Journal of Empirical Finance 30 (2015) 34–49

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Heuristic learning in intraday trading under uncertainty☆

Stelios D. Bekiros [1,2,3]

IPAG Business School, 184 Boulevard Saint-Germain, 75006 Paris, France
Department of Economics, European University Institute (EUI), Via della Piazzuola, 43, I-50133 Florence, Italy
Department of Accounting & Finance, Athens University of Economics and Business (AUEB), 76 Patission str, GR104 34, Athens, Greece

ARTICLE INFO

Article history:
Received 23 July 2013
Received in revised form 15 June 2014
Accepted 6 November 2014
Available online 13 November 2014

JEL classification:
F31
G17
G14
C45
C58

Keywords:
Intraday exchange rates
Market heterogeneity
Technical trading
Heuristic learning

ABSTRACT

Until recently economists focused on structural models that were constrained by a lack of high-frequency data and theoretical deficiencies. Little academic research has been invested in actually trying to build successful real-time trading models for the high-frequency foreign exchange market, which is characterized by inherent complexity and heterogeneity. The present work opens new directions for inference on market efficiency in an attempt to account for the use of technical analysis by practitioners over many years now. This paper presents a heuristic model that efficiently emulates the dynamic learning of intraday traders. The proposed setup incorporates agent beliefs, preferences and expectations while it integrates the calibration of technical rules by means of adaptive training. The study focuses on EUR/USD which is the most liquid and widely traded currency pair. The data consist of a very large tick-by-tick sample of bid and ask prices covering many trading periods to enhance robustness in the results. The efficiency of a technical trading strategy based on the proposed model is investigated in terms of directional predictability. The heuristic learning system is compared against many non-linear models, a random walk and a buy & hold strategy. Based on statistical testing it is shown that, with the inclusion of transaction costs, the profitability of the new model is consistently superior. These findings provide evidence of technical predictability under incomplete information and can be justified by invoking the existence of heterogeneity caused by many factors affecting market microstructure. Overall, the results suggest that the proposed model can be used to improve upon traditional technical analysis approaches.

1. Introduction

In conventional economics, markets are assumed to be efficient if all available information is reflected in current market prices. The efficient market hypothesis (EMH) assumes that markets are populated with rational agents and in the absence of transaction costs the market price fully reflects all available information (Fama, 1991). EMH offered a way to test predictions on real world markets, by identifying three sources of information, corresponding to three degrees of informational efficiency that can be tested separately, namely the weak, semi-strong, and strong-form efficiency. The EMH was supported during the late 1960s by a large body of empirical

☆ This research is supported by the Marie Curie Fellowship (FP7-PEOPLE-2011-CIG, No. 303854) under the 7th European Community Framework Programme. I am grateful to faculty members and seminar participants at the Department of Economics, European University Institute (EUI) for helpful comments and discussions. I also thank the staff of the EUI library and in particular Thomas Bourke, for providing the high-frequency data (collected by Olsen & Associates). The usual disclaimer applies.
E-mail addresses: stelios.bekiros@eui.eu, bekiros@aueb.gr.
[1] Tel.: +33 1 53 63 36 00; fax: +33 1 45 44 40 46.
[2] Tel./fax: +30 210 8203453.
[3] Tel./fax: +302108203453.

http://dx.doi.org/10.1016/j.jempfin.2014.11.002
0927-5398/

works mostly based on daily frequency data. The proposed market pricing models were random walks, Brownian motion or more general Lèvy processes. However, it proved to be debatable whether any observed departure from market efficiency was due to genuine market inefficiency or whether pricing model deficiencies caused a divergence between actual and theoretical prices. Moreover, the non-predictability doctrine of the EMH was based on the additional assumptions that news and events hitting the market are normally distributed, and that markets are composed of homogeneous agents. More recent works, questioned the validity of EMH (Dacorogna et al., 2001; LeBaron, 1996). Instead of assuming a homogenous market, in which all agents interpret and react to the news in the same way, a heterogeneous market is proposed where agents act in different time horizons and in differing ways. Liu et al. (1999) and Cont (2001) suggested that random walk models are unable to explain properties of real world markets such as volatility clustering and extreme correlations. Gouree and Hommes (2000) and Kurz (1994) present new evidence that all market agents have in fact bounded rationality. Similar conclusions are reached by Farmer and Lo (1999) in their discussion of market efficiency. They base their analysis on a comparison with the evolution of ecological systems. Farmer (1998) develops a market model inspired by ecological systems that contains agents with various trading strategies. Eventually, Vernon Smith and Daniel Kahneman were awarded the 2002 Nobel Prize in behavioral economics suggesting that human psychology, not always rational, is intertwined with price processes.

Heterogeneity in expectations can lead to market instability and complicated dynamics of prices, which are driven by endogenous market forces. Boundedly rational agents using simple rules-of-thumb for their decisions, provide a more realistic description of human behavior than perfect rationality with optimal decision rules. Instead, according to the EMH, fully rational agents are considered the driving forces of markets, which in turn operate in a way to aggregate and process the beliefs and demands of traders reflecting all available information. Opponents of the EMH e.g., La Porta et al. (1997) and Shiller (2002) argued that predictability reflects the psychological factors and “fads” of irrational investors in speculative markets. This irrational behavior was emphasized by Black (1986) and Shleifer and Summers (1990) in their exposition of noise traders who are described as acting on the basis of imperfect information and consequently cause prices to deviate from their equilibrium values.[4] Moreover, Black claimed that noise traders play a useful role in promoting market liquidity. More theoretical work on financial markets with heterogeneous agents has also gained an important momentum in the literature. Specifically, Brock and Kleidon (1992) show how bid–ask spreads fluctuate over the day by firm size categories as a measure of “thickness” of the market. Brock and Hommes (1997) build a general theory of expectation formation that nests rational expectations in an econometrically tractable system. Hommes (2000) offers a review of recent work on heterogeneous agent financial theory. While there are many ways to describe heterogeneous expectations the most promising approach seems to differentiate the expectations according to time dimension or time scale of the market participants. In fact any differences in planning horizons, trading frequency or institutional constraints are neglected in the rational context of the EMH. The work of Lux and Marchesi (1999) on agent based models revealed that the interaction of agents with different trading/investment horizons gives rise to certain properties such as volatility clusters, trend persistence, fat tails and scaling laws, thus reproducing most of the empirical regularities observed in the financial markets.

Shiller (1989) argues that most participants in financial markets are not “smart” — following the rational expectations model — but rather follow trends and fashions. However, variation over time in expected returns poses a challenge for asset pricing theory because it requires an explicit dynamic theory in contrast to the traditional static capital asset pricing model (CAPM). Interestingly, recent studies provide evidence that price movements of financial assets for short and medium-term horizons are to some extent predictable. It is shown by Gençay et al. (2002, 2003) and Pictet et al. (1992) that financial returns substantially depart from the random walk model and can be predicted with some success. Real-time trading models, for instance, have been proven relatively successful in capturing the inefficiencies of the currency or stock market. New evidence on the stylized facts of financial markets reveals long periods of apparent price inertia which give the impression of predictability, yet with sudden violent interruptions. At the same time markets present anti-persistent features with ultra-fast mean price reversions, which seem unpredictable but actually fluctuate within limited trading ranges with a gradual adjustment to new market equilibria.

The EMH was probably never conceived for short trading horizons and high-frequency markets, as it takes an unrealistic view of market response to new information. It is reasonable to assume that the markets need a finite time to adjust to any information as opposed to instantaneous integration of new price information according to the EMH. Specifically, the contradictions and inconsistencies of the EMH regarding the stylized facts of high-frequency financial data cannot be explained by conventional econometric modeling and the Markowitz mean–variance framework. Short or long-term traders and decision makers, treasurers and central bankers, interpret the same information differently.[5] In addition, new technology now enables efficient information flow, optimal identification of trading opportunities while it also contributes to higher market volume and liquidity. The improved liquidity also introduces a shift in perspective, with agents starting to focus on short-term trading horizons and higher trading frequencies.

Until recently economists focused on structural models that were constrained by a lack of high-frequency data and theoretical shortcomings. In this work an attempt is made towards predicting short-term price movements in particular for the high-frequency currency market, which has the highest volume of all financial markets. There is a growing body of literature on intraday foreign exchange markets, which comprises short-term transactions from traders of various geographical locations, with different time-horizons, risk-profiles, or regulatory constraints. Structural economic models have been utilized to test various forms of market efficiency in currency markets (Baillie and McMahon, 1989; MacDonald and Taylor, 1992). Meese and Rogoff (1983) carried out the first comprehensive out-of-sample forecasting studies of these models for exchange rates. However, the absence of any theory for the short-term movements of the foreign exchange rates makes the structural models irrelevant for these horizons. An approach based on

[4] Arbitrageurs are also assumed to dilute a minor part of these shifts in prices, yet the major component of deviation is tradable.

[5] This might be due to agents utilizing variant heterogeneous priors, as reported in Morris (1996).

linear and non-linear time series models frequently utilized by market practitioners is more suitable. Moreover, the heterogeneous structure of intraday FX markets can account for the fact that practitioners have effectively used methods of "technical analysis" over many years now. Overall, two types of agents are identified in FX markets: "fundamentalists", who base their expectations upon dividends, earnings or macroeconomic factors, and "chartists" (technical analysts) who instead base their trading strategies on historical patterns and heuristics and try to extrapolate trends in future prices (Brock and Hommes, 1998; Frankel and Froot, 1990). The present study focuses on the latter.

2. High-frequency exchange rates: market microstructure and stylized facts

The foreign exchange (FX) market has the highest market volume of all financial markets. Since the end of the 1990s, academic researchers have been gaining new insights into the behavior of the FX markets through analyzing intraday data. Daily or weekly data, which were much used in the literature (e.g., Baillie and McMahon, 1989; Hsieh, 1988), represent only a small subset of the information available at intraday frequencies, as they are only the average of a few intraday prices quoted by some large banks at a particular daytime. Considering that intraday operations account for more than 90% of the FX market volume, the analysis of intraday trader behavior leads to more insight into the market microstructure. High-frequency data are also less affected by structural breaks or shifts in the overall economy than low-frequency samples of many years. The growing volume of FX transactions has been increasingly made up of short-term transactions and results from the interaction of traders with different time-horizons, risk-profiles, or regulatory constraints. Non-financial corporations, institutional investors (mutual funds, pension funds, insurance companies) and hedge funds, have shifted their FX activities from long-term investment to short-term (profit-making) transactions. This movement is both enabled and enhanced by the development of real-time information systems and the decrease of transaction costs following the liberalization of cross-border financial flows. Research studies have shown that well-accepted empirical regularities of daily or weekly data do not always hold up in intraday analysis. In intraday frequencies, the homogeneity of market agents (which is a working hypothesis for studying daily, weekly, or low-frequency data) disappears. Indeed, the heterogeneous structure of intraday data may explain the fact that practitioners have effectively used methods of technical analysis over many years now (e.g., Gençay and Stengos, 1998; Taylor and Allen, 1992).

The FX market is decentralized and worldwide. However, trading is processed in a few particular and major markets: London, New York and Tokyo. The total trading activity is comprised of the geographic contributions of individual market centers with no business-hour limitations. Any institutional market agent can submit new bid–ask prices although these prices emanate from particular banks in particular locations and the deals are entered into dealers' books in particular institutions (Petersen and Fialkowski, 1994). One full tick contains the time stamp, a bid and an ask price and often some information on the origin of the tick (bank code, city, etc.).[6] The bid–ask spread reflects the transaction and inventory costs and the risk of the institution that quotes the price. On the part of the traders who buy or sell at a quoted price, the spread is the only source of costs. Thereby it can be considered as a good measure of the amount of friction between different market participants and thus a measure of market efficiency.

The stylized facts of intraday FX rates can be summarized in four features, namely autocorrelation of returns, distributional diversity, scaling properties and seasonality. FX rates are non-Gaussian, non-affine Markov processes and the moments of the distributions tend to vary over time (Calvet and Fisher, 2002; Loretan and Phillips, 1994). The literature presents a number of views regarding the distributions of intraday FX returns and the corresponding data-generating process. Westerfield (1997) claims returns to be close to Paretian stable, Boothe and Glassman (1987) to non-stable Student distributions and others reject any single distribution (Calderon-Rossel and Ben-Horim, 1982). Feinstone (1987) and Wasserfallen (1989) report that intraday returns are fat-tailed with substantial deviation from a Gaussian random walk model, while more recently Barndorff-Nielsen and Prause (2001) concluded that the normal inverse Gaussian distribution seems to capture some of the empirical features. Most researchers agree that a better description of the data generating process is in the form of a conditional heteroskedastic model rather than of an unconditional one. Baillie and Bollerslev (1991) and Engle et al. (1990) suggest that currency series are GARCH processes, albeit with mixed serial and long-term dependent processes. Stationary models cannot incorporate the observed long-term dependence, as they presume finite and integer lags. According to Attansio (1991) such non-linearity and non-stationarity could be caused by time-varying variance or by persistence in FX rates. Granger and Joyeux (1980) and Hosking (1981) propose fractionally integrated ARMA models (ARFIMA) to measure long-term dependence in combination with the conventional short-term serial correlations. However, these models were extensively analyzed and popularized by Baillie (1996) and Robinson (2003). Moreover, the research group of Olsen & Associates (Switzerland) published empirical studies of scaling properties in currency markets extending from a few minutes to a few years (Müller et al., 1990). They report scaling laws that describe mean absolute returns and mean squared returns as functions of their time intervals (varying from a few minutes to one or more years). Finally, Meese and Rogoff (1983) mention in their seminal work the failure of linear exchange rate models, while several more recent studies have provided further evidence of existing non-linearities (Black, 1986; De Long et al., 1990). However, most of these empirical studies were conducted on low frequencies.

[6] The trading system is based on a continuous double auction mechanism. An order book is kept, in which limit orders are sorted by price, time and volume and stored in two streams, i.e., the bid side for buy orders and the ask side for sell orders. When a new buy (sell) limit order enters the system, it triggers a trade if the price is higher than the best offer (or lower than the best bid), otherwise it is stored in the book. Market buy and sell orders are executed immediately from the top of the ask and bid stacks. Each realized transaction is represented by a tick, which is the most elemental level of the price process. As the arrival of trades is not uniform in time, tick data is asynchronous.

3. Computational intelligence and technical trading

Technical analysis is used among market practitioners as an effective technique to earn significant profits from financial trading. Recent works by Kaufman (1998), Kirkpatrick and Dahlquist (2007) and Murphy (1999) reveal that the application of technical analysis in trading can consistently lead to high profitability by exploiting trends and reversals. In Gençay (1998b), the Dow Jones index is studied and the results indicate that buy–sell signals from moving average models provide more accurate sign and mean squared prediction errors (MSPE) relative to random walk and GARCH models. Moreover, LeBaron (1997) and Levich and Thomas (1993b) use bootstrap simulations to demonstrate the statistical significance of the technical trading rules against well-known parametric models of exchange rates. Lo et al. (2000) evaluated the effectiveness of using chart patterns and found that over a thirty-year sample period several of them provided valuable information. In addition, Dacorogna et al. (1995) examined real-time trading models of exchange rates under heterogeneous trading strategies. They conclude that it is the identification of the heterogeneous market microstructure in a trading model which leads to an excess return after adjusting for market risk. However, profiting from technical analysis is still considered a controversial issue. Brock et al. (1992) show that trading rules outperform a buy & hold strategy on the Dow Jones Industrial Average Index, but without the inclusion of transaction costs in their analysis. Bessembinder and Chan (1998) replicate their work taking into consideration transaction costs and show that these subsume the profitability documented by Brock et al. (1992). A more comprehensive study of technical trading is surveyed in Canegrati (2008), with results that are mixed and market-dependent. Other studies on technical trading include Bekiros (2010), Gençay (1998a), Gençay et al. (2002), Gradojevic and Gençay (2010), Neftci (1991), and Taylor and Allen (1992).

While most studies have evaluated technical analysis at daily frequency, it may be most useful at higher frequencies, when fundamentals are changing the least. Lyons (2001) argues that speculators and arbitrageurs usually employ heuristic strategies based on technical analysis to profit from high-frequency trading. He provides evidence that most commonly hedge funds reside in those techniques usually during intra-day trading, although in liquid FX markets GMT-night time trading is also pursued.[7] Based on modern technology during mostly the last two decades, there have been attempts to implement high-frequency technical trading using machine learning approaches such as neural networks, genetic algorithms and expert systems. Traditionally, the primary means of detecting trends and patterns involved statistical methods such as linear regression analysis and autoregressive time-series models. However, because these models are linear they often produce inaccurate forecasts due to inherent non-linearities in the dependence structure of the real data. Specifically in high sampling frequencies, data are often characterized by chaotic behavior, fat tails and non-linear components. In heterogeneous markets chartists are often engaged in developing new models, or modifying existing methods, that would enhance forecasting ability and profitability particularly for data series with dynamic time-varying patterns. As a result, machine learning techniques are becoming widespread tools for analyzing markets as they are capable of dealing with non-linear modeling. Recent works include Nevmyvaka et al. (2006) that implement optimal execution strategies with artificial intelligence algorithms and Zhang et al. (1998) who utilize artificial neural networks in financial forecasting. Machine learning has been employed in control systems modeling, especially towards emulating human behavior. In financial applications, the modeling difficulty is inherent because the behavior of agents is complex and conflicting and the system environment cannot be described by linear analytically tractable models as in classic model-based control. However, conventional control has proven many times to be divergent and unfit to real-time demands. As a solution the application of intelligent control combines non-linear, adaptive, and stochastic methods to enhance system reliability and efficiency (Antsaklis and Passino, 1993). Powerful techniques from computational intelligence include evolutionary algorithms, neural networks, fuzzy logic, as well as hybrid approaches.

Artificial neural networks have been extensively used in learning literature in modeling non-linear time series and in function approximation. They comprise input and output vectors and processing units interconnected by adaptive connection weights, trained to store the "knowledge" of the network. Adya and Collopy (1998) demonstrated the advanced predictive ability of artificial neural networks for time series forecasting. Kuan and White (1994) and White (1989) suggested that the relationship between neural networks and conventional statistical approaches for time series forecasting is complementary. Additionally, the function approximation properties of neural networks have been thoroughly investigated by many authors such as Hecht-Nielsen (1989) and Hornik et al. (1990). Examples using neural networks in financial markets include Gençay (1998b), Green and Pearson (1994), Weigend (1991), Yao et al. (1996) and Zhang (1994). However, neural networks, in the same fashion as conventional time series models, incorporate as input variables only quantitative factors, such as returns, prices and other financial or economic variables. A number of qualitative factors, e.g., macroeconomic and political effects or trader psychology, may seriously influence the market trend. Hence, it is important to capture this unstructured expert knowledge.

Implemented initially in the area of control systems and decision theory, fuzzy logic has been recently utilized in economic and financial applications with highly promising results (Bekiros, 2010; Gradojevic and Gençay, 2010; von Altrock, 1997). Specifically, in a fuzzy system numeric variables (inputs and outputs) are translated into linguistic terms representing beliefs. Fuzzy learning rules represented by "if-then" statements are specified to associate fuzzy input to the output fuzzy set. 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. In general, fuzzy systems are widely applied in the areas of classification, decision support and process simulation, as an effective means of modeling human expert knowledge, experience and intuition (Jamshidi et al., 1997; Mamdani, 1974, 1977; Sugeno, 1988). Fuzzy learning could comprise an efficient mechanism of incorporating heterogeneous agent beliefs in the form of rules-of-thumb used by technical traders.

[7] Interestingly, Lyons (2001) reports that when inventory control and risk sharing are regulated and monitored in the FX markets, nearly 90% of intraday trading is carried out without the application of technical trading.

Recent works on behavioral finance revealed that the application of technical analysis in trading can consistently produce high returns, albeit it is still considered a "pseudoscience" compared to conventional fundamental analysis (Camillo, 2008; Irwin and Park, 2007; Lo et al., 2000). This is due to an inherent subjectivity and a general lack of established guidelines to systematically determine the amount of relevant historical information as well as the optimal parameters of the technical rules employed. Clearly, the entire set-up is heuristically determined and heavily dependent on traders' experience and beliefs (Bolger and Harvey, 1995; Lawrence and O'Connor, 1992). In this framework, fuzzy deduction seems to provide an ideal approach to mimic the learning mechanism under which technical analysts operate. The adaptability of artificial neural networks can also be utilized to produce a hybrid learning model which combines the inference mechanism of fuzzy modeling with adaptive training from neural networks (Bekiros, 2010; Buckley and Hayashi, 1994).

The main objective of this study is the development of a technical trading system that decodes the decision-making process of chartists in heterogeneous markets. In the next section a new learning model is presented. The proposed functionality integrates the calibration of heuristic technical rules by means of adaptive training. This model efficiently incorporates trader beliefs, preferences and adaptive expectations which are represented by fuzzy logic rules. Overall, this paper extends the literature on real-time technical trading systems, by presenting an adaptive fuzzy learning approach that leads to superior directional predictability. This is illustrated through a comparative investigation against other forecasting models. The rest of the paper is organized as follows; in Section 4 the new adaptive learning system is described. In Section 5 the other forecasting models used in this study are presented. Next, Section 6 describes the data. Section 7 provides the empirical results and Section 8 concludes.

4. Heuristic learning with adaptive expectations and beliefs: a real-time trading model

The key functionality of the heuristic learning system lies in the efficient mechanism of incorporating agents' heterogeneous beliefs, preferences and expectations under uncertainty. This is implemented via fuzzy deduction in the form of *if-then* inference rules. These rules represent the expert knowledge, behavioral patterns, experience and intuition of the agent and are specified to model the state space and associate the vectors of input and output variables of the trading system. This set-up is intended to provide a realistic model of the decision-making process under which rule-of-thumb technical analysts operate in the real-world market environment.

The learning system consists of the input, the rule layer and the output layer. In the input fuzzy layer all the input variables are translated into fuzzy terms whereas in the Boolean formalism inputs have a crisp numeric value. Each term is described by a fuzzy membership function, which models the *beliefs* of the agent. The fuzzy inference rules consist of two parts, the "if" part and "then" part. The "if" part utilizes an "and" association, proposed by Zimmerman and Thole (1978) and represents the minimum value among all the validity values of the "if" part. The output layer also utilizes membership functions for the output vector. In the defuzzification layer, the vector of output variables is converted from fuzzy back into crisp values. The aforementioned structure was initially introduced by Mamdani (1977) as the standard approach of fuzzy deduction. Alternatively, the approach by Sugeno (1985) and Takagi and Sugeno (1985) introduces linear dependences of each rule on the system's input variables, whereby no defuzzification process is required. The first-order Sugeno model has rules of the form "if x_1 is A and x_2 is B then $\pi = \alpha + \beta x_1 + \gamma x_2$", where (x_1, x_2) are the inputs, (A, B) are the fuzzy sets and (α, β, γ) are the parameters. This architecture is suited for modeling complex non-linear systems by interpolating multiple linear models. The parameter space comprises an antecedent part (parameters of the membership functions) and a corresponding consequent (polynomial parameters) for each rule. However, in these models the crucial issue is the choice of the fuzzy terms for each variable and the estimation of the consequent parameters, whereas the antecedent part is usually fixed and not "calibrated". In this way, the fuzzy deduction rules depend on ad-hoc assumptions and are usually subjectively constructed, and eventually the choice of membership functions depends on trial and error. Instead, in this study a training approach also used in artificial neural network applications will be utilized to "fine-tune" the fuzzy membership functions (Buckley and Hayashi, 1994; Nishina and Hagiwara, 1997). In the proposed setup, a first-order polynomial model with two inputs and two rules is utilized

$$\pi_{it} = \alpha_{it} + \beta_{it} x_{1t} + \gamma_{it} x_{2t} \tag{1}$$

where t the time index corresponding to the episodic learning (explained in details later), $i = 1, 2$ the rule number, x_j with $j = 1, 2$, the inputs participating in the ith rule, and π_{it} the first-order polynomial output of the ith rule. The two parameter sets, namely the membership function parameters and the polynomial parameters are all time-varying and adaptively updated. In the proposed architecture two membership functions are used for each input corresponding to two beliefs as being perceived by the boundedly rational agents, namely "low" and "high".[8] The definition of trader expectations is provided below.

Definition 1. The time-varying polynomial (consequent) coefficients $(\alpha_{it}, \beta_{it}, \gamma_{it})$ represent the adaptive expectations of the agents. These expectations are rule-dependent, i.e., each rule-of-thumb used in the trading decision utilizes another set of adaptive expectations.

[8] A natural addition to the current specification could be a third state e.g., "neutral", in which no strong belief is advocated. However, the current setup represents efficiently the fluctuation of the ultra high-frequency currency series. A third state added significant computational burden — marginally suitable — for real-time applications, while it proved qualitatively worse (or roughly equal) in terms of learning from the very large tick-by-tick data sample due to the curse of dimensionality.

The hybrid learning process uses a Widrow–Hoff steepest descent algorithm (Nguyen and Widrow, 1989; Widrow and Hoff, 1960) to optimize the beliefs' parameters, and a least squares-type algorithm to solve for the expectations' 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, the value function to solve for the optimal parameters is the mean squared error $E = 1/N[(y - y^o)^2]$, where y^o the target output and y the system output for N size sample. Initially, the agent beliefs for each input j of the state space and for every rule i are modeled as $\mu_{B_{it}}(x_j)$ where B_{it} is the linguistic label of x_j. Then, the weight by which each rule-of-thumb contributes to the agent decision-making process is attributed. It is not necessarily symmetric and corresponds to the *agent's preference* of the influence of the rule to the trading 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 and beliefs, instead of keeping it constant based on an ex-ante "rational" assumption. The definition of the trader's preferences is provided in correspondence to the expectation formation.

Definition 2. Agent's preferences p_{it} represent the influence of each rule-of-thumb utilized for the trading decision. They are also "flexible" (time-varying) and are estimated according to the episodic calculation of trader's beliefs.

A necessary condition for tractable parameter estimation is the normalization of the preferences (rule weights) $\overline{p}_{it} = p_{it}/p_{1t} + p_{2t}$, where $p_{it} = \prod_{j=1}^{2} \mu_{B_{it}}(x_j)$. Next, the learning rule outputs are calculated as $y_{it} = \overline{p}_{it}\pi_{it} = \overline{p}_{it}\ \ \alpha_{it} + \beta_{it}x_{1t} + \gamma_{i,t}x_{2t}$. Finally, all ruled-based preferences are aggregated producing the output of the system as a piecewise linear interpolating function, dynamically calibrated by the time-varying expectations and time-dependent beliefs

$$y_t = \sum_i y_{it} = \overline{p}_{1t}\left(\alpha_{1t} + \beta_{1t}x_{1t} + \gamma_{1t}x_{2t}\right) + \overline{p}_{2t}\ \ \alpha_{2t} + \beta_{2t}x_{1t} + \gamma_{2t}x_{2t.} \tag{2}$$

The output can be reformulated in a vector specification

$$y_t = \left[\overline{p}_{1t}x_{1t}\ \ \overline{p}_{1t}x_{2t}\ \ \overline{p}_{1t}\ \ \overline{p}_{2t}x_{1t}\ \ \overline{p}_{2t}x_{2t}\ \ \overline{p}_{2t}\right] \cdot \left[\beta_{1t}\ \ \gamma_{1t}\ \ \alpha_{1t}\ \ \beta_{2t}\ \ \gamma_{2t}\,\alpha_{2t}\right]^T = \mathbf{X}_t\mathbf{P}_t. \tag{3}$$

The following proposition (proof is shown in Appendix A), states the least squares-type algorithm to solve for the expectations' parameters in order to account for under/over-determinacy, serial autocorrelation, noise overfitting and numerical instabilities.

Proposition 1. *Given that the trader's decision signal formulated by Eq. (2) is based on time-dependent beliefs and flexible preferences, the adaptive expectations are updated and estimated as*

$$\mathbf{P}_t = \mathbf{V}_t\mathbf{D}_t^{-1}\mathbf{U}_t^{\mathbf{T}}\mathbf{Y}_t \tag{4}$$

where the input matrix $\mathbf{X}_t = \mathbf{V}_t\mathbf{D}_t^{-1}\mathbf{U}_t^{\mathbf{T}}$ *is decomposed into a diagonal matrix* $\mathbf{D}_t$ *that contains the eigenvalues, a matrix* $\mathbf{U}_t$ *of principal components, and an orthogonal normal matrix of right singular values* $\mathbf{V}_t$*. Thus,* via $\mathbf{P}_t$ *the adaptive expectations of the traders are estimated in each learning episode.*

Modeling the beliefs is performed via the fuzzification of the input variables into membership functions. Two symmetric triangular membership functions are applied, in order to optimize the training performance in terms of computational load (Ishibuchi et al., 1995).

Definition 3. The trader's belief is modeled via a symmetric triangular membership function as follows

$$\mu_{B_{it}}\left(x_j\right) = \begin{cases} 1 - \left(2\left|x_{jt} - b_{it}\right|/c_{it}\right), & \text{if} \quad 2\left|x_j - b_{it}\right| \le c_{it} \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

Definition 4. Given that beliefs are modeled as in Eq. (5), the b_{it} parameter represents the "intensity-of-belief" of the technical analyst while c_{it} corresponds to the "range-of-belief".

These parameters are implicitly well-suited for modeling trader's uncertainty perception mechanism. The adaptive learning is performed via a Widrow–Hoff gradient descent algorithm and for the belief parameters it is given as[9]

$$b_{it+1} = b_{it} - \tau_b(\partial E_t/\partial b_{it}) \tag{6}$$

$$c_{it+1} = c_{it} - \tau_c(\partial E_t/\partial c_{it}) \tag{7}$$

where τ_b, τ_c the learning rates (e.g., determines the change of parameter values and the convergence of the value error-function). The Widrow–Hoff learning minimizes the mean square error along the negative of the gradient of the performance function and is

[9] The Hessian can be also used for $\mathbf{u} = (b_{it}, c_{it})^T$, i.e., $\mathbf{u}_{it+1} = \mathbf{u}_{it} - \mathbf{H}_{\mathbf{k}}^{-1}\nabla E_t$, but it is computationally expensive. An approximation $\mathbf{H} = \mathbf{J}^{\mathbf{T}}\mathbf{J}$ with $\nabla E_t = \mathbf{J}^{\mathbf{T}}E_t$ where $\mathbf{J}$ the Jacobian could reduce the computational overhead.

based on an approximate steepest descent procedure. The following theorem, proved in Appendix A, provides the adaptive learning mechanism of the beliefs.

Theorem 1. *Given that beliefs are modeled as in Eq. (5) and updating is performed via Eq. (6) and Eq. (7), the adaptive learning of the intensity-of-belief is provided by*

$$b_{it+1} = b_{it} - \tau_b \left[4p_{it}\,\mathrm{sign}\left(x_j - b_{it}\right)\left(y_t - y_t^o\right)\left(\pi_{it} - y_t\right)\right] \Big/ \left(Nc_{it}\mu_{B_{it}}\sum_{i=1}^{2} p_{it}\right) \tag{8}$$

and of the range-of-belief by

$$c_{it+1} = c_{it} - \tau_c \left[2p_{it}\left(1 - \mu_{B_{it}}\right)\left(y_t - y_t^o\right)\left(\pi_{it} - y_t\right)\right] \Big/ \left(Nc_{it}\mu_{B_{it}}\sum_{i=1}^{2} p_{it}\right). \tag{9}$$

Learning could be repeated until there is no significant change (e.g., of order $\mathbf{10^{-3}}$) in the belief parameter space and a convergence goal criterion is met (e.g., $|E_{t+1} - E_t| \le 10^{-5}$). However, instead of using such a specification for the learning process, an alternative scheme is used in this study. Specifically, a "validation" set containing different observations from the original sample is applied to improve generalization. The validation error normally decreases, as does the in-sample error. Yet, when the algorithm loses its generalization power, the validation error will begin to rise. When this happens, the belief parameter space $\left(b_{it}, c_{it}{}^T\right)$ at the minimum of the validation error is returned.

To sum up, the heuristic learning rule that simulates trader decision process comprises two *deductive episodes* in a two-stage estimation algorithm. In the first time episode, the expectations and preferences are dynamically estimated using a singular value decomposition method, while the intensity- and range-of-belief remain static. Thereafter, the produced trading signal is using the previously calculated expectations and preferences and in the second deductive episode the estimation errors are backpropagated though the learning process to optimally determine the belief parameter updates, while the expectations and preferences are now kept fixed.

5. Forecasting models

The heuristic learning model is compared against a Markov-switching model, an ARFIMA model, a feedforward and a recurrent neural network in order to examine its relative predictability and profitability performance. These models are described below. In addition, a naïve strategy is comparatively used as well as a buy & hold strategy. The naïve strategy assumes that the best forecast for the future is the current price, in accordance with the random walk hypothesis. The buy & hold strategy involves buying the EUR/USD at the beginning of the out-of-sample period and sell it at the end. Hence, it assumes an efficient foreign exchange market, which does not allow excess returns because the conditional expectation of the high-frequency returns is zero, and gains only by exploiting the market trend.

5.1. Markov-switching model

In the non-linear time series, a class of regime-switching models assumes that the regime that occurs at time t cannot be observed and is determined by an unobservable process or state, denoted as s_t. In case of only two regimes (e.g., "low" and "high") and AR(1) specification in both regimes, the model is given by

$$y_t = \begin{cases} \varphi_{0,1} + \varphi_{1,1} y_{t-1} + \varepsilon_t & \text{if } s_t = 1 \\ \varphi_{0,2} + \varphi_{1,2} y_{t-1} + \varepsilon_t & \text{if } s_t = 2 \end{cases}. \tag{10}$$

The well-known model in this class is the Markov-switching model introduced by Hamilton (1989). In that s_t is assumed to be a first order Markov-process, i.e. s_t depends only on s_{t-1}. The transition probabilities are $P(s_t = 1|s_{t-1} = 1) = p_{11}$, $P(s_t = 2|s_{t-1} = 1) = p_{12}$, $P(s_t = 1|s_{t-1} = 2) = p_{21}$ and $P(s_t = 2|s_{t-1} = 2) = p_{22}$ with $p_{11} + p_{12} = 1$ and $p_{21} + p_{22} = 1$. Additionally, via the theory of ergodic Markov chains it can be shown (Hamilton, 1989) that for the two-state model, the unconditional probabilities are given by $P(s_t = 1) = (1 - p_{22})/(2 - p_{11} - p_{22})$ and $P(s_t = 2) = (1 - p_{11})/(2 - p_{11} - p_{22})$.

5.2. ARFIMA model

For a long-memory process, the autocovariance sequence at lag k satisfies $\gamma_k = Ck^{-\alpha}$ where C is the scaling parameter and $\alpha \in [0, 1]$. A common model is the fractional difference process for which $\alpha = 1 - 2d$ and d is the fractional difference parameter. Specifically, Baillie (1996) and Granger and Joyeux (1980) introduced fractional ARIMA models, which are a generalization of the standard ARIMA (p, d, q) models. Let x_t be a stochastic process whose dth order backward difference

$$(1-\Phi)^d x_t = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k x_{t-k} = \varepsilon_t \tag{11}$$

is a stationary process. For example, the first-order difference is $(1 - \Phi)x_t = x_t - x_{t-1}$ when $d = 1$. If ε_t is white noise process with variance σ_ε^2, then x_t is the simplest case of a fractional ARIMA process, a fractional ARIMA $(0, d, 0)$ or fractional difference process (FDP).

5.3. Feedforward and recurrent neural network

The output of a neural network is produced via the application of a transfer function. The functionality is to modulate the output space as well as prevent outputs from reaching very large values which can "block" training. Kao and Ma (1992) and Levich and Thomas (1993) found that hyperbolic sigmoid and tansigmoid transfer functions are appropriate for financial market data because they are non-linear and continuously differentiable, and these are both desirable properties for network training. 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 consists of the most recent observations that are processed for testing predictability. In selecting the optimum number of hidden neurons, despite the importance of such a choice, there is no explicit formula. Katz (1992) indicates that an optimal number of hidden neurons can be found between one-half to three times the number of inputs, whereas Ersoy (1990) proposes doubling the number of neurons until the network's RMSE performance deteriorates.

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

$$y_t = S\left[\beta_0 + \sum_{i=1}^{q} \beta_i G\left(\alpha_{i0} + \sum_{j=1}^{p} \alpha_{ij} x_{j,t}\right)\right] = f(\mathbf{x_t}, \mathbf{z}) \tag{12}$$

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

$$z_{t+1} = z_t + \tau \nabla f(\mathbf{x_t}, z_t)\left[y_t - f(\mathbf{x_t}, z_t)\right] \tag{13}$$

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

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

$$y_t = G\left[\beta_0 + \sum_{i=1}^{q} \beta_i g_{i,t}\right] \tag{14}$$

with

$$g_{i,t} = G\left(\alpha_{i0} + \sum_{j=1}^{p} \alpha_{ij} x_{j,t} + \sum_{h=1}^{q} \delta_{ih} g_{h,t-1}\right) \tag{15}$$

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

6. Data description and preliminary analysis

Euro was introduced (in a non-physical form) on 1 January 1999, when the national currencies of the Eurozone countries ceased to exist independently and were locked at fixed rates. According to the Bank of International Settlements Triennial Survey for 2007, the bulk of all transactions involving US dollar-Euro trading amounts approximately to 40% with a daily average turnover of more than 500 billion US dollars (BIS Triennial Survey, 2007). It is reported that on the spot market the US dollar (USD) was involved in 86.3% of transactions, followed by the euro (EUR) with 37.0%, the Japanese yen (JPY) with 17.0% and the Great Britain pound (GBP) with

15.0%.[10] The exact trading ratios represent EUR/USD, GBP/USD and USD/JPY. These are also the most liquid and widely traded currency pairs in the world with 27% market turnover share for EUR/USD, 13% for USD/JPY and 12% for GBP/USD (BIS Triennial Survey, 2007).[11] On an intraday basis, more than 10,000 ticks per business day are available for EUR/USD; that is an average of almost 10 ticks per minute which can rise to 30 or more ticks per minute during the busiest periods.[12] As the major part of all currency transactions involves US dollar-Euro, this study focuses on the EUR/USD pair.

In particular, the raw dataset consists of 5-minute bid and ask transaction prices for EUR/USD from 1/1/2004 to 10/1/2004, with a total of 112,898 data points.[13] It is one of the largest high-frequency tick-by-tick samples to be used in an academic study involving out-of-sample forecasting. The bid and ask prices at each 5-minute interval are obtained by interpolation over time, as in Dacorogna et al. (1993) and Müller et al. (1990). The weekend quotes from Friday 21:05 Greenwich mean time (GMT) to Sunday 21:00 GMT are removed as in Andersen and Bollerslev (1997), Andersen et al. (2001), and Jensen and Whitcher (2000).[14] Prices are calculated as the average of the logarithm of the bid and ask prices, i.e., $P_t = 1/2(\log P(bid) + \log P(ask))$ and continuously compounded 5-minute returns as $r_{t5} = 100(P_t - P_{t-1})$. Hence, a total sample of 56,449 5-minute EUR/USD return observations is used. However, Guillaume et al (1997) mentioned that price changes observed at very high frequencies can be overly biased by the buying and selling intentions of the quoting institutions. Gençay et al. (2009) argue that the news is assumed to arrive hourly and that it is unlikely in real-world trading applications the flow of information to be less frequent, i.e. over longer time intervals. Thus, in order to account for extremely high-frequency noise and no-activity periods in small time windows, the data is also aggregated over 1-hour intervals as $r_{t60} = \sum_{i=0}^{11} r_{t5-i}$. Gençay et al. (2010) also claim that working with 1-hour returns eliminates any price bias at lower frequencies and reduces computational burden. In this study both 5-minute and 1-hour frequencies are used in order to examine the relative model predictability and profitability performance as well as to test the robustness vis-à-vis different frequencies. In addition, the backtesting sample is equally divided in two sub-periods to enhance time-period robustness of the results. Specifically, the total sample spans the period 1/1/2004 (0:00 h) − 10/1/2004 (0:00 h) while the out-of-sample (backtesting) performance is investigated in the P_{Total} period from 9/1/2004 (0:00 h) to 10/1/2004 (0:00 h) at both frequencies with the use of a one-step-ahead moving window. The out-of-sample period is further segmented into two sub-periods, namely P_1: 9/1/2004 (0:00 h)–9/16/2004 (0:00 h) and P_2: 9/16/2004 (0:05 h)–10/1/2004 (0:00 h). The 1-hour aggregated frequency sample starts at 1:00 am on 9/1/2004 and ends similarly to the 5-minute one. The predictability of the heuristic learning model (denoted HLM) is examined against that of a Markov-switching model (MSR), a feedforward artificial neural network (FFANN), a recurrent artificial neural network (RANN) and an ARFIMA model, while a naïve (NV) and a buy & hold (B&H) strategy are also used as benchmarks.

The input space of the HLM comprises two lagged intraday returns. The input selection is conducted according to the methodology of Frances and van Dijk (2000) on non-linear models and neural networks. The output is the forecasted 5-minute-ahead and 1-hour-ahead return, i.e., $\hat{y}_{t5+1}$ and $\hat{y}_{t60+1}$ respectively. Specifically, the procedure for the selection of the lags involved initially the calculation of the Ljung–Box statistics for the first 10 lags of the EUR/USD return series in order to get a first indication at both frequencies. Significant autocorrelations of up to the second lag were identified for both frequencies. 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. Moreover, an analysis based on the RMSE conducted stepwise on the HLM for one to six endogeneous lags, revealed that the selected setup provided with the best results. All other topologies were found to be qualitatively worse or roughly equal compared with those finally presented in the results in terms of directional predictability, albeit with higher computational overhead. The input space of the artificial neural networks corresponds to past returns following Fernández-Rodriguez et al. (2000) and Gençay (1998a, 1998b). Moreover, based on the same methodology as for the HLM, the FFANN and the RANN models use two lagged intraday returns and a topology comprising 10 *g*-neurons in the hidden layers and a single-output layer. This empirical result also follows Ersoy (1990) and Katz (1992). The MSR model uses an AR(2) specification in both regimes as well as two beliefs (i.e. “low” and “high”) for the state variable, in direct association with the architecture of the HLM. Regarding the fractional ARIMA model, two lags for the AR part and zero for the MA part were also identified via the SIC on the price series. Andersen and Bollerslev (1997) estimated the fractional difference parameter to be $d = 0.36$ for intraday DEM/USD price series. Andersen et al. (2001) calculated six estimates of d from various measures for the DEM/USD and JPY/USD series at different frequencies. These six estimates of d varied from 0.346 to 0.448. Therefore in this study, d is set to 0.4 to represent the average of these six estimates. Thus a specification of an ARFIMA (2, 0.4, 0) was selected, which assumes that the 0.4th order backward difference is a stationary intraday return process.[15]

For the out-of-sample testing period the models utilize a rolling window of all previous observations (comprising the training in-sample) and produce forecasts for each day within the corresponding period. For instance, the first training sample is 1/1/2004 (0:00 h)–8/31/2004 (23:55 h). In case of the HLM, FFANN and RANN, a validation sample for each period covering the 25% of the training set is further utilized, in order to evaluate the generalization ability and avoid overfitting. The training set consists of the most recent observations that are processed in each period. The validation error normally decreases, as does the in-sample error, but when

[10] Volume percentages for all individual currencies should add up to 200%, as each transaction involves two currencies. Other currencies with high market turnover are the Swiss frank (6.8%), the Australian dollar (6.7%) and the Canadian dollar (4.2%).

[11] The six major currencies of Forex also include the Swiss frank (USD/CHF), Australian dollar (AUD/USD) and Canadian dollar (USD/CAD) in smaller market shares and transaction volumes. The six FX majors dominate the overall world market share with 76% of global trades having both currencies in the currency pair as a major, and more than 98% of all trades involving at least one major.

[12] The original form of market prices is tick-by-tick data: each “tick” is 1 unit of information, like a quote or a transaction price. The foreign exchange market generates hundreds or thousands of ticks per business day. Data vendors like Reuters transmit more than 275,000 prices per day for foreign exchange spot rates alone.

[13] The data are all collected by Olsen & Associates and provided by the European University Institute library database.

[14] Apart from this no further filtering was applied to the data, nor were any data points excluded.

[15] The results on the lag selection and the sensitivity analysis for all models are available upon request.

the algorithm losses its generalization power, the validation error will begin to rise. When this happens, the parameter space at the minimum of the validation error is estimated.

7. Empirical results

The rationalization of using non-linear models is based on the application of the well-known Brock, Dechert and Scheinkman (BDS) test (Brock et al., 1987) for the presence of non-linear dependence in the series (Table 1).[16] The test statistic under the null of i.i.d. is given by Brock et al. (1991)

$$W_{m,N}(\varepsilon) = N^{1/2}\left[C_{m,N}(\varepsilon) - C^{m}_{1,N}(\varepsilon)\right] \Big/ \sigma_{m,N}(\varepsilon). \quad (16)$$

where $C_{m,N}(\varepsilon)$ is the correlation integral from m dimensional vectors that are within a distance ε from each other, when the total sample is N, and $\sigma_{m,N}(\varepsilon)$ is the standard deviation of $C_{m,N}(\varepsilon)$. Under the null hypothesis, $W_{m,N}(\varepsilon)$ has a limiting standard normal distribution. The BDS test is applied on: (a) the original high-frequency 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 an AR(2)-GARCH-in-Mean(1,1) model in order to ensure that rejection of the null is not due to conditional heteroscedasticity (De Lima, 1996).

In all cases and for both frequencies 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.[17] The trading strategy of the technical analyst works as follows; at the end of each trading day the models are being re-estimated over a moving sample with a length equal to the training period. When the output of a model is greater than zero this is used as a buy signal and a value less than zero as a sell signal. The total return when transaction costs are not considered is estimated as

$$TR = \sum_{t=1}^{N_o} \lambda_t y_t \quad (17)$$

where N_o indicates the out-of-sample horizon, y_t is the realized return and λ_t is the recommended position which takes the value of (-1) for a short and $(+1)$ for a long position. A similar strategy has been employed, with considerable success, by a number of other researchers such as Gençay (1998a), Gençay (1998b), Fernández-Rodriguez et al. (2000) and Jasic and Wood (2004) among others. In order to evaluate the forecast accuracy of the models many measures could be utilized. Most standard measures rely on the MSE, RMSE etc. for each time horizon. However, these accuracy measures are all parametric in the sense that they rely on the desirable properties of means and variances, which occur when the underlying distributions are normal. In this study non-parametric methods are also used to analyze forecast accuracy as in Dacorogna et al. (2001). These "distribution free" measures do not assume normally distributed returns and can be used when this assumption is not valid. One measure that has this desirable property is the percentage of forecasts in the right direction. To a trader, for instance, it is more important to correctly forecast the direction (up or down) of any trend than its magnitude. The measure of directional quality (DQ) is defined as the percentage of correct predictions or correctly predicted signs

$$DQ = \kappa / N_o \quad (18)$$

where κ is the number of correct predictions. It should be noted that unlike predictions in other disciplines (e.g., weather forecasts), FX rate forecasts are valuable even when its direction quality is slightly above 50% and statistically significant. As no trader expects to be right all the time, in practice a sign rate "significantly" higher than 50% means that the forecasting model is better than the random walk (Dacorogna et al., 2001). However, in order to properly infer on the significance of the forecasts, a statistical measure of predictability should be applied. In this study the Henriksson–Merton (HM) statistic (Henriksson and Merton, 1981) is employed for the forecasted returns. Henriksson and Merton (1981) showed that the number of correct forecasts has a hypergeometric distribution under the null hypothesis of no market–timing ability. Additionally, the Sharpe ratio (SR) is considered as a measure of comparative profitability. It is the proportion of the mean return of the trading strategy over its standard deviation. The higher the SR, the higher the return and the lower the volatility

$$SR = \mu_{N_o} / \sigma_{N_o}. \quad (19)$$

The empirical results of the comparative implementation of all models are reported in Tables 2, 3 and 4. For all predictability and profitability indices, the HLM dominates the other forecasting models as well as the naïve and B&H strategy consistently in all periods and at both frequencies. Specifically, in all periods the total return for the trading strategy based on the HLM outperforms impressively the other models, while the same applies with the inclusion of transaction costs, which are estimated as 0.05% for each one-way trade,

[16] The general ARFIMA model although it is linear, given the dependence structure of the data, it seems likely that it could describe the high-frequency intraday FX rates adequately. Moreover, it was used in this study based on the evidence from a plethora of works, such as Baillie (1996) and Robinson (2003), that ARFIMA models could capture the strong non-linear dependence, long-memory and short-term serial correlations of high-frequency financial data.

[17] There is only one case of a less than 1% significant result, at the 5% level for $(m, \varepsilon) = (2, 2)$ at 1-hour frequency.

Table 1
BDS test.

FX rate	Correlation dim.	m = 2						m = 3						m = 4					
	Dim. distance	ε = 1		ε = 1.5		ε = 2		ε = 1		ε = 1.5		ε = 2		ε = 1		ε = 1.5		ε = 2	
	Frequency	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour
EUR/USD	Raw data	18.034	4.818	17.840	5.069	18.117	5.177	23.582	5.351	22.316	5.189	21.915	5.265	26.514	5.912	24.597	5.598	23.799	5.508
	AFR	18.912	4.840	16.754	4.705	17.284	4.702	24.699	5.544	21.298	4.890	21.211	4.811	27.750	6.086	23.659	5.358	23.238	5.109
	NLSSR	4.570	2.627	4.496	2.579	5.843	2.404	5.670	2.694	5.151	2.660	5.845	2.690	5.777	2.610	5.058	2.815	5.445	2.881

Notation
– Raw data = intraday EUR/USD 5-minute and 1-hour 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.
– The BDS test is applied on the total sample, P_{Total}: 9/1/2004-10/1/2004 for both frequencies.

Table 2
Out-of-sample tests for EUR/USD intraday returns (PTotal: 9/1/2004–10/1/2004).

Model frequency	NV		HLM		MSR		FFANN		RANN		ARFIMA	
	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour
TR	−11.550	0.387	8.541	2.615	2.267	−1.930	1.444	−1.148	−1.142	0.819	−11.801	−2.340
	(−13.118)	(0.251)	(7.285)	(2.480)	(1.230)	(−2.057)	(−0.152)	(−1.279)	(−2.739)	(0.670)	(−13.349)	(−2.476)
DQ	0.421	0.473	0.510	0.520	0.482	0.479	0.473	0.465	0.474	0.515	0.416	0.448
HM test	−8.545	−1.522	7.048***	1.991**	0.732	−0.929	0.440	−2.092	−0.881	1.281*	−8.991	−2.100
RMSE	0.040	0.126	0.024	0.094	0.039	0.086	0.073	0.120	0.033	0.086	0.025	0.086
SR	−1.174	0.134	0.854	0.909	0.226	−0.670	0.144	−0.398	−0.114	0.388	−1.204	−0.822
B&H return	0.020											

Notes
– HLM = heuristic learning model. RANN = recurrent artificial neural network. MSR = Markov-switching model. FFANN = feedforward artificial neural network. ARFIMA = Arfima model. NV = naïve strategy (RW hypothesis). B&H = buy and hold strategy.
– HT test = Henriksson and Merton (1981) test, asymptotically distributed as $N(0,1)$.
– TR measures the total return of the strategy. In parenthesis total return after transaction costs (0.05% average fixed cost for each one-way trade).
– The DQ 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 total sample (5-minute frequency) starts at 00:00 time on 9/1/2004 and ends at 00:00 on 10/1/2004 (6336 obs.). The 1-hour aggregated frequency sample starts at 1:00 am on 9/1/2004 and ends similarly (526 obs.). The two sub-periods P_1 and P_2 are equally divided on 9/16/2004 at 00:00 h.
*** Indicates significance at the one sided 1% level.
** Indicates significance at the one sided 5% level.
* Indicates significance at the one sided 10% level.

Table 3
Out-of-sample tests for EUR/USD intraday returns (P1: 9/1/2004–9/16/2004).

Model frequency	NV		HLM		MSR		FFANN		RANN		ARFIMA	
	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour
TR	−5.484	0.216	3.192	0.504	0.446	−0.665	0.108	1.087	−0.342	0.512	−5.868	−1.937
	(−6.266)	(0.146)	(2.606)	(0.436)	(−0.106)	(−0.722)	(−0.691)	(1.021)	(−1.142)	(0.438)	(−6.651)	(−2.002)
DQ	0.424	0.456	0.503	0.515	0.482	0.486	0.470	0.487	0.476	0.509	0.417	0.433
HM test	−5.380	−1.915*	3.871***	1.263	0.395	−1.508	−0.856	−0.461	−0.439	1.152	−6.413	−1.945
RMSE	0.040	0.123	0.025	0.092	0.042	0.087	0.072	0.117	0.033	0.087	0.025	0.087
SR	−1.112	0.148	0.637	0.346	0.088	−0.456	0.022	0.745	−0.068	0.357	−1.193	−1.354
B&H return	−0.002											

Notation as in Table 2.

following Fama and Blume (1966) and Hsu and Kuan (2005).[18] The HLM achieves a TR of 8.541 and 2.615 for 5-minute and 1-hour frequency respectively in P_{Total}, while the highest return of another model is 2.267 for MSR and 1.444 for the FFANN in P_{Total} at the 5-minute frequency.[19] It is also worth mentioning that the trading rules based on many other competing models present significant losses at the examined frequencies and periods. Instead, the HLM displays both frequency- and period-based robustness in the results. The fact that HLM outperforms the other strategies is also depicted in the proportion of correctly predicted signs (directional quality index), which is always higher compared to the aforementioned models. The sign rate is consistently above 50% ranging from 50.3% in P_1 for the 5-minute sample to 53.6% in P_2 for the 1-hour frequency (or 51% and 52% average sign rate for the P_{Total} respectively). The best directional quality for another model is that of 51.5% in P_{Total} for the RANN at the 1-hour sample frequency.[20] The HM test further corroborates on the statistical significance of the directional quality of the HLM. Highly significant values appear at the one-sided 1% level, such as 7.048 (in P_{Total} at 5-minute frequency) and 6.075 (in P_2 at 5-minute frequency), or at 5% level with 1.991 (in P_{Total} at 1-hour frequency). On the contrary, only the RANN model achieves statistical significance at the 10% level in case of 1-hour frequency in P_{Total}, while many models report non-significant or negative values. Additionally, the SR index, measuring the profitability per unit of risk, is much higher for HLM compared with the other models. MSR, FFANN and RANN report positive SRs, the first two for 5-minute and the latter for 1-hour frequency respectively, albeit significantly lower than that of HLM. In terms of the RMSE the results for the HLM are consistently lower compared with the FFANN, RANN, MSR and NV models and only relatively similar to that of ARFIMA. Interestingly, the fact that B&H strategy outperforms the FFANN and RANN models in most cases is not in accordance

[18] There is only one case where the return of another model is marginally higher than that of HLM, albeit with the inclusion of transaction costs it offsets. Specifically in P_1 at the 1-hour frequency, the return of the RANN is 51.2%, compared with 50.4% of the HLM. However, after transaction costs the return of RANN drops to 43.8% and that of HLM to 43.6%, while the HLM proves to be better in terms of statistical significance and directional quality.
[19] The reported results in Tables 2, 3 and 4 on the profitability (TR) may seem very high, but are absolutely justified as they correspond to a very large out-of-sample dataset (6336 observations out of a total sample of 56,449 5-min EUR/USD return observations). There are many studies in this line of literature that accord with these figures of profitability. For example, Fernández-Rodriguez et al. (2000) report 0.48 (or 48%) as total returns from a similar directional strategy in the period 10/2/91–10/2/92 including only 250 observations, or 1.5691 in the period 8/3/93 to 6/5/97 (approx. 1000 obs.) in another study (Fernández-Rodríguez et al., 1999). Similarly, Gençay et al. (2002) report an annualized total return of 9.630 for 1000 out-of-sample realizations of USD/DEM 5-min returns in the period 1/1/90–12/31/96.
[20] Please refer to the literature mentioned in the previous footnote also for comparable figures of directional quality.

Table 4
Out-of-sample tests for EUR/USD intraday returns (P2: 9/16/2004–10/1/2004).

Model frequency	NV		HLM		MSR		FFANN		RANN		ARFIMA	
	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour	5-minute	1-hour
TR	−6.066	0.172	5.349	2.111	1.821	−1.264	1.336	−2.235	−0.800	0.307	−5.932	−0.402
	(−6.849)	(0.106)	(4.680)	(2.045)	(1.335)	(−1.333)	(0.540)	(−2.300)	(−1.594)	(0.233)	(−6.694)	(−0.472)
DQ	0.418	0.490	0.515	0.536	0.481	0.471	0.477	0.444	0.472	0.502	0.415	0.463
HM test	−6.709	−0.815	6.075***	1.951**	0.640	−0.878	0.929	−2.324	−0.807	0.707	−6.301	−1.373
RMSE	0.039	0.129	0.024	0.096	0.036	0.086	0.074	0.123	0.033	0.085	0.025	0.086
SR	−1.237	0.120	1.073	1.487	0.364	−0.888	0.267	−1.576	−0.160	0.216	−1.216	−0.284
B&H return	0.022											

Notation as in Table 2.

with previous results derived by Fernández-Rodriguez et al. (2000) as well as with the conclusions reached by Christoffersen and Diebold (2006). It is also noticeable that in some cases an "active" trading strategy employed by sophisticated non-linear models compared with the naïve B&H, provides with worse even negative (loss) results. Nevertheless, the B&H strategy never outperforms HLM in terms of profitability.

8. Conclusions

The results validate the presence of technical predictability. It can be inferred that there is information content available in transaction data that can be used to predict high-frequency intraday trends. The forecast quality of the proposed model is evaluated on a very large tick-by-tick sample of 5-minute and 1-hour frequency. The robustness analysis, the extremely large number of observations and the statistical tests employed are convincing evidence that the heuristic learning model beats the trading rules based on many well-established non-linear models as well as the random walk in case of the EUR/USD intraday rates for short-term forecasting horizons. Technically, in terms of directional quality the proposed dynamically-adjusted learning model allows more precise and prompt identification of turning points. Model-wise, the dynamic update of expectations and preferences embedded in heuristic rules seems to efficiently mimic the decision-making mechanism of traders. Evidently, the incorporation of the trader's heterogeneous beliefs ("fads") and the adaptive calibration of the intensity- and range-of-belief parameters to match the agent's expectations, leads to optimal prediction under uncertainty. It seems also plausible that a B&H strategy would be the best in the extreme case of a pure trending market or in the absence of turning points in price movement, both of which would imply homogeneity and rationality in trader behavior. However, when there is uncertainty, turbulence and eventually heterogeneity caused by a number of factors that may affect market microstructure, the HLM will be better in terms of prediction performance. Overall, the results suggest that the heuristic learning system can be used to improve upon traditional technical analysis approaches. An interesting subject for future research would involve the incorporation of other sources of technical and fundamental economic signals to enhance learning.

Appendix A

Proof of Proposition 1. As it is mentioned, all ruled-based preferences are aggregated producing the output of the system as in Eq. (2)

$$y_t = \sum_i y_{it} = \overline{p}_{1t}\left(\alpha_{1t} + \beta_{1t}x_{1t} + \gamma_{1t}x_{2t}\right) + \overline{p}_{2t}\left(\alpha_{2t} + \beta_{2t}x_{1t} + \gamma_{2t}x_{2t}\right).$$

From Eq. (2) the output was re-formulated to a vector specification (Eq. (3))

$$y_t = \left[\overline{p}_{1t}x_{1t}\ \ \overline{p}_{1t}x_{2t}\ \ \overline{p}_{1t}\ \ \overline{p}_{2t}x_{1t}\ \ \overline{p}_{2t}x_{2t}\ \ \overline{p}_{2t}\right] \cdot \left[\beta_{1t}\ \ \gamma_{1t}\ \ \alpha_{1t}\ \ \beta_{2t}\ \ \gamma_{2t}\ \ \alpha_{2t}\right]^T = \mathbf{X}_t\mathbf{P}_t.$$

Direct matrix inversion techniques could be used on Eq. (3) to estimate the vector $\mathbf{P}_t$. However, this involves inverting input matrix $\mathbf{X}_t$ which is problematic when columns are dependent, or nearly independent. Direct inverting would mean that there are no serial autocorrelation and/or no noise in the input data, which is unrealistic especially in high-frequency financial applications. Other methods such as triangular or robust orthogonal decomposition are better in terms of under/over-determinacy, but often produce numerical instabilities and result in noise overfitting. In this study the singular value decomposition method (Golub and Reinsch, 1971; Golub and van Loan, 1989; Horn and Johnson, 1991) is proposed as a novel approach. This approach 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 $\mathbf{X}_t$ matrix is decomposed into a diagonal matrix $\mathbf{D}_t$ that contains the singular values, a matrix $\mathbf{U}_t$ of principal components, and an orthogonal normal matrix of right singular values $\mathbf{V}_t$. The eigenvalues in $\mathbf{D}_t$ are positive and arranged in a decreasing order. If there is heteroscedasticity, the number of significant principal components would be larger than one. Therefore to remove noise, the columns of $\mathbf{U}_t$ that correspond to small diagonal values in $\mathbf{D}_t$ are removed. The weight matrix is finally solved for as

$$\mathbf{P}_t = \mathbf{V}_t \mathbf{D}_t^{-1} \mathbf{U}_t^{\mathbf{T}} \mathbf{Y}_t$$

where $\mathbf{Y}_t$ is the vector of aggregated system output. Thus, via $\mathbf{P}_t$ the adaptive expectations of the traders are estimated in each learning episode. This completes the proof. ▪

Proof of Theorem 1. Based on Eqs. (6) and (7), the following chain rule is used to analyze the total derivative to its partial derivatives for the intensity-of-belief parameter

$$\partial E_t/\partial b_{it} = (\partial E_t/\partial y_t)(\partial y_t/\partial y_{it})(\partial y_{it}/\partial p_{it})\left(\partial p_{it}/\partial \mu_{B_{it}}\right)\left(\partial \mu_{B_{it}}/\partial b_{it}\right).$$

Similarly for the range-of-belief parameter the following is true

$$\partial E_t/\partial c_{it} = (\partial E_t/\partial y_t)(\partial y_t/\partial y_{it})(\partial y_{it}/\partial p_{it})\left(\partial p_{it}/\partial \mu_{B_{it}}\right)\left(\partial \mu_{B_{it}}/\partial c_{it}\right).$$

The partial derivatives are derived below

$$E_t = 1/N\left[(y-y^o)^2\right] \Rightarrow \partial E_t/\partial y_t = 2/N\left[(y_t-y_t^o)\right]$$

$$y_t = \sum_{i=1}^{2} y_{it} \Rightarrow \partial y_t/\partial y_{it} = 1$$

$$y_{it} = (p_{it}\pi_{it})\Big/\left(\sum_{i=1}^{2} p_{it}\right) \Rightarrow \partial y_{it}/\partial \pi_{it} = (\pi_{it}-y_t)/\sum_{i=1}^{2} p_{it}$$

$$p_{it} = \prod_{j=1}^{2} \mu_{B_{it}}\left(x_j\right) \Rightarrow \partial p_{it}/\partial \mu_{B_{it}} = p_{it}/\mu_{B_{it}}$$

$$\partial \mu_{B_{it}}\Big/\partial b_{it} = \begin{cases} 2sign\left(x_j-b_{it}\right)/c_{it}, & 2\left|\mathrm{x_j}-b_{it}\right| \le c_{it} \\ 0, & 2\left|\mathrm{x_j}-b_{it}\right| > c_{it} \end{cases}$$

$$\partial \mu_{B_{it}}\Big/\partial c_{it} = \left(1-\mu_{B_{it}}\right)\Big/c_{it}.$$

Substituting into the chain rule equations

$$\partial E_t\Big/\partial b_{it} = \left[4p_{it}sign\left(x_j-b_{it}\right)\left(y_t-y_t^o\right)\left(\pi_{it}-y_t\right)\right]\Big/\left(Nc_{it}\mu_{B_{it}}\sum_{i=1}^{2} p_{it}\right)$$

$$\partial E_t\Big/\partial c_{it} = \left[2p_{it}\left(1-\mu_{B_{it}}\right)\left(y_t-y_t^o\right)\left(\pi_{it}-y_t\right)\right]\Big/\left(Nc_{it}\mu_{B_{it}}\sum_{i=1}^{2} p_{it}\right)$$

where $sign(\arg) = 1$ if $\arg \succ 0$ and zero otherwise. Thus, the adaptive rule for the intensity-of-belief parameter is provided in the following recursive form (Eqs. (8), (9))

$$b_{it+1} = b_{it}-\tau_b\left[4p_{it}sign\left(x_j-b_{it}\right)(y_t-y_t^o)(\pi_{it}-y_t)\right]\Big/\left(Nc_{it}\mu_{B_{it}}\sum_{i=1}^{2} p_{it}\right)$$

whereas for the range-of-belief parameter:

$$c_{it+1} = c_{it}-\tau_c\left[2p_{it}\left(1-\mu_{B_{it}}\right)(y_t-y_t^o)(\pi_{it}-y_t)\right]\Big/\left(Nc_{it}\mu_{B_{it}}\sum_{i=1}^{2} p_{it}\right)$$

This completes the proof. ▪

References

Adya, M., Collopy, F., 1998. How effective are neural networks at forecasting and prediction? A review and evaluation. J. Forecast. 17, 481–495.
Andersen, T., Bollerslev, T., 1997. Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. J. Financ. 52, 975–1005.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. J. Am. Stat. Assoc. 96, 42–55.
Antsaklis, P.J., Passino, K.M., 1993. An Introduction to Intelligent and Autonomous Control. Kluwer Academic Publishers, Norwell, MA.
Attansio, O.P., 1991. Risk, time varying second moments and market efficiency. Rev. Econ. Stud. 58, 479–494.
Baillie, R.T., 1996. Long memory processes and fractional integration in econometrics. J. Econ. 73, 5–59.
Baillie, R.T., Bollerslev, T., 1991. Intraday and inter market volatility in foreign exchange rates. Rev. Econ. Stud. 58, 565–585.
Baillie, R.T., McMahon, P.C., 1989. The Foreign Exchange Market. Cambridge University Press, Cambridge.
Bank for International Settlements (BIS), 2007. Triennial Central Bank Survey: Foreign Exchange and Derivatives Market Activity in 2007, Basel, Switzerland.
Barndorff-Nielsen, O.E., Prause, K., 2001. Apparent scaling. Finance Stochast. 5, 103–113.
Bekiros, S., 2010. Heterogeneous trading strategies with adaptive fuzzy actor–critic reinforcement learning: a behavioral approach. J. Econ. Dyn. Control. 34 (6), 1153–1170.
Bessembinder, H., Chan, K., 1998. Market efficiency and the returns to technical analysis. Financ. Manag. 27 (2), 5–17.
Black, F., 1986. Noise. J. Financ. 41, 529–543.
Bolger, F., Harvey, N., 1995. Judging the probability that the next point in an observed time-series will be below, or above, a given value. J. Forecast. 14, 597–607.
Boothe, P., Glassman, D., 1987. The statistical distribution of exchange rates, empirical evidence and economic implications. J. Int. Econ. 22, 297–319.
Brock, W.A., Hommes, C.H., 1997. A rational route to randomness. Econometrica 65, 1059–1095.
Brock, W.A., Hommes, C.H., 1998. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. J. Econ. Dyn. Control. 22, 1235–1274.
Brock, W.A., Kleidon, A., 1992. Periodic market closure and trading volume: a model of intraday bids and asks. J. Econ. Dyn. Control. 16, 451–489.
Brock, W.A., Dechert, D., Scheinkman, J., 1987. A test for independence based on the correlation dimension. Social Science Research Working Paper, No. 8762. University of Wisconsin-Madison.
Brock, W.A., Hsieh, D., LeBaron, B., 1991. Nonlinear Dynamics, Chaos and Instability. The MIT Press, Cambridge, MA.
Brock, W.A., Lakonishok, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. J. Financ. 47, 1731–1764.
Buckley, J.J., Hayashi, Y., 1994. Fuzzy neural networks: a survey. Fuzzy Sets Syst. 66, 1–13.
Calderon-Rossel, J.R., Ben-Horim, M., 1982. The behavior of the foreign exchange rates, empirical evidence and economic implications. J. Int. Bus. Stud. 13, 99–111.
Calvet, L., Fisher, A., 2002. Multifractality in asset returns: theory and evidence. Rev. Econ. Stat. 84, 381–406.
Camillo, L., 2008. A combined signal approach to technical analysis on the S&P 500. J. Bus. Econ. Res. 6, 41–51.
Canegrati, E., 2008. A non-random walk down canary wharf. Technical Report. University Library of Munich, Germany.
Christoffersen, P., Diebold, F., 2006. Financial asset returns, direction-of-change forecasting, and volatility dynamics. Manag. Sci. 52, 1273–1287.
Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. Quant. Financ. 1, 223–236.
Dacorogna, M.M., Müller, U.A., Nagler, R.J., Olsen, R.B., Pictet, O.V., 1993. A geographical model for the daily and weekly seasonal volatility in the FX market. J. Int. Money Financ. 12, 413–438.
Dacorogna, M.M., Müller, U.A., Jost, C., Pictet, O.V., Olsen, R.B., Ward, J.R., 1995. Heterogeneous real-time trading strategies in the foreign exchange market. Eur. J. Financ. 1, 243–253.
Dacorogna, M.M., Gençay, R., Müller, U.A., Olsen, R.B., Pictet, O.V., 2001. An Introduction to High-frequency Finance. Academic Press.
De Lima, P.J.F., 1996. Nuisance parameter free properties of correlation integral based statistics. Econ. Rev. 15, 237–259.
De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Noise trader risk in financial markets. J. Polit. Econ. 98, 703–738.
Elman, J.L., 1990. Finding structure in time. Cogn. Sci. 14, 179–211.
Engle, R.F., Ito, T., Lin, W.L., 1990. Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. Econometrica 58, 525–542.
Ersoy, O., 1990. Tutorial at Hawaii International Conference on Systems Sciences.
Fama, E.F., 1991. Efficient capital markets: II. J. Financ. 46, 1575–1617.
Fama, E.F., Blume, M.E., 1966. Filter rules and stock-market trading. J. Bus. 39, 226–241.
Farmer, J.D., 1998. Market force, ecology, and evolution. Working Paper. Santa Fe Institute.
Farmer, J.D., Lo, A.W., 1999. Frontiers of finance: evolution and efficient markets. Proc. Natl. Acad. Sci. U. S. A. 96, 9991–9992.
Feinstone, L.J., 1987. Minute by minute: efficiency, normality and randomness in intra-daily asset prices. J. Appl. Econ. 2, 193–214.
Fernández-Rodríguez, F., Sosvilla-Rivero, S., García Artiles, M.D., 1999. Nearest neighbour forecast for the Nikkei index. Jpn. World Econ. 11, 395–413.
Fernández-Rodriguez, F., Gonzalez-Martel, C., Sosvilla-Rivero, S., 2000. On the profitability of technical trading rules based on artificial neural networks: evidence from the Madrid stock market. Econ. Lett. 69, 89–94.
Frances, P.H., van Dijk, D., 2000. Non-linear Time Series Models in Empirical Finance. Cambridge University Press, Cambridge.
Frankel, J.A., Froot, K.A., 1990. Chartists, fundamentalists, and trading in the foreign exchange market. Am. Econ. Rev. 80, 181–185.
Gençay, R., 1998a. Optimization of technical trading strategies with neural network models and evidence of profitability in security markets. Econ. Lett. 59, 249–254.
Gençay, R., 1998b. The predictability of security returns with simple technical trading rules. J. Empir. Financ. 5, 317–345.
Gençay, R., Stengos, T., 1998. Moving average rules, volume and the predictability of security returns with feed forward networks. J. Forecast. 17, 401–414.
Gençay, R., Ballocchi, G., Dacorogna, M.M., Olsen, R., Pictet, O.V., 2002. Real-time trading models and the statistical properties of foreign exchange rates. Int. Econ. Rev. 43 (2), 463–491.
Gençay, R., Dacorogna, M.M., Olsen, R., Pictet, O.V., 2003. Foreign exchange trading models and market behavior. J. Econ. Dyn. Control. 27, 909–935.
Gençay, R., Gradojevic, N., Sel uk, F., 2009. Profitability in an electronic foreign exchange market: informed trading or differences in valuation? Working Paper 25-09. The Rimini Centre for Economic Analysis, Italy.
Gençay, R., Gradojevic, N., Sel uk, F., Whitcher, B., 2010. Asymmetry of information flow between volatilities across time scales. Quant. Financ. 10 (8), 895–915.
Golub, G.H., Reinsch, C., 1971. Singular Value Decomposition and Least Squares Solutions: Handbook for Automatic Computation, Vol II: Linear Algebra. Springer, Heidelberg.
Golub, G.H., van Loan, C.F., 1989. Matrix Computations, 2nd ed. The Johns Hopkins University Press.
Gouree, J.K., Hommes, C.H., 2000. Heterogeneous beliefs and the nonlinear cobweb model. J. Econ. Dyn. Control. 24, 761–798.
Gradojevic, N., Gençay, R., 2010. Crash of '87 — was it expected? Aggregate market fears and long range dependence. J. Empir. Financ. 17 (2), 270–282.
Granger, C.W., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. J. Time Ser. Anal. 1, 15–29.
Green, H., Pearson, M., 1994. Neural nets for foreign exchange trading. Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets. Wiley, New York.
Guillaume, D.M., Dacorogna, M.M., Dave, R.D., Müller, U.A., Olsen, R.B., Pictet, O.V., 1997. From the bird's eye to the microscope: a survey of new stylized facts of the intradaily foreign exchange markets. Finance Stochast. 1, 95–129.
Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series subject to changes in regime. Econometrica 57, 357–384.
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, pp. 593–605.
Henriksson, R.D., Merton, R.C., 1981. On the market timing and investment performance II: statistical procedures for evaluating forecasting skills. J. Bus. 54, 513–533.
Hommes, C., 2000. Financial markets as nonlinear adaptive evolutionary systems. Quant. Financ. 1, 149–167.
Horn, R.A., Johnson, C.R., 1991. Topics in Matrix Analysis. Cambridge University Press, Cambridge.

Hornik, K., Stinchcombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Netw. 3, 551–560.
Hosking, J.R.M., 1981. Fractional differencing. Biometrika 68, 165–176.
Hsieh, D.A., 1988. The statistical properties of daily foreign exchange rates: 1974–1983. J. Int. Econ. 24, 129–145.
Hsu, P.-H., Kuan, C.-M., 2005. Reexamining the profitability of technical analysis with data snooping checks. J. Financ. Econ. 3, 606–628.
Irwin, S.H., Park, C.H., 2007. What do we know about the profitability of technical analysis? J. Econ. Surv. 21, 786–826.
Ishibuchi, H., Kwon, K., Tanaka, H., 1995. A learning algorithm of fuzzy neural networks with triangular fuzzy weights. Fuzzy Sets Syst. 71, 277–293.
Jamshidi, M., Titli, A., Zadeh, L., Boverie, S., 1997. Applications of Fuzzy Logic. Prentice Hall, New Jersey.
Jasic, T., 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. Appl. Financ. Econ. 14, 285–297.
Jensen, M.J., Whitcher, B., 2000. Time-varying long-memory in volatility: detection and estimation with wavelets. Working Paper.
Kao, G.W., Ma, C.K., 1992. Memories, heteroscedasticity and prices limit in currency futures markets. J. Futur. Mark. 12, 672–692.
Katz, J.O., 1992. Developing neural network forecasters for trading. Technical Analysis of Stocks and Commodities, pp. 58–70.
Kaufman, J.P., 1998. Trading Systems and Methods. John Wiley and Sons, New York.
Kirkpatrick, C.D., Dahlquist, J.R., 2007. Technical Analysis: The Complete Resource for Financial Market Technicians. Prentice Hall, Upper Saddle River, New Jersey.
Kuan, C.-M., White, H., 1994. Artificial neural networks: an econometric perspective. Econ. Rev. 13, 1–91.
Kurz, M., 1994. Endogenous economic fluctuations. Studies in the Theory of Rational BeliefsStudies in Economic Theory vol. 6. Springer, Berlin.
La Porta, R., Lakonishok, J., Shliefer, A., Vishny, R., 1997. Good news for value stocks: further evidence on market efficiency. J. Financ. 52, 859–874.
Lawrence, M., O'Connor, M., 1992. Exploring judgemental forecasting. Int. J. Forecast. 8, 15–26.
LeBaron, B., 1996. Technical trading rule profitability and foreign exchange intervention. Technical Report 5505. National Bureau of Economic Research.
LeBaron, B., 1997. Technical trading rules and regime shifts in foreign exchange. In: Acar, E., Satchell, S. (Eds.), Advanced Trading Rules. Butter-vorth-Heinemann.
Levich, R.M., Thomas, L.R., 1993. The significance of technical trading rule profits in the foreign exchange market: a bootstrap approach. J. Int. Money Financ. 12, 451–474.
Liu, Y., Gopikrishnan, P., Cizeau, Meyer, Peng, Stanley, H.E., 1999. Statistical properties of the volatility of price fluctuations. Phys. Rev. E. 60, 1390–1400.
Lo, A.W., Mamaysky, H., Wang, J., 2000. Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. J. Financ. 55, 1705–1765.
Loretan, M., Phillips, P.C.B., 1994. Testing the covariance stationarity of heavy-tailed time series. J. Empir. Financ. 1, 211–248.
Lux, T., Marchesi, M., 1999. Scaling and criticality in a stochastic multi-agent model of a financial market? Nature 397, 498–500.
Lyons, R.K., 2001. The Microstructure Approach to Exchange Rates. MIT Press, Cambridge, Massachusetts.
MacDonald, R., Taylor, M.P., 1992. Exchange rate economics. IMF Staff. Pap. 39, 1–57.
Mamdami, E., 1974. Applications of fuzzy algorithms for control of simple dynamic plant. Proceedings IEE. 121, pp. 1585–1588.
Mamdani, E., 1977. Application of fuzzy logic to approximate reasoning using linguistic systems. IEEE Trans. Comput. 26, 1182–1191.
Meese, R.A., Rogoff, J., 1983. Empirical exchange rate models of the seventies, do they fit out of sample? J. Int. Econ. 14, 3–24.
Morris, S., 1996. Speculative investor behavior and learning. Q. J. Econ. 111 (4), 1111–1133.
Müller, U.A., Dacorogna, M.M., Olsen, R.B., Pictet, O.V., Schwarz, M., Morgenegg, C., 1990. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. J. Bank. Financ. 14, 1189–1208.
Murphy, J.J., 1999. Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. Prentice Hall Press, New Jersey.
Neftci, S.N., 1991. Naive trading rules in financial markets and Wiener–Kolmogorov prediction theory: a study of technical analysis. J. Bus. 64, 549–571.
Nevmyvaka, Y., Feng, Y., Kearns, M., 2006. Reinforcement learning for optimized trade execution. ICML '06. Proceedings of the 23rd International Conference on Machine Learning. ACM Press, pp. 673–680.
Nguyen, D., Widrow, B., 1989. The truck backer-upper: an example of self-learning in neural networks. Proceedings of the International Joint Conference on Neural Networks. 2, pp. 357–363.
Nishina, T., Hagiwara, M., 1997. Fuzzy Inference Neural Network. Neurocomputing 14, 223–239.
Petersen, M.A., Fialkowski, D., 1994. Posted versus effective spreads, good prices or bad quotes? J. Financ. Econ. 35, 269–292.
Pictet, O.V., Dacorogna, M.M., Müller, U.A., Olsen, R.B., Ward, J.R., 1992. Realtime trading models for foreign exchange rates. Neural Netw. World 2, 713–744.
Robinson, P. (Ed.), 2003. Time Series with Long Memory. Oxford University Press, Oxford, UK.
Shiller, R.J., 1989. Market Volatility. The MIT Press, Cambridge, MA.
Shiller, R.J., 2002. From efficient market theory to behavioral finance. Cowles Foundation Discussion Paper No. 1385.
Shleifer, A., Summers, L.H., 1990. The noise trader approach to finance. J. Econ. Perspect. 4, 19–33.
Sugeno, M., 1985. Industrial Applications of Fuzzy Control. Elsevier Science Publications.
Sugeno, M., 1988. Fuzzy Control. Nikkan Kougyou, Shinbunsha.
Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. Syst. Man Cybern. 15, 116–132.
Taylor, M.P., Allen, H., 1992. The use of technical analysis in the foreign exchange market. J. Int. Money Financ. 11, 304–314.
Von Altrock, C., 1997. Fuzzy Logic and Neurofuzzy Applications in Business and Finance. Prentice Hall, Upper Saddle River, New Jersey.
Wasserfallen, W., 1989. Flexible exchange rates: a closer look. J. Monet. Econ. 23, 511–521.
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.
Westerfield, R., 1997. The distribution of common stock price changes: an application of transactions time and subordinated stochastic models. J. Financ. Quant. Anal. 12, 743–765.
White, H., 1989. Learning in artificial neural networks: a statistical perspective. Neural Comput. 1, 425–464.
Widrow, B., Hoff, M.E., 1960. Adaptive switching circuits. IRE WESCON Convention Record. IRE, New York, pp. 96–104.
Yao, J.T., Poh, H.-L., Jasic, T., 1996. Foreign exchange rates forecasting with neural networks. Proceedings of the International Conference on Neural Information Processing, Hong Kong, pp. 754–759.
Zhang, X.R., 1994. Non-linear predictive models for intra-day foreign exchange trading. Int. J. Intell. Syst. Account. Financ. Manag. 3, 293–302.
Zhang, G., Patuwo, B.E., Hu, M.Y., 1998. Forecasting with artificial neural networks: the state of the art. Int. J. Forecast. 14 (1), 35–62.
Zimmerman, H.-J., Thole, U., 1978. On the suitability of minimum and product operators for the intersection of fuzzy sets. Fuzzy Sets Syst. 2, 173–186.