofuzzy trading model compared to other well-established nonlinear trading models.

The remainder of this paper is organized as follows. Section 4 describes how the Neurofuzzy model for heuristic learning in heterogeneous markets is constructed. In Section 5 the other forecasting models used in this study are described. Section 6 provides a brief review of conditional volatility models. Finally, the empirical results are shown in Section 7 and Section 8 provides concluding remarks.

4. Heterogeneous beliefs and heuristic learning under incomplete information: the hybrid Neurofuzzy model

The superior functionality of the Neurofuzzy trading system lies in the efficient mechanism of incorporating the heterogeneous beliefs of the agent under uncertainty and imprecise knowledge, called *fuzzy learning rules* (or inference rules, using the fuzzy logic terminology). Fuzzy learning, represented in IF-THEN statements, is specified to associate input and output variables of a system, which in this case is a heterogeneous financial market, while modeling human expert knowledge, experience, psychology patterns and intuition of the agent. Consequently, the IF-THEN rules' set-up provides a very realistic model of the decision-making process under which rule-of-thumb traders operate.

Technically, the Neurofuzzy architecture consists of the input, the rule layer and the output layer. In the input fuzzy layer all the input variables are translated into fuzzy linguistic terms. Each term is described by fuzzy membership functions, which estimate the *"degree of belief"*, whereas in the Boolean formalism the particular variable would have a crisp numeric value. The type of membership functions is configured in this layer, whereas the parameters of these functions are processed and optimized via neural network learning. The fuzzy learning rules consist of two parts, the "IF" part and "THEN" part. The "IF" part utilizes an "AND" association. This operator proposed by Zimmermann et. al. (1978) represents the minimum value among all the validity values of the "IF" part. The output fuzzy layer incorporates the fuzzy membership functions for outputs. Finally, in the defuzzification layer, the output is converted from fuzzy variables back into crisp values. The aforementioned structure utilizes the Mamdani (1997) approach of fuzzy learning. Alternatively, Sugeno's (1985) approach introduces linear dependences of each rule on the system's input variables, whereby no defuzzification process is required. The more general first-order Sugeno fuzzy model has rules of the form *"IF x* is *A AND y* is *B THEN*

$z = h + c \cdot x_1 + d \cdot x_2$" , where A and B are fuzzy sets while *c, d,* and *h* are parameters. Because of the linear dependence of each rule on the system's input variables the Sugeno system is suited for modeling complex nonlinear systems by interpolating multiple linear models.

In order for agents to forecast the upward and downward trends of the financial market, a three-input, two rule first order Sugeno model is utilized, where the parameters *c, d, k* and *h* of the n^{th} rule contribute via the following first order polynomial:

$$z_n = h_n + c_n x_1 + d_n x_2 + k_n x_3 \qquad (1)$$

This model comprises two parameter sets, namely the membership function parameters and the polynomial parameters (*c, d, h, k*), which are all time-varying in order to account for dynamic persistence and structural changes in the input variables, and adaptively updated (the time index is dropped to improve readability). In the proposed architecture two membership functions are used for each input corresponding to two regimes, or *beliefs* as being perceived by the boundedly rational agents, namely "low" and "high". The hybrid learning process uses a Levenberg-Marquardt neural back-propagation algorithm (Hagan et. al., 1994) to optimize the membership parameters and a least squares-type algorithm to solve for the polynomial parameters. The polynomial parameters are updated first using a least squares-type algorithm and the membership parameters are then updated by backpropagating the errors. Finally, in order to solve for all parameters the squared error objective function is used:

$$E = \frac{1}{2}(y - y^o)^2 \qquad (2)$$

where y^o the target output and y the system output for N size sample.

The architecture of the proposed model consists of five sequential *"learning layers"* represented by $L_{l,i}$ where *l=1,…,5* the index of each layer, *i* the i^{th} node of layer $L_{l,i}$ and *j* the number of inputs. In the first layer the grades μ of the membership functions i.e. "*degrees of belief*" of each input *j* are generated:

$$L_{1,i} : \mu_{M_i}\left(x_j\right) \qquad (3)$$

In the second *"learning layer"* the rule weight coefficients are produced. In this layer a weight is attributed on each learning rule, which is not necessary symmetric, corresponding to the *agent's preference* of the influence of the rule to the final decision. Intelligently, as it is illustrated below, the system adaptively "learns by itself" and attributes the optimal weight to each rule in order to match the trader's expectations, instead of keeping it constant based on an ex-ante "rational" assumption. At the third layer the weights are normalized:

$$L_{2,i} : w_i = \prod_{j=1}^{m} \mu_{M_i} \left(x_j \right) \qquad (4)$$

$$L_{3,i} : \overline{w_i} = \frac{w_i}{w_1 + w_2} \qquad (5)$$

Within the fourth layer the learning rule outputs are calculated as follows:

$$L_{4,i} : y_i = \overline{w}_i z_i = \overline{w}_i \left(c_i x_1 + d_i x_2 + k_i x_3 + h_i \right) \qquad (6)$$

Finally, in the final *"learning layer"* all the inputs from the previous layer are aggregated producing the output of the system as a piecewise linear interpolating function, dynamically calibrated by the input-dependent normalized weights:

$$L_{5,i} : y = \sum_i y_i = \bar{w}_1 \left(c_1 x_1 + d_1 x_2 + k_1 x_3 + h_1 \right) + \bar{w}_2 \left(c_2 x_1 + d_2 x_2 + k_2 x_3 + h_2 \right) \qquad (7)$$

The last equation can be reformulated in the following matrix format:

$$y = \left[\overline{w}_1 x_1 \quad \overline{w}_1 x_2 \; \overline{w}_1 x_3 \quad \overline{w}_1 \quad \overline{w}_2 x_1 \quad \overline{w}_2 x_2 \; \overline{w}_2 x_3 \quad \overline{w}_2 \right] \cdot \left[c_1 \quad d_1 \quad k_1 \quad h_1 \quad c_2 \quad d_2 \; k_2 \quad h_2 \right]^T = \mathbf{X} \cdot \mathbf{W} \qquad (8)$$

The solution for the weight vector **W** to the above equation, if the **X** matrix was invertable and considering that the weights are known, could be the following:

$$\mathbf{W} = \mathbf{X}^{-1} \cdot \mathbf{Y} \qquad (9)$$

However, since this is not always applicable, other methods are used such as lower triangular or more robust orthogonal decompositions. In this study Singular Value Decomposition method (SVD) (Golub *et. al.,* 1971; Golub *et. al.,* 1989; Horn *et. al.,* 1991) is used. The SVD method has the advantage of using principal components to remove unimportant information related to white or heteroscedastic noise and thereby lessens the chance of overfitting. The **X** matrix is decomposed into a diagonal matrix **D** that contains the singular values, a matrix **U** of principal components, and an orthogonal normal matrix of right singular values **V**. The weight matrix is finally solved for using:

$$\mathbf{W} = \mathbf{V} \cdot \mathbf{D}^{-1} \cdot \mathbf{U}^{\mathrm{T}} \cdot \mathbf{Y} \qquad (10)$$

For the fuzzification of the input variables used by the technical trader, symmetric triangular membership functions are applied, in order to optimize the training performance in terms of computational load (Ishibuchi et. al., 1995). Additionally, because triangular functions contain two parameters, the a_i "peak" and the b_i "support" parameter, corresponding to the *"degree"* and *"range"* of the belief, are perfectly suited for modeling the agent's uncertainty perception mechanism. The membership function is as follows:

$$\mu_{M_i}\left(x_j\right)=\begin{cases}1-\dfrac{\left|x_j-a_i\right|}{b_i/2}, & \text{if } \left|x_j-a_i\right|\leq\dfrac{b_i}{2}\\ 0, & \text{else}\end{cases}\qquad(11)$$

The adaptive-expectations rule of the gradient descent algorithm for the *"degree-of-belief"* parameter is given below:

$$a_{i,t+1}=a_{i,t}-\frac{\eta_a}{p}\cdot\frac{\partial E}{\partial a_i}\qquad(12)$$

where p the training sample size and η_α the learning rate (e.g. determines the change of the a_i values and eventually the convergence of the square error function). A similar rule applies for the *"range-of-belief"* parameter. The following chain rule is used to analyze the total derivative to its partial derivatives:

$$\frac{\partial E}{\partial a_i}=\frac{\partial E}{\partial y}\cdot\frac{\partial y}{\partial y_i}\cdot\frac{\partial y_i}{\partial w_i}\cdot\frac{\partial w_i}{\partial \mu_{M_i}}\cdot\frac{\partial \mu_{M_i}}{\partial a_i}\qquad(13)$$

The partial derivatives are derived below:

$$E=\frac{1}{2}\left(y-y^o\right)^2\quad\Rightarrow\quad\frac{\partial E}{\partial y}=y-y^o=e\qquad(14)$$

$$y=\sum\nolimits_{i=1}^{n}y_i\quad\Rightarrow\quad\frac{\partial y}{\partial y_i}=1\qquad(15)$$

$$y_i=\frac{w_i}{\sum\nolimits_{i=1}^{n}w_i}z_i\quad\Rightarrow\quad\frac{\partial y_i}{\partial w_i}=\frac{\left(z_i-y\right)}{\sum\limits_{i=1}^{n}w_i}\qquad(16)$$

$$w_i=\prod_{j=1}^{m}\mu_{M_i}\left(x_j\right)\quad\Rightarrow\quad\frac{\partial w_i}{\partial\mu_{M_i}(x_j)}=\frac{w_i}{\mu_{M_i}(x_j)}\qquad(17)$$

$$\frac{\partial\mu_i}{\partial a_i}=\begin{cases}\dfrac{2\cdot\mathrm{sign}\left(x_j-a_i\right)}{b_i}, & \left|\mathrm{x_j}-a_i\right|\leq\dfrac{b_i}{2}\\ 0, & \left|\mathrm{x_j}-a_i\right|>\dfrac{b_i}{2}\end{cases}\qquad(18)$$

similarly:

$$\frac{\partial\mu_{M_i}}{\partial b_i}=\frac{1-\mu_{M_i}\left(x_j\right)}{b_i}\qquad(19)$$

Substituting into the chain rule equation:

$$\frac{\partial E}{\partial a_i} = \left(y - y^o\right) \cdot \frac{\left(z_i - y\right)}{\sum_{i=1}^{n} w_i} \cdot \frac{w_i}{\mu_{M_i}\left(x_j\right)} \cdot \frac{2 \cdot \text{sign}\left(x_j - a_i\right)}{b_i} \tag{20}$$

$$\frac{\partial E}{\partial b_i} = \left(y - y^o\right) \cdot \frac{\left(z_i - y\right)}{\sum_{i=1}^{n} w_i} \cdot \frac{w_i}{\mu_{M_i}\left(x_j\right)} \cdot \frac{1 - \mu_{M_i}\left(x_j\right)}{b_i} \tag{21}$$

The function sign(*arg*) takes the value of 1 if the argument is positive and zero otherwise. The adaptive rule for the *"degree-of-belief"* parameter is provided in the following recursive equation:

$$a_{i,t+1} = a_{i,t} - \frac{\eta_a}{p} \cdot \left[\frac{w_i}{\mu_{M_i}\left(x_j\right)} \cdot \frac{\left(z_i - y\right)}{\sum_{i=1}^{n} w_i} \cdot \frac{2 \cdot sign\ \left(x_j - a_i\right)}{b_i} \cdot \left(y - y^o\right) \right] \tag{22}$$

whereas for the *"range-of-belief"* parameter:

$$b_{i,t+1} = b_{i,t} - \frac{\eta_b}{p} \cdot \left[\frac{w_i}{\mu_{M_i}\left(x_j\right)} \cdot \frac{\left(z_i - y\right)}{\sum_{i=1}^{n} w_i} \cdot \frac{1 - \mu_{M_i}\left(x_j\right)}{b_i} \cdot \left(y - y^o\right) \right] \tag{23}$$

To sum up, the hybrid learning rule consists of two passes, in a way simulating the decision process of the trader. In each estimation step, during the forward pass the polynomial parameters and weights (*"adaptive expectations"* and *"preferences"*) are calculated using SVD method, while the membership parameters (*"degree"* and *"range"* of beliefs) remain fixed. Thereafter, the outputs are produced using the previously calculated polynomial parameters and in the reverse pass the errors are backpropagated within the learning layers to determine the membership parameter updates, while the polynomial parameters are kept fixed.

5. Other Forecasting models

The Neurofuzzy model is compared against a Markov-Switching model and a Recurrent Neural Network in order to examine its relative predictability and profitability performance. These models are described below.

Markov – Switching model

A well-established class of Regime-switching models assumes that the regime that occurs at time t cannot be observed, and is determined by an unobservable process, denoted as s_t. In case of only two regimes (e.g. "low" and "high") s_t can simply be assumed to take on 2 values, such that an $AR(1)$ model in both regimes is given by

$$y_t = \begin{cases} \varphi_{0,1} + \varphi_{1,1} y_{t-1} + \varepsilon_t & \text{if } s_t = 1, \\ \varphi_{0,2} + \varphi_{1,2} y_{t-1} + e_t & \text{if } s_t = 2, \end{cases} \tag{24}$$

The most popular model in this class, which was advocated by Hamilton (1989), is the Markov-Switching model, in which the process s_t is assumed to be a 1st order Markov-process (i.e. the current regime s_t only depends on the regime one period ago, $s_t - 1$). The model is completed by defining the transition probabilities of moving from one state to the other:

$$\begin{matrix} P(s_t = 1 | s_{t-1} = 1) = p_{11} \\ P(s_t = 2 | s_{t-1} = 1) = p_{12} \end{matrix} \text{ and } \begin{matrix} P(s_t = 1 | s_{t-1} = 2) = p_{21} \\ P(s_t = 2 | s_{t-1} = 2) = p_{22} \end{matrix} \tag{25}$$

where p_{ij} is equal to the probability that the Markov chain moves from state i at time $t-1$ to state j at time t. In order to define proper probabilities, they should be nonnegative, and it should also hold that $p_{11} + p_{12} = 1$ and $p_{21} + p_{22} = 1$. Additionally, in the Markov-Switching models the unconditional probabilities $P(s_t = i)$ for $i = 1, 2$ that the process is in each of the regimes are defined. Via the theory of ergodic Markov chains it is straightforward to show that for the 2-state MSW model these unconditional probabilities are given by Hamilton (1994):

$$P(s_t = 1) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}, \qquad P(s_t = 2) = \frac{1 - p_{11}}{2 - p_{11} - p_{22}} \tag{26}$$

Recurrent Neural Network

A single hidden layer feedforward network with sufficiently hidden units and properly adjusted parameters can theoretically approximate any function to any desired degree of accuracy[2]. The output of a neural network is produced via the application of a

[2] Despite the importance of selecting the optimum number of hidden neurons, there is no explicit formula for that matter. The geometric pyramid rule proposed by Masters (1993) considers $\sqrt{n \cdot m}$ neurons for a three-layer network with n inputs and m outputs. Katz (1992) indicates that an optimal number of hidden neurons can be found between one-half to three times the

transfer function. The functionality is to modulate the output space as well as prevent outputs from reaching very large values which can "block" training[3]. Learning typically occurs through training, where the training algorithm iteratively adjusts the connection weights. Common practice is to divide the sample into three distinct sets called the training, validation and testing (out-of-sample) sets; the training set is the largest and is used by the neural network to learn the patterns presented in the data, the validation set is used to evaluate the generalization ability in order to avoid overfitting and the training set should consist of the most recent observations that are processed for testing predictability[4]. Specifically, if $\mathbf{X_t} = (x_{1,t}, \ldots, x_{p,t})$ is the input of a single layer feedforward network with q hidden units, the output is given by:

$$y_t = S\left[\beta_0 + \sum_{i=1}^{q} \beta_i G\left(\alpha_{i0} + \sum_{j=1}^{p} \alpha_{ij} x_{j,t}\right)\right] = f\left(\mathbf{x_t}, \mathbf{z}\right) \qquad (27)$$

where $i=1,\ldots,q$ and $j=1,\ldots,p$. Consider $\mathbf{z} = \left(\beta_0, \ldots, \beta_q,\ \alpha_{11}, \ldots \alpha_{ij} \ldots, \alpha_{qp}\right)^T$ as the weight vector and *S, G* transfer functions. The solution of the network considers estimation of the unknown vector **z** with a sample of data values. A recursive estimation methodology, which is called backpropagation is used to estimate the weight vector, as follows:

$$z_{t+1} = z_t + \eta \nabla f(\mathbf{x_t}, z_t) \cdot \left[y_t - f(\mathbf{x_t}, z_t)\right] \qquad (28)$$

where $\nabla f(\mathbf{x_t}, z)$ is the gradient vector with respect to z and η the learning rate. The learning rate controls the size of the change of the weight vector on the *t-th* iteration. The z vector update is achieved via the minimization of the mean square error function.

Whilst feedforward neural networks appear to have no memory since the output at any time instant depends entirely on the inputs and the weights at that instant, Recurrent neural networks exhibit characteristics simulating short-term memory. In this study, Elman Recurrent neural networks (Elman, 1990) have been utilized. In Elman networks with a single hidden layer the lagged outputs of the hidden neurons are fed back into the hidden

number of inputs, whereas Ersoy (1990) proposes doubling the number of neurons until the network's RMSE performance deteriorates.

[3] Levich et. al. (1993) and Kao et. al. (1992) found that hyperbolic sigmoid and tansigmoid transfer functions are appropriate for financial markets data because they are nonlinear and continuously differentiable which are desirable properties for network learning

[4] The validation error starts decreasing until the network begins to overfit the data and the error will then begins to rise. The weights are calculated at the minimum value of the validation error.

neurons themselves. If $\mathbf{X_t} = (x_{1,t},...,x_{p,t})$ is the input with q hidden units and t the time index, the output of the network is given by:

$$y_t = G\left[\beta_0 + \sum_{i=1}^{q}\beta_i \cdot g_{i,t}\right] + \varepsilon_t \qquad (29)$$

where

$$g_{i,t} = G\left(\alpha_{i0} + \sum_{j=1}^{p}\alpha_{ij}x_{j,t} + \sum_{h=1}^{q}\delta_{ih}g_{h,\,t-1}\right) \qquad (30)$$

and $\mathbf{z} = \left(\beta_0,...,\beta_q,\ \alpha_{11},...\alpha_{ij}...,\alpha_{qp},\delta_{11},...\delta_{ih}...,\delta_{qq}\right)^T$ the weight vector and G the hyperbolic tangent sigmoid transfer function.

6. Volatility modeling

The basic assumption for the estimation of conditional variance is that $(r_t)_{t=1}^{n}$ which represents daily returns of the price of a financial asset, follows the stochastic process:

$$r_t = \mu_t + e_t = \mu_t + \sigma_t\varepsilon_t \qquad (31)$$

where $\mu_t = E(r_t / F_{t-1}), \sigma_t^2 = E(e_t^2 / F_{t-1})$ and $\varepsilon_t = (e_t / \sigma_t)$ has a conditional distribution function $F_t(\cdot)$. Under the parametric approach specific distributions for $F_t(\cdot)$ are considered such as the Gaussian *N(0,1)*, the Student-*t* or the Generalized Error Distribution *(GED)*. The conditional variance can be estimated by various models such as moving average (Alexander, 1998), given by the equation:

$$\sigma_t^2 = \frac{1}{m-1}\cdot\sum_{j=1}^{m}(r_{t-j} - \mu_t^m)^2 \qquad (32)$$

where $\mu_t^m = \frac{1}{m}\cdot\sum_{j=1}^{m}r_{t-j}$, the moving average of *m* trading days. Alternatively it can be estimated by one of the family of *GARCH* models (Hentschel, 1995). In particular, the GARCH(1,1) model is given by:

$$\sigma_t^2 = \alpha_0 + \alpha_1(r_{t-1} - \mu_t)^2 + \beta\sigma_{t-1}^2 \qquad (33)$$

where $\mu_t = (1/T)\sum_{i=1}^{T} r_{t-i}$. A special case of GARCH models is the Exponentially Weighted Moving Average *(EWMA)* specification, adopted by the *Riskmetrics (RM)* model of J.P.Morgan, under which:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)(r_{t-1} - \mu_t)^2 \qquad (34)$$

Riskmetrics has chosen λ=0.94 and λ=0.97 as the optimal decay factor for daily and monthly data respectively (Jorion, 2000).

7. Empirical results

Let P_t be the daily index price. The daily returns are then calculated by $r_t = \log(P_t) - \log(P_{t-1})$. The performance of the models is examined using logarithmic returns of the most prominent indices of U.S.A, Southeast Asia and Europe, namely NASDAQ Composite, NIKKEI255 and FTSE100 indices, from 2/8/1971 to 2/5/2002 (8087 observations) in case of NASDAQ and NIKKEI255, whereas from 1/2/1984 to 2/5/2002 (4722 observations) for FTSE100. The predictive performance of the models is examined in the period 4/8/1998 – 2/5/2002 (1000 observations) which has been reserved for out-of-sample testing purposes. This sample contains diverse regimes and several "problematical" events including the Asian crisis and the rise and fall of the *tech-market bubble*, which makes the analysis particularly interesting and applicable to other turbulent periods (e.g. the current financial crisis of 2007-2008 which leads to global recession and was caused by the credit insolvency of investment institutions and high oil prices).

Specifically, the inputs (x_j) of the Neurofuzzy model correspond to the returns (r_t) of the previous p days and the volatility daily changes $\delta\sigma = \sigma_t - \sigma_{t-1}$ while the output (y) is the forecasted one-day-ahead return ($\hat{r}_t$). The conditional volatility daily changes were calculated for the same periods via 20-day Moving Average, RiskMetrics (with 0.94 decay factor) and GARCH(1,1) models[5]. The inputs in the Recurrent neural network correspond to the daily returns over the previous p days, following Gençay (1998a, b) and Fernadez-Rodriguez et. al. (2000), while for the Markov-Switching model also p lagged daily returns are used as well as two beliefs (i.e. "low" and "high") in order to directly correspond to the membership function architecture of the Neurofuzzy model. For the test period the models utilize a moving window of all previous observations as a training sample and produce forecasts for each day within the corresponding period. The validation sample for each case

[5] The exponentially moving average corresponds to the approach adopted by *RiskMetrics* and for that reason it is denoted here as RM (0.94).

is the 30% of the training set, and is used to evaluate the generalization ability and avoid overfitting. The training set consists of the most recent observations that are processed in each case. The training and validation samples utilize a moving window of all previous observations in order to produce forecasts for each day within each backtesting period.

The volatility-based Neurofuzzy model (symbolized as VNF) corresponds to a specification where two lags of the returns and one lag of the conditional volatility changes appear in Eq.3 (j=3)[6]. The corresponding notation with the embedded volatility measure is VNF-MA(20), VNF-RM(0.94) and VNF-GARCH(1,1). In case of the Recurrent neural network (RNN) the best forecasting ability was derived empirically by a topology which incorporated 10 neurons (g) in the hidden layer and an output layer with a single neuron (y)[7]. The Markov-Switching model (MSW) model uses two lags of the returns corresponding to an AR(2) model in both regimes.

In order to account for the use of nonlinear models a test for the presence of non-linear dependence in the series is conducted. To that end, the well-known BDS test statistic was used, which under the null of i.i.d. is given by (Brock et.al., 1991):

$$W_{m,T}(\varepsilon) = T^{1/2}[C_{m,T}(\varepsilon) - C_{1,T}^{m}(\varepsilon)] \,/\, \sigma_{m,T}(\varepsilon) \qquad (35)$$

$C_{m,T}(\varepsilon)$ is the correlation integral from m dimensional vectors that are within a distance ε from each other, when the total sample is T, and $\sigma_{m,T}(\varepsilon)$ is the standard deviation of $C_{m,T}(\varepsilon)$. Under the null hypothesis, $W_{m,T}(\varepsilon)$, has a limiting standard normal distribution. The BDS test has been applied on: (a) the original data, (b) the residuals from an autoregressive filter AR(2) (based on the selected return lags), in order to ensure that the null is not rejected due to linear dependence, and (c) the natural logarithm of the squared standardized residuals from a AR(2) - GARCH-M (1,1) model in order to ensure that rejection of the null is not due to conditional heteroscedasticity (De Lima, 1996).

[6] The procedure for the selection of the lags involved the the calculation of the Ljung-Box statistics for the first 10 lags of the series. Significant autocorrelations of up to the second lag of the return series were identified. Additionally, the Akaike and Schwarz Information Criteria (AIC, SIC) that were estimated for the first six lags provided the minimum value at the second lag. As concerns the conditional volatility variable, sensitivity and RMSE analyses for different number of lags were conducted on the VNF but the results were not found to be qualitatively different from those presented in Table 2. Similar exercises were conducted for a different number of lagged returns but again the results are not better than those shown in Table 2.

[7] This empirical result follows Katz (1992) and Ersoy (1990).

Insert Table 1 here

In all three cases the null of i.i.d. at the 1% marginal significance level could be rejected and the evidence seemed to suggest that a genuine non-linear dependence is present in the data.

The trading rule works as follows; at the end of each trading day the models are being re-estimated over a rolling sample with a length equal to the training period. When the output of a model is greater than 0 this is used as a buy signal and a value less than 0 as a sell signal. The total return, when transaction costs are not considered, is estimated as:

$$R = \sum_{n+1}^{n+T+1} s \cdot r_t \qquad (36)$$

where T indicates the out-of-sample horizon, r_t is the realized return and $\hat{s}_t$ is the recommended position which takes the value of (-1) for a short and (+1) for a long position (e.g. Gençay 1998b, Jasic et. al. 2004). In order to evaluate the forecast accuracy of the models, the percentage of correct predictions or correctly predicted signs was calculated as follows:

$$\text{Sign Rate} = \frac{h}{T} \qquad (37)$$

where h is the number of correct predictions. Two other comparative profitability measures were also considered: the Ideal Profit (IP) and the Sharpe ratio (SR). The IP compares the forecasting system return against the perfect forecaster and is calculated by:

$$R = \frac{\sum_{t=n+1}^{n+T+1} s_t r_t}{\sum_{t=n+1}^{n+T+1} \left| r_t \right|} \qquad (38)$$

where the value $R = 0$ is considered as a benchmark to evaluate the performance of a trading strategy. When the direction indicator s_t takes the correct position for all observations in the sample, then $R = 1$, whereas if all forecasted positions are wrong, then the value is $R = -1$. The SR is the proportion of the mean return of the trading strategy over its standard deviation. The higher the SR is the higher the return and the lower the volatility:

$$SR = \frac{\mu_{R_T}}{\sigma_{R_T}} \qquad (39)$$

Finally, as a measure of the predictability the Henriksson-Merton (HM) statistic (Henriksson and Merton, 1991) was employed for the r_t (realized) and $\hat{r}_t$ (forecasted) returns. The statistic is based on the following contingency table:

$$\begin{array}{cc} & \begin{array}{cc} r_t > 0 & \qquad r_t \leq 0 \end{array} \\ \begin{array}{c} \hat{r}_t > 0 \\ \hat{r}_t \leq 0 \end{array} & \begin{bmatrix} n_1 & n \\ N_1 - n_1 & N_2 - n_2 \end{bmatrix} \end{array} \tag{40}$$

where n_1 is the number of correct forecasts in "up" markets, n_2 is the number of incorrect forecasts in "down" markets and N_1 , N_2 the number of up-market and down-market periods, respectively in the sample. Henriksson and Merton showed that n_1 has a hypergeometric distribution under the null hypothesis of no market–timing ability, which may be approximated by:

$$n_1 \sim N\left(\frac{nN_1}{N}, \quad \frac{n_1 N_1 N_2 (N-n)}{N^2(N-1)}\right) \tag{41}$$

where $N = N_1 + N_2$ and $n = n_1 + n_2$.

The empirical results of the comparative implementation of all models are reported in Table 2.

Insert Table 2 here

Considering total returns, a trading rule with any of the VNF models dominates the RNN and the MSW as well as the Buy & Hold (B&H) strategy consistently for all indices. Specifically, in case of NASDAQ the total return for the trading strategy based on the VNF-RM(0.94) model (173.6%) outperforms impressively RNN (29.2%), MSW (40.2%) and B&H strategy (4.5%) as well as the other volatility –based NF models (83.3% for VNF-MA(20) and 108.9% for VNF-GARCH(1,1)). The same applies with the inclusion of transaction costs, which are estimated as 0.05% for each one-way trade, following Hsu and Kuan (2005) and Fama and Blume (1966). Again, the trading rule remains significantly profitable (e.g. 150.4% return for VNF-RM(0.94) model). The fact that VNF-RM(0.94) model outperforms RNN, MSW (40.2%) and B&H strategy is also depicted in the proportion of correctly predicted signs (55.3%) which is higher compared to the aforementioned models. The HM test provides a further validation of the statistical significance of the sign rate with a 4.692 value at the one-sided 1% level. Additionally, the SR (annualized) and the IP are much higher than for RNN and MSW models and relatively higher that for the other two VNF models. In case of FTSE100 the VNF-MA(20) model provides the best results (102.6% or 78.7% after commissions and 52.8% sign rate) compared to other VNF models but again with an indicating superiority relatively to the RNN and MSW models (i.e. a total loss of 6.4% for the RNN model and a gain of 31.4% for the RNN model). The statistical significance (HM test) of the optimal VNF predictor (3.772), the SR (1.297) and the IP (0.108) scores confirm the above results. Finally, in the NIKKEI255 case the VNF-GARCH(1,1) model provided the best out-of-sample performance with a total return of 87.5% (57.4% after costs), sign rate of 50.1% (statistically significant at the 1% level), SR of 0.901 and IP of 0.078. It is noticeable that the B&H strategy in the examined period for NIKKEI255 produced a 53.2% loss. The significant profitability of the VNF-GARCH(1,1) model may be compromised with the marginal improvement of the sign rate (barely over 50%), yet it is due to the substantial improvement of the quantitative importance of the correctly forecasted signs. In terms of RMSE the results for the VNF models for all indices are consistently lower compared to the RNN and relatively the same for the MSW model, with one exception for the VNF-GARCH(1,1) model in case of NASDAQ index. The overall return of VNF models compared to the B&H policy is always superior for all indices. The fact that all models outperform B&H strategy accords with previous results derived by Fernández et. al. (2000) as well as with the conclusions reached by Christoffersen and Diebold (2003).

The comparison between the three different specifications for the volatility-based NF models show that simple methods of historical volatility measurement, like the equally weighted and the exponentially weighted moving averages, can produce sign forecasts that are no worse than those obtained from more complicated econometric models that are often used to model conditional volatility. This has not been surprising since it is documented that forecasts of volatility, e.g for the NASDAQ index from MA rules (Schwert, 2002) closely approximate those from GARCH (1,1) models. Simon (2003) also reports that the Glosten-Jagannathan-Runkle (1993) GARCH volatility forecasts of the NASDAQ 100 average 3.0 percentage points higher that the actual when the EWMA volatility forecasts are only 1.5 percentage points below actual volatility.

Overall, the predictive ability of the VNF models is significantly higher compared to the other models. A possible explanation is that a B&H strategy would be the best for the stock indices in the extreme case with no turning points in the testing period. However, when there are many turning points during a period and the more turning points occur, the better the VNF model will be in prediction performance. Technically, in terms of sign prediction, the proposed model acting as a dynamically-adjusted piecewise linear interpolator compared to the static nonlinear neural predictor or the probabilistic regime-switching Markov model, leads to more precise identification of turning points. Model-wise, it is the dynamic update of the of "*expectations*" and "*preferences*" (polynomial parameters) of the *fuzzy learning rules* indicating the efficient decision-making mechanism of incorporating the heterogeneous beliefs ("fads") of the trader under uncertainty, and the adaptive "calibration" of the "*degree- and range-of-belief*" membership functions to match the agent's expectations of "low" and "high" regime, that leads to more precise and prompt identification of market turning points and optimal prediction.

8. Conclusions

The present paper expands the literature on heterogeneous learning rules in speculative markets where heuristic strategies reflect the rules-of-thumb of boundedly rational investors, by introducing a trading system with the incorporation of beliefs, preferences and idiosyncratic behavioral patterns represented by *fuzzy learning rules*. Moreover, beyond the existing practice that has utilized separately neural networks or fuzzy logic systems in financial forecasting applications, it presents a volatility-based hybrid Neurofuzzy model that leads to superior predictions upon the direction-of-change of the market.

The results suggest that with the inclusion of transaction costs, the performance of the proposed Neurofuzzy models (VNF) in case of returns of NASDAQ Composite, FTSE100 and NIKKEI255, is consistently superior to the Markov-switching models, Recurrent Neural Networks as well as the Buy & Hold strategy for all indices. The daily volatility changes of the conditional volatility which are embedded in the Neurofuzzy model and were produced from alternative estimating techniques generated a substantial improvement of the profitability per unit of risk over the investigated market period, as it provided valid information for a potential turning point on the next trading day. Specifically, these results seem to indicate that the volatility-based Neurofuzzy model has been optimally "trained" to correctly relate changes in conditional volatility with the "sign" of the market one day ahead. This may be attributed to a number of reasons. The first associates increases in volatility to higher expected returns. In the case when increases in volatility are generated from "bad" news lower prices will occur the next trading day. However, when increases in volatility are generated from "good" news it is not clear what the net effect on prices will be. The second reason associates increases in volatility with trigger strategies followed by many portfolio managers. Every time volatility rises, the risk limit is being hit for some portfolios and then liquidation follows, thus putting a pressure on the market. The results are in broad accordance to the conclusions reached from a "statistical" perspective according to which there is a close relationship between asset return signs and asset return volatilities.

Overall what leads to higher predictability is the dynamic update of the "*expectations*" and "*preferences*" of the heuristic learning rules combined with the adaptive calibration of the "*degrees-of-belief*" that match agent's "fads".

References

Adya, M. and Collopy, F., 1998, "How effective are neural networks at forecasting and prediction? A review and evaluation", J. Forecasting, 17, 481-495.

Alexander C., 1998, "Volatility and Correlation: Measurement Methods and Applications", Risk Management and Analysis, 1, John Wiley, 125-168.

Al-Shammari, M. and Shaout, A., 1998, "Fuzzy logic modeling for performance appraisal systems: A framework for empirical evaluation", Expert Systems with Applications, 14, 323-238.

Altrock, C., 1997, "Fuzzy Logic and Neurofuzzy Applications in Business and Finance", Prentice Hall, Upper Saddle River, New Jersey.

Bekaert, G. and G. Wu, 2000, Asymmetric Volatility and Risk in Equity Markets, The Review of Financial Studies 13, No 1, 1-42.

Black, F., 1986, Noise, Journal of Finance 41, 529-543.

Bollerslev, T., 1986, "Generalized Autoregressive Conditional Heteroskedasticity", Journal of Econometrics, 31, 307-327.

Box, G. and Jenkins, G., 1970, "Time Series Analysis: Forecasting and Control", Holden-day, Inc., San Francisco, CA.

Brock, W., D. Hsieh and B. LeBaron, 1991, Nonlinear Dynamics, Chaos and Instability, The MIT Press.

Brock, W.A., Hommes, C.H., 1998, Heterogeneous Beliefs and Routes to Chaos in a Simple Asset Pricing Model. Journal of Economic Dynamics and Control 22.

Buckley, J. J. and Hayashi, Y., 1994 "Fuzzy Neural Networks: A Survey", Fuzzy Sets and Systems, 66, 1-13.

Christie,A. A., 1982, The Stochastic Behavior of Common Stock Variances-Value, Leverage and Interest Rate Effects, Journal of Financial Economics 10, 407-432.

Christoffersen, P., and F.Diebold, 2003, Financial Asset Returns, Direction-of-change Forecasting, and Volatility Dynamics, working paper 10009, NBER.

Cybenko, G., 1989, "Approximation by Superposition of a Sigmoidal Function", Mathematics of Control, Signals and Systems 2, 303 -314.

De Lima, P.J.F., 1996, Nuisance parameter free properties of correlation integral based statistics, Econometric Reviews 15, 237-259.

Elman, J. L. 1990, Finding structure in time, Cognitive Science 14, 179-211.

Ersoy O., 1990. Tutorial at Hawaii International Conference on Systems Sciences

Fama, E. F., 1970, Efficient Capital Markets: A Review of Empirical Work. Journal of Finance. 25, pp. 383-417.

Fama, E. F., 1991, Efficient Capital Markets II. Journal of Finance 46(5):1575-1617.

Fama, E. F. and French, K. R., 1995, Size and book-to-market factors in earnings and returns. Journal of Finance 50, 131-155.

Fama, E.F., and M.E. Blume, 1966, “Filter rules and Stock-Market Trading”, Journal of Business, 39, 226-241

Fernández-Rodriguez, F., C. Gonzalez-Martel and S. Sosvilla-Rivero, 2000, On the Profitability of Technical Trading Rules based on Artificial Neural Networks: Evidence from the Madrid Stock Market, Economic Letters 69, 89-94.

Funahashi, K., 1989, "On the Approximate Realization of Continuous Mappings by Neural Networks", Neural Networks 2, 183 -192.

Gallant, A.R. and White, H., 1988, "There exists a neural network that does not make avoidable mistakes", Proceedings of the Second Annual IEEE Conference on Neural Networks, San Diego, CA. IEEE Press, New York, I, 657-664.

Gallant, A.R. and White, H., 1992, "On learning the derivatives of an unknown mapping with multiplayer feedforward networks", Neural Networks 5, 129-138.

Gençay, R., 1998a, The Predictability of Security Returns with Simple Technical Trading Rules, Journal of Empirical Finance 5, 347-359.

__________, 1998b, Optimization of Technical Strategies and the Profitability in Security Markets, Economic Letters 59, 249-254.

Giot, P., 2005, Relationships Between Implied Volatility Indexes and Stock Index Returns, The Journal of Portfolio Management, 92-100.

Glosten, L., R. Jaganathan and D. Runkle, 1993, On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, The Journal of Finance 48, 1779-1801.

Golub, G. H., and Reinsch C. (1971). Singular value decomposition and least squares solutions: Handbook for automatic computation Vol II: Linear algebra, Heidelberg, Springer.

Golub, G. H., and Charles F. Van Loan (1989). Matrix Computations, The Johns Hopkins University Press. 2nd Edition.

Green, H., Pearson, M., 1994, "Neural nets for foreign exchange trading, in: Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets", Wiley, New York.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series subject to changes in regime, Econometrica 57, 357-84

__________, 1994. Time Series Analysis, Princeton: Princeton University Press

Harvey, A.C., 1989, "Forecasting Structural Time Series Models and the Kalman Filter", Cambridge University Press, Cambridge.

Hagan, M. T., and M. Menhaj, 1994, “Training feedforward networks with the Marquardt algorithm,” IEEE Transactions on Neural Networks, vol. 5, no. 6, pp. 989–993.

Hecht-Nielsen, R., 1989, "Theory of the backpropagation neural networks", Proceedings of the International Joint Conference on Neural Networks, Washington, DC, IEEE Press, New York, I, 593-605.

Henriksson, R.D. and R.C. Merton, 1981. On the Market timing and Investment Performance II: Statistical procedures for evaluating forecasting skills, Journal of Business 54, 513-533.

Hentschel L., 1995, "All in the Family: Nesting Symmetric and Asymmetric GARCH Models", Journal of Financial Economics, 39, 71-104.

Hommes, C.H., 2001, “Financial markets as complex adaptive evolutionary systems”, Quantitative Finance 1, 149-167.

Hommes, C.H, 2005, Heterogeneous Agents Models: two simple examples, In: Lines, M. (Ed.), Nonlinear Dynamical Systems in Economics, CISM Courses and Lectures no. 476, Springer-Verlag: Vienna-New York, 131-164.

Hommes, C.H., 2006, Heterogeneous Agent Models in Economics and Finance, In: L. Tesfatsion, K.L. Judd (eds.), Handbook of Computational Economics, Volume 2: Agent- Based Computational Economics, Elsevier Science B.V, 1109-1186.

Horn, R. A., and Johnson, C. R. (1991). *Topics in Matrix Analysis*, Cambridge University Press.

Hornik, K., 1991, "Approximation Capabilities of Multilayer Feedforward Net- works", Neural Networks, 4, 251 -257.

Hornik, K., Stinchcombe, M., and White, H., 1989, "Multilayer Feedforward Networks are Universal Approximators", Neural Networks 2, 359 -366.

____________________,1990, "Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks", Neural Networks, 3, 551-560.

Hsu P-H., Kuan C-M., 2005, “Reexamining the Profitability of Technical Analysis with Data Snooping Checks”, Journal of Financial Econometrics, 3, 606-628

Ishibuchi, H., Kwon, K. and Tanaka, H., 1995, "A learning algorithm of fuzzy neural networks with triangular fuzzy weights", Fuzzy Sets and Systems, 71, 277-293.

Jamshidi, M., Titli, A., Zadeh, L. and Boverie, S., 1997, edts, "Applications of Fuzzy Logic", Prentice Hall, New Jersey.

Jasic, T. and Wood, D. (2004) The profitability of daily stock market indices trades based on neural network predictions: case study for the S&P 500, the DAX, the TOPIX, and the FTSE in the period 1965-1999, Applied Financial Economics, 14, 285-297

Jorion P., 2000, "Value at Risk", 2ed., McGraw-Hill, New York.

Kaneko, Takaomi, 1996, "Building a financial diagnosis system based on fuzzy logic production system", Computers Industrial Engineering, 31, No.3/4, 743-746.

Kao G.W. and Ma C.K., 1992, "Memories, heteroscedasticity and prices limit in currency futures markets", Journal of Futures Markets, 12, 672-692.

Katz J.O., 1992, "Developing neural network forecasters for trading", Technical Analysis of Stocks and Commodities, 58-70.

Klir, G. J. and Yuan, B., 1995, "Fuzzy Sets and Fuzzy Logic: theory and applications", Prentice Hall, Upper Saddle River.

Kosko, B., 1992, "Neural Networks and Fuzzy Systems", Prentice-Hall, Englewood Cliffs.

Krugman, P., 1987, Trigger Strategies and Price Dynamics in Equity and Foreign Exchange Markets, NBER Working Paper No 2459.

Kuan, C.-M. and White, H., 1994, "Artificial neural networks: An econometric perspective", Econometric Reviews 13, 1-91.

La Porta R., Lakonishok, J., Shliefer, A. and R. Vishny, 1997, Good news for value stocks: Further evidence on market efficiency. Journal of Finance 52, 859-874.

Levich R.M. and Thomas L.R., 1993, "The significance of technical trading rule profits in the foreign exchange market: A bootstrap approach; in Strategic Currency Investing - Trading and Hedging in the Foreign Exchange Market", Probus, Chicago, 336-365.

Lin, C.-T and Lee, C. G., 1996, "Neural Fuzzy Systems. A Neuro-Fuzzy Synergism to Intelligent Systems", Prentice Hall, New York.

Mamdami, E., 1977 "Application of Fuzzy Logic to approximate reasoning using linguistic systems", IEEE Transactions on Computers, 26(12), 1182-1191.

____________, 1974, "Applications of fuzzy algorithms for control of simple dynamic plant", Proceedings IEE, 121, 1585-1588.

Manger, R., 1994, "Using holographic neural networks for currency exchange rates prediction", Proceedings of the 16th International Conference on Information Technology Interface, Pula, Croatia.

Nishina, T. and Hagiwara, M., 1997, "Fuzzy Inference Neural Network", Neurocomputing, 14, 223-239.

Poddig, A., 1993, "Short term forecasting of the USD/DM exchange rate", Proceedings of First International Workshop on Neural Networks in Capital Markets, London.

Rawani, A., 1993, "Forecasting and trading strategy for foreign exchange market", Information Decision Technology, 1, 19.

Refenes, A.N., Zapranis, A. and Francis, G., 1994, "Stock performance modeling using neural networks: a comparative study with regression models", Neural Network, 5, 961-970.

Schwert, W., 2002. Stock Volatility in the new millennium: how wacky is Nasdaq?, Journal of Monetary Economics 49, 3-26.

Shiller, R. J., 1987, Investor behavior in the October 1987 stock market crash: survey evidence, NBER working paper No. 2446, November 1987, published in: Shiller, R.J., Market Volatility, MIT Press, Cambridge, 1989, chapter 23.

Shiller, R. J., 2002, From Efficient Market Theory to Behavioral Finance. Cowles Foundation Discussion Paper No. 1385.

Shleifer, A. and Summers, L.H., 1990, The noise trader approach to finance, Journal of Economic Perspectives 4 (2), 19-33.

Simon, D., 2003, The Nasdaq Volatility Index During and After the Bubble, The Journal of Derivatives, 9-23.

Simon, H.A., 1957, Models of man, Wiley, New York, NY.

Sugeno, M., 1985, "Industrial applications of fuzzy control", Elsevier Science Publications Co.

__________, 1988, "Fuzzy Control", Nikkan Kougyou, Shinbunsha.

Weigend, A.S., 1991, "Generalization by weight-elimination applied to currency exchange rate prediction", Proceedings of the IEEE International Joint Conference on Neural Networks, Singapore.

Whaley, Robert, 2000, The Investor fear gauge, The Journal of Portfolio Management 26, 12-27.

White, H., 1989, "Learning in artificial neural networks: A statistical perspective", Neural Computing, 1, 425-464.

Yao, J.T., Poh, H.-L. and Jasic, T., 1996, "Foreign exchange rates forecasting with neural networks", Proceedings of the International Conference on Neural Information Processing, Hong Kong, 754-759.

Zhang, X.R., 1994, "Non-linear predictive models for intra-day foreign exchange trading", International. Journal Intelligent Systems Accounting and Finance Management, 3 (4), 293-302.

Zimmerman, H.-J. and Thole, U., 1978, "On the Suitability of Minimum and Product Operators for the Intersection of Fuzzy Sets", Fuzzy Sets and Systems, 2, 173-186.

TABLE 1: BDS test

Index	Correlation dim.	*m*=2		*m* =3		*m* =4	
	Dim. distance	*ε=1*	*ε=1.5*	*ε=1*	*ε=1.5*	*ε=1*	*ε=1.5*
NASDAQ Composite	Raw data	37.671	36.145	45.860	44.472	51.296	48.581
	AFR	34.736	34.449	42.998	43.019	48.344	47.151
	NLSSR	13.433	10.892	15.011	12.514	15.114	12.638
FTSE100	Raw data	9.930	11.916	13.000	15.373	15.186	17.434
	AFR	9.785	11.871	12.665	15.159	14.817	17.216
	NLSSR	5.517	5.338	5.108	4.917	4.348	4.157
NIKKEI255	Raw data	24.704	23.491	33.338	30.759	39.566	35.119
	AFR	24.513	23.237	33.289	30.615	39.568	35.021
	NLSSR	11.618	10.363	13.657	12.042	13.983	12.035

Notes

- Raw data = daily index returns, AFR = residuals from an autoregressive filter AR(2), NLSSR = natural logarithm of the squared standardized residuals from AR(2) - GARCH-M (1,1) model.
- m = dimension, ε = number of standard deviations of the data.
- Significance at the 1% level corresponds to the critical value 2.58.

TABLE 2: Out-of-sample performance of the trading models

Index	*Model*	Total Return (%)	B&H Return (%)	Sign Rate	HM test	RMSE	Sharpe Ratio (ann.)	Ideal Profit
NASDAQ Composite	**VNF-RM(0.94)**	**173.6** (150.4)	4.5	0.553	4.692*	0.025	1.170	0.098
	VNF-MA(20)	**83.3** (59.1)		0.527	2.401**	0.024	0.553	0.047
	VNF-GARCH(1,1)	**108.9** (86.1)		0.540	3.364*	0.039	0.727	0.061
	RNN	**29.2** (3.2)		0.525	1.993**	0.029	0.190	0.016
	MSW	**40.2** (17.8)		0.536	2.920**	0.024	0.268	0.022
FTSE100	**VNF-RM(0.94)**	**52.9** (29.1)	-17.3	0.518	3.124*	0.012	0.664	0.056
	VNF-MA(20)	**102.6** (78.7)		0.528	3.772*	0.013	1.297	0.108
	VNF-GARCH(1,1)	**47.3** (25.3)		0.505	1.436***	0.013	0.601	0.049
	RNN	**-6.4** (-32.0)		0.477	-0.188	0.015	-0.079	-0.007
	MSW	**31.4** (5.6)		0.505	1.591***	0.013	0.395	0.033
NIKKEI255	**VNF-RM(0.94)**	**36.2** (9.5)	-53.2	0.483	1.324***	0.015	0.364	0.033
	VNF-MA(20)	**26.4** (2.3)		0.485	2.344**	0.015	0.274	0.018
	VNF-GARCH(1,1)	**87.5** (57.4)		0.504	3.582*	0.015	0.901	0.078
	RNN	**5.7** (-19.4)		0.462	-1.217	0.019	0.063	0.005
	MSW	**23.5** (-5.0)		0.482	1.191***	0.015	0.237	0.021

Notes

- VNF=Volatility-based Neurofuzzy Model. RNN = Recurrent Neural Network. MSW=Markov-Switching Model
- MA(20) = Moving Average with a 20 days window, RM(0.94) = RiskMetrics' exponentially weighted MA rule (decay factor = 0.94), GARCH(1,1)= Bollerlsev GARCH model
  HT test = Henriksson and Merton (1981) test, asymptotically distributed as *N(0,1)*.
- In parenthesis Total Return after transaction costs (0.05% average fixed cost for each one-way trade)
- The sign rate measures the proportion of correctly predicted signs. The Sharpe ratio is defined as the ratio of the mean return of the strategy over its standard deviation (it has been annualized by multiplying it with the squared root of 250). The Ideal Profit is the ratio of the returns of the trading strategy over the returns of a perfect predictor.
- (*), (**), (***) indicate significance at the one sided 1%, 5% and 10% levels.