Article:

Manahov, V. and Zhang, H. orcid.org/0000-0002-8727-4906 (2019) Forecasting financial markets using high-frequency trading data: Examination with strongly typed genetic programming. International Journal of Electronic Commerce, 23 (1). pp. 12-32. ISSN 1086-4415

https://doi.org/10.1080/10864415.2018.1512271

Forecasting Financial Markets Using High-Frequency Trading Data: Examination with Strongly Typed Genetic Programming

by

Viktor Manahov* and Hanxiong Zhang

Abstract

Market regulators around the world are still debating whether or not high-frequency trading (HFT) plays a positive or negative role in market quality. We develop an artificial futures market populated with high-frequency traders (HFTs) and institutional traders using Strongly Typed Genetic Programming (STGP) trading algorithm. We simulate real-life futures trading at the millisecond timeframe by applying STGP to E-Mini S&P 500 data stamped at the millisecond interval. A direct forecasting comparison between HFTs and institutional traders indicate the superiority of the former. We observe that the negative implications of high-frequency trading in futures markets can be mitigated by introducing a minimum resting trading period of less than 50 milliseconds. Overall, we contribute to the e-commerce literature by showing that minimum resting trading order period of less than 50 milliseconds could lead to HFTs facing a queuing risk resulting in a less harmful market quality effect. One practical implication of our study is that we demonstrate that market regulators and/or e-commerce practitioners can apply artificial intelligence tools such as STGP to conduct trading behaviour-based profiling. This can be used to detect the occurrence of new HFT strategies and examine their impact on the futures market.

Keywords: Evolutionary Computation, Artificial Intelligence, High-Frequency Trading, Algorithmic Trading, Big Data Analytics, Financial Econometrics.

JEL Classification: D04, D53, G12, G14, G15, G17

*Corresponding author: Viktor Manahov, PhD, The University of York, York, Heslington East, YO10 5GD, United Kingdom, tel: +44 (0) 1904325847, email: viktor.manahov@york.ac.uk

Hanxiong Zhang, PhD, The University of Lincoln, Brayford Pool, Lincoln, United Kingdom, LN6 7TS, tel: +44 (0) 1522835639, email: hzhang@lincoln.ac.uk

Short bio of authors

Viktor Manahov

After completing his BA (Hons) in Business Studies at the Open University, UK, Viktor Manahov went on to study an MSc in Finance and Investment Management at the University of Aberdeen, UK. He received his PhD titled 'An investigation of the behaviour of financial markets using agent-based computational models' from Newcastle University, UK.

He is a member of the UK Higher Education Academy and is currently teaching modules related to finance and stock market trading at the University of York, UK. Viktor is also a member of the editorial board of the *Review of Behavioral Finance.*

His research focuses on agent-based modelling and artificial stock markets; genetic programming trading algorithms; stock market forecasts and valuation of securities; high frequency trading techniques; analysis of financial markets behaviour; empirical properties of asset returns; and stylized facts and statistical issues.

Hanxiong Zhang

Dr Hanxiong Zhang is a Senior Lecturer in Banking and Finance at Lincoln International Business School, UK. Hanxiong joined the University of Lincoln as a Lecturer in Financial Economics in April 2014. He holds a PhD from University of Newcastle upon Tyne, UK. His research interests cover time-series modelling, market efficiency, trading strategy and asset pricing. He is a CFA charter holder.

Introduction

Today's trading in futures markets is more complex and often involves little human intervention. The dramatic increase in the variety of trading platforms combined with significant technological advancements, makes the process by which trading orders are processed and executed much more complex than it was ten years ago. Significant technological investments resulted in latency improvements, where computer algorithms execute trading orders based on electronically received information before human traders. Speed races in today's trading are occurring even at nanosecond (billionths of a second) intervals [6]. High-frequency traders (HFTs) are able to anticipate future trading order flows because they process intraday trading messages faster than other market participants [16]. Several studies highlight that such anticipatory or front-running trading generates negative externalities, such as limited liquidity provision, forcing other slower market participants to abandon trading, or facilitating overinvestments in technological bases [3, 18, 21, 34].

However, most studies on the topic are lacking the ability to identify which trades and quotes come from HFT. This research obstacle makes it difficult to investigate how HFT affects the market and other market participants [14, 20, 24]. This is due to the fact that no publicly available dataset, including NASDAQ 120, allows researchers to directly identify all HFT [2]. Egginton et al. [14] argues that it is hardly possible to identify orders generated by computer algorithms in the U.S. equities markets and all previous studies used proxies to measure the level of algorithmic trading and HFT[1]. To investigate the implications of HFT on market efficiency, most of the extant research up to date proceeds after somehow identifying via proxy measures or a combination of variables such as trading volume, cancellations, inventory turnover and order-to-execution ratios the trades generated by HFTs

[1] Frino et al. [19] use several proxies to identify algorithmic trading in futures markets.

[16]. Moreover, the huge number of variables and very complicated cause-effect relationships among these variables and potential outcomes imposes another research obstacle [15].

In contrast, this study uses a special adaptive form of the Strongly Typed Genetic Programming (STGP) and millisecond data of E-Mini S&P 500 to demonstrate how HFTs front-run other market participants' orders and generate significant trading profits. There are several reasons for selecting the E-Mini S&P 500. On the one hand, this particular financial instrument is the world's most actively traded stock index futures contract, with over 2.2 million contracts traded per day [17].

In addition, Baron et al. [2] suggest that the E-Mini S&P 500 is a favourable setting for examining HFT because this is a highly liquid market with a high number of HFT companies regularly trading. Moreover, this particular financial instrument is only traded on the Chicago Mercantile Exchange, and there is no concern about unobserved trading orders executed on other trading venues.

The STGP (described in Appendix A) is an extremely suitable sophisticated trading algorithm that successfully replicates HFT scalping strategies. While, Dunis et al. [13] suggest that Genetic Programming (GP) models perform remarkably well even in simple trading exercises, Paddrik et al. [38] report that a zero-intelligence agent-based model of the E-Mini S&P 500 futures market enables close examination of the market microstructure[2]. Östermark [37] suggests that genetic algorithms provide a powerful supplement to traditional econometric techniques, while Chatterjee et al. [8] notes that many statistical and mathematical restrictions can be avoided by employing genetic algorithms. Lensberg et al.

[2] While Python and Java programming languages are suitable for trading at the minute timeframe, C++, ASIC and FPGA languages are appropriate for trading at the very low latencies of microseconds and nanoseconds. Machine learning languages such as Genetic Programming, Strongly Typed Genetic Programming and Genetic Algorithms are appropriate for trading signal research and statistical analysis. All of these programming languages are interconnected in HFT.

[29] suggest that GP is extremely powerful financial tool which minimises the amount of a priory structure associated with traditional functional forms and statistical selection procedures. Mendes et al. [33] employ GP in the foreign exchange market to achieve positive forecasting results. More recently, Chen and Wang [9] point out that GP has the advantage of systematic random search and derivative-free optimisation. We reproduce the HFT scalping strategies in an artificial futures market environment where the impact of these strategies can be examined and new regulations can be evaluated to maintain the overall health of the financial system. Using STGP, we replicate the interactions between HFTs and institutional traders and compare their performance under the same underlying trading order streams. In other words, we replicate real-life futures trading sessions which allow us to avoid the obstacles in the studies discussed above.

Our empirical findings have important implications for market regulators, academics and the general public. To summarise, the contribution of this study is two-fold.

First, this is the first study to use an innovative trading algorithm and millisecond data to provide empirical evidence of how HFT front-running scalping strategies operate in futures markets, imposing discriminatory disadvantages on other market participants. We observe that HFTs frequently cancel recently placed orders from around the best quotes leading to a substantial reduction in the certainty of execution prices making prices more transient. Moreover, the process of placing a large number of trading orders in a short time span creates a false sense of the supply and demand for the E-Mini S&P 500 and hence adversely impacts market quality. Second, we estimate in precise quantitative terms the daily profits of HFT, providing an advantage over existent studies, such as that of Brogaard [4] which observed HFT activities in the aggregate data only, thus preventing them from calculating the exact profitability of HFT. We also propose new regulatory measure such as a minimum resting trading order period of less than 50 milliseconds to mitigate the negative consequences of HFT scalping strategies in futures markets.

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

Related literature

Brunnermeier and Pedersen [5] show that front-running of trading orders leads to price 'overshooting' and amplifies a large trader's liquidation cost and default risk. Moreover, front-running trading strategies make the market illiquid when liquidity is most needed. More recently, Baron et al. [2] estimates that HFTs collectively accumulate over $23 million in trading profits in the E-Mini S&P 500 futures contract during August 2010. Hirschey [24] uses a year of the trader-level data from the NASDAQ to examine return and trade patterns around periods of aggressive buying and selling by HFTs. The author demonstrates that HFTs earn profits by identifying patterns in trade and order data that actually allow them to front-run the order flow and trade ahead of other market participants. Li [30] attempts to model the front-running HFTs and show that they effectively levy a speed tax on traditional traders, making markets less liquid and prices ultimately less informative. Moreover, when infinitely front-running HFTs compete, their negative implications on market quality persist and such negative implications are more severe when HFTs possess more heterogeneous speeds. In another computational experiment Leal et al. [28] build an agent-based model to analyse the interplay between low-and high-frequency trading and its implications on market dynamics. On the one hand, the authors observe that an increase in trading order cancellations leads to higher volatility levels and more intense flash crashes. On the other hand, they also lead to faster price-recoveries which reduce the duration of flash crashes.

Egginton et al. [14] examine all trades and quotes for NYSE and NASDAQ listed stocks for all trading days in 2010 and suggested that order cancellation is a pervasive process with several hundred events occurring during a trading day. They argue that during periods of intense order cancellation financial instruments experience decreased liquidity, higher trading costs and increased short-term volatility.

Sun et al. [40] use tick level data of 105 stocks in the US market from January 2008 to October 2010 to show that HFT can reduce execution costs when supplying liquidity. Jarnecic and Snape [25] analyse the order submission strategies by HFTs and traditional traders in the limit order book by using the sample period from April 1, 2009 to June 30, 2009 for FTSE-100 stocks and confirm our empirical results. Their evaluations suggest that HFTs cancel orders of all durations from around the best quotes, thereby reducing the certainty of execution prices and making trading more difficult for non-HFT participants, by making prices more transient. Similarly, Han et al. [21] construct a simple model of market making in which high-frequency market makers rapidly cancel orders after receiving an adverse signal and observed that low frequency market makers widen the bid-ask spreads, thus leading to liquidity erosion.

In a recent study, Fishe et al. [16] use WTI crude oil futures contract traded on the CME/Nymex exchange from September to December, 2011 to investigate whether there is a class of market participants who follow strategies that appear to anticipate local price trends. The authors demonstrate that there are anticipatory traders capable of processing information prior to the overall market and systematically act before other participants. Kumaresan and Krejic [27] examine the trading trajectories for atomic orders in an environment consisting of several trading venues and carry out an optimization procedure to find the most optimal order placement solution for algorithmic trading orders. Although the authors claim that this is a significant computational breakthrough, they implement execution window measured in minutes, which does not seems to correspond with real-life HFT. In contrast to all other studies on the topic which typically rely on econometric tests only, we use an innovative STGP trading algorithm and millisecond data to demonstrate how HFT front-running scalping strategies operate in practice.

Experimental design.

Twenty years ago, the process in which financial instruments were traded was of simple nature: an investor deciding to buy or sell and transmitted this information to a broker, who then sends the order to an exchange, where bid and ask orders were matched and executed. All market participants had access to the same information about the bid-ask spread.

Today's brokers use trading algorithms to route different segments of an order to different exchanges at super human speed of milliseconds, microseconds and even nanoseconds.

We use a special adaptive form of the Strongly Typed Genetic Programming (STGP), which enables us to choose and adjust different parameters to suit our specification, such as the minimum price increment, number of participants 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 1. We create simulated futures market, which is a hypothetical market with real-world market price data. Each market participant in our experiment represents an artificial trader who is equipped with their own trading rule, where 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 Appendix A. The main idea is that the trader's trading rule will improve by a natural selection process based on the survival of the fittest [31]. Hence, 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, STGP trading algorithm evolves the model step-by-step by feeding it with millisecond quotes the E-Mini S&P 500, and therefore the forecasting models evolve mimicking the real-life futures market.

The process of developing trading rules

Initially, each individual trader has only one trading rule which is created randomly which enables 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. This process is repeated until at least one trading rule in the population achieves the desired level of fitness, measured by a trader's investment return over a specified period. It should be noted that this initial random nature can result in the creation of meaningless trading rules or trading rules which cannot be evaluation thoroughly since they do not return the value that function needs. Nevertheless, as Montana [35] notes, these programming issues can be resolved by the introduction of STGP, where the process requires the definition of a specific set to fit the problem.

Each trading rule in our artificial futures market setting take historical millisecond prices of the E-Mini S&P 500 and generate advice which consists of the desired position which is estimated as a percentage of the trader’s wealth and an order limit price for buying and selling the financial instrument[3].

The trading rules logic comprises of 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 contrasts, 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. [39] 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 stability [31].

Another important feature of the STGP is that each trader discovers the intrinsic value of the E-Mini S&P 500 individually without any communication between traders, ensuring individuality and that the level of intelligence of each artificial trader is not affected by other traders. This allows the development of more meaningful trading rules for both HFTs and institutional traders.

Structure of the artificial futures market and the differences between HFTs and institutional traders.

We examine HFT front-running scalping strategies within the context of artificial futures market populated by 100,000 boundedly rational traders. All artificial traders in the model are

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

not orientated towards a predetermined formation of strategies and therefore are free to develop and continually evolve new and better trading rules through time. Our artificial futures market is populated by 80,000 institutional traders and 20,000 HFTs (20 per cent of the total population based on the continuous Breeding Fitness Return).

Both HFTs and institutional traders trading rules are created using STGP programming technique explained in Appendix A. However, the main difference between the two trading groups is that the HFTs' group consists of the traders that momentarily perform best in terms of the continuous Breeding Fitness Return, and therefore they possess lower latency. Although the institutional traders and HFTs both observe the same millisecond data of the E-Mini S&P 500 and generate trading orders, HFTs are able to access and process the data first due to their low latency features. In other words, HFTs are able to foresee the quotes of the E-Mini S&P 500 and submit trading orders before institutional traders. The Breeding Fitness Return is a trailing return of a wealth moving average which determines the fitness rules of traders. This return is calculated over the last *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 250. In the case where the age is less than *n*, no value is calculated. This particular type of return is used to measure the fitness criterion for the selection of traders to breed.

Breeding is, in essence, a process of creating new artificial traders to replace poor performing ones based on the values derived from Equation (1) below. Both HFTs and institutional traders operate in the same market and accumulate wealth by investing in two financial instruments that are available in the artificial stock market – the risky E-Mini S&P 500 and the risk-free instrument represented by cash. Because our artificial futures market continuously evolve, traders with trading rules that perform well become wealthier, positively influencing the forecasting accuracy of the model. In each period, an artificial trader has wealth given by the following formula:

$$W_{i,t} = M_{i,t} + P_t h_{i,t} \tag{1}$$

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 the E-Mini S&P 500 held by artificial trader i respectively, in period t, and P_t is the price of the E-Mini S&P 500 in period t.

The clearing mechanism and order generation for the artificial futures market.

Our artificial futures market is a simulated double auction market, where all the buy and sell orders are collected. The artificial traders receive historical quotes of the E-Mini S&P 500 and evaluate their trading rule and subsequently calculate the number of contracts they need to buy or sell. If contracts 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 1,000 contracts of the E-Mini S&P 500 which is priced at $38.50 and has $80,000 in cash, their wealth is $118,500 and their position in E-Mini S&P 500 is 32.5%. If the trading rule generates a signal of a position of 50% and a limit price of $38.50, the limit order will be produced to purchase 539[4] additional E-Mini S&P 500 contracts with a price of $38.50. The artificial futures market 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 contracts purchased by traders is always equal to the number sold by other traders and if the number of

[4] 50% * (118,500/38.50) – 1000 = 539 contracts.

contracts offered and the number of contracts 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 futures market price will be the price of the previous quote [31]. As in real-life E-mini market, there is no designated market maker and there are no short-selling constraints.

Description of data and transaction costs.

The dataset used in this study consist of real-life millisecond data of the E-Mini S&P 500 from February, 2014 to December, 2014. We obtained the data from Nanex (www.nanex.net). We only select the front-month dataset for each month – the contract with the nearest expiration date. The E-Mini S&P 500 expirations months are March, 2014; June, 2014; September, 2014; and December, 2014. For empirical investigation we focus on February, 2014, which has a March, 2014 expiration; May, 2014, which has a June, 2014 expiration; August, 2014, which has a September, 2014 expiration and November, 2014, which has a December, 2014 expiration. This allows us to exclude months in which the leading contract expires in order to eliminate the rollover effect. Although the E-Mini S&P 500 trades virtually round the clock, we only consider millisecond data during regular market hours when the markets of the underlying equities of the S&P index are open and before the daily halt in trading: weekdays from 8:30 a.m. to 3:15 p.m. Central Standard Time (CST). Each contract has a multiplier of $50 times the value of the underlying S&P 500 index, and therefore a contract with an index value of 1,000 suggest that the futures contract is valued at $50,000. The tick size in E-Mini S&P 500 is 0.25 index points. Hence, considering the $50 multiplier, a one tick change is equal to $12.50.

The STGP trading algorithm processed 18,655,490 trading messages stamped at the millisecond interval for the E-Mini S&P 500 in February, 2014; 22,878,525 trading messages in May, 2014; 27,368,175 trading messages in August, 2014; and 16,282,009 trading messages in November, 2014.

Baron et al. [2] report that the cost of exchange fees per contract is $0.15[5]. We employ transaction costs of $0.20 per contract for our profit calculations. Although slightly higher than the current standards, the level of transaction costs takes into account the costs of HFT companies. These include software platforms, labour and risk management systems but does not include co-location of services (Aitken et al.[1] argue that the presence of HFT leads to the introduction of co-location services).

Experimental results

Traders' activity on artificial futures market.

The aim of this section is to investigate artificial traders' activity on our futures market, which has been designed to run in parallel with real-life futures market. All empirical tests below are based on data generated by the STGP trading algorithm for HFTs and institutional traders.

First, we examine what happens to the limit orders of the E-Mini S&P 500 after they are submitted to the artificial futures market. Let τ denote the time between order submission and cancellation. The probability of cancellation in the interval $(0, t]$ is represented by the distribution function:

[5] The E-Mini S&P 500 futures market does not involve marker – taker transaction costs unlike most equity markets.

$$P_{Cancel}(t) = \Pr(\tau \leq t) \quad (2)$$

We extract all trading activity generated by the STGP trading algorithm for the E-Mini S&P 500 to estimate the distribution function using the life-table method, and taking execution as the censoring event. In contrast to all other studies we are able to observe and count the number of executed and cancelled orders by extracting generated data from the STGP trading algorithm.

Table 2 shows that a large number of limit orders submitted by HFTs are cancelled almost immediately after submission. Table 2 reports that, $P_{Cancel}(50)$, the probability of cancellation within 50 milliseconds is 0.918. By the time 500 milliseconds have elapsed, this probability dramatically decreases to 0.056. At the same time the probability of cancellation for institutional traders measured at 50 milliseconds is 0.001 increasing to 0.028 at 500 milliseconds. A comparison of cancelled orders by HFTs and institutional traders indicates that HFTs cancel substantially larger proportion of orders after a very short duration. The extremely high level of cancelled orders indicates the high level of aggressiveness of HFTs in exploiting the orders placed by institutional traders on the artificial futures market. This is consistent with Hasbrouck and Saar [22] who suggests that over one third of limit orders are cancelled within two seconds by HFTs.

In addition, Leal et al. [28] highlight that this type of empirical results stimulates the occurrence of high bid-ask spreads in the market, thus increasing the likelihood of a significant decrease in the price of any financial instrument. A direct comparison between cancelled and executed orders indicates that execution seems the less probable event for all time intervals. Furthermore, Jarnecic and Snape [25] report that short duration orders contribute to the difficulty of trading by non-HFT participants by lifting quotes and inhibiting the certainty of long-term investors when attempting to demand liquidity. This finding motivates us to examine the exact location of short duration orders to find out whether HFTs operate with those orders inside the spread. This type of analysis is important due to the fact that frequent removal of orders located within the spread can reduce the certainty of the execution price for institutional traders trying to demand liquidity.

Table 3 report that order cancelations are present inside or at the best quotes, and this finding is substantially more pronounced for HFTs measured up to 50 milliseconds. HFTs generate 71.24% of order cancellations within the quoted prices, and a further 18.49% at the best quotes. The empirical results in Table 4 reveals that for orders that are cancelled with greater than 50 millisecond frequency, the percentage for orders cancelled by HFTs decreases to 63.28% inside the best price and a further 15.01% at the best quote. We observe the opposite trend with institutional traders.

They increase the amount of cancelled limit orders within the best quote from 8.10% measured up to 50 milliseconds to 12.53% with frequencies greater than 50 milliseconds. These findings indicate that HFTs frequently cancel trading orders of different durations but more often cancel recently placed orders from around the best quotes. As a result, the certainty of execution prices has been substantially reduced making prices more transient and imposing trading obstacles for non-HFT participants. Moreover, frequent order cancellation creates a false sense of supply and demand for the E-Mini S&P 500. By quickly cancelling a large number of orders within the 50 millisecond interval, HFTs could create potentially exploitable latency arbitrage opportunities.

Budish et al. [6] report that there are about 800 such arbitrage opportunities per day in the two largest securities that track the S&P 500 index alone – the E-Mini S&P 500 and the iShares SPDR S&P 500 exchange traded fund, totalling $75 million per year.

Traders' profitability in artificial futures market.

The aim of this section is to measure the level of profitability of artificial traders operating in futures markets. All empirical tests below are based on data generated by the STGP trading algorithm for HFTs and institutional traders.

One of the most important characteristics of high frequency millisecond data is the high presence of no price changes in the E-Mini S&P 500. We take into account this market inactivity by modifying the Student's *t* distribution associated with the standardized residuals:

$$f\left\langle\frac{\in_t}{\sigma_t}\middle|\delta_t\right\rangle = \{p_0 \text{ if } \delta_t = 1 \tag{3}$$

Or

$$f\left\langle\frac{\in_t}{\sigma_t}\middle|\delta_t\right\rangle = \begin{cases}\dfrac{g_{v(\in_t/\sigma_t)}}{1-p_0}\end{cases} \text{ if } \delta_t = 0 \tag{4}$$

where $g_{v(g)}$ measure the Student's density function; $\in_t$ represents the residuals of the time series; σ_t is the standard deviation of the time series; p_0 represents the probability of a sequence of two zero returns; δ_t measure market inactivity as follows:

$$\delta_t = \begin{cases}1\\0\end{cases} \text{if, otherwise} \tag{5}$$

If $\delta_t = 1$, the forecast $x_{t+i|t} = 0$ for $i = 1,2 \cdots$ [32].

Given the large amount of millisecond trading messages, an important issue that arises is the Lindley's paradox. This phenomenon can potentially lead to overstatement of statistical significance and a tendency to reject the null hypothesis even when the posterior odds favour the null.

Connolly [10] proposes the following equation to overcome the issue and estimate sample size adjusted critical values for *t* statistics:

$$t^* = \left[(T-k)\left(T^{1/T}-1\right)\right]^{1/2} \tag{6}$$

where *T* is the sample size; *k* measure the number of estimated parameters. The null hypothesis is the posterior probability, which is the statistical probability that a hypothesis is true computed in the light of relevant observations. We implement large-sample adjustments to the critical *t*-values in order to avoid overstatement of statistical significance. If a calculated test statistic exceeds the appropriate critical value from Equation 6, the sample evidence is said to favor the alternative hypothesis. First, in order to evaluate statistically the forecasting abilities of HFTs and institutional investors, we estimate the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE). For all three of the error statistics retained, the lower the output, the better the forecasting accuracy of the model:

$$RMSE = \sqrt{\frac{1}{n}\sum_{\tau=t+1}^{t+n}(Y_\tau - Y_t)^2} \tag{7}$$

$$MAE = \left(\frac{1}{n}\right)\sum_{\tau=t+1}^{t+n}|Y_\tau - Y_t| \tag{8}$$

$$MAPE = \frac{1}{n}\sum_{\tau=t+1}^{t+n}\left|\frac{Y_\tau - Y_t}{Y_t}\right| \tag{9}$$

where Y_t represents the actual values of the price of the E-Mini S&P 500; Y_τ is the forecasted values of the price of the E-Mini S&P 500. When difference between actual and forecasted values of the price of the E-mini S&P 500 is far off, RMSE, MAE and MAPE are bigger values and therefore the forecasts are less accurate. A direct forecasting comparison between

HFTs and institutional traders indicate the superiority of the former. The RMSE, MAE and MAPE for HFTs are significantly smaller than the errors produced by institutional traders (Table 5).

Table 5 illustrate that the HFTs presents the best statistical results in all four months under investigation for the E-Mini S&P 500. HFT scalping strategies trading the E-Mini S&P 500 in August, 2014 outperform the other three months measured by the lowest forecasting errors. We compute the modified Diebold-Mariano (MDM) test, which is an extension of the Diebold and Mariano [12], test to verify the statistical significance of the above analysis. Under the null hypothesis of the test is the equivalence in forecasting accuracy between several models:

$$MDM = T^{-1/2}[T + 1 - 2k + T^{-1}k(k-1)]^{1/2}DM \quad (10)$$

where T represents the number millisecond trading messages for the E-Mini S&P 500 in February, 2014; May, 2014; August, 2014; and November, 2014; k the number of the one-step-ahead forecasts; *DM* represents the Diebold-Mariano test which compares the forecast accuracy of two forecast methods. The null hypothesis under the test is that they have the same forecast accuracy. A negative value of the MDM test suggests that the first forecast is more accurate than the second forecast. We apply the MDM test to measure the predictive abilities of HFTs vs. institutional traders.

The test measures each period with student distribution of T-1 degrees of freedom (MSE and MAE are used as loss functions). Table 6 presents the statistics for the four months under investigation, comparing the performance of HFTs with institutional traders. Table 6 indicates that the null hypothesis of the modified Diebold-Mariano test of equal forecasting accuracy has been rejected for all comparisons and for both loss functions at the 1% level of significance. Moreover, the statistical superiority performance of HFTs' strategies is confirmed by the negative values of the MDM statistic for both loss functions.

The RMSE, MAE and MAPE are all important error measures, yet they may not correspond to profits. We therefore calculate the daily profits generated by HFTs and institutional traders for the most traded five days in each month (full trading volume reported in Appendix B). 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 HFT and institutional trader begins each trading day with a zero inventory position.

Baron et al. [2] suggest that a marking-to-market modelling mechanism is an appropriate profit measuring tool for market participants who end the trading day with a zero inventory. We estimate the end of the day profits for both artificial trading groups 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 the E-Mini S&P 500 at close of trading:

$$\pi_{i,t} = \sum_{n=1}^{N_{i,T}} p_n y_{i,n} + p_T y_{i,T} \tag{11}$$

where $n = 1, \cdots, N_{i,T}$ denotes the trades for trader i from the start of the trading day to the end of the trading day; p_n 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. Transaction costs of $0.20 per contract are taken into account. Table 7 shows that HFTs generate significantly higher profits than institutional traders for the E-Mini S&P 500 in all four months.

We observe that HFTs' profits are higher due to their higher speed friction γ. Here, higher γ means more severe HFTs' front-running. Li [30] defines the speed friction as:

$$\gamma = \sum_{j=1}^{J} \left(\alpha_j (1 - \alpha_j) \prod_{i=0}^{j-1} (1 - \alpha_1)^2 \right) \tag{12}$$

where $\alpha_j = \frac{n_j}{n_{j+1}}$ is the equilibrium fast trading intensity for all j. The speed friction is not affected by the other parameters and is determined entirely on the fast traders' speed profile $\{n_1, n_2, \cdots, n_j\}$. The profits of HFTs come from the price impact of institutional trades in our artificial futures market. This finding is consistent with Baron et al. [2], who claim that trading the E-Mini S&P 500 is a zero-sum gain: one trader's profit comes directly at the expense of another trader. Furthermore, Li [30] argues that in the presence of more fundamental uncertainty, the price impact of trades is higher and front-running an order is more profitable. At the same time when there is more noise trading on the market, the trading volume is higher and there are more trading orders available for front-running.

While HFT scalping strategies are very profitable, they might carry some risk on a day-to-day basis. The standard deviation of the profits (Table 8) reports a wide variety of different values, with the highest variation of profits ($391) generated by HFTs in August, 2014. We estimate the probability of default for both groups of traders as an arithmetic Brownian motion with constant drift α and constant volatility σ. Considering the fact that the daily profits for August, 2014 are normally distributed with mean $\alpha = \$391$, standard deviation $\sigma = \$7.50$, and the initial wealth (V_0) of all artificial traders is $100,000, we can estimate the probability of HFTs' default in August by implementing the following formula based on the theory of hitting times by Karlin and Taylor [26].

$$p(default) = \exp\left(\frac{-2\alpha V_0}{\sigma^2}\right) \tag{13}$$

By calibrating Equation (13) to the values of α, σ^2, and V_0 listed in Table 8, we find that HFTs' probability of default is virtually zero.

Although trading profits reported in Table 7 give us an idea of the real magnitude of HFTs' profits, risk-adjusted performance is of prime importance. The monthly Sharpe ratio for HFTs and institutional traders has been calculated as:

$$SR_{i,t} = \frac{r_{i,t} - r_f}{\sigma_i} * \sqrt{252/12} \tag{14}$$

where $r_{i,t}$ represents the average daily return estimated from the daily profit; σ_i is the standard deviation of trader $i's$ return over the sample period; r_f is the risk-free rate set at the value of the daily continuously compounded rate converted from the annualised investment yield on a one-month US Treasury bill (data up to 31st December, 2014 has been downloaded from the Federal Reserve statistical release website at www.federalreserve.gov/releases/h15). Table 9 illustrates that HFTs have the highest risk-return tradeoff, generating a Sharpe ratio of 1.99 in August, 2014. Hence, we conclude that while HFTs bear some minimum risk, their risk-adjusted returns are much higher than institutional traders within artificial futures market settings.

To examine the trading horizon of HFTs in the most profitable month, we follow Hasbrouck and Sofianos [23] and decompose their profits in August, 2014 (based on most traded five days) over different time frames by applying spectral analysis. The timeframe over which HFTs generate their profits provides more specific details about their trading strategies. Spectral analysis view marked-to-market profits as a function of two different time series such as prices and the level of inventory, which can vary at different frequencies.

Similar to Baron et al. [2], 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 (HFTs buy before the price of the E-Mini S&P 500 increases) they generate profits. If the two time series are not in a phase (HFTs buy before the price of the E-Mini S&P 500 decreases) they experience losses. Marked-to-market profits for HFTs can be expressed as:

$$\pi_\tau = \sum_{t=0}^{\tau} x_t (p_t - p_{t-1}) = \sum_{t=0}^{\tau} x_t \cdot \Delta p_t \tag{15}$$

where x_t represents the inventory holdings of HFTs at time t and p_t is the price of E-Mini S&P 500 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 HFTs' 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. [2] and develop the following two functions:

$$x(\omega) \sum_{t=0}^{T} x_t e^{2\pi i t\omega/T} \tag{16}$$

$$\Delta p(\omega) = \sum_{t=0}^{T} \Delta p_{t+1} e^{2\pi i t\omega/T} \tag{17}$$

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 (17) and obtain the following:

$$\pi_T = \frac{1}{T} \sum_{\omega=1}^{\infty} \hat{x}(\omega)\, \Delta p(\omega) = \frac{1}{T} \sum_{\omega=1}^{\infty} 2 * Real(\hat{x}(\omega)\Delta p(\omega)) \tag{18}$$

Where, *Real* represents a function that takes a real part in a complex number; $2 * Real(\hat{x}(\omega)\Delta p(\omega))$ is the component of the marked-to-market profits generated by HFTs at frequency ω. The second equality in Equation (18) is a result based on the fact that an imaginary part of $\hat{x}(\omega)\Delta p(\omega)$ is equal to zero.

Table 10 shows that in August, 2014 HFTs make the largest profits of $630 at the very short interval between 0 and 50 milliseconds and the smallest profits of $42 at the longest time scale between 3,501 and 4,000 milliseconds. Therefore, the HFTs do not try to infer the long-term fundamental value of the E-Mini S&P 500 but emphasize entirely on capturing short-term price dynamics. We have found that HFTs' 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.

The results of spectral analysis are consistent with the notion that HFTs generate profits by anticipating and front-running the order flow. Narang [36] have estimated that front-running generates \$1.5 to \$3 billion in annual profits for HFTs in the US equity market alone. To examine the actual persistence of HFTs' profits in August 2014, we investigate whether profits from a previous day's trading are a good predictor of the current day's profits. This is an important robust exercise because persistent profits distributed over time indicate that HFTs will extend their strong performance in the future at the expense of institutional traders.

Baron et al. [2] propose the following OLS regression which we implement in our examination for persistence of HFTs' profits:

$$Profit_{i,t} = \alpha + \beta_1 Profit_{i,t-1} + \beta_2 Aggressiveness_{i,s} + \beta_3 Volume_{i,t} + \beta_4 Volatility_{s,t} + \varepsilon_{i,t} \tag{19}$$

where $Profit_{i,t}$ represents modified log profits such as $sign(profits) * \log(1 + |profits|)$ to incorporate any negative profits; $Volume_{i,t}$ is the log of each artificial trader's trading volume for day t; $Volume_{i,t}$ denotes the price volatility for day t defined as the volume-weighted standard deviation of the price process for the same day; $Aggressiveness_{i,s}$ represents the trader $i's$ volume-weighted aggressiveness ratio. The univariate results for the HFTs' profits in August, 2014 (Table 11) reveal statistical significance indicating that one-day lagged performance is a good predictor of the current day's performance. Similarly, the statistical significance of the control variables $Volume_{i,t}$, $Volatility_{s,t}$ and $Aggressiveness_{i,s}$ demonstrates the persistence of HFTs' profitability because the specification with control variables maintains the statistical significance from the univariate regression results. This finding indicates that profitability is persistent even after controlling for time effects. This is in line with the findings of Baron et al. [2] but opposite to Chae et al. [7], who point out that algorithmic traders incur losses by trading.

The ever-increasing demand for speed and technological improvements creates an arms race issue and raises questions whether the speed of incorporating information into the market at the millisecond timeframe has any social value. In 2010, an American company named Spread Networks invested $300 million in a new high-speed fiber optic cable in order to reduce round-trip communication time between New York and Chicago from 16 milliseconds to 13 milliseconds. In 2015, several HFT companies invested in microwaves rather than fiber optic cable due to the fact that the light travels faster through air than glass. The new microwave technology helps decreasing transmission time from 13 milliseconds to 8.1 milliseconds. Similar speed races in financial markets occur on a regular basis, often measured at microsecond (millionth of a second) and even nanosecond (billionth of a second) timeframes. As a benchmark to this superhuman speed of trading we would like to highlight that the blink of a human eye lasts approximately 400 milliseconds. Delaney [11] uses techniques from real options analysis to provide insights into the optimal time traders should invest in high frequency technologies. From a social welfare perspective, in order to be socially optimal, traders should wait longer when the cost of technology is very high and the level of and HFT is also high. Furthermore, the author shows that the level of HFT always exceeds the socially optimal welfare level.

Biais et al. [3] provide an analysis of the implications of a Pigovian tax (a tax applied to market activity that is generating negative externalities) on HFT and demonstrate that the socially optimal level of HFT would be reached if the tax imposed is equal to the externalities generated by HFT. On 6th of May, 2010 the front-month of June E-Mini S&P 500 experienced dramatic decline of 5.1% within a 13 minute period. A cascade of executed orders decreased further the price of the E-Mini S&P 500 to 6.4%. The next executed order triggered the CME Globex Stop Logic Functionality, which pauses execution of all orders for 5 seconds, if the next transaction were to execute outside the price range of 6 index points. During this pause of 5 seconds (named the 'Reserve State') the market is still open and market participants are allowed to submit, modify or cancel trading orders. However, execution of pending trading orders is delayed until actual trading resumes after 5 seconds.

To mitigate the negative consequences of HFTs and eliminate front-running, we propose the following regulatory measures. First, based on our empirical findings, we propose a cooling-off period of less than 50 milliseconds rather than 'Reserve State' of 5 seconds[6]. The current regulatory debates include a cooling-off period of 500 milliseconds. To minimize the number of cancelled orders, market regulators worldwide are currently discussing a so-called minimum resting trading order period. This would require an order to stay on an order book for 500 milliseconds eliminating traders who operate at much faster speeds. However, both the proposed minimum resting trading order period of 500 milliseconds and the current 'Reserve State' practice of 5 seconds does not seem to be efficient when compared to our empirical findings.

Robustness checks.

To examine the robustness of our empirical findings we modified some of the artificial market parameters. Panel A of Table 12 shows the rate of cancellation and execution of limit orders by 10,000 HFTs (10% of the total population, genome depth of 10 and genome size of 2,048) and 90,000 institutional investors (90% of the total population, genome depth of 10

[6] Our policy recommendation is based on trading orders executed at the millisecond interval only. With recent technological improvements in software and hardware trading orders are executed at the microsecond and even nanosecond intervals. Therefore our policy recommendation may not be efficient at these timeframes.

and genome size of 2,048). The probability of cancellation within 50 milliseconds, $P_{Cancel}(50)$ is 0.702, compared to the probability of cancellation of 0.918 in a market populated with 20,000 HFTs and 80,000 institutional investors.

Panel B of Table 12 illustrate the rate of cancellation and execution of limit orders by 40,000 HFTs (40% of the total population, genome depth of 40 and genome size of 8,192) and 60,000 institutional investors (60% of the total population, genome depth of 40 and genome size of 8,192). The probability of cancellation within 50 milliseconds in this case is 0.994, compared to the probability of cancellation of 0.918 in a market populated with 20,000 HFTs and 80,000 institutional investors.

Our robustness profitability checks in Table 13 suggest that reduced number of HFTs (10% of the total population) operating in the artificial futures market is leading to decreased profitability in all four front-months of the E-Mini S&P 500 in 2014. We observed increased profitability in all front-months under investigation when we introduced more HFTs to the market. Our profit-based estimations in Table 14 shows that the greater the number of HFTs in the market, the higher the profits. These findings indicate that greater presence of HFTs in the market is leading to cancellations of larger proportions of trading orders and greater profitability. This is in line with our initial empirical results.

Conclusions

Rapid improvements in the technological base for generating and executing trading orders dramatically increased the speed and sophistication of the trading tools available to market participants. Making an accurate bid or ask call in the futures markets is no longer a sufficient condition for generating profits. Determining a fast-moving opportunity in front of the other market participants seems to have the greatest influence. However, the practice of computers running futures markets raised concerns among investors and regulators around the world.

In this study, we simulate real-life trading within artificial futures market settings and observe that HFTs generate a large number of cancelled orders within 50 milliseconds which may make trading more difficult and costly for institutional traders who lack access to sophisticated software platforms for HFT. We have found that HFTs are a major user of very short duration orders that are frequently cancelled from inside or at the best quotes. This particular trading behaviour reduces the certainty of execution and imposes trading obstacles for institutional traders by making the price of the E-Mini S&P 500 more transient. A direct forecasting comparison between HFTs and institutional traders indicate the superiority of the former. Our spectral analysis confirms that HFTs generate profits by front-running the order flow. If one group of market participants such as HFTs generates faster access to the order flow than institutional traders, those participants with their lower latency would have an unfair advantage in the marketplace. Overall, a high level of cancelled orders combined with scalping strategies could impose severe picking-off risks for undisclosed trading orders and may make them very inefficient.

In terms of market regulation, we think that the introduction of a minimum resting trading order period of less than 50 milliseconds could impose an obstacle for profit generation of HFTs. Minimum resting trading order period of less than 50 milliseconds could lead to HFTs facing a queuing risk resulting in a less harmful market quality effect. One practical implication of our study is that we demonstrate that market regulators can apply artificial intelligence tools such as STGP to conduct trading behaviour-based profiling. This can be used to detect the occurrence of new HFT strategies and examine their impact on the futures market.

References

1. Aitken, M., Cumming, D., and Zhan, F. Trade size, high-frequency trading, and colocation around the world. *European Journal of Finance*, (2014), 1-21.

2. Baron, M., Brogaard, J. and Kirilenko, A. The trading profits of high frequency traders. *SSRN Electronic Journal*, (2012).

3. Biais, B., Foucault, T. and Moinas, S. Equilibrium fast trading. *Journal of Financial Economics*, 116, 2 (2015), 292-313.

4. Brogaard, J. A. High frequency trading and its impact on market quality. *Working Paper*, Kellogg School of Management, 2010.

5. Brunnermeier, M. K. and Pedersen, L. H. Predatory trading. *Journal of Finance*, 60, 4 (2005), 1825-1863.

6. Budish, E., Cramton, P. and Shim, J. The high-frequency trading arms race: Frequent batch auctions as a market design response. *Quarterly Journal of Economics*, 130, 4 (2015), 1547-1621.

7. Chae, J., Khil, J. and Lee, E. J. Who makes markets? Liquidity providers versus algorithmic traders. *Journal of Futures Markets*, 33, 5 (2013), 397-420.

8. Chatterjee, S., Laudato, M. and Lynch, L. A. Genetic algorithms and their statistical applications: An introduction. *Computational Statistics and Data Analysis*, 22, 6 (1996), 633-651.

9. Chen, Y. and Wang, X. A hybrid stock trading system using genetic network programming and mean conditional value-at-risk. *European Journal of Operational Research*, 240 3 (2015), 861-871.

10. Connolly, R. A. An examination of the robustness of the weekend effect. *Journal of Financial and Quantitative Analysis*, 24, 2 (1989), 133-169.

11. Delaney, L. An examination of the optimal timing strategy for a slow trader investing in a high frequency trading technology. *Working Paper*, City University of London, 2015.

12. Diebold, F. X. and Mariano, R. S. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13 (1995), 253-265.

13. Dunis, C. L., Laws, J. and Karathanasopoulos, A. GP algorithm versus hybrid and mixed neural networks. *European Journal of Finance*, 19, 3 (2013), 180-205.

14. Egginton, J. F., Van Ness, B. F. and Van Ness, R. A. Quote stuffing. *Financial Management*, 45, 3 (2016), 583-608.

15. Felker, T., Mazalov, V. and Watt, S. M. Distance-based high-frequency trading. *Procedia Computer Science*, 29 (2014), 2055-2064.

16. Fishe, R. P. H., Haynes, R. and Onur, E. Anticipatory traders and trading speed. *Journal of Financial and Quantitative Analysis*, Forthcoming (2018).

17. Folger, J. *Beginners guide to E-Mini futures contracts: E-Mini characteristics*, 2015. http://www.investopedia.com/university/how-to-trade-e-mini-futures-contracts/ (Accessed on April 20 2015).

18. Foucault, T., Kozhan, R. and Tham, W. W. Toxic arbitrage. *Review of Financial Studies*, 30, 4 (2017), 1053-1094.

19. Frino, A., Mollica, V. and Webb, R. I. The impact of co-location of securities exchanges and traders computer servers on market liquidity. *Journal of Futures Markets*, 34, 1 (2014), 20-33.

20. Goldstein, M. A., Kumar, P. and Graves, F. C. Computerized and high-frequency trading. *Financial Review*, 49, 2 (2014), 177-202.

21. Han, J., Khapko, M. and Kyle, A. S. Liquidity with high-frequency market making. *Swedish House of Finance Research Paper No. 14-06*, 2014.

22. Hasbrouck, J. and Saar, G. Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets*, 12, 2 (2009), 143-172.

23. Hasbrouck, J. and Sofianos, G. The trades of market makers: An empirical analysis of NYSE specialists. *Journal of Finance*, 48, 5 (1993), 1565-1593.

24. Hirschey, N. Do high-frequency traders anticipate buying and selling pressure?. *Working Paper*, London Business School, 2017.

25. Jarnecic, E. and Snape, M. The provision of liquidity by high-frequency participants. *Financial Review*, 49, 2 (2014), 371-394.

26. Karlin, S. and Taylor, H. M. *A First Course in Stochastic Processes*. 2 edn. New York: Academic Press, 1975.

27. Kumaresan, M. and Krejić, N. Optimal trading of algorithmic orders in a liquidity fragmented market place. *Annals of Operations Research*, 229, 1 (2015), 521-540.

28. Leal, S. J., Napoletano, M., Roventini, A. and Fagiolo, G. Rock around the clock: An agent-based model of low- and high-frequency trading. *Journal of Evolutionary Economics*, 26, 1 (2016), 49-76.

29. Lensberg, T., Eilifsen, A. and McKee, T. E. Bankruptcy theory development and classification via genetic programming. *European Journal of Operational Research*, 169, 2 (2006), 677-697.

30. Li, W. High frequency trading with speed hierarchies. *Working Paper*, John Hopkins University, 2017.

31. Manahov, V. The rise of the machines in commodities markets: New evidence obtained using Strongly Typed Genetic Programming. *Annals of Operations Research*, 260, 1 (2018), 321-352.

32. Meade, N. A comparison of the accuracy of short term foreign exchange forecasting methods. *International Journal of Forecasting*, 18, 1 (2002), 67-83.

33. Mendes, L., Godinho, P. and Dias, J. A Forex trading system based on a genetic algorithm. *Journal of Heuristics*, 18, 4 (2012), 627-656.

34. Menkveld, A. J. and Zoican, M. A. Need for speed? Exchange latency and liquidity. *Review of Financial Studies*, 30, 4 (2017), 1188-1228.

35. Montana, D. J. Strongly Typed Genetic Programming. *Evolutionary Computation*, 3, 2 (1995), 199-230.

36. Narang, R. K. *Inside the Black Box: A Simple Guide to Quantitative and High Frequency Trading*. New Jersey: Wiley, 2013.

37. Östermark, R. Genetic modelling of multivariate EGARCH processes: Evidence on the international asset return signal response mechanism. *Computational Statistics and Data Analysis*, 38 (2001), 71−93.

38. Paddrik, M., Hayes, R., Todd, A., Yang, S., Beling, P. and Scherer, W. An agent based model of the E-Mini S&P 500 applied to flash crash analysis. *Computational Intelligence for Financial Engineering and Economics*, (2012), 1-8.

39. Sermpinis, G., Stasinakis, C., Theofilatos, K. and Karathanasopoulos, A. Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms—Support vector regression forecast combinations. *European Journal of Operational Research*, 247, 3 (2015), 831-846.

40. Sun, E. W., Kruse, T. and Yu, M.-T. High frequency trading, liquidity, and execution cost. *Annals of Operations Research*, 223, 1 (2014), 403-432.

Appendices

Artificial stock market parameters	
Total population size (traders)	100,0000
HFTs' size(percentage of the total population)	20%
Initial wealth(equal for all traders)	100,000
Transaction costs	$0.20 per contract
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	4096*
Maximum genome depth	20**
Minimum initial genome depth	2
Maximum initial genome depth	5
Breeding cycle frequency (quotes)	1
Minimum breeding age (quotes)	80***
Initial selection type	random
Parent selection (percentage of initial selection that will breed)	5%****
Mutation probability (per offspring)	10%
Total number of millisecond quotes processed- E-Mini S&P 500, February, 2014	18,655,490
Total number of millisecond quotes processed- E-Mini S&P 500, May, 2014	22,878,525
Total number of millisecond quotes processed- E-Mini S&P 500, August, 2014	27,368,175
Total number of millisecond quotes processed-E-Mini S&P 500, November,2014	16,282,009
Creation of unique genomes	Yes
Offspring will replace the worst performing traders of the initial selection	Yes

* 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.

** Maximum genome depth measures the highest number of hierarchical levels that occurs in a trader's genome (trading rule). The depth of a trading rule can be an indicator of its complexity.

*** This 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 over the last 80 quotes will be compared.

**** 5% of the best performing traders of the initial selection that will act as parents in crossover operations for creating new traders.

Table 1. Artificial futures market parameter settings.

Time (milliseconds)	Cancellation	Execution
HFTs		
0-50	0.918	0.069
51-100	0.824	0.040
101-200	0.667	0.023
201-300	0.211	0.017
301-500	0.056	0.005
Institutional traders		
0-50	0.001	0.016
51-100	0.014	0.099
101-200	0.018	0.174
201-300	0.021	0.388
301-500	0.028	0.779

This table presents cumulative probabilities of cancellation and execution within the millisecond interval. Data has been generated and extracted from the STGP trading algorithm for the front-month of the E-Mini S&P 500 (February, 2014; May, 2014; August, 2015 and November, 2014). The probabilities are estimated as $1 - S(t)$, where $S(t)$ represent the survival function of cancellation and execution. In order to calculate the survival function we extracted all trading activity for E-Mini 500 S&P 500 generated by STGP trading algorithm and used the life-table method.

Table 2. The rate of cancellation and execution of limit orders by HFTs and institutional traders generated by STGP trading algorithm for the E-Mini S&P 500's front-month (February, 2014; May, 2014; August, 2014 and November, 2014).

Cancellations with durations $\leq$ 50 milliseconds	
HFTs	
Order location	**Percentage of orders**
Inside best	71.24*
At best	18.49*
Institutional traders	
Inside best	28.31*
At best	8.10*

This table reports the location and percentage of cancelled limit orders by HFTs and institutional traders for E-Mini S&P 500. The significance of the differences between HFT s and institutional traders are estimated using z-statistics for comparing two proportions. * indicates statistical significance at the 1% level.

Table 3. Location of cancelled limit orders ($\leq$50 milliseconds) generated by STGP trading algorithm for the E-Mini S&P 500's front-month (February, 2014; May, 2014; August, 2014 and November, 2014).

Cancellations with durations > 50 milliseconds	
HFTs	
Order location	**Percentage of orders**
Inside best	63.28*
At best	15.01*
Institutional traders	
Inside best	19.18*
At best	12.53*

This table reports the location and percentage of cancelled limit orders by HFTs and institutional traders for E-Mini S&P 500. The significance of the differences between HFTs and institutional traders are estimated using z-statistics for comparing two proportions. * indicates statistical significance at the 1% level.

Table 4. Location of cancelled limit orders (> 50 milliseconds) generated by STGP trading algorithm for the E-Mini S&P 500's front-month (February, 2014; May, 2014; August, 2014 and November, 2014).

Forecasting error	RMSE	MAE	MAPE
February 2014			
HFTs	0.0008	0.0007	14.21%
Institutional traders	0.0029	0.0025	44.17%
May 2014			
HFTs	0.0009	0.0006	19.88%
Institutional traders	0.0037	0.0031	38.24%
August 2014			
HFTs	0.0005	0.0003	10.06%
Institutional traders	0.0027	0.0022	47.99%
November 2014			
HFTs	0.0010	0.0008	20.11%
Institutional traders	0.0039	0.0020	49.57%

Table 5. Summary of HFTs and institutional traders forecasting statistical performance of the E-Mini S&P 500's front-month (February, 2014; May, 2014; August, 2014 and November, 2014).

Diebold-Mariano	$\mathbf{MDM_1}$	$\mathbf{MDM_2}$
February 2014		
Institutional traders	-9.14*	-10.07*
May 2014		
Institutional traders	-8.63*	-9.99*
August 2014		
Institutional traders	-4.22*	-5.78*
November 2014		
Institutional traders	-7.54*	-8.10*

$\mathbf{MDM_1}$ and $\mathbf{MDM_2}$ are the statistics estimated for the MSE and MAE loss functions. While MSE and MAE are used as loss functions, the modified Diebold-Mariano (MDM) test follows the student distribution with T-1 degrees of freedom. The table represents the application of the MDM test to the two forecasting models: HFTs vs. institutional traders. Negative values of the MDM test suggest that the first forecasting model (HFTs) is more accurate than the second model. The lower the negative value the more accurate are the HFTs' forecasts. ***** indicates rejection of the MDM null hypothesis of equal forecasting accuracy.

Table 6. Summary results of modified Diebold-Mariano statistics for the E-Mini S&P 500's front-month (February, 2014; May, 2014; August, 2014 and November, 2014).

Date	HFTs	Institutional traders
February 2014		
05/02/2014	$317	$26
11/02/2014	$303	$47
19/02/2014	$299	$33
21/02/2014	$300	$30
27/02/2014	$289	$29
May 2014		
01/05/2014	$287	$22
08/05/2014	$280	$20
09/05/2014	$301	$29
15/05/2014	$293	$34
19/05/2014	$299	$28
August 2014		
04/08/2014	$390	$30
12/08/2014	$397	$32
21/08/2014	$386	$23
26/08/2014	$381	$31
28/08/2014	$399	$35
November 2014		
06/11/2014	$251	$20
12/11/2014	$247	$19
20/11/2014	$238	$23
24/11/2014	$240	$18
25/11/2014	$233	$25

This table reports the daily profits by HFTs and institutional traders. All daily profits are estimated as the difference between the prices at which HFTs and institutional traders bought and sold shares of E-Mini S&P 500. We follow Baron et al. [4]:

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

where $n = 1, \ldots, N_{i,T}$ denotes the trades for trader i from the start of the trading day to the end of the trading day; p_n represent the price of the trade; $y_{i,n}$ measure 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. Transaction costs of $0.20 per contract are taken into account.

Table 7. Daily profits based on the most traded five days in each month by HFTs and institutional traders generated by STGP trading algorithm for the E-Mini S&P 500's front-month (February, 2014; May, 2014; August, 2014 and November, 2014).

Statistics	HFTs	Institutional traders
February 2014		
Mean	$301***	$33***
Median	$300	$30
Standard Deviation	$10.09	$8.21
Skewness	0.44	1.16
Kurtosis	2.46	2.82
May 2014		
Mean	$292***	$27***
Median	$293	$28
Standard Deviation	$8.66	$5.64
Skewness	-0.33	0.06
Kurtosis	1.68	1.66
August 2014		
Mean	$391***	$33***
Median	$390	$32
Standard Deviation	$7.50	$1.92
Skewness	-0.08	0.39
Kurtosis	1.54	1.99
November 2014		
Mean	$242***	$21***
Median	$240	$20
Standard Deviation	$7.19	$2.91
Skewness	0.13	0.41
Kurtosis	1.65	1.60

*** indicates that the mean profit value is statistically different from zero.

Table 8. Distribution of profits by HFTs and institutional traders generated by STGP trading algorithm for the E-Mini S&P 500 (based on most traded five days in each month).

Month	HFTs	Institutional traders
February 2014	1.52	0.61
May 2014	1.43	0.59
August 2014	1.99	0.75
November 2014	1.27	0.57

Table 9. Monthly Sharpe ratios of the E-Mini S&P 500 generated by STGP trading algorithm for HFTs and institutional traders (based on most traded five days in each month).

Time length (milliseconds)	HFTs profit for August 2014
0-50	$630
51-100	$409
101-200	$246
201-300	$128
301-500	$90
501-1,000	$79
1,001-1,500	$71
1,501-2,000	$67
2,001-2,500	$58
2,501-3,000	$55
3,001-3,500	$53
3,501-4,000	$42

This table examines trading profits over different time lengths for the most traded days in August, 2014 for E-Mini S&P 500, implementing the methods of Hasbrouck and Sofianos [6].

Table 10. Spectral analysis associated with HFTs' trading profits for the most traded five days in the most profitable month (August, 2014) for the E-Mini S&P 500.

Univariate regressions	
Variables	**HFTs' profits in August 2014**
$Profit_{i,t}$	0.19*
Control variables	
$Aggressiveness_{i,s}$	1.589*
$Volume_{i,t}$	0.903*
$Volatility_{s,t}$	0.409**
R^2adj	0.09

This table examines the consistency of HFTs' profits in August 2014 by investigating whether HFTs' profit yesterday is a good predictor for their profits today. We use the following OLS regression:

$$Profit_{i,t} = \alpha + \beta_1 Profit_{i,t-1} + \beta_2 Aggressiveness_{i,s} + \beta_3 Volume_{i,t} + \beta_4 Volatility_{s,t} + \varepsilon_{i,t}$$

where $Profit_{i,t}$ represent modified log profits such as $sign(profits) * \log(1 + |profits|)$ to incorporate any negative profits; $Volume_{i,t}$ is the log of each artificial trader's trading volume for day t; $Volatility_{s,t}$ denote the price volatility for day t defined as the volume-weighted standard deviation of the price process for the same day; $Aggressiveness_{i,s}$ represent the trader $i's$ volume – weighted aggressiveness ratio. * indicates significance at the 1% level; ** indicates significance at the 5% level.

Table 11. Consistency of HFTs profits for the E-Mini S&P 500 in the most profitable month (August 2014).

Panel A*	**Cancellation**	**Execution**
Time (milliseconds)	**HFTs**	
0-50	0.702	0.283
51-100	0.688	0.133
101-200	0.619	0.101
201-300	0.322	0.047
301-500	0.099	0.015
	Institutional investors	
0-50	0.009	0.221
51-100	0.038	0.255
101-200	0.066	0.317
201-300	0.080	0.509
301-500	0.089	0.822
Panel B**	**HFTs**	
0-50	0.994	0.011
51-100	0.873	0.010
101-200	0.698	0.007
201-300	0.295	0.004
301-500	0.085	0.001
	Institutional investors	
0-50	0.001	0.010
51-100	0.006	0.037
101-200	0.015	0.061
201-300	0.020	0.080
301-500	0.026	0.093

* Panel A shows the rate of cancellation and execution of limit orders by 10,000 HFTs (10% of the total population, genome depth of 10 and genome size of 2,048) and 90,000 institutional investors (90% of the total population, genome depth of 10 and genome size of 2,048). ** Panel B shows the rate of cancellation and execution of limit orders by 40,000 HFTs (40% of the total population, genome depth of 40 and genome size of 8,192) and 60,000 institutional investors (60% of the total population, genome depth of 40 and genome size of 8,192). Millisecond data has been generated and extracted from the STGP trading algorithm for the front-month of the E-Mini S&P 500 (February, 2014; May, 2014; August, 2015 and November, 2014). The probabilities are estimated as $1 - S(t)$, where $S(t)$ represents the survival function of cancellation and execution. In order to calculate the survival function we extracted all trading activity for E-Mini 500 S&P 500 generated by STGP trading algorithm and used the life-table method.

Table 12. Robustness checks related to the rate of cancellation and execution of limit orders by HFTs and institutional traders generated by STGP trading algorithm for the E-Mini S&P 500's front-month (February, 2014; May, 2014; August, 2014 and November, 2014).

Date	HFTs	Institutional traders
	February 2014	
05/02/2014	$203	$31
11/02/2014	$184	$55
19/02/2014	$155	$42
21/02/2014	$190	$44
27/02/2014	$172	$37
	May 2014	
01/05/2014	$124	$35
08/05/2014	$138	$26
09/05/2014	$166	$38
15/05/2014	$199	$42
19/05/2014	$201	$40
	August 2014	
04/08/2014	$257	$51
12/08/2014	$244	$64
21/08/2014	$218	$43
26/08/2014	$248	$50
28/08/2014	$261	$72
	November 2014	
06/11/2014	$111	$23
12/11/2014	$122	$27
20/11/2014	$103	$30
24/11/2014	$100	$22
25/11/2014	$119	$41

All daily profits are estimated as the difference between the prices at which HFTs and institutional traders bought and sold shares of E-Mini S&P 500. We follow Baron et al. [4]:

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

where $n = 1, \dots, N_{i,T}$ denotes the trades for trader i from the start of the trading day to the end of the trading day; p_n represent the price of the trade; $y_{i,n}$ measure 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. Transaction costs of $0.20 per contract are taken into account.

Table 13. Robustness checks for the E-Mini S&P 500 front-month's daily profits based on the most traded five days in each month by 10,000 HFTs (10% of the total population, genome depth of 10 and genome size of 2,048) and 90,000 institutional investors (90% of the total population, genome depth of 10 and genome size of 2,048).

Date	HFTs	Institutional traders
	February 2014	
05/02/2014	$421	$11
11/02/2014	$399	$34
19/02/2014	$338	$25
21/02/2014	$404	$19
27/02/2014	$316	$20
	May 2014	
01/05/2014	$330	$12
08/05/2014	$318	$10
09/05/2014	$396	$18
15/05/2014	$411	$26
19/05/2014	$377	$18
	August 2014	
04/08/2014	$499	$21
12/08/2014	$495	$15
21/08/2014	$481	$14
26/08/2014	$490	$29
28/08/2014	$503	$33
	November 2014	
06/11/2014	$314	$10
12/11/2014	$306	$12
20/11/2014	$299	$14
24/11/2014	$322	$11
25/11/2014	$315	$23

All daily profits are estimated as the difference between the prices at which HFTs and institutional traders bought and sold shares of E-Mini S&P 500. We follow Baron et al. [4]:

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

where $n = 1, \dots, N_{i,T}$ denotes the trades for trader i from the start of the trading day to the end of the trading day; p_n represent the price of the trade; $y_{i,n}$ measure 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. Transaction costs of $0.20 per contract are taken into account.

Table 14. Robustness checks for the E-Mini S&P 500 front-month's daily profits based on the most traded five days in each month by 40,000 HFTs (40% of the total population, genome depth of 40 and genome size of 8,192) and 60,000 institutional investors (60% of the total population, genome depth of 40 and genome size of 8,192).

Appendix A

Genetic Programming and Strongly Typed Genetic Programming

Although, Schlereth et al. [17] designed agent-based models that provide a promising link to individual behaviour, most existing techniques to agent-based system design fail to deal with the complexity of design [9]. To address this issue, we implement the Strongly Typed Genetic Programming (STGP). The STGP is a more sophisticated version of Genetic Programming (GP) whose application of generic functions and data types makes it more sophisticated than GP. GP can be considered an extension of Genetic Algorithms (GAs). GAs are techniques based on evolution and natural selection [16]. Under GAs' approach the economy is seen as evolving complex system in which artificial traders perform the activities of the real-life economy [19]. The procedure enhances search output by performing different solutions with genetic operators [11]. A benefit of GP over traditional GA is that players evolve not just the values of variables but also the structure of their models [5]. GP represents a machine-learning method to automate the development of computer programs in terms of natural evolution [2], which works by defining a goal in the form of quality criterion [1]. If there are inputs X and outputs Y, a program p is generated which satisfies $Y = p(X)$. GP uses variable-length tree structures for representing candidate solutions [20]. Opposite to neural networks, decision-tree structures represent specific rules that can be expressed in English [10]. The leaf nodes of the tree are the terminals whereas the non-leaf nodes are known as non-terminals. Terminals are usually inputs to the program with no argument and the non-terminals are functions often represented with at least one argument. The parse trees represent the trading rules of 20,000 HFTs and 80,000 institutional traders in our experiment. The typical genetic structure of the trading rule consists of hundreds of nodes and is rather unwieldy to actually write out, however, it can be simplified to equivalent algorithmic trading rules, as shown below.

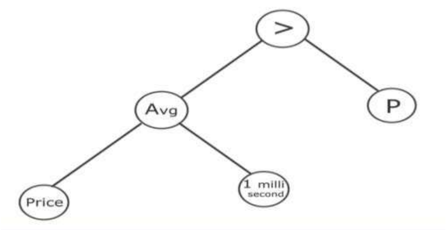

Figure 1. Example of time – dependent trading rule for institutional traders.

Figure 1 illustrates that the trading rule for institutional traders sends a buy signal if the average futures price over the past 1 millisecond is greater than the current price. A sell signal is sent otherwise. Institutional traders do not momentarily perform best in terms of the continuous Breeding Fitness Return, and therefore they possess higher latency in trading operations. Therefore, they are unable to foresee the quotes of the E-Mini S&P 500 and submit trading orders before HFTs.

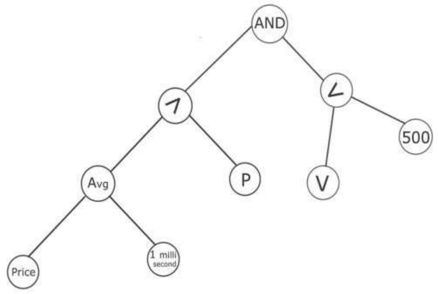

Figure 2. Example of time-dependent trading rule for HFTs.

Figure 2 indicates that the trading rule of HFTs sends a buy signal if the average futures price over the past 1 millisecond is greater than the current price and the current volume is less than 500. A sell signal is sent otherwise. The current volume function protects HFTs from sweep risk exposure. Large losses caused by sweeps (adverse price movements against HFTs' transient positions) can substantially reduce or even eliminate their profitability, so the management of sweep risk is of paramount importance for HFTs. HFTs 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.

The main steps in developing a genetic program are as follows:

1. Create initial randomly generated population of trading rules (trees). The random generation of trees enables the whole range of possible trading rules to be studied. The only requirement for trading rules is that they be well defined and produce output appropriate to the problem of interest. These trading rules apply the fundamental principles of biological evolution to create a new and improved population of trading rules. The creation and development of this new population is based on a domain-independent system governed by the Darwinian theory of natural selection under the principle of *survival of the fittest.*
2. Calculate the fitness of each trading rule in the initial population with accordance to appropriate criterion.
3. Create a new population by implementing the following operations:
   (i) Copy existing traders into the new population (crossover).
   (ii) Randomly select a pair of existing trading rules and recombine subtrees from them to produce a new trading rule (mutation). While crossover mixes subtrees of the population, mutation replaces subtrees with new subtrees. The operations of crossover and mutation are performed with the probability of selection for the operations, and skewed towards selecting traders with higher levels of fitness.
4. Calculate the fitness of each trader in the new population.
5. Repeat these operations, recording the overall fittest traders.

In the crossover process, randomly selected subtrees are swapped (exchanged). More specifically, a crossover point in the tree is randomly selected within each parent. Trading rules are selected on the basis of their fitness, with the crossover allocating future trials to areas of the search space whose trading rules contain parts from the superior trading rules. The best performing trading rules from the initial selection are selected based on the Breeding Fitness Return to act as parents in the crossover process. The Breeding Fitness Return process represents a trailing return of a wealth moving average and is an integral part of the latency of HFTs. This is in fact the return over the last n quotes of an exponential moving average of a trader's wealth, where n could potentially have the maximum breeding value of 250. Each pair of parents generates two offspring trading rules, so the number of parents and the number of offspring are equal at all times. In this innovative programming process the newly created trading rules replace those that performed poorly in the initial selection based on the replacement Fitness Return. This type of return represents the average return of a wealth moving average per millisecond quote since the creation of the very first trading rule. In other words, this is the cumulative return of an exponential moving average of a trader's wealth, divided by the trader's breeding value.

In the process of mutation a pair of trading rules has been randomly selected from the population, with probability weighted in favor of higher fitness trading rules. The subtrees of the two parent rules are then randomly selected. One of the selected subtrees is subsequently discarded and replaced by another subtree to generate the offspring rule. The GP searches areas of the solution space by evolving a population of trading rules, with the trading rules in each successive generation tending to become adept at solving the problem. As full technical explanation of crossover and mutation is beyond the scope of this paper, the reader may refer to Koza [12] for more details.

Strongly Typed Genetic Programming (STGP) is a version of GP whose application of generic functions and data types makes it more sophisticated than GP (in STGP each node is connected to a particular return). STGP is specifically suited to optimize structural or functional form. To create a parse tree, one needs to take into account important additional programming criteria such as when the root node of the tree returns a value of the type required by the problem, and when each non-root node returns a value of the type required by the parent node as an argument [15]. While GP can be written in any programming language, the STGP is typically written in a specific programming language, which is a combination of Ada [3] and Lisp [18]. The concept of generics as a method of developing strongly typed data is the critical component adopted from Ada. Additionally, Lisp incorporates the concept of having programs represented by actual parse trees [14]. While in conventional GP, one needs to specify all the programs and variables that can be used as nodes in a parse tree and deal with the search space of the order of 10^{30} - 10^{40}. STGP however reduces the searching state-space size to a greater degree [13]. On the other hand, the STGP search space composes the set of all legal parse trees, which means that all functions have the correct number of parameters of the correct type. On most occasions, the

STGP parse tree is limited to a certain maximum depth (Table I illustrates that 20 is the maximum depth in the artificial futures markets featured in this study). We set the maximum depth to 20 in order to keep the search space finite and manageable, while not allowing the trees to grow to an extremely large size. The critical concepts in STGP are generic functions (a mechanism for specifying a class of functions), and the process of assigning generic data types for these functions [8]. STGP has the flexibility to allow all variables, constraints, arguments and returned values of any type. The only strict requirement is that the type of data for each element has to be specified in the early stage of the programming process. The resulting initialization process and the various genetic operators associated with it are enabled to create syntactically correct trees. Those trees on the other hand are beneficial to the entire programming process because the search space can be significantly reduced [7].

Appendix B

E-Mini S&P 500 daily trading volume generated by STGP trading algorithm.

February 2014	
Date	**E-Mini S&P 500 trading volume**
03/02/2014	880,273
04/02/2014	889,011
05/02/2014	984,471
06/02/2014	888,250
07/02/2014	870,369
10/02/2014	880,376
11/02/2014	999,035
12/02/2014	799,937
13/02/2014	820,111
14/02/2014	890,255
15/02/2014	900,004
16/02/2014	856,551
19/02/2014	1,003,578
20/02/2014	868,480
21/02/2014	985,844
22/02/2014	883,999
23/02/2014	858,045
27/02/2014	979,932
28/02/2014	901,808
29/02/2014	845,431
30/02/2014	799,897
May 2014	
01/05/2014	1,207,009
02/05/2014	1,090,371
05/06/2014	1,103,484
06/06/2014	1,089,677
07/05/2014	1,110,080
08/05/2014	1,382,644
09/05/2014	1,299,388
12/05/2014	1,080,304
13/05/2014	1,005,989
14/05/2014	1,199,213
15/05/2014	1,463,011
16/05/2014	1,085,888
19/05/2014	1,344,656
20/05/2014	1,090,756
21/05/2014	1,117,353
22/05/2014	1,125,089
23/05/2014	1,006,355
27/05/2014	1,110,748
28/05/2014	1,190,377
29/05/2014	1,004,023
30/05/2014	1,102,378
August 2014	
01/08/2014	1,303,087
04/08/2014	1,674,926
05/08/2014	1,299,737
06/08/2014	1,384,211
07/08/2014	1,406,309
08/08/2014	1,337,743
11/08/2014	1,487,008
12/08/2014	1,700,380
13/08/2014	1,307,100
14/08/2014	1,299,878
15/08/2014	1,311,529

18/08/2014	1,388,141
19/08/2014	1,400,272
20/08/2014	1,294,988
21/08/2014	1,505,020
22/08/2014	1,352,090
25/08/2014	1,381,309
26/08/2014	1,609,999
27/08/2014	1,402,613
28/08/2014	1,524,530
29/08/2014	1,289,830
November 2014	
03/11/2014	775,338
04/11/2014	718,474
05/11/2014	800,099
06/11/2014	861,033
07/11/2014	745,940
10/11/2014	788,975
12/11/2014	880,737
13/11/2014	799,954
14/11/2014	767,834
17/11/2014	798,121
18/11/2014	756,110
19/11/2014	766,989
20/11/2014	901,006
21/11/2014	804,828
24/11/2014	890,342
25/11/2014	885,050
26/11/2014	812,736
28/11/2014	777,902

References

1. Archetti, F., Giordani, I., and Vanneschi, L. Genetic programming for anticancer therapeutic response prediction using the NCI-60 dataset. *Computers and Operations Research*, 37, 8 (2010), 1395-1405.

2. Banzhaf, W., Nordin, P., Keller, R. E., and Francone, F. D. *Genetic Programming: An Introduction on the Automatic Evolution of Computer Programs and Its Applications*. San Francisco, CA, USA: Morgan Kaufmann Publishers, 1998.

3. Barnes, J. G. P. *Programming in ADA*. Reading, MA: Addison-Wesley, 1982.

4. Baron, M., Brogaard, J. and Kirilenko, A. The trading profits of high frequency traders. *SSRN Electronic Journal*, (2012).

5. Dworman, G., Kimbrough, S. O. and Laing, J. D. On automated discovery of models using Genetic Programming: Bargaining in a three-agent coalitions game. *Journal of Management Information Systems*, 12, 3 (1995), 97-125.

6. Hasbrouck, J. and Sofianos, G. The trades of market makers: An empirical analysis of NYSE specialists. *Journal of Finance*, 48, 5 (1993), 1565-1593.

7. Haynes, T., Schoenefeld, D. and Wainwright, R. Type Inheritance in Strongly Typed Genetic Programming, In Kinnear, K. and Angeline, P. (eds.) *Advances in Genetic Programming 2*. MIT Press, 1996.

8. Haynes, T., Wainwright, R., Sen, S. and Schoenefeld, D. Strongly Typed Genetic Programming in evolving cooperation strategies. *Proceedings of the sixth international conference on Genetic Algorithms*, 1995, 271-278.

9. Karageorgos, A., Thompson, S. and Mehandjiev, N. Agent-based system design for B2B electronic commerce. *International Journal of Electronic Commerce*, 7, 1 (2002), 59-90.

10. Kim, J. W., Lee, B. H., Shaw, M. J., Chang, H.-L. and Nelson, M. Application of decision-tree induction techniques to personalized advertisements on internet storefronts. *International Journal of Electronic Commerce*, 5, 3 (2001), 45-62.

11. Kim, K.-J. and Ahn, H. Collaborative filtering with a user-item matrix reduction technique. *International Journal of Electronic Commerce*, 16, 1 (2011), 107-128.

12. Koza, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, 1992.

13. Montana, D. J. Strongly Typed Genetic Programming, Technical report 7866. Bolt Beranek and Newman, Inc, 1994.

14. Montana, D. J. Strongly Typed Genetic Programming. *Evolutionary Computation*, 3, 2 (1995), 199-230.

15. Montana, D. J. *Strongly Typed Genetic Programming*. 2002, Available at: http://personal.d.bbn.com/~dmontana/papers/stgp.pdf (Accessed on: May 01 2015).

16. Oliver, J. R. Artificial agents learn policies for multi-issue negotiation. *International Journal of Electronic Commerce*, 1, 4 (1997), 49-88.

17. Schlereth, C., Barrot, C., Skiera, B. and Takac, C. Optimal product-sampling strategies in social networks: How many and whom to target?. *International Journal of Electronic Commerce*, 18, 1 (2013), 45-72.

18. Steele, G. L. *Common Lisp: The Language*. Burlington, MA: Digital Press, 1984.

19. Wu, D. J. Artificial agents for discovering business strategies for network industries. *International Journal of Electronic Commerce*, 5, 1 (2000), 9-36.

20. Wyns, B. and Boullart, L. Efficient tree traversal to reduce code growth in tree-based genetic programming. *Journal of Heuristics*, 15, 1 (2009), 77-104.