

High frequency trading from an evolutionary perspective: financial markets as adaptive systems

Article

Accepted Version

Manahov, V., Hudson, R. and Urquhart, A. ORCID:
<https://orcid.org/0000-0001-8834-4243> (2019) High frequency trading from an evolutionary perspective: financial markets as adaptive systems. *International Journal of Finance and Economics*, 24 (2). pp. 943-962. ISSN 1099-1158 doi:
<https://doi.org/10.1002/ijfe.1700> Available at
<https://centaur.reading.ac.uk/79183/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/ijfe.1700>

Publisher: Wiley

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

High frequency trading from an evolutionary perspective: Financial markets as adaptive systems

Abstract

The recent rapid growth of algorithmic High Frequency Trading (HFT) strategies makes it a very interesting time to revisit the long standing debates about the efficiency of stock prices and the best way to model the actions of market participants. To evaluate the evolution of stock price predictability at the millisecond timeframe and to examine whether it is consistent with the newly formed Adaptive Market Hypothesis (AMH), we develop three artificial stock markets using a Strongly Typed Genetic Programming (STGP) trading algorithm. We simulate real-life trading by applying STGP to millisecond data of the three highest capitalised stocks: Apple, Exxon Mobil and Google and observe that profit opportunities at the millisecond timeframe are better modeled through an evolutionary process involving natural selection, adaptation, learning and dynamic evolution than by using conventional analytical techniques. We use combinations of forecasting techniques as benchmarks to demonstrate that different heuristics enable artificial traders to be ecologically rational, making adaptive decisions that combine forecasting accuracy with speed.

Keywords: Genetic Programming, Evolutionary Computation, Market Efficiency, Adaptive Market Hypothesis, High Frequency Trading.

JEL Classification: F3, G10, G12, G14, E47.

1.Introduction

Due to the advances in technology and the rapid growth of high frequency trading (HFT), advanced financial markets have substantially eliminated human intermediation in the trading process and replaced it with automated electronic limit order books which have allowed the growth of trading algorithms as one of the main investment tools. Some of the trading algorithms generated imitate the behaviour of humans in the trading process and over the last few years, these trading algorithms have substantially improved their speed to match the incidence of bid and ask orders (McGowan, 2010).

Concerns over financial market stability and the equitable treatment of all market participants have prompted renewed interest in market quality and market regulators are still debating whether or not HFT is beneficial or harmful to market efficiency¹. Since the seminal work of Fama (1965, 1970) who introduced the efficient market hypothesis (EMH), there has been a plethora of studies analysing market efficiency and adaptability. Recently this topic has been brought into new focus by the rapid growth of profitable HFT (Baron *et al.* 2012). However, many studies of market efficiency have significant deficiencies such as data-snooping (Sullivan *et al.* 1999), ex-post selection of profitable trading rules (Zakamulin 2014), false discoveries (Harvey and Liu 2014), inattention to transaction costs (Park and Irwin (2007) as well as studies basing their conclusions on econometric tests or theoretical hypotheses which treat market properties as essentially static, failing to consider the evolutionary processes of adaptation, learning and survival of market participants (Campbell *et al.* 1997). Lo (2004) argues that market outcomes are obtained not in an analytical way, but through an evolutionary process of trial-and-error and natural selection. The process of natural selection enables survival-of-the fittest and determines the composition of market participants and their trading strategies. In contrast, this study implements a special adaptive form of the Strongly Typed Genetic Programming (STGP) utilising historical millisecond data of the most capitalised companies: Apple, Exxon Mobil and Google. The STGP (described in the Appendices) represents a sophisticated trading algorithm that successfully replicates real-life trading strategies performed at the millisecond timeframe. Using STGP, we compare the forecasting performance of our artificial traders with several combined forecasting methods. In other words, we simulate real-life trading sessions which allow us to avoid the obstacles in the studies discussed above. Given the arguments of Lo (2004; 2005) that financial markets are governed by evolutionary processes, STGP is an extremely suitable approach to investigate market efficiency. This is due to the fact that all 100,000 artificial traders in each of the three stock markets in our experiment compete, learn, adapt, evolve and try to survive. The environment of heterogeneous traders where stock prices and traders' beliefs co-evolve over time provides an appropriate laboratory platform to

¹ While Zhang (2010), Kirilenko *et al.*, (2011), Biais *et al.*, (2013), Foucault *et al.*, (2013) show that HFT can have harmful effects on market quality, Domowitz (2010), Brogaard *et al.*, (2014), Boehmer *et al.*, (2012), Jovanovic and Menkveld (2012) Martinez and Rosu (2013) Carrion (2013), Hansbrouck and Saar (2013) argue that HFT play a positive role on market quality.

investigate market efficiency. Hommes (2011) suggests that adaptation and learning in heterogeneous structures are important tools for analysis of financial market behaviour. Moreover, the random nature of the initial trading rules of all traders and their subsequent evolution allows us to observe directly the processes described by Lo (2004; 2005). Therefore, the aim of this study is to evaluate the evolution of stock price predictability at the millisecond timeframe and to examine whether this evolution is consistent with the notion of adaptive markets.

The contribution of this study is three fold. First, this is the first study to use an innovative trading algorithm and real-life millisecond data to provide concrete empirical evidence of market adaptability at the millisecond timeframe. We estimate in precise quantitative terms the daily profits of high frequency traders (HFTs) after taking into account realistic transaction costs, providing an advantage over existent studies, such as that of Brogaard (2013) and Carrion (2013) which observed the activities of HFT using only aggregate data, thus preventing them from calculating the exact profitability. In order to measure the statistical accuracy and trading efficiency, we compare the predictive ability of our artificial traders to benchmarks of forecast combination approaches such as Support Vector Regression (SVR), the Least Absolute Shrinkage and Selection Operator (LASSO) and the Kalman Filter (KF).

Second, we take into account all the issues in previous studies as potentially affecting the reliability of trading results. The presence of 100,000 artificial traders in our experiment ensures forecasting model stability and lower sensitivity to random factors. All artificial traders learn from their experience, evaluating the profitability of trading rules based on their predictive power rather than in-sample fit. We also avoid data-snooping biases by ensuring all trading rules are evaluated and executed by artificial traders.

Third, we observe that various heuristics enable artificial traders to be ecologically rational, making adaptive decisions that combine forecasting accuracy with speed. We have found that market participants learn, compete, adapt, survive and evolve toward a higher degree of sophistication. These findings suggest that our artificial stock markets can be modelled as evolving ecological systems which consist of large numbers of heterogeneous traders competing for profits on each market. Such ecological systems not only experience varying degrees of efficiency but also cycles of efficiency and adaptability as changes in profit opportunities lead to shifts in the composition of market participants. Therefore, we find strong evidence to show that market participants can be modeled in a way supportive of the Adaptive Market Hypothesis.

The remainder of this paper is organised in the following way: Section 2 comprises the literature review, while Section 3 presents the experimental design of the three artificial stock markets, the forecasting models and data utilised in this study. Section 4, reports the artificial agents' trading activity and profitability, while Section 5 presents the conclusions. Additional clarifying and technical material can be found in the Appendices.

2. Literature review

2.1. The Adaptive Market Hypothesis (AMH)

The AMH, formulated by Andrew Lo (2004, 2005), argues that many of the behavioural biases in finance are in fact consistent with an evolutionary model of investors learning and adapting to a changing environment. It is the impact of these evolutionary forces on financial institutions and market participants that determines the efficiency of markets, and the performance of investments, businesses and industries. The principles of AMH are that; (1) Individuals act in their own self-interest, (2) Individuals make mistakes, (3) Individuals learn and adapt, (4) Competition drives adaptation and innovation, (5) Natural selection shapes market ecology and (6) Evolution determines market dynamics. These principles offer a number of practical implications within finance. Firstly, the risk premium varies over time due to the stock market environment and demographics of investors in that environment. For example until recently, US markets were populated with investors who had never experienced a genuine bear market, which no doubt has shaped their aggregated risk preference. Thus irrespective of whether prices fully reflect all available information, the particular path dependency that market prices have taken over the past few years influences their current aggregate risk preferences. The second implication that is contrary to the EMH is that arbitrage opportunities do exist from time to time in the market. Lo (2004) cites Grossman and Stiglitz (1980) who observe that without such profit opportunities, there would be no incentive to gather information, and the price discovery aspect of financial markets would collapse. Thus from an evolutionary viewpoint, active liquid financial markets imply that profit opportunities must exist. However as they are exploited, they disappear. But new opportunities are continually being created as certain species of trader die out and rather than move towards a higher efficiency, the AMH implies that complex market dynamics such as trends, panics, bubbles and crashes are continually witnessed in natural market ecologies. The third implication is that investment strategies are successful or unsuccessful, depending on the particular market environment. Contrary to the EMH, the AMH implies that strategies may decline for a time, and then return to profitability when environmental conditions become more conducive to such trades. The final implication of the AMH is that characteristics such as value and growth may behave like 'risk factors' from time to time, that is, stocks with these characteristics may yield higher expected returns during periods when those attributes are in favour. Lo (2005) argues that convergence to equilibrium is neither guaranteed nor likely to occur, and that it is incorrect to assume that the market must move towards some ideal state of efficiency. Instead, the AMH relies on more complex market dynamics, such as cycles, trends, crashes and bubbles.

The AMH has gained substantially attention in the recent literature, with a number of papers supporting the hypothesis in developed stock markets (for example Ito and Sugiyama 2009; Kim

et al. 2011; Urquhart and Hudson 2013; Urquhart and McGroarty 2014 & 2016), developing stock markets (Smith 2012; Dyakova and Smith 2013a & 2013b; Hull and McGroarty 2014; Smith 2016), foreign exchange markets (Charles *et al.* 2012; Levich and Poti 2015) and even precious metal markets (Charles *et al.* 2015; Urquhart 2016).

Relatively little research has applied Genetic Programming to questions related to market efficiency and the AMH. In terms of testing market efficiency using genetic algorithms, Neely *et al.* (1997) provide evidence of the effectiveness of genetic algorithms in the foreign exchange markets although a subsequent paper by Neely (2003) is less supportive of their value in the stock market. Allen and Karjalainen (1999) use genetic algorithms to find technical trading rules using daily data for the S&P 500 index from 1928 to 1995. They find little evidence of economically significant rules although they do consider that more advanced genetic algorithms are worthy of investigation as are more liquid markets. The rapid advances in genetic algorithm design and reductions in market trading costs seen in recent years make revisiting this area seem appropriate. Manahov and Hudson (2014a) develop artificial stock markets using a special adaptive form of the Strongly Typed Genetic Programming based learning algorithm and apply it to data from the FTSE 100, S&P 500 and Russell 3000. They show that the stock market dynamics are consistent with the evolutionary process of the AMH since trader populations behave as an efficient adaptive system over time.

2.2. Forecasting with Genetic Programming (GP)

A number of studies have used GP to investigate trading rules and forecast stock markets with varying degrees of success. Allen and Karjalainen (1999) find that out-of-sample trading rules do not earn excess returns compared to a buy and hold strategy after taking into account transaction costs although they suggest their results were based on a very basic algorithm. Iba and Sasaki (1999) demonstrate that GP possess significant forecasting accuracy compared to neural networks in an experiment using data of Nikkei 225. Kaboudan (2000) implement GP to generate one-day-ahead forecasts and design a successful single day-trading strategy in which decisions are based on GP forecasts of lowest and highest daily prices. Becker and Seshadri (2003) used monthly data to design specific trading rules that were able to outperform a buy-and-hold strategy when dividends are not included in stock returns. Potvin *et al.* (2004) adopted GP to investigate the Canadian TSE 300 index and show that GP trading rules are accurate when the market falls or when it is stable.

Other studies apply GP to foreign exchange and derivative markets. Neely *et al.* (1997) examine the foreign exchange market with GP to find strong evidence of economically significant *ex ante* excess returns for six different exchange rates between 1981 and 1995.

In another study, Neely and Weller (2003) investigate the performance of GP using half-hourly high-frequency data of four currency pairs and found no profitability. Chen *et al.* (1998) and Chidambaran *et al.* (1998) use GP to develop option pricing mechanisms which perform much better than the Black-

Scholes formula. Bhattacharyya *et al.* (2002) obtained one-hour data for the USD/DEM currency pair and GP mechanism to find reliable predictive performance of the model. Banzhaf *et al.* (1997), Chen and Yeh (1997) and Farnsworth *et al.* (2004) provide additional evidence that forecasting and trading can be performed successfully using GP.

In summary, recent empirical studies in a wide variety of markets find strong evidence consistent with the AMH, which supports the theoretical discussion of Soufian *et al.* (2014) who argue that the AMH gives a theoretical basis for a new financial paradigm which better describe the behaviour of financial markets. In contrast, GP has, as yet, been little used to explore these issues.

3. Experimental design

We use a special adaptive form of the Strongly Typed Genetic Programming (STGP), similar to Manahov *et al.* (2014b) and Manahov (2016), which enables us to choose and adjust different parameters to suit our specification, such as the minimum price increment, number of participants (agents) and their wealth, the level of transaction costs, and differing trading preferences. The exact number of evolutionary parameters that we can specify is listed in Table 1². Each market participant represents an artificial trader who is equipped with their own trading rule. Each trader has initial wealth of \$100,000 making it meaningful to consider subsequent profits in monetary terms. The selection of the best performing traders and the production of the new genomes is conducted through the recombination of the parent genomes by crossover and mutation operations, which are further elaborated in the Appendices. The basic idea is that the trader's trading rule will improve by a natural selection process based on the survival of the fittest. Therefore the evolutionary nature of the trading process and price dynamics enable the artificial traders to recognize, learn and exploit profit opportunities while continually adapting to the changing market conditions. Consequently, the STGP trading algorithm evolves the model step-by-step by feeding it with real time millisecond quotes of Apple, Exxon Mobil and Google stocks, and therefore the forecasting models evolve by referencing to the characteristics of the actual markets studied.

3.1. The process of developing the initial trading rules

Each individual trader has only one trading rule (examples of trading rules for both trading groups can be found in the Appendices) which are initially created randomly enabling the whole range of possible trading rules to be studied. To create later generations, we apply the crossover recombination technique and mutation operation, where the crossover recombination technique randomly chooses parts of two trading rules to exchange in order to create two new trading rules, and the mutation operation randomly changes a small part of a trading rule (Dempster and Jones, 2001). This process is repeated until at least

² We used Altreva Adaptive Modeler (<http://www.altreva.com>) as our software platform to create artificial traders and generate their trades.

one trading rule in the population achieves the desired level of fitness, measured by a trader's investment return over a specified period. Each trading rule in our artificial stock market setting take real-time millisecond prices of Apple, Exxon Mobil and Google and generate advice which consists of the desired position which is defined as a percentage of the trader's wealth and an order limit price for buying and selling the financial instrument³. The logic of the trading rules uses information on price and volume, minimum, maximum and average functions related to millisecond price and trading volume data, and different logical and comparison operators. In the conventional Genetic Programming (GP) procedure, trading rules are evaluated by the same fitness function in each generation. In contrast, the STGP evaluates the fitness of traders through a dynamic fitness function, which enables the return estimation period to move forward and include the most recent quotes in the markets. Sermpinis *et al.*, (2015) notes that having a novel fitness function is crucial in financial modelling, where statistical accuracy does not always correspond to financial profitability of the deriving forecasts. Also, while the GP replaces the entire genetic population through crossover and mutation techniques at a time, STGP only replaces a small proportion of the entire population which ensures a gradual change in population and thus greater model stability⁴ (Manahov, 2016).

Another important feature of the STGP is that each trader discovers the intrinsic value of each stock individually without any communication between traders, ensuring individuality and that the level of intelligence of each artificial trader is not affected by other traders. To more realistically model the features of HFT we consider two groups of traders with different properties: scalpers and strategic informed traders as discussed in sub-section 3.2.

3.2. Structure of the artificial stock market and the differences between HFT scalpers and strategic informed traders

We examine the profitability of HFT strategies and their relationship with the Adaptive Market Hypothesis (AMH) within the context of a number of markets which are populated by up to 100,000 bounded rational traders. We develop different artificial stock markets for each of the three financial instruments and all artificial traders are created using the STGP programming technique documented in the Appendices. None of the traders are orientated towards any particular strategy and are therefore free to develop and continually evolve better trading rules over time.

All of the traders constantly learn, adapt and evolve and the survival-of-the-fittest principle in place means the elimination of the worst performing traders, where performance is measured by their Breeding Fitness Return

³ This process is further explained in subsections 3.2 and 3.3.

⁴ Trading models are built incrementally, which means that there is no overfitting on historical data. Trading signals are based on the activities of all agents and not just on a single trading rule. Forecasting models are constantly evolving and adapting to market changes instead of being static.

The Breeding Fitness Return is a trailing return which is calculated over the n quotes of data of an exponential moving average of traders' wealth, where n is set to the minimum breeding age with a maximum of 250⁵. This return is used to measure the fitness criterion for the selection of traders to breed, where breeding is a process of creating new artificial traders to replace poor performing ones. All artificial traders accumulate wealth by investing in the financial instruments available, namely the three risky stocks and the risk-free instrument represented by cash. Since the markets continuously evolve, artificial traders who perform well will become wealthier which will positively influence the forecasting accuracy of the model. We follow Manahov (2016) to estimate the artificial trader wealth:

$$W_{i,t} = M_{i,t} + P_t h_{i,t} \quad (1)$$

where $W_{i,t}$ is the wealth accumulated by trader i in period t ; $M_{i,t}$ and $h_{i,t}$ represents the money and the amount of security held by artificial trader i respectively, in period t , and P_t is the price of the asset in period t . We incorporate two different types of trader in our analysis: scalpers and strategic informed traders. The main difference between the two trading groups is that the HFT scalpers group consists of traders that momentarily perform best in terms of the continuous Breeding Fitness Return, and therefore possess lower latency. Although the strategic informed traders and HFT scalpers both observe the same millisecond data of the three stocks and generate trading orders, HFT scalpers are able to access and process the data first due to their low latency features. In other words, HFT scalpers are able to foresee the quotes of the three financial instruments and submit trading orders before strategic informed traders. Another difference between the two trading groups is that HFT scalpers are designed to capture and avoid sweep risk, which is the risk related to trading against large informed toxic orders (for instance, large institutional orders) positioned at multiple levels of the order book (Manahov,2016). This means that HFT scalpers are highly risk adverse and will not hold what they perceive as potentially risky positions, and aim for high frequency turnover to neutralise unanticipated adverse flows. Large losses due to sweeps can mitigate any profitability, so management of sweep risk is of paramount importance for HFT scalpers.

3.3. The clearing mechanism and order generation for the three artificial stock markets

The three artificial stock markets are simulated double auction markets, where all the buy and sell orders are collected. The artificial traders receive real-time quotes of Apple, Exxon Mobil and Google and evaluate their trading rule and subsequently calculate the number of shares they need to buy or sell. If shares need to be bought or sold, an order is generated to buy or sell the required amount determined by the specified limit price. For example, if a trader holds 2,000 shares of Google which is priced at \$22 and has \$70,000 in cash, their wealth is \$114,000 and their position in Google is 38.5%. If the

⁵ In the case where the age is less than n , no value is calculated.

trading rule generates a signal of a position of 60% and a limit price of \$22, the limit order will be produced to purchase 1,109⁶ additional Google shares with a price of \$22. The three artificial markets then calculate the clearing price and all trading orders are executed at the clearing price which is where the highest trading volume from limit orders can be matched. In cases when the clearing price can be matched at multiple price levels, then the clearing price is the average of the lowest and highest of those prices. The number of shares purchased by traders is always equal to the number sold by other traders and if the number of shares offered and the number of shares asked are not equal, the remaining orders will be partially executed. Therefore the orders at the clearing price will be selected for execution with priority for market orders over limit orders, and then on a first-in-first-out (FIFO) basis. In the unlikely event of no matching limit orders, no market orders are executed and the artificial stock market price will be the price of the previous quote (Manahov,2016).

3.4. Benchmark forecasting methods.

We follow Sermpinis *et al.*, (2014) in our selection of benchmark forecasting methods. We implement hybrid forecasting models including Support Vector Regression (SVR), Least Absolute Shrinkage and Selection Operator (LASSO) and the Kalman Filter as benchmark forecasting models because they offer linear as well as non – linear modelling features. The idea of combining different forecasting models to improve prediction accuracy originates from Bates and Granger (1969), who suggest that forecast combinations can be viewed as a simple and effective way to improve and justify the forecasting ability over that provided by individual models.

3.4.1. Support vector regression (SVR).

SVR is based on the statistical learning theory which is related to the properties of learning machines enabling them to generalise effectively in response to data. The associated learning algorithms are able to analyse data used for classification and regression analysis. In simple terms, when used for regression, the method aims to find a function that differs from the dependent variable by less than a certain amount whilst being as simple (‘flat’) as possible. SVR functionality is based on the computation of a linear regression function in a high dimensional feature space where the input data are mapped via a nonlinear function (Basak *et al.*2007). The main advantage of SVR is that instead of minimising the observed training error, SVR attempts to minimise the generalized error bound so as to achieve enhanced forecasting performance. Suykens *et al.*, (2002) argue that SVR forecasting models do not experience multiple local minima and are capable of dealing with the high – dimensional, noisy and complex problems that are found with millisecond data. Vapnik (1995) introduced the ϵ -sensitive loss function making SVR a robust technique for constructing data – driven and non – linear regression

⁶ 60% * (114,000/22) – 2000 = 1,109 shares.

models. According to Sermpinis *et al.*, (2014) SRV forecasts offer a balance between model accuracy and model complexity.

Every SVR consists of ε -SVR and ν -SVR algorithms. To describe the ε -SVR algorithm we consider the training data $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where $x_i \in X \subseteq R$, $y_i \in Y \subseteq R$, $i = 1, \dots, n$ and n is the total number of trading samples. The SVR function can be expressed as:

$$f(x) = w^T \varphi(x) + b \quad (2)$$

where w and b represent the two regression parameter vectors of SVR and $\varphi(x)$ represents the nonlinear function that positions the input data vector x to a feature space where the data used for training is linear. Equation (2) can be represented as a dial problem and its solution is based on the introduction of two Lagrange multipliers a_i, a_i^* and mapping with the kernel function $K(x_i, x)$:

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) K(x_i, x) + b \quad \text{where } 0 \leq a_i, a_i^* \leq C \quad (3)$$

Factor b in Equation (3) has been estimated following the Karush – Kuhn – Tucker conditions (Vapnik, 1995).

The ν -SVR algorithm consists of the ε parameter and the new parameter $\nu \in (0, 1)$. In terms of ν -SVR we can transform the optimization problem to:

$$\text{Minimize } C \left(\nu \varepsilon + \frac{1}{n} \sum_{i=1}^n (\xi_i + \xi_i^*) \right) + \frac{1}{2} \|w\|^2 \quad (4)$$

where ξ_i and ξ_i^* are two slack variables and $\|w\|^2$ represent the complexity (flatness) of the model. The solution of the quadratic optimization problem described in Equation (4) can be represented in the following way:

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) K(x_i, x) + b \quad \text{where } 0 \leq a_i, a_i^* \leq \frac{C}{n} \quad (5)$$

Parameter selection is an important issue in SVR, with the lack of information related to the noise of the trading datasets making priori ε -margin setting a difficult task (Sermpinis *et al.*, 2014). Prior studies

have highlighted four different approaches to deal with this issue, with Trafalis and Ince (2000) suggesting setting ϵ as a non-negative constant ($\epsilon=0$ or alternatively equal to a very small number), whereas Smola *et al.* (1998) recommend choosing ϵ based on the statistical efficiency of a location parameter maximization. Cao *et al.* (2003) advise that we need to determine ϵ by implementing the cross – validation technique, while Basak *et al.* (2007) argue that ϵ can be controlled via ν -SVR. In our study we use the kernel function K , and therefore we need to estimate two kernel – independent parameters (ν and C) and the RBF kernel parameter (γ):

$$K(x_i, x) = \exp\left(-\gamma\|x_i - x\|^2\right), \quad \gamma > 0 \quad (6)$$

We have implemented the RBF kernels in our experiment because they effectively overcome the issue of overfitting of data and are very precise in directional accuracy. In order to determine the optimal value for the two kernel – independent parameters (ν and C), we follow Cherkassky and Ma (2004) standard parameterization of the SVR solution:

$$|f(x)| \leq \left| \sum_{i=1}^{n_{sv}} (a_i - a_i^*) K(x_i, x) \right| \leq \sum_{i=1}^{n_{sv}} |a_i - a_i^*| |K(x_i, x)| \leq \sum_{i=1}^{n_{sv}} C |K(x_i, x)| \quad (7)$$

For $K(x_i, x) = \exp\left(-\gamma\|x_i - x\|^2\right) \leq 1$, we estimate the upper level of the SVR function as $|f(x)| \leq C n_{sv}$. Hence, we calculate C as:

$$C = \max\left(\left|\bar{y} + 3\sigma_y\right|, \left|\bar{y} - 3\sigma_y\right|\right) \quad (8)$$

where \bar{y} , σ_y represent the mean and the standard deviation of the millisecond data. Similarly to Sermpinis *et al.*, (2014) we use 5-fold cross – validation to calculate the values of ν and γ .

3.4.2. Least absolute shrinkage and selection operator (LASSO).

LASSO is a regression analysis technique that implements variable selection and regularisation in order to improve the prediction accuracy and interpretability of the prediction model it produces. In broad terms, it modifies the model fitting process to select only a subset of the possible independent variables compared to traditional regression methods. LASSO improves prediction error by shrinking large regression coefficients which reduces overfitting, which is a major pitfall in forecasting. Essentially LASSO works by forcing the sum of the absolute value of the regression coefficients to be less than a fixed value.

Consequently certain coefficients become set to zero and LASSO chooses a simpler predictive model that does not include those coefficients. We estimate the LASSO coefficients by using Breiman's non-negative garrotte minimization process (Yuan and Lin, 2007):

$$\beta_{LASSO} = \arg \min_{\beta} \left\{ \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^d \beta_j x_{ij} \right)^2 \right\} \text{ subject to } \sum_{j=1}^d |\beta_j| \leq k, \quad k > 0 \quad (9)$$

where k is the tuning parameter which monitor the amount of shrinkage applied to the coefficients.

3.4.3. Kalman Filter (KF).

The KF is very powerful tool for data filtering and predicting the future state of a system based on the previous ones. The KF uses different measurements of a group of variables observed over time, partly consisting of statistical noise and other inaccuracies, and generates estimates of unknown values of the variables that are more precise than those based on a single measurement alone, by computing a joint probability distribution over the variables for each timeframe. The KF is essentially a set of mathematical equations that consists of predictor-corrector type estimator that is optimal when some presumed conditions are met. The KF infers parameters of interest from indirect and uncertain observations and generates good results in practice due to optimality and structure. The KF in our experiment consists of the following measurement and state equations:

$$f^t_{c_{NNS}} = \sum_{i=1}^3 a_i^t f_i^t + \varepsilon_t, \quad \varepsilon_t \square NID(0, \sigma_{\varepsilon}^2) \quad (10)$$

$$a_i^t = a_i^{t-1} + n_t, \quad n_t \square NID(0, \sigma_n^2) \quad (11)$$

where $f^t_{c_{NNS}}$ represent the dependent variable or the combination forecast at time t ; f_i^t ($i=1,2,3$) denote the independent variables (individual forecasts) at time t ; a_i^t ($i=1,2,3$) describe the time – varying coefficients at time t ; ε_t and n_t measure the two noise terms.

We calculate all a_i^t in time, with the log-likelihood of the model taking into account observations up to time t . We then apply a numerical optimization algorithm based on σ_n^2 to maximize the likelihood function.

3.5. Description of data and transaction costs.

We obtained millisecond data from Nanex for the top three global companies based on market capitalization: Apple Inc, Exxon Mobil Corp and Google Inc (Bloomberg and PwC as of July 2014). To enhance the forecasting accuracy of our experiment, we divided the dataset into an in-sample period from 3 February, 2014 to 30 May, 2014 and an out-of-sample period from 2 June, 2014 to 30 June, 2014.

The STGP trading algorithm processed 742,008 (543,189 in-sample and 198,818 out-of-sample) trading quotes stamped at the millisecond interval for Apple; 511,378 (402,993 in-sample and 108,385 out-of-sample) trading quotes at the same frequency for Exxon Mobil; 489,572 (391,657 in-sample and 97,721 out-of-sample) millisecond trading quotes for Google.

When presenting out-of-sample results, the sample split which identify the beginning of the out-of-sample period is a choice variable. Hansen and Timmermann (2012) argue that there are no broadly accepted recommendations on how to define the sample split. Forecasters have adopted different practical approaches. Hyndman and Athanasopoulos (2013) suggest that the in-sample dataset should be approximately 80% of the entire data with the remaining 20% allocated to out-of-sample evaluation, although the values depend on how long the entire sample is and how far ahead one would like to forecast. Another popular approach is to select the initial estimation sample to have a minimum length with the remaining data used for out-of-sample evaluation. For instance, Marcellino, Stock and Watson (2006) and Pesaran, Pick and Timmermann (2011) implement the first 20 years of the dataset, when available, to compute the forecasting models for different macroeconomic variables. Another approach is to do the opposite and select a specific sample length for the out-of-sample period (Inoue and Kilian, 2008). Alternatively, Welch and Goyal (2008) and Rapach, Strauss and Zhou (2010) suggest that forecasters may examine multiple out-of-sample periods and observe the forecasting performance of each.

However, all of the above approaches depend on ad-hoc choices of the individual split points. In addition, Hansen and Timmermann (2012) argue that multiple split points could lead to data-mining issue because the reported values could be those that favor a given model. Therefore, we adopted the approach of Hyndman and Athanasopoulos (2013) to divide the dataset. Our in-sample dataset consist of 85 trading days (from 3rd of February, 2014 to 30th of May, 2014) with the remaining 21 trading days (from 2nd of June, 2014 to 30th of June, 2014) devoted to out-of-sample evaluation. As Hyndman and Athanasopoulos (2013) suggest, the out-of-sample period depends on how long the entire sample is and what the forecasting horizon is. Since our sample consists of 1,742,958 observations in total for the three financial instruments, we allocated slightly more than 20% of the millisecond dataset to the out-of-sample evaluation period. Apple Inc, the most traded of the three stocks in our study has 26.8% or 198,818 out of 742,008 observations for out-of-sample evaluation, followed by Exxon Mobil with 21% and Google with 20%.

Narang (2013) suggest that the typical round trip transaction costs in HFT are \$0.003 per share. We employ transaction costs of \$0.004 for our profit calculations. Although slightly higher than the current standards, the level of transaction costs used is conservative and takes into account the operational costs of HFT companies such as investments in technology, data and collection fees, and salaries. These include software platforms, labour and risk management systems but does not include co-location of services due to price differences in various trading venues.

4. Experimental results

4.1. HFT predictability and profitability.

In this section we measure the forecasting accuracy of our models⁷. Table 2 reports the in – sample and out – of – sample descriptive statistics for the three financial instruments. It is evident that stock prices are slightly skewed for nearly all in – sample and out-of-sample observations. Due to the fact that both sample sizes for the three stocks are large, an important issue that occurs is Lindley’s paradox (Lindley, 1957). This paradox can lead to overstatement of statistical significance and a tendency to reject the null hypothesis even when the posterior odds favour the null. A number of studies since Lindley have recognised that as sample size increases, there is a tendency to reject the null hypothesis unless the significance level is adjusted downwards. Connolly (1989) proposes the following solution to the Lindley’s paradox and estimation of sample size adjusted critical values for t statistics⁸:

$$t^* = (T - k)^{0.5} (T^{1/T} - 1) \quad (12)$$

where T is the sample size; k measure the number of estimated parameters. The corresponding Bayesian sample size adjusted critical F -values are:

$$F^* = \left[(T - k_1) / p \right] \left[T^{p/T} - 1 \right] \quad (13)$$

where T is the sample size; k is the number of coefficients estimated and p is the number of coefficients restricted under the null hypothesis. If the estimated test statistic exceeds the appropriate critical value from (12) or (13), the sample evidence is in favour of the alternative hypothesis. According to Connolly, if the prior odds are, for example, 3 to 1, the right-hand side of (12) or (13) must be multiplied by 3, because the posterior to favour the alternative hypothesis should take at least three times the F -statistic. One of the properties of high frequency millisecond data is the high presence of no price changes even in very frequently traded stocks like Apple, Exxon Mobil and Google. We account for this market inactivity by modifying the Student’s t distribution associated with the standardized residuals:

$$f \left\langle \frac{\epsilon_t}{\sigma_t} \middle| \delta_t \right\rangle = \{ p_0 \quad \text{If } \delta_t = 1 \quad (14)$$

Or

⁷ We extracted the data from STGP algorithm. To avoid the statistics being affected by the initialisation process, the first 2500 quotes of all three financial instruments were omitted from empirical testing. We consider the first 2500 quotes of historical data as a training period during which the model may show initially chaotic behavior.

⁸ Connolly (1989) used US data to examine the robustness of the day-of-the-week (DOW) effect and weekend effect to alternative estimation and testing procedures. The results suggested that sample size can distort the interpretation of classical test statistics unless the significance level is adjusted downward.

$$f \left\langle \frac{\epsilon_t}{\sigma_t} \middle| \delta_t \right\rangle = \begin{cases} g_{v(\epsilon_t/\sigma_t)} & \text{If } \delta_t = 0 \\ 1 - p_0 & \end{cases} \quad (15)$$

where $g_v(\cdot)$ measure the Student's density function; δ_t measure market inactivity as follows:

$$\delta_t = \begin{cases} 1 & \text{If } |x_t| + |x_{t-1}| = 0, \text{ otherwise} \\ 0 & \end{cases} \quad (16)$$

If $\delta_t = 1$, the forecast $x_{t+i/t} = 0$ for $i = 1, 2, \dots$ (Meade, 2002).

In order to evaluate statistically the forecasting abilities of our models, we compute the RMSE, the MAE and the MAPE. For all three of the error statistics, the lower the output, the better the forecasting accuracy of the model:

$$RMSE = \sqrt{\frac{1}{n} \sum_{\tau=t+1}^{t+n} (Y_t - Y_\tau)^2} \quad (17)$$

$$MAE = \left(\frac{1}{n} \right) \sum_{\tau=t+1}^{t+n} |Y_t - Y_\tau| \quad (18)$$

$$MAPE = \frac{1}{n} \sum_{\tau=t+1}^{t+n} \left| \frac{Y_t - Y_\tau}{Y_\tau} \right| \quad (19)$$

where Y_t represent the actual values and Y_τ is the forecasted values.

A direct in-sample and out-of-sample comparison between traditional forecast combination models and models designed using the STGP indicate the superiority of the latter. The RMSE, MAE and MAPE for STGP are significantly smaller than the errors produced by the SVR, LASSO and Kalman Filter (Table 3, Panels A and B). From Table 3, we note that that STGP presents the best statistical results in all in-sample and out-of-sample periods for the three stocks. Artificial agents trading Apple's shares are the best performer both in-sample and out-of-sample periods as measured by the lowest forecasting errors. Moreover, we employ Giacomini and White (2006) unconditional predictive ability approach for out-of-sample predictive ability testing and forecast selection. This approach investigates which forecasting *method* rather than which *model* is more accurate on average. The main advantage of Giacomini and White's approach is the unified treatment of both nested and nonnested models. The comparison between nested models is important because it is often of interest to tests whether forecasts generated by a given model can outperform those generated by a nested benchmark model. The unconditional predictive ability approach is the Diebold and Mariano (1995) test extended to an environment that

allows estimation of parameters. For each financial instrument we perform pairwise tests of equal predictive ability of the four forecasting methods using a squared error loss. For $\tau=1$ millisecond, we test the null hypothesis:

$$H_0 : E \left[\left(Y_{t+\tau} - \hat{f}_{t,m_f} \right)^2 - \left(Y_{t+\tau} - \hat{g}_{t,m_g} \right)^2 \mid \zeta_t \right] \equiv E \left[\Delta L_{t+\tau} \mid \zeta_t \right] \text{ for } \zeta_t = \{\emptyset, \Omega\} \quad (20)$$

Equation (20) indicates the comparison of the accuracy of forecasts \hat{f}_{t,m_f} and \hat{g}_{t,m_g} for the τ -steps-ahead variable $Y_{t+\tau}$, using a loss function $L_{t+\tau}(\square)$ and information set ζ_t .

In Table 4, the numbers within parentheses are the ratios for mean squared forecast errors (MSEs) for the method in the column relative to the method in the row and a plus (minus) sign indicates that the method in the column outperforms (underperforms) the method in the row at the 10% significance level, as evidenced by a relative MSE less (greater) than 1. The rows labelled ‘Bound’ accommodate Hochberg’s (1988) modified Bonferroni p -value bounds for testing the multiple hypothesis that all pairwise comparisons are zero for a specific column reference method⁹.

The square brackets in Table 4 represent Hochberg’s (1988) p -value bound for the hypothesis that all pairwise comparisons are zero for that panel. Table 4 shows the statistical superiority of the STGP method in all comparisons with the SVR, LASSO and Kalman Filter methods for the three financial instruments. For instance, equal unconditional forecasting ability of the STGP method and the LASSO method is rejected with a p -value of 0.003 and the STGP method outperforms the LASSO method with a MSE ratio of 0.6 for Apple. In contrast, the Kalman Filter is characterised by the worst forecasting performance for Apple, Exxon Mobil and Google. Hochberg-Bonferroni modified p -values does not change the conclusions that emerge from the pairwise testing.

The RMSE, MAE and MAPE are all important error measures, yet they may not correspond to attainable profits from trading rules. We therefore calculate the daily profits generated by the best statistical performers – artificial traders designed on the STGP basis for the most traded five days in each month from February to June, 2014¹⁰. Daily profits for each market participant, i , are estimated for each trading day, t , based on marked-to-market accounting, taking into consideration the fact that every artificial trader begins each trading day with a zero inventory position. All 100,000 artificial traders end the trading day with a zero inventory and therefore a marking-to-market technique at the end of the day is an appropriate method for estimation of daily profits (Baron *et al.*, 2012). More specifically, we calculate the end of the day profits for all traders as the cumulative cash received from selling short positions minus the cash gained from buying long positions, plus the value of any outstanding positions at the end of the trading day, marked to the market price of Apple, Exxon Mobil and Google at close of trading:

⁹ Hochberg (1998) suggested ordering the p -values from testing r hypotheses as $p_{(1)}, \dots, p_{(r)}$ and estimating the bound as $\text{Bound} = \min_{j=1, \dots, r} (r-j+1)p_{(j)}$.

¹⁰ Due to space constraints the full trading volume report is available upon request from the authors.

$$\pi_{i,t} = \sum_{n=1}^{N_{i,T}} p_n y_{i,n} + p_T y_{i,T} \quad (21)$$

where $n = 1 \dots N_{i,T}$ denotes the trades for trader i from the start of the trading day ($t = 0$) to the end of the trading day ($t = T$); p_s represents the price of the trade; $y_{i,n}$ measures the quantity of the n -th trade generated by trader i ; $p_T y_{i,T}$ is the value of any outstanding positions at the end of the day.

Round trip transaction costs of \$0.004 per share are taken into account. Table 5 indicates that STGP forecasting models generate out-of-sample daily profits ranging from \$57 on 2nd of February, 2014 to \$224 recorded on 25th of June, 2014. We observe substantial out-of-sample profits produced by our artificial traders indicating high external validity of our STGP evolving forecasting models. This finding suggests that the out-of-sample tests are free from data over-fitting, which represents the biggest forecasting pitfall, providing a more powerful framework to evaluate the performance of the competing forecasting models.

4.2. HFT and the AMH.

The AMH postulates that learning, competition, and evolutionary selection pressures govern financial markets. Market participants are no longer the rational beings of the standard paradigm, but rather boundedly rational ‘satisfiers’ (in the terminology of Simon (1995)). The main advantage of our artificial stock market is that we are able to directly observe traders’ behavior at the individual level and demonstrate how market participants compete, learn, adapt, survive and evolve through time. The AMH has several relevant predictions for the present study. First, we provide evidence suggesting that profit opportunities exist in financial markets. Our findings indicate that profit opportunities are better modelled through an evolutionary process involving natural selection rather than using analytical econometric models. On the one hand, the process of natural selection ensures the survival-of-the-fittest, that is, the need to better cope with changing market conditions, market dynamics and trading opportunities. On the other hand, natural selection in our experiment determines the number of market participants (the worst performing traders are eliminated from the experiment and replaced by new ones) and trading strategies. In other words, natural selection is in place to select the fittest market participants within an evolutionary framework in which artificial traders interact and evolve dynamically to use the best trading strategies.

Under these circumstances our artificial traders compete, learn, adapt and evolve. Second, market participants learn from trial – and – error, compete and adapt to a constantly changing market environment, survive and evolve. Artificial traders evolve from their initial random trading strategies

towards trading strategies with a higher degree of sophistication. To examine this process, we would need to provide a definition of complexity. Since the behavior of all artificial traders in our experiment is characterized by their forecasting models, which are in the format of STGP, we can easily formulate two definitions of traders' complexity. While the first definition is based on the number of nodes within the GP tree, the second definition is based on the depth of the tree. On each trading day from 3rd of February to 30th of June, 2014, we have a profile of the evolved STGP – trees for 100,000 traders $\{f_{i,t}\}$. Since all forecasting models are in the format of STGP parse trees, the sophistication of traders' strategies can be measured by the associated depth. According to Chen and Yeh (2002) the depth of the parse tree can be measured through the length of the longest path from root to endpoint. The length of a parse tree can be estimated by counting the number of nodes that appear in the tree (this is the actual number of elements used in the program). Let $d_{i,t}$ measure the depth (the length of the longest path) of the model $f_{i,t}$ and $l_{i,t}$ measure the length (the number of nodes) of $f_{i,t}$. To determine the exact degree of trader sophistication, we use the following equations:

$$d_t = \frac{\sum_i^{100,000} d_{i,t}}{100,000} \quad (22)$$

$$l_t = \frac{\sum_i^{100,000} l_{i,t}}{100,000} \quad (23)$$

Our profitability comparisons show that a higher degree of traders' sophistication is associated with higher profitability of trading rules. For instance, artificial traders generated profits of \$57 at the initial stage of the experiment, while more sophisticated trading rules returned \$224 on the last day of the experiment highlighting the existence of learning and that evolutionary processes are in place. We take naïve strategy such as martingale ($f_{i,t} = 0$) to be a benchmark for comparison with the observed behaviour of our artificial traders. For both measures of sophistication, the complexity degree of the martingale model ($f_{i,t} = 0$) is 1. Consistent with Chen and Yeh (2002), we find no evidence that traders' behaviour will converge to the simple martingale model.

The AMH suggest that trading rules became less profitable as markets became aware of their existence. We empirically test this assumption within the millisecond timeframe by applying spectral analysis to the out-of-sample profits for the three stocks. We follow Hasbrouck and Sofianos (1993) and decompose the out-of-sample profits over different time frames.

Spectral analysis sees marked-to-market profits as a function of two different time series, prices and the level of inventory, which can vary at different frequencies. Similar to Baron *et al.* (2012), we implement Fourier analysis to decompose prices and inventories into groups of different frequencies. In the case when the two time series, prices and inventories are in the same phase (artificial traders buy before the price of the three financial instruments increases) they generate profits. If the two time series are not in a phase (artificial traders buy before the price of Apple, Exxon Mobil and Google decreases) they experience losses. Marked-to-market profits for HFT scalpers can be expressed as:

$$\pi_{\tau} = \sum_{t=0}^{\tau} x_t (p_t - p_{t-1}) = \sum_{t=0}^{\tau} x_t \cdot \Delta p_t \quad (24)$$

where x_t represents the inventory holdings of HFT scalpers at time t and p_t is the price of Apple, Exxon Mobil and Google at time t .

One of the requirements of the spectral analysis is the stationarity of x_t and Δp_t . This requirement has been satisfied because HFT scalpers' inventories (x_t) is a mean-reverting process and the first difference of the prices process denoted as Δp is a martingale difference sequence. We follow Baron *et al.* (2012) and develop the following two functions:

$$\hat{x}(w) = \sum_{t=0}^T x_t e^{2\pi i t w / T} \quad (25)$$

$$\Delta p(\omega) = \sum_{t=0}^T \Delta p_{t+1} e^{2\pi i t \omega / T} \quad (26)$$

where ω represents the frequency of different groups; $\hat{x}(\omega)$ and $\Delta p(\omega)$ are the two spectral densities of the x_t and p_t . We apply Fourier analysis to Equation (26) and achieve the following:

$$\pi_T = \frac{1}{T} \sum_{\omega=1}^{\infty} \hat{x}(\omega) \Delta p(\omega) = \frac{1}{T} \sum_{\omega=1}^{\infty} 2 * \text{Real}(\hat{x}(\omega) \Delta p(\omega)) \quad (27)$$

where *Real* represents a function that takes a real part in a complex number; $2 * \text{Real}(\hat{x}(\omega) \Delta p(\omega))$ is the component of the marked-to-market profits generated by the artificial traders at frequency ω . The second equality in Equation (27) is a result based on the fact that an imaginary part of $\hat{x}(\omega) \Delta p(\omega)$ is equal to zero.

Table 6 shows that in the out – of – sample period under investigation, the artificial traders make the largest profits of \$301 at the very short interval between 0 and 500 milliseconds and the smallest profits

of \$10 at the longest time scale between 2,801 and 3,000 milliseconds. Hence, we can conclude that market efficiency measured at millisecond timeframes varies over time and trading rules became less profitable as markets became aware of their existence, consistent with the AMH. Moreover, the timeframe over which artificial traders generate their profits provides more specific details about their trading strategies. It seems that the artificial traders do not try to infer the long-term fundamental value of the four financial instruments but emphasize entirely capturing short-term price dynamics.

The artificial traders' profits are not determined by the difference between their entry price and the fundamental value of the three assets, but by the difference between their entry and exit prices. Consistent with Carrion (2013), we observe that the fastest HFT scalpers generate a larger fraction of profits, indicating that they possess better models than strategic informed traders for predicting intraday price evolution. Our empirical findings suggest several broad conclusions. We observe that profitable opportunities exist and they persist for considerable periods and are not the result of data mining. Empirical findings based on spectral analysis indicate that trading rules became less profitable as markets became aware of their existence. The concept of 'ecological rationality' developed by Todd and Gigerenzer (2003) offers a reasonable explanation as to why artificial traders equipped with relatively simple decision making mechanism perform well in forecasting exercises. It seems that this forecasting success is due to the fact that the structure of the mechanism is adapted to the structure of the information in the stock market environment. Different heuristics enable artificial traders to be ecologically rational, making adaptive decisions that combine forecasting accuracy with speed. The 'ecological rationality' phenomenon describes the structure of the market environment as well as the structure of the heuristics and the match between them. All 100,000 artificial traders in our experiment use relatively simple heuristics for forecasting purposes in an ecologically rational way, using rather limited information based on the historical quotes of the three financial instruments. Moreover, we have found that market participants learn, compete, adapt, survive and evolve toward a higher degree of sophistication. These findings suggest that our artificial stock markets can be viewed as co – evolving ecological systems which consist of 100,000 heterogeneous traders competing for profits in each market. Such ecological systems not only experiences varying degrees of efficiency but also cycles of efficiency and adaptability as changes in profit opportunities are leading to shifts in the composition of market participants. All these findings are consistent with Blume and Easley (1992) who argue that the 'market selection hypothesis' which is governed by natural selection and survival of economic agents provide a better representation of market efficiency.

Furthermore, the concept of 'ecological rationality' indicate that the relationship between market efficiency and the rational 'economic agent' postulated by the traditional economic theory is rather inappropriate.

4.3. Robustness checks.

Robustness checks are necessary for our main results. We obtain historical millisecond data for Russell 1000 (large-cap index) and Russell 2000 (small-cap index) from 3 February, 2014 to 30 June, 2014. Similarly to the three individual stocks, we divided the dataset into an in-sample period from 3 February, 2014 to 30 May, 2014 (85 trading days) and an out-of-sample period from 2 June, 2014 to 30 June, 2014 (21 trading days). The STGP trading algorithm processed 498,128 (325,010 in-sample and 101,118 out-of-sample) trading quotes stamped at the millisecond interval for Russell 1000, and 415,803 (300,412 in-sample and 85,391 out-of-sample) trading quotes at the same frequency for Russell 2000.

In another robustness check we increased the out-of-sample evaluation period to 66 trading days (from 31 March, 2014 to 30 June, 2014) and decreased the in-sample period to 40 trading days (from 3 February, 2014 to 28 March, 2014) for the two indices. The STGP trading algorithm processed 189,289 in-sample and 308,839 out-of-sample trading quotes stamped at the millisecond interval for Russell 1000, and 158,005 in-sample and 257,797 out-of-sample trading quotes at the same frequency for Russell 2000.

A direct in-sample and out-of-sample comparison between traditional forecast combination models and models designed using the STGP indicate the superiority of the latter (Tables 7 and 8, Panels A and B). All forecasting errors for STGP are significantly smaller than the errors produced by the SVR, LASSO and Kalman Filter. Similarly to our experiment with the three individual stocks, we note that that STGP presents the best statistical results in all in-sample and out-of-sample periods for Russell 1000 and Russell 2000, even when we increased the out-of-sample period. This demonstrates the stability of our empirical findings.

5. Conclusion

Despite the large number of studies on the EMH both from developing and developed markets, a consensus on whether financial markets are efficient continues to be elusive. Although there is striking evidence that stock prices do not follow random walks and possess some degree of predictability, there is a lack of strong alternative theoretical explanations to the EMH. Lo (2004) suggests a new model, the Adaptive Market Hypothesis, where the degree of stock price predictability varies under changing market conditions.

However, Lo neither proposes any specific indicators of market conditions nor any refutable factors about the direction of the relationship between stock price predictability and the indicators of market conditions. We aim to fill this gap by evaluating the evolution of stock price predictability at the millisecond timeframe and examining whether its evolution is consistent with the notion of the AMH. Our forecasting exercise shows that a direct in-sample and out-of-sample comparison between traditional forecast combination models and models designed using STGP indicate the superiority of the latter. The RMSE, MAE and MAPE for STGP are significantly smaller than the errors produced by the benchmark forecasting models such as SVR, LASSO and the Kalman Filter. STGP gives the best statistical results in all in – sample and out – of – sample periods for Apple, Exxon Mobil and Google measured by the lowest forecasting errors. Moreover, we have found that STGP forecasting models generate in –sample daily profits ranging from \$57 on 2nd of February, 2014 to \$177 recorded on 29th of May, 2014. We also observe substantial out-of-sample profits generated by our artificial traders. Our findings indicate that profit opportunities are not best identified analytically, but through an evolutionary process involving natural selection. The process of natural selection ensures the survival-of-the- fittest, that is, the ability to cope better with changing market conditions, market dynamics and trading opportunities. Natural selection in our experiment also determines the nature of market participants (the worst performing traders are eliminated from the experiment and replaced by new ones) and trading strategies. In other words, natural selection is in place to select the fittest market participants within an evolutionary framework in which artificial traders interact and evolve dynamically. We observe that our artificial traders evolve their initial random trading strategies toward trading strategies with a higher degree of sophistication. Different heuristics enable artificial traders to be ecologically rational, making adaptive decisions that combine forecasting accuracy with speed. These findings suggest that our artificial stock markets can be viewed as evolving ecological systems which consist of 100,000 heterogeneous traders competing for profits on each market. Therefore, our empirical findings indicate that the relationship between market efficiency and the rational ‘economic agent’ governed by the traditional economic theory is rather inappropriate and financial markets should be viewed from a Darwinian survival – of – the fittest perspective and, more specifically within an evolutionary framework in which markets, financial instruments, and traders interact, compete, learn, adapt and evolve dynamically, as suggested by the AMH.

However, it should be noted that there is a difference between human intelligence and artificial trader intelligence. Artificial traders are programmed to obey orders and perform certain tasks as per the

commands given to them. Moreover, artificial traders are lacking feelings and emotions while human beings can feel various emotions and also express these emotions to others. The main limitation of this particular experiment comes from the fact is that the computational memory of artificial traders is rather short-term in nature. Short-term memory is the type of memory that is used to manipulate information. Engle *et al.* (1999) demonstrates a weak relationship between short-term memory and human intelligence indicating the limitations of cognitive abilities of artificial traders.

Tables

Table 1: Artificial stock market parameter settings. ^a denotes the maximum genome size measure the total number of nodes in a trader's trading rule. A node is a gene in the genome such as a function or a value. ^b denotes the maximum genome depth measuring the highest number of hierarchical levels that occurs in an trader's genome (trading rule). The depth of a trading rule can be an indicator of its complexity. ^c is the minimum age required for agents to qualify for potential participation in the initial selection. The age of a trader is represented by the number of quotes that have been processed since the trader was created. This measure also specifies the period over which agent performance will be compared. Our minimum breeding age is set to 80, which means that the trader's performance after the last 80 quotes will be compared. ^d is where 5% of the best performing traders of the initial selection that will act as parents in crossover operations for creating new traders.

Parameter	Parameter Setting
Total population size (traders)	100,000
Initial wealth(equal for all traders)	\$100,000
Transaction costs	\$0.004/share
Significant Forecasting range	0% to 10%
Number of decimal places to round quotes on importing	2
Minimum price increment for prices generated by model	0.01
Minimum position unit	20%
Maximum genome size ^a	4096
Maximum genome depth ^b	20
Minimum initial genome depth	2
Maximum initial genome depth	5
Breeding cycle frequency (quotes)	1
Minimum breeding age (quotes) ^c	80
Initial selection type	random
Parent selection (percentage of initial selection that will breed) ^d	5%
Mutation probability (per offspring)	10%
Total number of quotes processed- Apple Inc	742,008
Total number of quotes processed- Exxon Mobil	511,378
Total number of quotes processed- Google	489,572
Seed generation from clock	Yes
Creation of unique genomes	Yes
Offspring will replace the worst performing traders of the initial selection	Yes

Table 2: Summary statistics for the in-sample and out-of-sample data for Apple, Exxon Mobil and Google.

	In-Sample	Out-of-Sample
Panel A: Apple		
Mean	93.06	94.75
Std.dev	1.76	1.24
Skewness	0.99	0.67
Kurtosis	2.98	2.18
Jarque-Bera	14236.22	9006.14
Panel B: Exxon Mobil		
Mean	90.64	91.02
Std.dev	0.63	0.58
Skewness	0.80	0.71
Kurtosis	2.97	1.99
Jarque-Bera	12549.72	10872.01
Panel C: Google		
Mean	553.44	554.37
Std.dev	2.62	2.07
Skewness	0.81	0.68
Kurtosis	2.96	2.19
Jarque-Bera	9106.94	8254.77

Table 3: The summary of the in-sample and out-of-sample forecasting statistical performance for Apple, Exxon Mobil and Google. ‘SVR’ denotes the Support Vector Regression, ‘LASSO’ denotes the Least Absolute Shrinkage and Selection Operator while ‘KF’ denotes the Kalman Filter.

	In-Sample			Out-of-Sample		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Panel A: Apple						
STGP	0.0021	0.0011	39.67%	0.0009	0.0007	32.13%
SVR	0.0049	0.0038	64.83%	0.0031	0.0028	58.01%
LASSO	0.0054	0.0051	81.32%	0.0050	0.0037	67.90%
KF	0.0078	0.0066	90.11%	0.0068	0.0053	74.25%
Panel B: Exxon Mobil						
STGP	0.0030	0.0027	42.74%	0.0014	0.0008	36.76%
SVR	0.0058	0.0044	69.99%	0.0039	0.0027	50.01%
LASSO	0.0070	0.0069	84.87%	0.0063	0.0039	60.92%
KF	0.0086	0.0083	88.20%	0.0075	0.0057	71.05%
Panel C: Google						
STGP	0.0035	0.0033	47.99%	0.0019	0.0014	41.06%
SVR	0.0062	0.0061	78.26%	0.0044	0.0033	63.33%
LASSO	0.0074	0.0077	81.10%	0.0059	0.0041	69.89%
KF	0.0093	0.0079	85.00%	0.0070	0.0066	71.46%

Table 4: The out-of-sample unconditional forecasting ability tests for Apple, Exxon Mobil and Google. We employ the Giacomini and White (2006) test of equal unconditional predictive ability. The entries are the p -values of the tests for the four forecast methods in the corresponding row and column. The forecasting horizon τ is 1millisecond in all tests. The numbers in parentheses are the ratios of MSEs for the method in the column relative to the method in the row. A plus (minus) sign indicates that the method in the column outperforms (underperforms) the method in the row at the 10% significance level, as evidenced by a relative MSE less (greater) than 1. For instance, equal unconditional forecasting ability of the STGP method and the SVR method is rejected with a p -value of 0.003 and the STGP method outperforms the SVR method with a MSE ratio of 0.6 for Apple. The rows labelled ‘Bound’ shows the Hochberg-Bonferroni (HB) multiple hypothesis p -value bound for the method in the column relative to all other methods. The square brackets shows the HB p -value bound for the hypothesis that all pairwise comparisons are zero for that panel.

Benchmark	STGP	SVR	LASSO	KF
Panel A: Apple				
Bound	0.011	0.053	0.069	0.083
SVR	0.003 (0.60)			
LASSO	0.014 (0.71)	0.083 (0.92)		
KF	0.026 (0.79)	0.116 (0.96)	0.318 (0.98)	[0.006]
Panel B: Exxon Mobil				
Bound	0.022	0.063	0.072	0.088
SVR	0.018 (0.73)			
LASSO	0.028 (0.80)	0.108 (0.94)		
KF	0.036 (0.83)	0.210 (0.97)	0.588 (0.82)	[0.009]
Panel C: Google				
Bound	0.033	0.085	0.099	0.103
SVR	0.029 (0.69)			
LASSO	0.044 (0.86)	0.166 (0.82)		
KF	0.058 (0.91)	0.300 (0.98)	0.603 (2.81)	[0.012]

Table 5: The out-of-sample daily profits based on the most traded days generated by the STGP trading algorithm for Apple, Exxon Mobile and Google. All daily profits are estimated as the difference between the prices at which artificial traders bought and sold shares of Apple, Exxon Mobil and Google.

February 2014		March 2014		April 2014		May 2014		June 2014	
Date	Profit	Date	Profit	Date	Profit	Date	Profit	Date	Profit
02/12/2014	\$57	03/03/2014	\$88	04/01/2014	\$122	05/09/2014	\$150	06/05/2014	\$199
02/14/2014	\$65	03/11/2014	\$93	04/15/2014	\$128	05/14/2014	\$158	06/09/2014	\$205
02/18/2014	\$71	03/12/2014	\$101	04/23/2014	\$133	05/21/2014	\$166	06/17/2014	\$214
02/24/2014	\$79	03/20/2014	\$109	04/24/2014	\$139	05/27/2014	\$169	06/23/2014	\$220
02/27/2014	\$84	03/26/2014	\$117	04/30/2014	\$144	05/29/2014	\$177	06/25/2014	\$224

Table 6. Out-of-sample daily profits spectral analysis for Apple, Exxon Mobil and Google. This table examines out-of-sample STGP trading profits over different millisecond time lengths for Apple, Exxon Mobil and Google, implementing the methods of Hasbrouck and Sofianos (1993).

Time length (milliseconds)	Out-of-sample STGP profit
0-500	\$301
501-700	\$199
701-900	\$107
901-1000	\$98
1001-1200	\$75
1201-1400	\$64
1401-1600	\$46
1601-1800	\$40
1801-2000	\$33
2001-2200	\$30
2201-2400	\$24
2401-2600	\$20
2601-2800	\$15
2801-3000	\$10

Table 7. The summary of the in-sample (80%) and out-of-sample (20%) forecasting statistical performance for Russell 1000 and Russell 2000. ‘SVR’ denotes the Support Vector Regression, ‘LASSO’ denotes Least Absolute Shrinkage and Selection Operator, while ‘KF’ denotes the Kalman Filter.

	In-sample			Out-of-sample		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Panel A: Russell 1000						
STGP	0.0044	0.0032	42.11%	0.0021	0.0018	29.99%
SVR	0.0068	0.0051	68.03%	0.0038	0.0030	45.02%
LASSO	0.0079	0.0068	75.19%	0.0052	0.0041	64.13%
KF	0.0086	0.0074	81.93%	0.0066	0.0058	72.09%
Panel B: Russell 2000						
STGP	0.0049	0.0028	48.95%	0.0031	0.0028	33.10%
SVR	0.0069	0.0044	63.18%	0.0049	0.0049	50.03%
LASSO	0.0073	0.0061	72.22%	0.0066	0.0060	69.91%
KF	0.0081	0.0075	83.28%	0.0079	0.0074	79.95%

Table 8. The summary of the in-sample (38%) and out-of-sample (62%) forecasting statistical performance for Russell 1000 and Russell 2000. ‘SVR’ denotes the Support Vector Regression, ‘LASSO’ denotes Least Absolute Shrinkage and Selection Operator, while ‘KF’ denotes the Kalman Filter.

	In-sample			Out-of-sample		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Panel A: Russell 1000						
STGP	0.0056	0.0025	50.04%	0.0016	0.0013	22.14%
SVR	0.0066	0.0032	59.95%	0.0028	0.0029	40.27%
LASSO	0.0080	0.0044	67.86%	0.0046	0.0048	66.38%
KF	0.0094	0.0061	78.38%	0.0058	0.0065	81.11%
Panel B: Russell 2000						
STGP	0.0060	0.0029	53.11%	0.0020	0.0018	25.08%
SVR	0.0071	0.0035	61.35%	0.0030	0.0032	43.93%
LASSO	0.0083	0.0050	70.79%	0.0049	0.0053	70.04%
KF	0.0096	0.0066	80.99%	0.0061	0.0068	84.38%

Appendix

Genetic Programming (GP) and Strongly Typed Genetic Programming (STGP)

GP is a machine-learning method to automate the development of computer programs in terms of natural evolution process (Banzhaf *et al.*, 1998). One population of programs is replaced by a next generation of programs that are better suited to performing the designated task. This is done by using genetic algorithms that are based on evolutionary principles. In the GP process, individuals in a population compete and mate with each other in order to generate increasingly stronger individuals (Lien *et al.*, 2003). In GP parse trees (also known as tree genomes) are implemented, rather than lines of program code, as the basis of the algorithms. For instance, Figure 1 illustrates a tree that uses the input variables $x = (a,b,c)$ to calculate the mathematical expression $a - 2b + 9/c$.

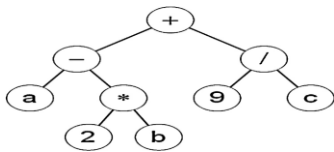


Figure 1. Example of GP program tree genomes (Copied from Wappler and Wegener, 2006).

The parse trees represent the trading rules of HFT scalpers and strategic informed traders in our experiment. The typical genetic structure of the trading rule consists of hundreds nodes and is rather complex to write out, but it can be simplified to an equivalent algorithmic trading rules as shown below.

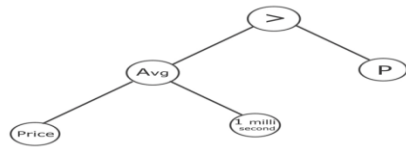


Figure 2. Example of time-dependent trading rule of strategic informed traders.

Figure 2 illustrates that the trading rule of strategic informed traders sends a buy signal if the average stock price over the past 1 millisecond is greater than the current price. A sell signal is sent otherwise.

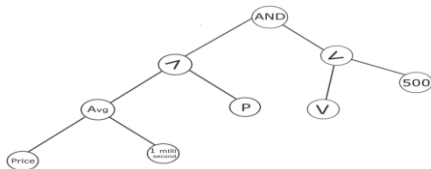


Figure 3. Example of time-dependent trading rule of HFT scalpers.

Figure 3 indicates that the trading rule of HFT scalpers sends a buy signal if the average stock price over the past 1 millisecond is greater than the current price and the current volume is less than 500 while a sell signal is sent otherwise. The current volume function protects HFT scalpers from sweep risk exposure. Large losses caused by sweeps (adverse price movements against HFT scalpers transient positions) can substantially reduce or even eliminate their profitability, so the management of sweep risk is of paramount importance for HFT scalpers. HFT scalpers use the market microstructure to capture and avoid sweep risk, which is the risk related to trading against large informed toxic orders (for instance, large institutional orders) positioned at multiple levels of the order book (Manahov, 2016).

STGP is an advanced version of GP with more sophisticated application of generic functions and data types. In conventional GP there are no intrinsic constraints on the data types that can be used as nodes. This means the user must specify all the programs and variables that can be used as nodes in a parse tree and deal with a search space

of the order 10^{30} – 10^{40} (Montana, 1995). STGP is set up so that there is an allotted type for each constant and each variable. Thus, there is a requirement in STGP that the type of data for each element is specified in advance. As a result the initialization process and the different genetic operators will only support the construction of syntactically correct (i.e. legitimate) trees, and this is beneficial to the process because the search space is reduced (Haynes *et al.*, 1996). STGP effectively reduces the state-space size which must be searched. The STGP search space is represented by the set of all legitimate parse trees, which means that all functions have the correct number of parameters of the correct type. The STGP parse tree is also usually limited to some maximum depth (in our experiments this is set to 20) which keeps the search space finite and manageable by preventing trees from growing to an excessive size.

Both GP and STGP processes follows a series of steps:

1. An initial population of programs is created randomly each one represented by a different parse tree. The random nature of the initial rules ensures a large variety of possible trading rules is investigated in the experiment.
2. These programs are then run and the programs producing the best solutions in the first generation are permitted to ‘breed’ and produce the next generation of programs. The best solutions are those that most closely match a specified fitness criteria. Because every pair of parents creates two offspring trading rules, the number of parents and the number of offspring is equal.
3. The newly created trading rules replace poorly performing trading rules in the initial selection based on the replacement fitness return. The breeding of next generations of programs continues until the solution reached is deemed sufficiently adequate.

The STGP process produces syntactically correct programs or trading rules as the generations progress through crossover and mutation. Crossover is the outcome of two parent programs having been selected as suitable to procreate. After the selection of two parent programs, a cross over point is determined by randomly selecting a functional or terminal node within each parent and producing a child program which consists of random elements of the parent programs (How *et al.*, 2009). Mutation occurs when an element of the parent program is randomly modified during breeding with resultant effect upon the child program. In our experiment, the crossover process can be described in the following way (Figure 4). The trading strategies S_i and S_j are the two parents and a one-point crossover has been used to create children (new trading rules) S_k and S_l . Therefore, a breaking point has been randomly selected and parts of the two parent trading rules are exchanged to create two new trading rules.

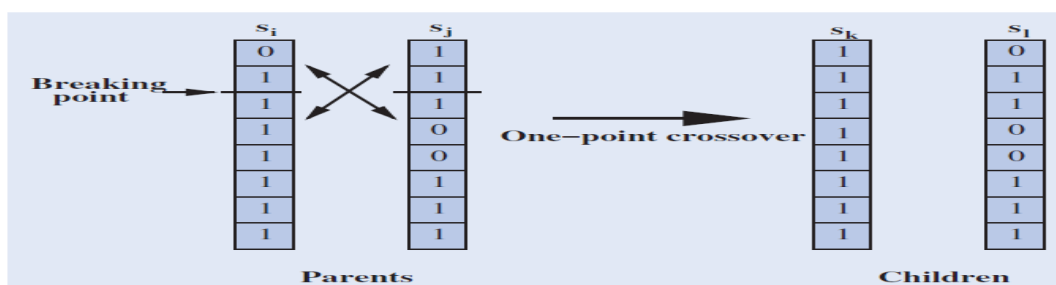


Figure 4. The process of crossover in STGP for generating new trading rules (Copied from Chakraborti *et al.* 2011).

References

- Allen, F., Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. *Journal of Financial Economics*, 51, 245-271.
- Banzhaf, W., Nordin, P., Keller, R. E., Francone, F. D. (1998). Genetic Programming: an introduction: on the automatic evolution of computer programs and its applications. Morgan Kaufmann Publishers Inc: San Francisco.
- Barnes, J. (1982). *Programming in Ada*. Reading, MA: Addison-Wesley.
- Baron, M., Brogaard, J., Kirilenko, A. (2012). The trading profits of high frequency traders. Working paper - http://faculty.chicagobooth.edu/john.cochrane/teaching/35150_advanced_investments/Baron_Brogaard_Kirilenko.pdf.
- Basak, D., Pal, S., Patranabis, D. C. (2007). Support vector regression. *International Journal of Neural Information Processing*, 10, 203-224.
- Basak, D., Pal, S., Patranabis, D. C. (2007). Support Vector Regression. *Neural Information Processing-Letters and Reviews*, 11(10): 203-224.
- Bates, J. M., Granger, W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, 20(4): 451-468.
- Becker, I. A., Seshadri, M. (2003). GP-evolved technical trading rules can outperform buy and hold. In: 3rd International Workshop on Computation.
- Biais, B., Foucault, T., Moinas, S. (2013). Equilibrium fast trading. HEC Paris research paper 968/2013.
- Blume, L., Easley, D. (1992). Evolution and market behaviour. *Journal of Economic Theory*, 58, 9-40.
- Bhattacharyya, S., Pictet, O. V., Zumbach, G. (2002) Knowledge-intensive genetic discovery in foreign exchange markets. *IEEE Transactions of Evolutionary Computation*, 6(2):169-178.
- Brogaard, J. (2011). High frequency trading and volatility. Working paper, available at http://www.researchgate.net/publication/228273562_High_Frequency_Trading_and_Volatility_Working_Paper.
- Brogaard, J. (2013). High frequency trading and market quality. Working paper, available at http://www.researchgate.net/publication/228261470_High_Frequency_Trading_and_Market_Quality.
- Boehmer, E., Fong, K., Wu, J. (2012). International evidence on algorithmic trading. *SSRN Working Paper* (available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2022034).
- Brogaard, J., Hendershott, T., Riordan, R. (2014). High frequency trading and price discovery. *The Review of Financial Studies*, 27(8): 2267-2306.
- Cao, L. J., Chua, K. S., Guan, K. L. (2003). *C-ascending support vector machines for financial time series forecasting*. In: Computational Intelligence for Financial Engineering Proceedings, 317-323.
- Campbell, J. Y., Lo, A. W., MacKinlay, A. C. (1997). The econometrics of financial markets. Princeton, Princeton University Press.
- Carrion, A. (2013). Very fast money: high frequency trading on NASDAQ. *Journal of Financial Markets*, 16(4), 680-711.
- Charles, A., Darné, O, Kim, J. H. (2012). Exchange-rate return predictability and the adaptive markets hypothesis: Evidence from major foreign exchange rates. *Journal of International Money and Finance*, 31, 1607-1626.
- Charles, A., Darné, O, Kim, J. H. (2015). Will precious metals shine? A market efficiency perspective. *International Review of Financial Analysis*, 41, 284-291.
- Chen, S-H., Yeh, C-H. (1997). Modeling speculators with Genetic Programming. *Evolutionary Programming VI. Lecture Notes in Computer Science*, 1213:137-147.
- Chen, S-H., Yeh, C-H. (1998). Option pricing with genetic programming. In: Proceedings of the 3rd Annual Genetic Programming Conference.
- Chen, S-H., Yeh, C-H. (2002). On the emergent properties of artificial markets: the efficient market hypothesis and the rational expectation hypothesis. *Journal of Economic Behaviour and Organization*, 49, 217-239.
- Cherkassky, V., Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, 17, 113-126.
- Chidambaram, N. K., Lee, C. H. J., Trigueros, J. R. (1998). Adaptive evolutionary approach to option pricing via genetic programming. In: Proceedings of the 3rd Annual Genetic Programming Conference.
- Connolly, R. A. (1989). An examination of the robustness of the weekend effect. *Journal of Financial and Quantitative Analysis*, 24, 133-169.
- Dempster, M. A. H., Jones, C. M. (2001). A real-time adaptive trading system using genetic programming. *Quantitative Finance*, 1, 397-413.
- Diebold, F. X., Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253-263.

- Domowitz, I. (2010). Take heed the lessons from the 1962 flash crash. Available at <http://advancedtrading.com/exchanges/225700888>. Accessed 01/03/2015.
- Dyakova, A., Smith, G. (2013a). The evolution of stock market predictability in Bulgaria. *Applied Financial Economics*, 23(9), 8005-816.
- Dyakova, A., Smith, G. (2013b). Bulgarian stock market relative predictability: BSE-Sofia stocks and South East European markets. *Applied Financial Economics*, 23(15), 1257-1271.
- Engle, R.F., Tuholski, S.W., Laughlin, J.E., and Conway, A.R.A. 1999. Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3): 309-331.
- Fan, W., Gordon, M.D., Pathak, P. (2005). Genetic programming-based discovery of ranking functions for effective web search. *Journal of Management Information Systems*, 21(4):37-56.
- Fama, E. F. (1965). The behaviour of stock market prices. *The Journal of Business*, 38, 34-105.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25, 383-417.
- Farnsworth, G.V., Kelly, J.A., Othling, A., Pryor, R. (2004). Successful technical trading agents using Genetic Programming. Sand Report, SAND2004-4774.
- Foucault, T., Hombert, J., Rosu, I. (2013). News trading and speed. HEC Paris research paper 975/2013.
- Giacomini, R., White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6): 1545-1578.
- Hansen, P.R. and Timmermann, A. (2012). Choice of sample split in out-of-sample forecast evaluation [online]. Available from http://cadmus.eui.eu/bitstream/handle/1814/21454/ECO_2012_10.pdf?sequence=1. Accessed 27/06/2018
- Hansbrouck, J., Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, 16(4), 741-770.
- Harvey, C., Lui, Y. (2014). Evaluating Trading Strategies. *Journal of Portfolio Management*, 40(5), 108-118.
- Haynes, T. D., Schoenfeld, D. A., Wainwright, R. L., (1996). *Type inheritance in Strongly Typed Genetic Programming*. In *Advances of Genetic Programming 2* (Eds. Kinnear, K., Angeline, P). MIT Press, Cambridge, USA
- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 75:800-802.
- Hommel, C. (2011). The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic Dynamics and Control*, 35(1):1-24.
- How, J., Ling, M., Verhoeven, P. (2009). Does size matter? A genetic programming approach to technical trading. *Quantitative Finance*, 10(2)-131-140.
- Hull, M., McGroarty, F. (2014). Do emerging markets become more efficient as they develop? Long memory persistence in equity indices. *Emerging Markets Review*, 18, 45-61.
- Hyndman, R.J. and Athanasopoulos, G. (2013). Forecasting: principles and practice [online]. Available from <https://robjhyndman.com/uwafiles/fpp-notes.pdf>. Accessed 28/06/2018.
- Iba, H., Sasaki, T. (1999). Using genetic programming to predict financial data. In: *Evolutionary Computation. CEC99. Proceedings of 1999 IEEE*.
- Inoue, A. and Killian, L. (2008). How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation. *Journal of The American Statistical Association*, 103:511-522.
- Ito, M., Sugiyama, S. (2009). Measuring the degree of time varying market inefficiency. *Economics Letters*, 103(1), 62-64.
- Jovanovic, B., Menkveld, A. J. (2012). Middleman in limit-market orders. *SSRN Working Paper*, (available at http://www.researchgate.net/publication/228261470_High_Frequency_Trading_and_Market_Quality).
- Kaboudan, M.A. (2000). Genetic programming prediction of stock prices. *Computational Economics*, 16:207-236.
- Kim, J. H., Shamsuddin, A., Lim, K. (2011). Stock return predictability and the adaptive markets hypothesis: Evidence from century-long U.S. data. *Journal of Empirical Finance*, 18, 868-879.
- Kirilenko, A., Kyle, A. S., Samadi, M., Tuzun, T. (2011). The Flash crash: the impact of high frequency trading on an electronic market. *SSRN Working Paper* (available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1686004).
- Levich, R. M., Poti, V. (2015). Predictability and 'good deals' in currency markets. *International Journal of Forecasting*, 31(2), 454-472.
- Lien, D., Tse, Y.K., Zhang, X. (2003). Structural change and lead-lag relationship between the Nikkei spot index and futures price: A genetic programming approach. *Quantitative Finance*, 3, 136-144.
- Lindley, D.V. (1957) A statistical paradox. *Biometrika*, 44:187-192.
- Lo, A.W. (2004) The Adaptive Markets Hypothesis. *Journal of Portfolio Management*, 30, 15-29.
- Lo, A. W. (2005). Reconciling efficient markets with behavioural finance: the adaptive markets hypothesis. *Journal of Investment Consulting*, 7, 21-44.
- Manahov, V. (2016). Can High-Frequency Trading Strategies Constantly Beat the Market? *International Journal of Finance and Economics*, 21, 167-191.

- Manahov, V., Hudson, R. (2014a). A note on the relationship between market efficiency and adaptability – New evidence from artificial stock markets. *Expert Systems with Applications*, 41(16), 7436-7454.
- Manahov, V., Hudson, R., Gebka, B. (2014b). Does high frequency trading affect technical analysis and market efficiency? And if so, how? *Journal of International Financial Markets, Institutions and Money*, 28, 131-157.
- Marcellino, M., Stock, J. H. and Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting economic time series. *Journal of Econometrics*, 135:499-526.
- Martinez, C. H., Rosu, I. (2013). High frequency traders, news and volatility. Working paper.
- McGowan, M. J. (2010). The rise of computerized high frequency trading: use and controversy [online]. Available from: <http://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1211&context=dltr>. Accessed on 20/09/2015.
- McKee, T. E., Lensberg, T. (2002). Genetic Programming and rough sets: A hybrid approach to bankruptcy classification. *European Journal of Operational Research*, 138:436-451.
- Montana, D. J. (2002). Strongly Typed Genetic Programming [online]. Available at <http://personal.d.bbn.com/dmontana/papers/stgp.pdf>. Accessed on 15/01/2015.
- Montana, D. J. (1995). Strongly Typed Genetic Programming. *Evolutionary Computation*, 3(2), 199-230.
- Montana, D. J. (1994). *Strongly Typed Genetic Programming*. Technical Report 7866.
- Narang, R. K. (2013). *Inside the black box. A simple guide to quantitative and high frequency trading*. John Wiley & Sons, New Jersey.
- Neely, C. (2003). Risk-adjusted, ex ante, optimal technical trading rules in equity markets. *International Review of Economics and Finance*, 12(1): 69-87.
- Neely, C., Weller, P., Dittmar, R. (1997). Is technical analysis in the foreign exchange market profitable? A genetic programming approach. *Journal of Financial and Quantitative Analysis*, 32(4), 405-426.
- Neely, C. J., Weller, P. A. (2003). Intraday technical trading in the foreign exchange market. *Journal of International Money and Finance*, 22: 223-237.
- Park, C-H., Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys*, 21, 786-826.
- Pesaran, M. H. and Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance*, 50:1201-1228.
- Potvin, J., Soiano, P., Valee, M. (2004). Generating trading rules on the stock market with genetic programming. *Computational Operations Research*, 31:1033-1047.
- Rapach, D. E., Strauss, J. K. and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23:821-862.
- Sermpinis, G., Stasinakis, C., Dunis, C. (2014). Stochastic and genetic neural network combinations in trading and hybrid time-varying leverage effects. *Journal of International Financial Markets, Institutions and Money*, 30, 21-54.
- Sermpinis, G., Stasinakis, C., Theofilatos, K., Karathanasopoulos, A. (2015). Modelling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms-Support Vector Regression forecast combinations. *European Journal of Operational Research*, 1-16.
- Simon, H. (1955). A behavioural model of rational choice. *Quarterly Journal of Economics*, 69, 99-118.
- Smith, G. (2012). The changing and relative efficiency of European emerging stock markets. *The European Journal of Finance*, 18(8), 689-708.
- Smith, G. (2016). The Relative Predictability of Stock Markets in the Americas. *International Journal of Finance and Economics*, 21, 131-142.
- Smola, A., Murata, N., Scholkopf, B., Muller, K. R. (1998). *Asymptotically optimal choice of ϵ -loss for support vector machines*. In: Niklasson, L., Boden, M., Ziemke, T. (Eds.). *Proceedings of the International Conference on Artificial Neural Networks, Perspectives in Neural Computing*, Springer, pp.105-110.
- Soufian, M., Forbes, W., Hudson, R. (2014). Adapting financial rationality: Is a new paradigm emerging? *Critical Perspectives on Accounting*, 25(8), 724-742.
- Steele, G. (1984). *Common Lisp*. Burlington, MA: Digital Press.
- Sullivan, R., Timmermann, A., White, H. (1999). Data-snooping, technical trading rule performance and the bootstrap. *The Journal of Finance*, 54(5), 1647-1691.
- Suykens, J. A. K., Brabanter, J. D., Lukas, L., Vandewalle, L. (2002). Weighted last squares support vector machines: robustness and sparse approximation. *Neurocomputing*, 48:85-105.
- Todea, A., Ulici, M., Silaghi, S. (2009). Adaptive market hypothesis: Evidence from Asia-Pacific financial markets. *The Review of Finance and Banking*, 1(1), 7-13.
- Todd, P. M., Gigerenzer, G. (2003). Bounding rationality to the world. *Journal of Economic Psychology*, 24, 143-165.

- Trafalis, T. B., Ince, H. (2000). Support vector machine for regression and applications to financial forecasting. *Proceedings of International Joint Conference on Neural Networks*. 1, 348-353.
- Urquhart, A. (2016). How predictable are precious metal returns? *The European Journal of Finance*, forthcoming.
- Urquhart, A., Hudson, R. (2013). Efficient or adaptive markets? Evidence from major stock markets using very long run historic data. *International Review of Financial Analysis*, 28, 130-142.
- Urquhart, A., McGroarty, F. (2014). Calendar effects, market conditions and the Adaptive Market Hypothesis: Evidence from long-run US data. *International Review of Financial Analysis*, 35, 154-166.
- Urquhart, A., McGroarty, F. (2016). Are stock markets really efficient? Evidence of the adaptive market hypothesis. *International Review of Financial Analysis*, 47, 39-49.
- Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York.
- Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455-1508.
- Yuan, M., Lin, Y. (2007). On the non-negative garrotte estimator. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 143-161.
- Zakamulin, V. (2014). The real-life performance of market timing with moving average and time-series momentum rules. *Journal of Asset Management*, 15, 261-278.
- Zhang, F. (2010). High frequency trading, stock volatility, and price discovery. *SSRN Working Paper* (Available at <http://ssrn.com/abstract=1691679>).