An investigation of the behaviour of financial markets using agent-based computational models.

by

Viktor Manahov

A Thesis submitted for the degree of Doctor of Philosophy at

Newcastle University Business School

May, 2014

Signature of Author……

Newcastle University Business School

Supervisor: Professor Robert Hudson, Head of Economics Department, Newcastle University Business School.

Co-supervisor: Dr. Bartosz Gebka, Director of MSc Finance programmes, Newcastle University Business School.

Abstract

This thesis aims to investigate the behaviour of financial markets by using agent-based computational models. By using a special adaptive form of the Strongly Typed Genetic Programming (STGP)- based learning algorithm and real historical data of stocks, indices and currency pairs I analysed various stylized facts of financial returns, market efficiency and stock market forecasts.

This thesis also sought to discuss the following: 1) The appearance of herding in financial markets and the behavioural foundations of stylised facts of financial returns; 2) The implications of trader cognitive abilities for stock market properties; 3) The relationship between market efficiency and market adaptability; 4) The development of profitable stock market forecasts and the price-volume relationship; 5) High frequency trading, technical analysis and market efficiency.

The main findings and contributions suggest that: 1) The magnitude of herding behaviour does not contribute to the mispricing of assets in the long run; 2) Individual rationality and market structure are equally important in market performance; 3) Stock market dynamics are better explained by the evolutionary process associated with the Adaptive Market Hypothesis; 4) The STGP technique significantly outperforms traditional forecasting methods such as Box-Jenkins and Holt-Winters; 5) The dynamic relationship between price and volume revealed inconclusive forecasting picture; 6) There is no definite answers as to whether high frequency trading is harmful or beneficial to market efficiency.

Acknowledgments

I am entirely grateful to Professor Hudson for his invaluable help troughout my studies at Newcastle University Business School. This work could not be possible without his help. I am also grateful to my co-supervisor Dr. Bartosz Gebka whose accurate comments and suggestions made this research richer. I hope that we can find time and space to continue working together on future research topics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

1.1.1 The Rationale for Agent-Based Modelling.

The types of phenomena that are interesting in finance and still difficult to explain analytically involve the complex nature of heterogeneous boundedly rational agents and their interactions within the constraints imposed by financial institutions and authorities (Tay, 2013). Traditional finance theories describe what agents' actions, expectations and strategies are in equilibrium with the outcome these create behaviours aggregate (Arthur, 2005).

The equilibrium story begun more than a century ago when Leon Walras introduced a model of equilibrium pricing in his pioneering 1874 work *Elements of Pure Economics*. The desire to develop powerful Walrasian general equilibrium models has been greater ever since. Economists are not only investigating the existence of equilibria within these models, but the uniqueness and stability of models. The Arrow-Debreu (1959, 1964) framework represents the modern embodiment of Walrasian models. However, this framework is still far too complex to be solved for millions of heterogeneous consumers and companies (Ehrentreich,2008). Moreover, static equilibrium approach cannot easily distinguish between multiple equilibria and cannot answer how individual investors' behaviour, strategies, and expectations might react to or change with the patterns they create, which is the main topic of my research.

It is well known fact that macroeconomic models are often described in terms of three different vectors: y_t represent the endogenous variables, x_t is the vector of exogeneous forcing variables, and $\in_t$ the set of random shocks (Ehrentreich,2008). All fixed parameters are subsumed in a vector θ:

$$y_{t+1} = F\left(y_t, x_t, \theta, \in_t\right) \quad (1)$$

Lucas (1976) made criticism that it is likely that some of the parameters in vector θ may change because of a shift of policy regime λ . Lucas (1976) argued that aggregate quantities and prices could react in a different way than forecasted because agents might change their behaviour and this process cannot be captured in the equations. For instance, rational individuals could change their expectations even before a planned policy shift is introduced. The author suggested that an attempt to exploit a potential trade-off between unemployment and the inflation rate via an expansionary monetary policy may thus be foiled transforming equation (1.0) into:

$$y_{t+1} = F\left(y_t, x_t, \theta(\lambda), \mu, \in_t\right) \quad (2)$$

Where $\theta(\lambda)$ consist of regime dependent parameters and μ consist of truly invariable taste and technology parameters (Ehrentreich,2008). Lucas (1976) was unable to offer a solution to this fundamental problem imposing a standard that no conventional macroeconomic model can probably ever fulfil (Ehrentreich,2008).

More recently another substantial hurdle occurred in the economist's pathway. The *fallacy of composition* phenomenon states that what is true for individual agents may not hold for the aggregate economy (Caballero, 1992; Hartley, 1997). This phenomenon emphasises the tension between microeconomics and macroeconomics. We all know that the economy is formed by a large number of consumers and companies whose interactions could cause emergent behaviour at the macroeconomic level. However, appropriate policy recommendations for individual economic entities may be inefficient for the aggregate economy or vice versa.

Jacobs *et al.* (2010) consider how conventional economic models deal with so-called liquidity black holes. On Monday, 19 October 1987, liquidity disappeared from the U.S. stock market as numerous investors all attempted to sell at the same time. Similar liquidity black holes occurred when the hedge fund Long-Term Capital Management collapsed in 1998 and, more recently, during the 2008-09 economic downturn (Jacobs 1999, 2004, 2009). In these liquidity black hole cases, the price process wasn't fixed. However, traditional economic models assume the contrary intuition providing a solid platform for misleading results.

Some scholars argue that financial market models should be based on market efficiency and rational market participants (Muth, 1961; Fama, 1970; Fama, 1991). Some others suggest that investors are not-so-rational beings who are swayed by fads, fashions, and other cognitive biases (Bikhchandani *et al.,* 1992). Financial markets are populated by a mixture of these investors and others with completely different investment patterns. As a result of this diversity behavioural finance emerged as a research programme, which allowed for explanations of financial phenomena based on irrational investor behaviour and provided two main pedagogical goals such as what kind of mistakes to avoid while investing and what strategies in financial markets are more appropriate for earning abnormal returns (Subrahmanyam, 2007). Conventional economic models assume random processes modeled as a Brownian motion or as a function of a Brownian motion that does not allow the investigation of whether a given combinations of investor behaviours lead to particular price patterns (Jacobs *et al.,* 2010).

Many macroeconomic models claims that the choices of all different agents in one sector can be as the choices of a single 'representative' standard utility maximizing agent whose rationality significantly coincides with the aggregate rationality of all heterogeneous agents. The notion of the representative agent was born in Marshall's *Principles of Economics.* Kirman (1992) argued that this assumption is wrong because such models are inappropriate in studying issues of unemployment and places where no trade at all takes place (Varian, 1987). Kirman (1992) pointed out that there is no direct relation between individual and collective rationality and there is no formal justification for the assumption that the aggregate of individuals act like an individual maximizer. The author went further suggesting that even we assume that the rationality of the aggregate can be matched to those of a maximizing individual, the reaction of the representative agent to some aspects of the initial model may not be the same as the aggregate reaction at individual level (for instance, a change in government policy). Also, the personal preferences of the representative agent cannot be used to decide whether a particular economic model is more appropriate than another.

Moreover, the aggregate behaviour is characterised by complex dynamics and developing a representative agent whose behaviour has these dynamics could potentially lead to constrained results. Kirman (1992) concluded that the choices and rationality of the representative agent may be substantially different to those of the entire society, and therefore the representative agent should have no future. In a very recent experiment, Tay (2013) considered the role of the representative agent in decision-making processes when the decisions have to be made by all participating heterogeneous decision makers, each with their own unique preferences and private information that are not directly observable by the other decision makers, and concluded that this serious issue cannot be solved by conventional analytical modelling techniques.

Furthermore, conventional models of financial markets do not indicate the dynamic process that will need to happen in order to achieve the equilibrium condition. This is especially valid for models that produce multiple eqilibria, and therefore it is unclear which equilibrium among the multiple eqilibria representative agents would converge on (Tay, 2013).

Shleifer and Summers (1990) also put in doubt the future of the representative agents and their role in rational arbitrages. Rational arbitrageurs play a central role in standard finance by assuring that security prices follow their fundamental values and there are no riskless arbitrage opportunities even if there are many irrational noise traders. The authors demonstrated, contrary to the EMH, that arbitrage does not fully counter the responses of prices to fluctuations in uninformed demand and introduced the notion of limited arbitrage where movements in investor sentiment are an important indicator of asset prices.

Significant evidence further undermined the role of the representative agent in financial markets. Individual agents in financial markets tend to be overconfident, which makes them take on more risk (Alpert and Raiffa, 1982). Andreassen and Kraus (1998) claims that agents tend to extrapolate past time series resulting in chasing specific trends and patterns. Tversky and Kahneman (1982) found evidence that agents tend to underestimate base rates and emphasise acquiring new information leading to overreaction to news.

Financial analysts were surprised by the financial crisis of 2007-2009. Geanakoplos *et al.* (2012) asked whether the bubble was generated by low interest rates and lending criteria, irrational exuberance, too much refinancing or too much leverage? Although leverage was the main variable that went up and down along with housing prices, how can researchers disregard the other explanations, or determinate which is more important? Conventional models attempt to answer these questions by developing equilibrium models with a representative agent. However, the household sector comprises hundreds of millions of individuals, with great heterogeneity, and therefore traditional models are not capable of accurately calibrating this heterogeneity and the role that it played in the bubble.

The above economic and finance issues posed a number of important questions such as why fully rational and perfectly informed agents have essentially no ability to exercise free choice and are incapable of generating emergent economic phenomena? (Lawson,1997; Ehrentreich,2008). Conventional economic approaches use equational representation to enable a global markets characterisation, but why have they failed to explain the link between individual investor behaviour and the global market dynamics and trends that emerged? (Mathieu and Secq, 2012).

Why do traditional economic and finance models find it very difficult to explain the empirical features of real-world financial markets such as fat-tails and their associated asset return distributions, volatility clustering, and the existence of cross correlations between asset returns, volume and volatility.

Why traditional economic and finance models still cannot fulfil the Lucas (1976) critique, cannot solve the Arrow-Debreu (1954) framework, cannot justify the role of the representative agents, cannot satisfy the *fallacy of composition* phenomenon and failed to foresee the recent economic crisis? Are there any better economic modelling techniques available?

For the purpose of developing meaningful assumptions about anticipated market activities, researchers need models capable of representing the strategies of different market participants and changes in the composition of those participants. Hartley (1997) pointed out the existence of such models.

However, despite various anomalies, financial markets are important and increasingly powerful institutions. From the social point of view financial markets are still rational in the way our society is organised because humans preferences, though bounded in their rationality still represent some measure of social value. The intrinsic indifference of financial markets to non-financial matters such as social and environmental factors becomes a very important link between the financial reality and the 'lifeworld' (Thielemann, 2000; Mackintosh *et al.,* 2000). Rubinstein (2001) suggested that even if markets are not rational, abnormal profit opportunities still may not exist, indicating that financial markets may be said to be 'minimally rational'.

1.1.2 Methodological Aspects of Agent-Based simulations.

Rapid advances in computing technologies have changed society nowadays. The increased availability of powerful computers in the last ten years has been used as a platform for developing new models and forecasting techniques. Scientific advances both in software and hardware stimulated financial field research that was impossible before.

Huge computational capacity at the beginning of 21st century enabled the development of an entirely new generation of intelligent computing models such as agent-based computation, neural networks, fuzzy logic, genetic algorithms and genetic programming. The main aim of these intelligent computing techniques is to perform actions that simulate human decision-making processes. They are characterised by learning from past experience and high levels of flexibility allowing them to adapt to new circumstances. Moreover, the new computational techniques can find better solutions to issues unsolvable by traditional mean. The development of Agent-Based Computational Economics (Tesfatsion, 2002) and Computational Finance (Tsang and Martinez-Jaramillo, 2004) provide some tools to tackle the limitations of traditional economics and finance. The agent-based approach is more akin to reality than other modelling approaches and provides an opportunity to represent and test theories which cannot easily be described using mathematical formulae (Axelrod, 1997). Agent-based models (ABM) reveal much richer behavioural structure that is embedded in a financial entity which may otherwise be overlooked by conventional equation-based models (Tay, 2013).

For example, Parunak *et al.* (1998) compared ABM and equation-based modelling of a supply network and reported that equation-based models failed to capture quite a number of rich effects, such as memory effect of backlogged orders, amplification of order variation and transition effects, which were preset in ABM. Moreover, ABM does not require fully rational market participants capable of generating emerging phenomena. The ability to cope with heterogeneous and boundedly rational individuals makes it a perfect tool to study decentralized markets (Ehrentreich, 2008). Kirman (1992) argued that heterogeneity of agents is an important condition in solving the standard economic model and researchers should study aggregate activity from the perspective of the direct interaction between different individuals.

The new computational techniques can find better solutions to issues unsolvable by traditional means. For instance, ABM offer an easy escape from the Lucas (1976) critique and if a policy change is announced or even implemented, artificial agents recalculate their optimization problem, given their objective function and budget constraints. This also satisfies the desire for a microfoundation in classical terms because agents' behaviour is generated from a utility maximinization problem (Ehrentreich, 2008).

Also, artificial agents are capable of finding the competitive equilibrium allocation for the Arrow-Debreu framework (Sargent, 1979) and do not commit the *fallacy of composition* by ignoring valid aggregation concerns (Kirman, 1992). ABM are characterised by transparent dynamics that can be recorded allowing the researcher to reverse the time line of evolution to examine how a particular equilibrium has been reached (Tay, 2013). Geanakoplos *et al.* (2012) developed an agent-based computational model which mimicked the behaviour of more than two million potential homeowners in more than ten years. Remarkably, their model included details on each homeowner such as race, income, age, wealth, and marital status. The purpose was to investigate how these characteristics correlate with the home buying behaviour of individuals. Their model detected processes such as how rising prices and the possibility of refinancing encourage individuals to engage in speculative operations by purchasing more-expensive houses than they can afford to buy. These processes cannot be captured by conventional dynamic stochastic models.

Parunak *et al.* (1997) argue that ABM often map more naturally to the structure of the problem than equation-based models by specifying simple behavioural and transition rules attached to well defined entities, therefore providing a medium for the infusion of any economic theory or methodology into the model. Furthermore, the implementation of agent interactions can be governed by space, networks, or a combination of structures (Alam *et al.,* 2012). This would be far more complex task to achieve by mathematics or any other conventional economic models.

Agent-based computational modeling is an interesting bottom-up approach to studying economics and finance, which allows the disaggregation of systems into individual components that can potentially have their own characteristics and rule sets. Tesfatsion (2009) describe agent-based computational economics (ACE) as 'evolving systems of autonomous interacting agents'. As advocated by many during the financial crisis of 2007-2008 (e.g. Bouchard (2008), Farmer and Foley (2009), and Lux and Westerhoff (2009)), agent-based models (ABM) should play a large role in future financial modeling. Researchers use computers in order to study the evolution of investors' behavior under controlled experimental conditions. An important point to highlight is that ABM generate investment agent interactions which evolve further over time without any additional manipulations from the modeler.

The core purpose of ABM encompasses the process of simulating simultaneous interactions of multiple investment agents in an attempt to determinate complex behavior. Computational finance simulates the actions and interactions of multiple autonomous individuals and investigates their impact on the whole system.

ABM consists of a number of agents which interact both with each other and with their environment, and can make decisions and change their actions as a result of their interactions (Ferber, 1999). The behaviour of the whole system depends on the aggregated individual behaviour of each agent. Agents can interact either indirectly through a shared environment and/or directly with each other through markets, social networks and/or institutions (Matthews *et al.,* 2007).

ABM are able to capture complex emergent behaviour (unanticipated behaviour shown by a system) observed in the interactions of multiple artificial agents equipped with simplistic commanding rules. Agents can be very diverse ranging from ordinary consumers and investors to global policy-makers. Thus, the computerised environment can simulate anything from local shops to governmental structures.

Some well documented examples of ABM include: voting behaviours in elections (Kollman *et al.*1992); size-frequency distributions for traffic jams (Nigel and Rasmussen, 1994); identifying and exploring behaviours in battlefields (Ilachinski, 1997); company size and growth rate distributions (Axtell, 1999); price variations within stock-market trading (Bak *et al.* 1999); understanding theories of political identity and stability (Lustick, 2002); social networks of terrorist groups (North *et al.*2004); business coalitions over industry standards (Axelrod, 2006); multi-cellular level interaction and behaviour (Athale and Deisboeck, 2006); modeling economic processes as dynamic systems of interacting agents (Tesfatsion, 2006), to name but a few.

ABM generates simulation models of multiple investors with different investment targets. Computer programs record all interactions of artificial agents in order to investigate their behaviour over time. Agent-based computation enables the modelling of individual heterogeneity and agent positioning in financial, geographical or environmental types. ABM successfully possesses a wider range of nonlinear behaviour in comparison with traditional equilibrium models.

For instance, LeBaron (2005) introduced a series of different financial models which provide a plausible explanation for economic downturns and simulate liquidity stagnation that never occurs in conventional equilibrium models. One of the main advantages of the ABM is that they effectively overcome the restrictive assumptions required by analytical models for tractability. For example, all agents within a specific setting could be modelled as heterogeneous with respect to their preferences, forecasting rules or trading strategies. Table 1.0 illustrate the main characteristics of traditional and agent-based models.

Traditional models	Agent-based models
Deterministic (one future)	Stochastic (multiple futures)
Allocative (top-down)	Aggregative (bottom-up)
Equation based formulas	Adaptive agents
Do not give explanations	Explanatory power
Few parameters	Many parameters
Spatially coarse	Spatially explicit
Environment given	Environment created
You react to them	You learn from them

Table 1.0 Main characteristics of traditional and agent-based models (source: Barnard, 1999).

Some of the other advantages over traditional economic models include:

•There is no global mechanism that controls the interactions of agents in ABM. Instead agents compete with each other and coordinate their actions. Heterogeneity also allow for the specification of traders with varying degrees of rationality. This offers advantages over traditional economic approaches that assume perfectly rational individuals, if they consider individuals at all (Crooks and Heppenstall, 2012).

•Different behavioural patterns and strategies are continuously adapted as agents constantly learn. Learning enables agents to switch their strategies in a better way.

•In financial terms, the computerized economy works far away from global equilibrium with constant possibilities for improvement.

•ABM has the ability to capture emergent environmental dynamics. Agents exhibit complex behaviour, including learning and adaptation. Furthermore, the ability of ABM to describe the behaviour and interactions of a system allows for system dynamics to be directly implemented into the model. This represents a movement away from the static nature of traditional economic models (Crooks and Heppenstall, 2012).

There are two general issues in the development of ABM- the design of the agents and the design of the environment. An important issue concerns the development of decision making rules and their incorporation into the model. Even when the modeller uses artificial neural networks, genetic algorithms or genetic programming, he/she need to apply a predefined set of informational variables.

Although these rules can successfully mimic real-life decision making rules, it is not known if any of these rules indeed faithfully represent the inductive reasoning behaviour demonstrated by actual human decision-makers. Another related concern is that the modeller has to decide whether the agents should be enabled to learn only from their own experiences or from the collective experience of all agents in the model (social learning). Also, the design of the ABM environment is much more complicated as many details at the agent and institutional level need to be defined (Tay, 2013). On the other hand, models may have far too many parameters, and the impact of many of these parameters may not be well understood or tractable (LeBaron, 2004a).

1.1.3 Comparison between agent-based models and human subject experiments

Two computer-based technologies, the computational laboratory and the experimental, have begun to have implications on financial research. This section explores the relationship between artificial markets populated by computer-based agents and experimental markets with human participants. Both approaches exploit controlled laboratory conditions as means of isolating the resources of aggregate phenomena.

Most of the papers combining the two approaches have used primary the agent-based models to analyse the results obtained from laboratory experiments with human subjects. There are only a few exceptions where researchers have not sought to analyse findings from agent-based simulations with follow-up experiments involving human subjects (Duffy, 2006). This is based on the assumption that human subject experiments impose more constraints on what a researcher can do than agent-based modelling simulations. There are several important reasons for this apparent pattern. For instance, experiments performed with human participants generally need to be short in nature, both to prevent boredom among participants (one has to worry about human subject becoming bored) and to avoid the bankruptcy of the investigators who reward the participants with monetary payments. In sharp contrast, agent-based computational experiments can be run for many generations to diminish dependence on initial conditions (Tesfatsion, 2003).

Moreover, there are limits to the number of different agent characteristics that researchers can hope to induce in an experimental laboratory and time and budget constraints limit the number of simulations that can be performed. Furthermore, the number of agents used in agent-based modelling is quite different from that adopted in experiments with human subjects. For example, all experiments in this thesis involve 10,000 traders, a number which is significantly higher than the number of participants in experiments with human subjects. Agent-based artificial stock markets allowed me to run experiments for many more periods than are possible in human subject experiments (I have run 1,250 experiments in Chapter 6). Moreover, agent-based simulations enable indirect communication between traders (via their use of Strongly Typed Genetic Programming to update trading rules) – a feature that is not present in experiments with human subjects. In agent-based artificial stock market environment a number of different agent characteristics can be build, starting from traders with zero memory and random behaviour (Chapter 6) and going through markets populated by different groups of traders (Chapters 4, 5 and 6).

Gode and Sunder (1993) demonstrated that zero intelligence computational traders are able to achieve allocative efficiency that exceeds the efficiency achieved by human actors operating in the same experimental environment. Gode and Sunder (1993) and Chan *et al.* (1999) demonstrated that agent-based artificial markets allow greater control over the preferences and information-processing capabilities of agents than experimental markets with human subjects. According to the authors, human subjects vary in their learning abilities, habits and preferences (attitude towards risk for example), even after careful efforts to control most of the differences by experimenters. Smith (1982) argues that the external validity of human subject behaviour can be questioned because humans are often inexperienced with the task under examination.

Although human subject experiments cannot control subject behaviour, agent-based models have many degrees of freedom imposing computational difficulties for researchers. Agent-based researchers are often unable to provide external validity for the simple rules they assign to their artificial agents (Duffy, 2006). Also, sometimes evolutionary algorithms used in the development of agent-based models act as black boxes and the evolutionary algorithms may not be suitable for modelling decision-making processes (Duffy, 2006).

According to Duffy (2006) evolution is often a slow process and so algorithms that mimic this process tend to operate better under unchanging conditions. However, financial systems are often developed as state dependent, and may also experience temporary shocks. Under such circumstances, the performance of evolutionary algorithms may need modification relative to less volatile conditions for which they were initially created. Duffy (2006) argues that the processes by which artificial agents in agent-based models form expectations and adapt to constantly changing environmental conditions is not typically based on any specific micro evidence. The author went even further suggesting that the empirical comparisons that most researchers perform are between the simulated aggregate results and the empirical issue under investigation. LeBaron (2001) pointed out another dilemma in this area – where to set the bounds for boundedly rational traders. The trading strategies can be allowed to become far too complex in the evolutionary process leading to over specialised rules.

Unlike experiments with human subjects, in which the dynamics of the subjects' behaviour over many trading periods are almost never modelled exactly, agent-based artificial stock markets can easily accommodate complex learning behaviour, heterogeneous preferences, asymmetric information, and ad hoc heuristics (Poggio *et al.,* 2001). Financial markets are particularly appealing applications for agent-based models for many different reasons. The first reason is that the key debates in finance about market efficiency (Chapter 4) and rationality (Chapter 6) are still unresolved. The second reason is that financial markets contain many curious puzzles such as herding behaviour and various stylised facts (Chapter 5), the 'size effect' and volume-return relationship (Chapter 8) that are not well understood. The third reason is that financial markets provide a wealth of pricing and volume data that can be analysed (Chapter 9). The fourth reason is that when considering evolution (Chapter 4), financial markets provide a good approximation to a crude fitness measure through wealth or return performance (Chapters 7, 8 and 9). Finally there are strong links to relevant experimental results that in some occasions operate at the same time scales as actual financial markets (examples can be found in all chapters of my thesis.

1.2 Motivations for the thesis

The Efficient Market Hypothesis (EMH) postulates that asset prices should always be consistent with their fundamental values because security prices reflect all available information. Since its creation, the concept of efficient markets has been applied to many theoretical models and empirical studies of asset prices, generating several controversial debates.

Unfortunately conventional financial models are not capable of replicating the complexity of the stock markets. In Chapter 4, I used the STGP computational technique and several traditional econometric methods to examine the emergent properties of the stock markets. Agent-based modelling and artificial stock markets in particular, provide an appropriate environment for testing the Efficient Market Hypothesis (EMH) and the Adaptive Market Hypothesis (AMH). My experiments performed under artificial laboratory stock market settings generated rich varieties of market dynamics to investigate the formation of stock market dynamics and to measure the level of market efficiency.

Many scholars express concerns that herding behavior causes excess volatility, destabilises financial markets and increase the likelihood of systemic risk. Investigating herding behavior in financial markets is of particular interest because it might provide an explanation of excess volatility and bubbles.

The economic crises of the 1980's and 1990's suggested that herding behavior caused excess volatility and financial system fragility (Corcos *et al.,* 2002). However it is difficult to test all these theoretical assumptions directly. For instance, the research related to herding behavior focused mainly on statistical measures of clustering. The empirical difficulty in testing comes from the fact that there is no database of private information available and therefore it is hardly possible to prove whether market participants disregard their own information and engage in herding behavior.

This serious research obstacle can be eliminated in experimental settings such as agent-based artificial stock market where the information possessed by market participants can be strictly controlled.

My experiments in Chapter 5 represent extremely suitable environment to investigate the occurrence of 'spurious herding' (Bikchandani and Sharma, 2001) or 'investigative herding' (Froot *et al.,* 1992; Hirshleifer *et al.,* 1994) where a group of traders such as 'Best Agents' and 'All Agents' make similar decisions because they face similar information (historical data of Dow Jones, IBM and GE). I measure herding behaviour as the average tendency of a group of traders ('Best Agents' and 'All Agents') to buy (sell) Dow Jones, IBM and GE at the same time. This herding behaviour measure accesses the correlation in trading patterns for a particular group of traders and their tendency to buy and sell the same set of financial instruments.

Market structure and individual rationality remain at the centre of a debate as to which is the main driving force in market performance. The empirical research regarding whether individual rationality or market structure most influences market performance and market efficiency has been controversial. So far academics have been divided into two main camps. While one camp believes that market structure is the main driving force behind market performance (Gode and Sunder, 1993; Sunder, 2006a and 2006b; Sunder 2007) the other camp claims that individual rationality plays a significant role in the formation of market properties (Cliff and Bruten, 1997; Brewer *et al.,* 2002; Yeh, 2007 and 2008).

It is very difficult, however, to determinate the role of trader's individual rationality in a general stock market environment where thousands of individual investors with different risk attitudes, expectations, wealth preferences and even pleasure in trading operate. Investors can implement a fundamentalist strategy, a technical strategy or a combination of both to form their forecasting expectations. The nature of their trading strategies changes over time, and traders often adopt combined strategies coupled with gut feelings.

Obviously it is difficult to control the behavior of investors and investigate their decision making processes. Oberlechner (2011) argue that academics experience great difficulties in interpreting how groups of these strategies behave in a general market environment. The agent-based markets and zero intelligence markets, in particular, is an appropriate tool for examining the market mechanism in isolation from the traders who populate the market.

In stock market simulation experiments that I developed in Chapter 6, artificial traders receive information about the value of a real financial instrument and observe the history of past trades. Artificial traders in my experiments are able to make independent decisions creating a heterogeneous market structure which helped me to determine whether trader cognitive abilities or the market structure itself play a dominant role in market efficiency.

Stock market predictability remains a topic of continuous controversy. Predicting stock market returns is a difficult task which depends on economic and political factors as well as investors expectations. Some researchers suggest that stock market predictability is due to noise trading activities or speculative bubbles, which force stock prices to deviate from their intrinsic values. Researchers and investors constantly make efforts to forecast the future price movements and develop profitable trading rules.

The application of Genetic Programming (GP) to financial forecasts has been rather scarce with few papers found in the literature. The application of the Strongly Typed Genetic Programming (STGP) technique developed by Montana back in 1995 to stock market forecasting has not yet been done. The STGP technique is an enhanced version of genetic programming that enables the population to change gradually, which is an essential factor in maintaining a certain degree of model stability.

In Chapter 7, I implemented the STGP technique to develop forecasting models which present new evidence suggesting less adherence to random walks in financial time series. I used real-life historical data of the S&P 500, IBM and General Electric (GE) to compare four different in-sample and out-of-sample STGP models to traditional econometric forecasting models such as Box-Jenkins and the Holt-Winters exponential smoothing.

There are two main motivations behind my forecasting experiments. Firstly, I wanted to determinate whether, to what extent, and in which form the stock returns of the three financial instruments in excess of the risk free rate were indeed predictable and profitable. My second motivation was to provide rare experimental tests of the Optimism Principle developed by Picard and Cook (1984) and the Adaptive Expectations Hypothesis (Weigand *et al.,* 2004).

The ‘size effect’ is the anomalous pricing of the size factor as indicated by a significant risk premium in the conventional capital asset pricing model. There is substantial evidence in academia that size is an appropriate variable for explaining the cross-sectional variation of long-run asset returns (Horowitz *et al.,* 2000). Since the pioneering work of Banz (1981) and Reinganum (1981, 1983), the application of firm size (shares outstanding multiplied by stock price) to measure abnormal asset returns became widespread in the finance research.

In Chapter 8, I used the STGP technique to shed light on the ‘size effect’ equity puzzle and empirically demonstrate the impact of trading volume of the future direction of stock returns. I also performed experimental tests of the errors-in-expectation or extrapolation hypothesis developed by Lakonishok *et al.* (1994) and LaPorta *et al.*(1997).

I used real-life historical quotes of three indices- the Russell 1000, the Russell 2000 and the Russell 3000 to investigate the predictability of small-cap stocks and large-cap stocks and the dynamic causal relationship between trading volume and index returns.

Academic economists argue that with a turnover of 2,000 billion US dollars per day (BIS, 2005), it is highly improbable that the FX market is not at least weakly efficient and systematic profitable opportunities based on past price analysis should rapidly diminish by arbitrage (Curcio, *et al.,*1997). Hence, the puzzle still remains as to why technical analysis is widely used by investors and, consequently, whether the currency markets are informationally weakly efficient.

Traditionally, studies on foreign exchange (FX) technical trading profitability published so far failed to account for transaction costs, trading rule optimisation over time, out-of-sample verification, and data snooping issues (Park and Irwin, 2007).

Also, there are some common misunderstandings about how high frequency trading (HFT) affects the financial markets. Is HFT beneficial or harmful to market efficiency? What is the role of HFT on price discovery and efficiency processes? To what extend the policy makers should control HFT? The existing literature on the topic is scarce and does not provide clear evidence about the claims.

The literature mostly demonstrates the positive effects of HFT on market properties (Gomber *et al.,* 2011). Moreover, discussions on this topic often lack sufficient and precise information of how HFT trading algorithms work in practice.

A remarkable gap between the results of academic research on HFT and its perceived impact on markets in the public, media and regulatory discussions can be observed. To a certain extent the HFT research has been biased because many studies are administrated by investment banks or investors who use HFT strategies (Gomber *et al.,* 2011). The apparent lack of conclusive evidence enabled HFT to operate with limited regulatory understanding. Policymakers around the world are still debating whether to introduce limits on HFT or even completely ban it. Hence, I think than further research on the topic is highly desirable.

In my last experiment I applied the STGP technique to one-minute high frequency data of the six most traded currency pairs to examine whether the FX markets need time to process information and could be inefficient at very high frequencies but efficient at lower, e.g., daily horizons.
I also assessed the impact of HFT on market quality and integrity and discussed whether policymakers around the world should introduce stricter rules on HFT or even completely ban it. I am not aware of any study utilizing minute-by-minute price data in this context. Secondly, I don't ignore any of the issues identified in the literature as potentially affecting the reliability of trading results and inference based on them: I control for transaction costs, allow agents to learn from their experience and to switch to more profitable rules, evaluate the profitability of rules based on their predictive power rather than in-sample fit, and avoid data-snooping biases by allowing all potential rules and their combinations to be traded on and evaluated by agents. Thirdly, I am the first to apply genetic programming techniques to analyse the impact of HFT on market quality.

1.3 Main findings and contributions

The artificial traders in Chapter 4 of the thesis can be considered to be agents that adapt, learn, evolve and try to survive. The random nature of the trading rules of the agets allowed me to observe how they learn, adapt and survive. The presentce of 10,000 heterogeneous and interactive adaptive traders, rich in dynamics, provides the opportunity to study the stock market as complex adaptive system. Artificial traders are, by definition, capable of adapting, learning, and evolving, which makes them extremely suitable for the analysis of market efficiency and adaptability, because adaptation and learning in heterogeneous structures are known as important tools for analysisng financial market behaviour (Hommes, 2001).

My research contributions of Chapter 4 include the following:

• The introduction of increased heterogeneity and greater genetic diversity leads to higher market efficiency in terms of the AMH.

•The presence of different market sizes suggests that the market is more efficient when the population size increases.

• Stock market dynamics and nonlinearities are better explained by the evolutionary process associated with the Adaptive Market Hypothesis proposed by Lo (2004).

•Market efficiency seems to exist simultaneously with the need for adaptive flexibility.

• Individual trader learning, adaptation and evolution reinforced the motion of efficient markets.

Chapter 5 of the thesis provides evidence of the emergent properties of herding behaviour, stock market efficiency and stylised facts of financial returns such as leptokurtosis, non-IIDness and volatility clustering that characterise real world financial markets. The main contribution of Chapter 5 is the trade-off that I developed between reality (real historical data of the three financial instruments) and the explanatory power of the stylised facts analysed through the STGP technique. The other findings and contributions include:

• In line with previous research, I found some evidence of more herding in a group of stocks than in individual stocks.

• The magnitude of herding in my experiment is far from dramatic and does not exhibit the long-run mispricing of assets and bubble formation.

• My experimental tests demonstrate that an artificial stock market populated by a small fraction of best performing agents behaves differently from a market composed of a greater number of less well performing agents.

•The market based on the behaviour of the entire population of 10,000 artificial traders exhibit less herding and is more efficient in terms of the EMH than the segmented market populated by best performing agents (5 per cent of the total population).

Chapter 6 of my thesis investigates the implications of trader cognitive abilities on stock market properties. This chapter provides a new perspective to the research in this field by examining and explaining the reason for the divergence of the results in the literature.

•The main contribution of Chapter 6 is associated with a mixture of positive and negative impacts from individual intelligence and market performance.

This finding is achieved through:

•Using real-life historical data of S&P 500 and the Coca-Cola Company (most studies used artificial data).

•I implemented 10,000 artificial traders (most studies used up to 2,000).

•I used Strongly Typed Genetic Programming technique (all studies used Genetic Programming).

•I created greater genome depth of the most intelligent traders (compared to genome depth up to 15 in other studies).

Chapter 7 uses Strongly Typed Genetic Programming technique to find profitable stock market forecasts. The main findings and contributions of this chapter are:

•Small-cap stocks are more predictable than large-cap stocks.

• Out-of-sample tests demonstrated STGP outperformance resulting in statistically and economically significant excess returns after taking into account appropriate transaction costs.

• I performed rare tests of the Optimism Principle (Picard and Cook, 1984) and empirically demonstrated that some forecasters believe that their predictions are more accurate than they are in reality.

• I found support for the Adaptive Expectations Hypothesis (Weigand *et al.,*2004), where decision-makers rely predominantly on recent trends in forming their future forecasts.

Chapter 8 of this thesis presents a new evidence of small-cap stocks profitability and the price-volume relation. This chapter contributes to the existing literature by investigating whether small-cap stocks are more profitable than large-cap stocks and whether trading volume has any predictive power for stock returns. After implementing the STGP technique and analysing historical quotes of the Russell 1000, Russell 2000 and the Russell 3000, I found the following:

• Substantial evidence of small-cap stocks dominance which is not period-specific as some researchers argue.

•The investigation of the dynamic relationship between trading volume and index returns revealed an inconclusive picture that does not allow me to develop volume-based trading strategies.

• My findings are consistent with the errors-in-expectations hypothesis (Lakonishok *et al.,*1994 and LaPorta *et al.*,1997), which posits that excess returns of growth stocks are driven by more optimistic forecasts compared to those of value stocks.

Chapter 9 offers in depth analysis of high frequency trading profitability and market regulation. In my last experiment I applied the STGP technique to one-minute high frequency data of the six most traded currency pairs and I have found that:

• The search for answers to the puzzle described above should be conducted on high frequency rather than daily data, as the trading and resulting price adjustments take place on an intraday basis. The most academic studies related to technical trading in the FX market were not consistent with the real-life practice of technical analysis because they largely limited their trading strategies to daily data observations (Brabazon and O'Neil, 2004; Qi and Wu, 2006; Reitz and Taylor, 2006).

• My empirical results suggest that the STGP technique applied to six of the most traded currency pairs significantly outperform the traditional econometric parametric and non-parametric forecasting models leading to statistically and economically significant profits in the presence of transaction costs.

• I found evidence of positive impact of HFT on price efficiency and market dynamics.

•However, I think that the results are inconclusive and the need for regulatory intervention cannot be neglected.

1.4 Summary of the structure of the thesis.

In summary, the remainder of the thesis is organised in the following way: In Chapter 2, I review the notion of artificial stock market modelling. My intention is to provide a comprehensive literature review of the broad topic of my thesis. Chapters 4 to 9, which include the main analysis of my simulation results provides topic-specific literature review by examining the most important works that have been done in the past and review the most recent research in the area.

In Chapter 3, I describe in detail the computational model and the software platform I use in my experiments. Chapter 4 provides a direct test of the Efficient Market Hypothesis within artificial stock market settings. Chapter 5 analyse the occurrence of herding behaviour in financial markets and the behavioural foundations of the stylised facts of financial returns. Chapter 6 investigates the impact of trader cognitive abilities on stock market properties. Chapter 7 demonstrates whether the use of the Strongly Typed Genetic Programming (STGP) technique can lead to profitable trading strategies. Chapter 8 compares the predictive ability of the STGP technique and traditional econometric forecasting models to determine the profitability of small-cap stocks and large-cap stocks. Chapter 9 describes a novel approach to high frequency technical trading and its profitability. It measures the level of profitability and investigates whether high frequency trading has beneficial or harmful effect on market efficiency. Chapter 10 concludes my work and summarises the research findings. Finally, I point out some promising areas of future research.

Chapter 2

Literature Survey

2.1 Artificial Stock Market Modelling

2.1.1 Introduction

Stock markets are the main source for companies to obtain money by providing a marketplace to facilitate the exchange of assets. Market participants are individual and institutional investors, hedge funds and publicly traded corporations. Stock markets represent an important and interesting challenge for agent-based modellers for two main reasons. First, stock markets worldwide still pose many open questions that standard modelling approaches have not been able to solve so far. The second reason is that a huge amount of financial data is available for experimental agent-based modelling purposes.

The recent financial crisis highlighted the need for new research tools capable of dealing with the high level of complexity of the financial world. A natural method to analyse a complex system such as the stock market is to implement an agent-based modelling technique which entails simulating the stock market from the bottom up with a huge number of interacting heterogeneous boundedly rational traders that are designed to copy the behaviour of real-life traders (Tay, 2013). Financial markets offer an important field of agent-based modelling application, since agent objectives are clearly identified (Brandouy, 2013) and *Bayesian learning* (machine learning) can be used by agents to incorporate all available information into the decision making process (Mitchell, 1997).

One possibility to examine the properties of real markets is to build artificial markets, whose dynamics are represented by computer programmes designed to simulate different behaviours. Some of these computer programmes may attempt to replicate naïve behaviour, others may generate intelligence. Since the behaviour of all agents is under the researchers' control, the experiments have means to control different experimental factors and relate market behaviour to observed phenomena (Tsang and Martinez-Jaramillo, 2004).

Artificial stock markets are grounded on an individual-based approach with local interactions, heterogeneous traders, distributed knowledge and resources, agent autonomy- features that cannot be used or done in conventional, aggregate models (Mathieu and Brandouy, 2012). Artificial stock markets propose a powerful platform to test new regulations, new exchange structures or new investment strategies in completely controlled environment (Brandouy, 2013). The development of artificial stock markets has thus become a major application for the agent-based paradigm.

2.1.2 Early studies and small artificial stock market models

Most of the earliest simulations of financial markets intended to design an entire functioning financial market, which carefully analysed a small number of strategies used by agents to trade a risky asset. The first artificial stock markets were rather analytic than computational and are characterised by high level of tractability (LeBaron, 2004a).

In general, artificial stock market models are divided into two categories such as 'few-type models' which consist of small fixed sets of trading strategies and 'many-type models' in which artificial traders are allowed to choose from large and evolving number of trading strategies. 'Few-type models' encompass the early years of artificial stock market modelling (Ehrentreich, 2008). Although in the 'few-type models' artificial traders are allowed to choose from among fixed sets of trading strategies, they provide an important dimension of tractability and clear connections between model parameters. This feature cannot be seen in the more complex 'many-type models'.

The first artificial stock market structure was built by Stigler in 1964, who generated trading orders as random variables to investigate the effect of the regulations of the SEC on American stock markets using empirical data from the 1920s and 1950s. Garman (1976) developed two basic models, 'dealership' and 'auction' markets to examine the moment-to-moment trading activities in asset markets. The author presented some new explanations of observed empirical phenomena such as leptokurtosis, movement of asset prices around their inventory positions and the dependence of probability functions on previous transaction prices.

A few years later, Figlewski (1978) developed another 'few-type' artificial financial market model. His market model investigates the impact of distributing wealth among market participants with different levels of information. In order to establish price formation processes across the market, Figlewski assumed that all agents know the wealth levels of each other (unrealistic in real financial markets). While better informed agents outperformed the efficient benchmark by 14 per cent, simulations suggests that the overall market efficiency has been reduced by the introduction of agents with inferior information.

Although the disadvantage of Figlewski's experiment is the presence of very limited information on the dynamics of prices and trades, it is still an important early investigation on how wealth dynamics affect the convergence to an efficient market (this is quite similar to my experiment in Chapter 4).

Frankel and Froot (1986, 1986) were the first to simulate foreign exchange markets by combining a standard monetary model of open economy macroeconomics with chartist-fundamentalist approach to expectation formation in order to provide a possible explanation of the dollar bubble over the first half of the eighties.

DeGrauwe *et al.*(1993) claimed that Frankel and Froot's model leads to chaotic behaviour of exchange rates and proposed a model of artificial stock market which was the first to explain some stylised facts other than the mere deviation from the fundamental value recorded in the literature so far.

Kim and Markowitz (1989) developed their artificial stock market to investigate the reasons for the stock market crash in 1987 when the U.S. stock market decreased by more than 20 per cent. Since this dramatic decrease could not be explained by the emergence of new information, research orientated around factors other than information-based trading was necessary in determining stock price volatility. In their experiment the authors tried to analyse the relationship between the share of agents pursuing portfolio insurance strategies and the volatility of the market. The simulated market contained two different types of traders, 'rebalancers' and 'portfolio insurers', and two assets, stocks and cash. The wealth w of each trader at time t was quantified by:

$$w_t = q_t p_t + c_t \tag{3}$$

Where q_t is the number of assets the trader holds at time t, p_t is the price of the asset at time t and c_t denotes the cash traders hold at time t. 'Rebalancers' keep half of their wealth in stocks and the other half in cash, i.e.

$$\text{Target of rebalancers: } q_t p_t = c_t = 0.5_{wt} \tag{4}$$

Hence, the rebalancing strategy has a stabilizing effect on the market because increasing prices induce the rebalancers to increase their supply or reduce their demand; decreasing prices have the opposite effect. ‘Portfolio insurers’ adopted a strategy that guarantee minimum wealth at a specified insurance expiration date.

The actual experiment started with rebalancers’ portfolios in disequilibrium- ‘rebalancers’ initially have either too many or too few assets. The overall result of this particular approach was the demonstration of the destabilizing potential of portfolio insurance strategies. The authors concluded that trading volume, price volatility, and the size of price changes significantly increased when the fraction of portfolio insurance traders increased. Kim and Markowitz simulated a market composed by traders who pursue strategies found in real-life markets and provided a detailed description at the microscopic level. In contrast to this specific model, more recent simulations of stock markets deal with much more stylised facts and descriptions of traders’ behaviour. Although the trading strategies were well defined allowing assessment of their impact, a market populated by many traders using portfolio insurance strategies can be very unstable (LeBaron, 2006).

In a simulated financial market of Kirman and Teyssiere (2001) technical and fundamental traders followed a finite set of well-defined portfolio rules and shifted between a limited number of strategies according to an epidemiological process of contagion. Kirman and Teyssiere compared the long memory properties of the simulated market with actual volatility in various foreign exchange series, and reported a very good quantitative match. The major drawback of this artificial stock market is that traders have to follow a finite set of preliminary trading rules restricting their ability to evolve and learn.

Although more recently created, the artificial stock market of Farmer and Joshi (2002) is a small market with fewer trading strategies designed to replicate a benchmark set of real trading strategies. The authors used U.S. aggregate real dividends interpolated to daily frequencies to simulate traders that follow trend following and value strategies. This artificial stock market generated long swings away from fundamental values and uncorrelated daily returns. It also generated fat tails, volatility clustering and trading volume persistence. This artificial stock market provide clear tractability since it was based on realistic trading strategies.

The study of Westerhoff (2003c) investigated the impacts of price limits on trading, which is one of the most interesting questions for market policy makers. In general terms, price limits prevent the wide movement of asset prices and when the price reaches that limit, no trades occur above or below the specified limit. The trading process does not terminate, but orders are executed at or within the price limit. This artificial stock market model consisted of technical traders whose demand was dependent on recent price trends, and fundamental traders. The pool of fundamental traders increased when the distance between the asset price and fundamental values increased generating natural fundamental reverting dynamics. When the price limit increased from zero, the distance of the traded price from fundamental values fell, and when the price limit passed 0.5 percent the asset prices deviated from their intrinsic values. The author concluded that setting optimal price limits critically depends on asset price movements around their fundamental values. Although interesting, the empirical results generated by artificial traders in this experiment does depend critically on very specific behavioural assumptions made for the market maker.

The minority game generated important results for financial researchers (LeBaron, 2004a). The game involved two different doors, denoted 0 and 1, and an odd number of individuals choosing a door in each round of the game. After the choices are made, the door with the smaller number of individuals wins the race. The purpose of the game is to be contrarian in nature, and not to follow everyone else (Arthur, 1994; Challet and Zhang, 1997).

Savit *et al.* (1999) presented a relatively simple artificial market model to examine how agents develop their codings and dynamics and how they adapt their behaviour in an endogenously changing market environment. Their experiment compared the volatility to a specific benchmark where agents were allowed to choose their strategies in a completely random fashion. For this purpose the authors studied the impact of changing the length of the historical data set (m) on volatility and market dynamics. Small changes in the length of m resulted in larger observed volatility than the benchmark indicating that artificial agents adapt against each other in a manner which generates substantial volatility. Interestingly, large changes in m enabled random volatility, highlighting that strategies based on bigger historical data set could be associated with a random number generator in their strategy choices.

Although famous for their tractability and simple trading strategies the 'few-type models' are often biased towards any particular trading strategy adopted ex ante by the modeller. Apart from the ad hoc nature of their trading strategies and the lack of learning algorithms engaged in searching the space in order to find new trading opportunities, the 'few-type models' are limited in their abilities to seek out and take advantage of any inefficiencies in the search space. The 'many-type models' discussed in the next section aims to fill this gap.

2.1.3 Many-type artificial stock market models

The so called 'many-type models' include larger sets of trading strategies constructed by using a wide variety of computational techniques. These computerised models enabled researchers to examine the processes of emergence, survival and co-evolution of trading strategies over time. They also attempted to perform in depth analysis of market efficiency. If the artificial market occasionally moves to a state where inefficiency appears, then the evolutionary algorithm of the many-type models will search for new strategies.

Levy *et al.*(1994) offered one of the first of 'many-type models' approaches in a collaborative work at Hebrew University including both economists and physicists. Their artificial stock market consisted of an ensemble of interacting speculators whose behaviour was derived from a traditional utility maximization scheme. At the beginning of every period each trader i divided up his entire wealth $W(i)$ into stocks and bonds (cash wasn't included). With $X(i)$ denoting the amount of stocks in the portfolio of investor i, his wealth was quantified as follows:

$$W_{t+1} = \underbrace{X(i)W_t(i)}_{sumofshares} + \underbrace{(1 - X(i))W_t(i)}_{sumofbonds} \quad (5)$$

Where the superimposed boundaries were between 0.01< $X(i)$<0.99. Interestingly, the authors have found chaotic motion in stock prices and lack of empirical scaling laws in any of their experiments. This artificial stock market was criticised by Zschischang and Lux (2001) on the basis of sensitivity of the original results to the initial conditions of the experiment.

The authors imposed further critiques by stating that the empirical results may be sensitive to the number of traders in the experiment. This critique is interesting, but it was based on three different memory lengths only- 10, 141, and 256. It remains to be seen if it has other implications on longer memory lengths.

Arifovic (1996) produced an artificial stock market with richer and extensive structure based on the foreign exchange model of Kareken and Wallace (1981) to analyse the dynamics of the market under different parameters. The results showed that the consumption level in the first period converges to near its optimum value but the exchange rate constantly moves over time, and never settles to any constant value. Remarkably, in equilibrium the return of the two assets was the same, and therefore the learning of traders was indifferent between holding the two currencies. Hence, fractions of traders moved to holding one currency or another, shifting the exchange rate around as they change demand between currencies. This clearly demonstrated that equilibrium can be achieved if traders terminate their learning and exploration processes in the market. However, the experiment has a major drawback in that, if there exists one price series and exchange rate paring that corresponds to the equilibrium level, then there will be many such equilibria.

Mahfoud and Mani (1996) presented a new genetic-algorithm-based market for inductive machine learning. The model was applied to financial forecasting in individual stocks. The genetic algorithm model was benchmarked against a neural network model on 5000 stock-prediction experiments. Consistent with my forecasting evaluations in Chapters 7, 8, and 9, the authors concluded that both artificial models significantly outperformed the market, with the genetic algorithm model generating more accurate forecasts.

Lettau (1997) developed his artificial stock market to examine the portfolio decisions of boundedly rational agents in order to determine how investors in mutual funds move money into and out of four different groups of mutual funds. The model was rather a simple one in which agents used a genetic algorithm to update their portfolio decisions. The author demonstrated that mutual fund investors follow the same pattern as the adaptive agents in the model in terms of investing funds. Furthermore, Lettau (1997) claim that a model with entry and exit of artificial agents is able to match quite closely the mutual fund data sets.

Lettau's experimental results demonstrate that in different specifications the genetic algorithm can learn the optimal parameter for the portfolio policy, nevertheless, there are some important caveats. This is a very stylised and simplified artificial stock market. The author made no attempt to model the price formation process at all. Hence, this cannot beviewed as an attempt to simulate real-life financial markets, in which the dependence between today's price and trader's strategies is the most critical aspect for the agent-based modelling approach (Chapters 7, 8 and 9). However, this experiment represents a good learning tool because the setup is quite straightforward (LeBaron, 2006).

Youssefmir and Huberman (1997) investigated the issue of volatility clustering in financial markets by analysing a simple resource allocation model where traders were allowed to choose between different resources with the payoff of every single resource dependent on the actual number of traders using it. The authors observed that the volatility clustering in their market was reminiscent of real financial markets and in the competitive market dynamics they analysed, traders payoffs were decreasing in the number of traders using the resource. The other connection between this experiment and real financial markets is that traders were allowed to implement various trading strategies to optimize their behaviour.

Tsang *et al.* (1998) proposed a genetic programming based artificial market named EDDIE (Evolutionary Dynamic Data Investment Evaluator) to forecast horse races as well as financial instruments. Given a set of variables, EDDIE attempted to find interactions among variables and discover non-linear functions. When experimented on 180 handicap races based on real data in the UK, the model outperformed all common strategies used in horse race betting by great margins. EDDIE was then applied to historical dataset of the S&P 500 and achieved a decent annual rate of return over a three and a half year period.

Li (2001) expanded the EDDIE artificial stock market by implementing a special function called FGP-2, which provided the researcher with a handle for turning the precision against the missing forecasting opportunities. This valuable addition enabled Li to pick investment opportunities with greater confidence.

Arifovic and Gencay (2000) constructed an artificial stock market based on the previous work of Arifovic (1996) and presented significant evidence that the generated return series from the model are not only nonlinearly dependent, but are most likely chaotic in nature. Their estimation was conducted by applying the Lyapunov exponent (quantity that characterises the rate separation of infinitesimally close trajectories) altogether with various phase diagrams on the dynamics.

Routledge (2001) examined the effect of agents' learning by enabling them to buy a costly signal about a future dividend payout of an asset. The actual learning took place when traders were allowed to convert the noisy signal into a future dividend forecasts and traders who didn't want to buy the signal used their current price to infer the future dividend payout. The author observed that the original equilibrium was unstable and analysed the dynamics of this instability.The equilibrium instability was due to a change in the market proportions of informed versus uninformed traders and when the number of uninformed traders decreased, the ability of their fraction to learn decreased too due to small sample sizes. Grossman and Stiglitz (1980) demonstrate that there is an equilibrium in which a number of traders will purchase the signal of future dividend payout of a stock. Routledge (2001) argue that this can be achieved within the GA learning environment. There are, however, sets of parameters for which the original equilibrium proves to be unstable. For instance, the change in market segments of informed versus uninformed traders means that the parameters of the linear forecasts are wrong. In particular, the uninformed traders have to have the ability to interpret the price with fewer of their type in the experiment. At the same time, if the number of uniformed traders decreases, the ability to learn decline based on the small sample size leading to convergence to a situation in which all traders are informed (Routledge, 1999).

Johnson *et al.* (2001) presented an interesting application of the minority game to artificial stock market forecasting. Real financial time series were implemented in a binary string dependent on price movements and the continued model dynamics were used in out-of-sample forecasting evaluation. Artificial traders were able to generate some insignificant profits in real high frequency historical data. However, the main criticism of this model is that it does not have any natural role for prices in its basic version.

LeBaron (2006) argues that it would appear that the contrary principles of the minority game is somewhat forced, and real-life stock markets it may be better to follow the herding behaviour (explained in Chapter 5) for a short period of time. Moreover, it remains to be seen how robust and reliable the empirical results generated by artificial traders in this experiment are.

In another minority game framework, Hart *et al.* (2003), recommended possible governmental market policies aimed at mitigating bubbles and crashes and stabilising financial markets. The authors orientated their artificial market towards simulating a financial crash before it happened in reality and observed a relieve of the crash pressure and stabilised financial market dynamics. Their policy recommendation was associated with a kind of a release valve in market dynamics which lets off steam before the real big crash occurs.

All the artificial stock market models described in this section were among the first attempts at microscopic simulations of financial markets and their aims were more to provide mechanisms for bubbles and crashes than investigating statistical properties of the generated time series. However, all these artificial stock markets often suffer from the lack of precise definition of their far too many parameters. Also, the impact of many of these parameters is not well understood by researchers. Moreover, there are still relatively few general principles that researchers can apply to the development of the 'many-type models' of artificial stock markets (LeBaron, 2006). Nevertheless, this criticism did not stop all future experimentations. There are directions in which the artificial stock markets are moving giving these markets a more solid foundation discussed in the next section.

2.1.4 Emergence behaviour and many agent models

Artificial stock markets described in this section move farther from examining model dynamics, and more toward analysing the occurrence of trading strategies and examining their co-evolution over time. Simulated stock markets in this category aim at determining which trading strategies will appear and survive thorough self-reinforcement, and which will fail (LeBaron, 2004a).

The model of Rieck (1994) represents one of the earliest attempts to examine the emergence and evolution of trading strategies. He replicated actual trading strategies using evolutionary techniques and concluded that fundamental strategies are not able to take over the market and drive asset prices to their fundamental values. In artificial markets populated by fundamentalist based trading strategies the asset prices moved away from their intrinsic values, but eventually returned. These results are suggestive that artificial stock market models could be replicated with more microstructure orientated trading mechanisms.

The Santa Fe Artificial Stock Market (SF-ASM) developed by LeBaron *et al.* (1999) was one of the most influential artificial stock market models, which provided a solid base for other artificial stock market experiments. It was initially written in the C programming language and imported into the Swarm language later on, generating much interest in science. The main aim of the model was to understand the trading environmental behaviour in a market where trading strategies evolve and compete against each other. There were two assets that artificial agents trade- a risky asset with a random dividend, d_t and a risk-free bond which provide constant level of interest, r. The dividend was associated with the following autoregressive process:

$$d_t = \overline{d} + p\left(d_{t-1} - \overline{d}\right) + \varepsilon_t \qquad (6)$$

where ε_t is Gaussian and independent, and p equals 0.95.

Under constant absolute risk aversion (CARA) utility and Gaussian distributions for dividends and prices, the actual demand for holding shares of the risky stock by trader i, was quantified in the following form:

$$s_{t,i} = \frac{E_{t,i}\left(p_{t+1} + d_{t+1}\right) - p_t\left(1+r\right)}{\gamma\sigma^2_{t,i,p+d}} \tag{7}$$

Where p_t is the price of the risky stock at t, $\sigma^2_{t,i,p+d}$ is the conditional variance of $p+d$ at time t, for trader i, γ is the absolute risk aversion coefficient, and $E_{t,i}$ is the expectation of a trader i at time t. The authors assumed a fixed number of traders, N ,and a number of stocks equal to the number of traders:

$$N = \sum_{i=1}^{N} s_i \tag{8}$$

Where s_i represents the optimal linear functional form for trader i. The authors successfully implemented the most important processes of learning and forecasting in the model through a classifier forecasting system which matches current information with conditional forecasting of future prices. Artificial stock market agents forecast future prices by matching precise trading rules to current informational fundamental or technical sets. A linear price and dividend forecast was presented in the following form:

$$E_{t,i}\left(p_{t+1} + d_{t+1}\right) = a_j\left(p_t + d_t\right) + b_j \tag{9}$$

Where p_t is the price of the risky asset at time t, $E_{t,i}$ is the expectation for agent i at time t, and d_t is a state variable. The parameters a_j and b_j are initially set to random values distributed informally about the rational expectation equilibrium (REE). P measure the probability and it is a crucial parameter which estimate the number of learning periods of all artificial agents as a function $K = K\left(p\right)$, where measures the learning rate. All matched trading rules were evaluated according to their accuracy in forecasting price and dividends. Every single trading rule kept a record of its squared predicting error:

$$\sigma^2_{t,i,j} = \beta\sigma^2_{t-1,i,j} + \left(1-\beta\right)\left(\left(p_{t+1} + d_{t+1}\right) - E_{t,i,j}\left(p_{t+1} + d_{t+1}\right)\right)^2 \tag{10}$$

The worst performing 15 per cent of the trading rules in their experiment were eliminated from the market participant's rule setting and replaced by completely new rules via genetic algorithm crossover and mutation techniques.

The SF-ASM model successfully replicated stylized facts or empirical puzzles of the real-world financial time series such as fat tails (non-normal return distributions) and persistent volatility. Also, artificial agents were able to collectively learn a homogeneous rational expectations equilibrium (REE) for certain parameters resulting in time series and individual forecast values consistent with the equilibrium parameter values.

One of the major disadvantages of the SF-ASM is that the software is not easy to read or use. Also, the software was developed before objective programming languages were popular, and was only adapted to objective form. Although some of the general statistical properties and features can be replicated, exact simulation of trajectories across different computing platforms proved difficult. LeBaron (2005) argue that the classifier system of SF-ASM has proved to be very complicated for modelling purposes. This is due to the fact that many parameters have to be defined, and it is rather unclear which of these is important. Moreover, the implementation of the classifier and the lack of technical trading rules are often criticised. Finally, another important critique of SF-ASM is that by assuming constant absolute risk aversion utility functions, the software ignores the wealth dynamics of traders. Fortunately, this serious design issue is not present in Altreva Adaptive Modeler (there is no constant absolute risk aversion utility function and the software takes into account the wealth level of traders, Section 3.2 of the thesis).

Arifovic and Masson (1999) created a model to simulate the emergence of foreign exchange crises. The artificial market model aimed at better understanding of periodic currency crises in developing countries, because recent studies have shown that developing markets are vulnerable to periodic crises as well as significant capital outflows leading to overall macroeconomic instability. The agents were allowed to make simple risk neutral portfolio decisions between developing and developed country dept and then compared the returns of developed country debt with the expected returns of developing country dept. Initial values from the Argentina's currency crisis were imported to the model to observe the process of emergence of the crisis.

Chakrabarti and Roll (1999) produced an artificial stock market where traders mimicked other large traders' behaviour in an attempt to achieve more sophisticated forecasting results. This was done by observing and adjusting trading behaviour towards better performing agents. The authors observed that intensified trading activity lead to sharp price movements and a enhanced learning process. Another important conclusion is related to the threshold level at which a trade conducted between two artificial market participants is noticed by all other traders. Chakrabarti and Roll purposefully reduced the threshold level in their experiment resulting in reduced price volatility and increased forecasting accuracy. This fact indicated the presence of highly effective learning processes in the market. The authors implemented a novel approach to investigate the impact of various parameters and performed different simulations at randomly chosen parameter values. To analyse this huge dataset, they performed multiple linear regressions on the parameter values in order to investigate their implications on empirical findings. This, however, is a lengthy and indirect to examine parameter sensitivity.

Chakrabarti (1999) replicated one of the most well documented features in intra-day data- the U-shaped pattern in bid-ask spreads. The bid-ask spreads tend to be wide at the opening of trading, narrow during the day, and wide before the close, forming the U-shaped pattern. Simulated foreign exchange traders received random order flow during trading hours providing them with information about the aggregate order flow, and the value of the currency. Traders were risk averse and their reservation prices were estimated in a Bayesian learning framework. The empirical results revealed the presence of the U-shaped spreads and volatility presence. Surprisingly, there was more unexplained variation in the variables during afternoon trading highlighting the importance of price path dependence in trades executed at different time slots. Moreover, Chakrabarti (1999) detected the presence of significant nonlinear effects in both quadratic and cross terms indicating the complex relationship between the underlying information and preference parameters.

The artificial stock market model of Lux and Marchesi (1999) has its roots in earlier attempts of economists at implementing heterogeneity into stochastic models of speculative stock markets (Samanidou *et al.,*2007). This particular simulation of a stock market was inspired by the analysis of herd behaviour in ant colonies (Kirman, 1993), and earlier applications of statistical mechanics to different issues in sociology and political sciences.

Their artificial stock market model was capable of generating bubbles with over-or undervaluation of the stocks and periodic oscillations with repeated market crashes. Traders were allowed to switch between a chartist and fundamentalist trading strategy creating more complicated market dynamics than previous models found in the literature. The complicated dynamics generated chaotic patterns in mean values of the relevant state variables represented by the number of traders in each group plus the market price. Statistical analysis of these chaotic patterns revealed that they are characterised by fat tails providing a possible explanation of one of the ubiquitous stylised facts in financial time series.

In a series of more recent papers Huang and Solomon (2000, 2001, 2002) extended the Levy *et al.* (1994) simulation of artificial stock market and implemented a random multiplicative process for the wealth of each trader, with different traders coupled through their average wealth in a similar way to predator-prey models (based on Lotka-Volterra systems frequently used to describe the dynamics of biological systems in which two species interact, one a predator and one its pray). The authors assumed that all traders start with the same wealth but later each of them speculates differently on the market and gains or losses funds proportional to his current wealth state quantified by:

$$w_i\left(t+1\right)=\lambda w_i\left(t\right) \quad \left(i=1,2,.....,N\right) \tag{11}$$

Where λ is a number fluctuating in a small interval D centred about unity. The important conclusion was that due to the random nature of the multiplicative process a few hundred professional speculators and not the millions of non-speculative families dominate most of the market movements.

Cont and Bouchaud (2000) proposed one of the simplest artificial stock market models, having only a few free parameters and a cluster (group) of traders making joint decisions in order to simulate the herding behaviour of traders. According to their model at each time step, each cluster either trades with probability $2a$ or sleeps with probability $1-2a$. During the trading process agents either buy or sell an amount proportional to the size s of the cluster i. Hence, a represent the probability of the member of a cluster to be a buyer.

The market price is driven by the difference between the total supply and demand and the logarithm of the price changes proportionally to this difference. Similar to my experimental findings in Chapter 5, the authors observed and explained important stylised facts of financial markets such as volatility clustering, positive correlations between trading volume and price fluctuations and asymmetry between sharp peaks and flat valleys.

Ehrenstein (2002) and Westerhoff (2003) applied Cont and Bouchaud's model to investigate the implications of a small Tobin tax on all transactions and to estimate the exact amount of tax needed. Working independently both authors discovered that depending on parameters, either the total tax revenue has a maximum in the region of up to 1 per cent taxation, or it increases monotonically. After taking into account the tendency of governments to overexploit available sources of tax income, they recommended the introduction of the Tobin tax for the first case, but not for the second. Moreover, the authors claimed that this would potentially reduce the amount of speculation, but not by an order of significant magnitude (Ehrenstein *et al.,*2005).

Lux and Marchesi (2000) investigated the occurrence of stylised facts of financial returns within an artificial multi-agent framework of speculative activity where artificial traders with both chartist and fundamentalist strategies operated. The authors presented a possible explanation for stylised facts such as unit root in levels together with heteroscedasticity and leptorkurtosis of returns. They observed that volatility clustering phenomenon is a consequence of the market being subject to occasional temporal instability.

De Fontnouvelle (2000) implemented models with artificial agents to investigate whether the forces driving volatility and trading volume in financial markets are the same as those driving asset returns. The author concluded that the driving forces behind return volatility and trading volume differ from asset returns. However, the simulated time series for return volatility and trading volume showed a copersistence very similar to that observed in actual financial markets.

Darley *et al.* (2000) simulated realistic market trading mechanism to investigate the impact of changing the tick size with different populations of parasite artificial traders. The evaluation was done through traders' ability to closely track fundamental values, and price volatility.

The effect of changing the tick size was insignificant when there were just a few parasite traders. However, when the number of parasites increased, the impact of the parasite strategies also increased leading to reduced market ability to track intrinsic values. This particular finding is contrary to the perceived assumption that any reduction in the tick size is leading to efficient market functioning.

In 2000, Chen and Yeh constructed artificial stock market model similar to the SF-ASM. The slight difference was that the price adjustment process occurs in response to excess demands based on forecasting abilities of future prices. The authors implemented genetic programming technique in order to modify traders learning function. An important point is that Chen and Yeh enabled all traders to evolve their prediction functions. The actual goal of each artificial trader was to maximize the one-period expected utility function:

$$E_{i,t}\left(U\left(W_{i,t+1}\right)\right) = E\left(-\exp\left(-\lambda W_{i,t+1}\right)\middle| I_{i,t}\right\rangle\right) \quad (12)$$

Subject to:

$$W_{i,t+1} = \left(1+r\right)M_{i,t} + h_{i,t}\left(P_{t+1} + D_{t+1}\right) \quad (13)$$

Where $W_{i,t+1}$ is the total wealth of trader i at time period t; P is the price of the stock at time period t; D_t is per-share cash dividends; $E_{i,t}(.)$ represent trader i conditional expectation of W_{t+1} given her information up to t; $I_{i,t}$ is the information set and r is the riskless interest rate.

Chen and Yeh designed an innovatory pool of various naturally occurring trading rules called a ‘business school’. In essence the ‘business school’ included a group of artificial agents (school members) competing with each other for the purpose of determining the best possible forecasting models. Both the ‘business school’ participants and traders continually co-evolved over time. While the success of the ‘business school’ members was measured by their current forecasting abilities, traders’ success was observed by the accumulation of wealth resources.

The interesting feature is that each trader individually chooses between two options- whether to trade in the artificial stock market or to request some time off to attend the 'business school' in order to experiment with specific forecasting models developed and provided by school participants. The rationale behind 'business school' attendance is to determine a forecasting model that is superior to the one currently adopted by the other artificial traders. Remarkably, Chen and Yeh concluded that individual artificial traders behave as non-believers of the Efficient Market Hypothesis (EMH), although experimental results suggested that the artificial stock market is efficient. This outcome is consistent with my experimental findings presented in Chapter 4. The other conclusion they made is that market behaviour experienced constant changes. For instance, the initial success of some trading strategies was no longer effective as they were adopted by a large number of artificial agents. Although simulations of this artificial stock market displayed some features of the real-life return time series, there are several features that disagree with the real data. For instance, there is a large level of positive skew and the return series are independent, which indicates there may be no persistent volatility pattern. Moreover, the authors tested for the presence of a unit root and couldn't reject a unit root. This seems surprising considering the stationary nature of the dividend process.

A year later, Chen and Yeh (2001), applied genetic programming technique to evolve a population of traders learning over time. Their experiment was based on the SF-ABM and aimed to understand the emergent properties of the EMH and the REE hypotheses. They inquired whether the macrobehaviour depicted by those two hypotheses is consistent with academic theories of microbehaviour. The authors applied a series of econometric tests and concluded that the results cannot be interpreted as a simple scaling-up of individual behaviour and proposed a conjecture based on sunspot-like signals to explain why macrobehaviour can be very different from microbehaviour.

Chen *et al.* (2001) introduced a stochastic simulation model of a financial market populated by noise traders and fundamentalist speculators to explore the behaviour of the model when testing for the presence of chaos or non-linearity in the simulated data. All artificial traders in their experiment were allowed to switch between optimistic and pessimistic moods as well as between noise trading and speculative trading. The authors achieved very unstable results in their tests for non-linearity and dependence.

They reported both acceptance and strong rejection of identically and independently distributions in different realisations of the model. However, when tested for independence in second moments and estimating GARCH structures, the results appear much more robust and the computed GARCH model closely assembled the typical outcomes of various empirical studies.

The Genoa artificial stock market was created by Raberto *et al.* (2001). This artificial stock market consisted of heterogeneous agents trading one single asset through a realistic trading mechanism for price formation. Initially, agents were endowed with a finite amount of cash and in each period they made random buy and sell decisions constrained by available resources and dependent on the volatility of previous periods. The most important aspect of the Genoa artificial market was its ability to generate the key stylised facts such as fat tails and volatility clustering using simple trading rules. The disadvantage of this particular artificial stock market is that traders are generally fairly unsophisticated. Buyers and sellers group into larger dependent groups, which then move together creating herding behaviour by design (this is similar to my empirical outcomes discussed in Chapter 5).

Bullard and Duffy (2001) simulated a traditional macroeconomic model for asset prices by enabling traders to predict future price levels using a recursive regression technique, which exhibited excess volatility. The purpose of this particular simulation was to capture market features similar to real-life stock markets. The authors reported parameters that were able to give them reasonable volatility in stock returns and low volatility in per capita consumption growth consistent with U.S. macro economic data.

LeBaron (2001a) and LeBaron (2002) presented artificial stock market models capable of mimicking different real-life portfolio strategies. Traders' memory lengths followed Levy *et al.* (1994), and while some of them evaluated their strategies by using a small past history of returns, some others used longer histories. Traders' preferences were based on constant relative risk aversion, and therefore more wealthy traders possessed a larger part of the market. Agents had the freedom to choose optimal strategies according to their history length and traded a risky and a risk free asset.

The author compared all simulated results with real historical data for the S&P 500, and reported replication of a large range of financial market phenomena such as long memory volatility persistence, volatility cross correlations and the tendency for volatility to generally lead trading volume. However, the main ctiticism of this artificial stock market is that predictability in the early stages of the market is unrealistically high. Regressions of returns on simple lagged returns can generate R-squared values as high as 0.7.

Stock markets are often dependent on the actual behaviour of a broker dealer who controls liquidity (LeBaron, 2004a). Chan and Shelton (2001) study the role of this dealer, and its ability to learn optimal strategies. The dealer in their experiment was positioned within a market with random order flow generated by both informed and uninformed traders. Uninformed traders operated in a random manner and informed ones received a signal about the true value of an asset. The authors tested different learning mechanisms and reinforcement learning algorithms and captured the optimal policy function for the market maker.

Audet *et al.* (2001) addressed the important question of order book versus dealer markets to investigate which type of market is preferrable and whether there is a general optimal stock market structure. The latter question is a puzzling one, since many real-life stock markets have both trading systems working together. The order book was designed to simulate most modern electronic trading systems and allowed all traders to know the current depth and liquidity in the market.

The dealer markets also mimicked actual features of this type of market disclosing less information on the state of the market, but enabling traders to reveal less information related to their current demands. Interestingly, when order flow increased and become more correlated across customers, dealer markets seemed more desirable destination for trading. This particular finding indicated that when investors move large orders through a trading system, the anonymity of the dealer system outperformed the transparency of the order based system. Moreover, the authors demonstrated that dealer markets are more attractive when the number of dealers is increased and their risk aversion is reduced. It was evident that the dealer market was able to absorb significantly more risk as answering a critical policy question.

Yeh (2003) based his artificial stock market principles on Chen and Yeh (2001) to examine market performance under different tick sizes. Similarly to most models in this section of my thesis, traders used forecasting information based on future prices and dividends to establish a reservation price. Orders were executed in random ordering of traders, within an order book. In cases when traders' reservation price does not cross the current book, traders issued a limit order through a simplistic algorithm consisted of the reservation price and the data quotes. His experimental results demonstrated that the tick size plays an important role in market performance. Reduced tick sizes has led to decreased spreads positively affecting market liquidity. This finding is consistent with some of the high frequency trading outcomes of Chapter 9 of this thesis.

In a similar fashion, Bottazzi *et al.* (2003) compared the Walrasian trading protocol with a batch action procedure (similar to Altreva Adaptive Modeler, Section 3.2.4), and a limit order book to determine the optimal trading structure. The structure of the model was similar to the SF-ASM, enabling traders to produce relatively simple price and volatility forecasts as weighted averages over the recent past. Based on random principles the traders decided between executing a limit order stored in the order book, or a market order implemented immediately. The batch market was a preferable destination for trading when the probability of a market order was large, but when this probability was low, the order book market was preferred. This finding indicated that in markets where traders are not able to reveal any information in the limit book, a batch action is required to aggregate the sparse quantity of information received through order flow.

Although the vast majority of artificial stock markets cited so far described empirical replication of stylized facts, the model developed by Winker and Gilli (2001) attepted to formally estimate parameters based on the Kirman (1991) model, where agents switched chartist and fundamentalist strategies and a weighted average of their forecasts was used by portfolio managers trading in the foreign exchange markets. Winker and Gilli (2001) fitted two features of actual financial returns such as kurtosis and the first order volatility coefficient in an ARCH (1) model and provided a detailed study of the sensitivity of the results to different parameter specifications.

Joshi *et al.*(2002) applied game theory to the SF-ASM to examine the optimal frequency for traders to revise their market forecasting rules. The authors reported a unique strategic Nash equilibrium, and this particular equilibrium is sub-optimal in the sense that artificial agents' earnings are not maximized and the market is inefficient. Joshi *et al.* (2002) concluded that this strategic equilibrium is based on the prisoner's dilemma principles and suggested that financial markets can end up in situations similar to this dilemma in which frequent revisions of forecasting rules and extensive technical trading enhance price variability and reduce earnings.

Aoki (2002) composed an artificial stock market framework for market participation with infinitely many strategies or trading rules. The author derived the number of types or clusters of agents from a rather general specification of the entry and exit dynamics to demonstrate that often the sum of the fractions of agents in the two largest groups will be close to one. This finding provided a theoretical rationale for the confinement to two trader groups in many models of speculative dynamics.

Iori (2002) presented an interesting artificial stock market where traders received a signal that combined information on choices of local neighbours, and this signal was used as an input into trader's decision to generate a bid or ask order for an asset. Then the market maker covered all imbalances in orders and adjusted the asset prices by:

$$p_{t+1} = p_t \left(\frac{D_t}{Z_t} \right)^{\alpha} \tag{14}$$

Where D_t and Z_t represented the number of positive and negative values of the signal. Asset returns were estimated by log difference of equation (14), and the volatility was calculated by the absolute values of these returns. Remarkably, this particular artificial stock market generated nearly uncorrelated returns with volatility close to long memory and simulated behaviour of a long hyperbolic decay pattern in the autocorrelations. The model also showed strong positive correlation between trading volume and volatility and indicated that the threshold part of the simulations caused volatility to occur. LeBaron (2006) pointed out that the only disadvantage of this particular artificial stock market is that the thresholding of the signal is vitally important for volatility clustering to appear.

Similar to my experiment in Chapter 8, Chen and Liao (2003) simulated the stock market environment to examine the possible explanations for the presence of the casual relation between stock returns and trading volume. The authors argue that conventional assumptions such as information and reaction asymmetry, noise traders and tax factors are not explicitly required. Their simulation findings suggested that the stock price-volume relation could be seen as a genetic property of financial markets in the light of autonomous interacting traders.

Cincotti *et al.* (2003) developed an artificial stock market with a finite amount of cash resources and number of stocks. The initial training of the model consisted of a random trading strategy constrained by the finiteness of resources available and by the level of market volatility. Three different trading strategies were introduced and experimented in two different market conditions: a steady market and a growing market with asset inflation. The stock market properties successfully generated stylised facts such as volatility clustering, fat tails and reversion to the mean (I detected the same stylised facts in Chapter 5). The authors demonstrated that the profitability of each trading strategy depends on the periodicity of portfolio reallocation and on the market condition. However, only a strategy that fully exploited the reversion to the mean process gave satisfactory results in all cases.

Lawrenz and Westerhoff (2003) simulated a foreign exchange market populated by heterogeneous, boundedly-rational traders which rely on a mix of technical and fundamental trading rules evaluated by genetic algorithm learning. Their model simultaneously generated numerous stylised facts such as unit roots, fat tails, volatility clustering, a fuzzy relationship between news and exchange-rate movements, co-integration between the exchange rate and its fundamental value, a declining kurtosis under time aggregation and weak evidence of mean reversion.

The artificial stock market created by Noe *et al.* (2003) represents the first model to consider corporate finance related issues. Their experiment examined how companies raise capital and issue securities that maximize their profits. The model also investigated how investors evaluate the securities issued by companies based on past information.

This particular simulation included a company with an investment project that needs to be financed by choosing from six different securities that can be issued, and two potential external investors. The six securities included debt and equity as well as convertible and subordinated dept. When the evolutionary learning process was absent from their experiment, companies competed against a fixed set of investors who knew the appropriate pricing operations and equity and subordinated debt dominated the market in contrast to the real-life markets where straight debt is used. Surprisingly when an evolutionary learning process was introduced in the model, debt was the most commonly used financial instrument with subordinated debt placed second, and equity third. The authors demonstrated that the popularity of debt is associated with the ability of artificial traders to learn via co-evolutionary learning dynamics and evaluate different financial instruments. An important conclusion of this experiment is that investors tend to underprice securities in cases with or without learning. Although this is an interesting first attempt to connect artificial stock markets with corporate finance and the coevolution of traders' behaviour along with the institutions that guide that behaviour is interesting, the results of this study need to be explored under different learning specifications and investment institutions.

Ehrentreich (2004) reconsidered and reassessed the evidence of technical trading in the SF-ABM framework in light of an upwardly biased mutation operator and an arbitrary mutation operator. While the arbitrary mutation operator failed to establish technical trading activities, two tests other than the inappropriate investigation of the aggregate bit level established the existence of technical trading beyond that level. When the selective forces are rather weak, Ehrentreich found that the continued existence of technical trading can be reconciled with the EMH.

Gulyas *et al.* (2004) extended the early version of the SF-ABM model by combining the realm of agent-based modelling and participatory simulations where some agents are artificial, while a human subject imposed control on others. The authors demonstrated that blending models of experimental economics and agent-based modelling tools can enhance the processes of testing human economic behaviour as well as theoretical assumptions embedded in computational models. Interestingly, their experimental results indicated that technical trading could result in market deviations such as bubbles and crashes.

Markose *et al.* (2003) investigated the Red Queen principle within artificial stock market settings. The Red Queen theory places constraints on performance enhancement of all individuals if each is to maintain the status quo in relative fitness measured by an index relative to aggregate performance. In artificial stock market settings agents have to individually learn and adapt in a multi population genetic programming environment, retraining of genetic programmes is performed in *ad hoc* way. However, Markose *et al.* (2003) argue that in terms of the Red Queen principle, performance enhancement should be of an endogenous constrainted type, which is different to *ad hoc* and exogenously imposed by the researcher. Moreover, the authors argue that the rate of retraining is associated with the extent to which a lower bond constraint on agent wealth relative to aggregate wealth is satisfied or not.

Korczak and Lipinski (2004) presented an artificial stock market based on an evolutionary algorithm that used technical analysis trading rules. The authors combined two approaches to investigate whether replacing the original set of rules by a set of linear combinations of them causes a significant decline in performance. The first approach included 350 trading rules and the second one consisted of 150 trading rules. After running numerous experiments with real-life data from the Paris Stock Exchange, the authors demonstrated that the original set of trading rules can successfully be replaced with the smaller set of linear combinations of them without a major efficiency loss, while achieving a significant reduction in computing time. This finding enabled optimization of trading strategies in real-life systems because computing time is the major constraint in real-time data processing.

Similar to Altreva Adaptive Modeler stock market structure (Section 3.2.3), Raberto and Cincotti (2005) composed a double-action artificial stock market populated by agents with heterogeneous beliefs who trade one risky asset in exchange for cash. All agents were designed to issue initial random orders subject to budget constraints in order to investigate two significant stylised facts of the limit order book: the distribution of waiting times between two consecutive transactions and the instantaneous price impact function. Raberto and Cincotti demonstrated both theoretically and through simulations that when the order waiting times are exponentially distributed, trading waiting times are also exponentially distributed.

Amilton (2005) implemented the efficient method of moments and maximum likelihood methods to examine earlier simulations of financial markets. The author compared the results of his models to the results of real data and conventional econometric models. He claimed that previous findings in artificial market modelling are based on unrealistic computing of the noise term and when the stochastic process is properly estimated the models are still capable of generating some stylised facts, but the fit is quite poor.

Mannaro *et al.* (2008) studied the effects of the Tobin tax within artificial stock market settings. The microstructure of their market was constructed of four different types of traders: random traders, fundamentalists, momentum traders, and contrarians with limited resources. There were two separate artificial markets with different transaction taxes and traders were allowed to choose in which market they would like to trade. An attraction function that drives traders' choice based on perceived profitability was introduced. After performing extensive simulations the authors concluded that the Tobin tax increases volatility and decreases trading volumes.

The artificial stock market created by Beltratti and Margarita (1992) and developed further by Beltratti *et al.,* 1996 is very different from the other artificial markets presented so far. The model was designed to search for emergent pattern in the trading behaviour of agents, but unlike the other markets, this one didn't have any organised trading institutions. Traders operated in completely random fashion bumping into potential trading patterns. All traders developed forecasts based on assumptions and past local and global information. Then traders compared their predictions and the one with the largest forecast was allowed to buy 1 share from the trader with the smallest forecast. The actual trade was executed at the average of the two forecasts and the market kept record of the average execution price across the random pairings.

A common concern about all artificial stock markets discussed so far is validation. The main criticism is that there are too many degrees of freedom. Agent-based modellers are able not just to move freely through large parameter environments, but can also change entire markets in the attempt to fit different sets of stylised facts of financial returns. Moreover, researchers are starting to move from the more stylised earlier artificial stock markets towards advanced simulations of financial markets discussed in the next section.

The latter try to replicate very explicitly the actual trading mechanisms that are being used in real-life trading (for example, Altreva Adaptive Modeler, Section 3.2) rather than building a stylised trading framework. The advanced artificial stock markets are well designed to describe the construction and design of actual stock markets.

2.1.5 Advanced artificial stock market modelling trends

Vytautas and Ramanauskas (2010) proposed an artificial stock market based on the interaction of heterogeneous traders that is characterised by forward-looking behaviour governed by a reinforcement-learning algorithm combined with an evolutionary selection mechanism. This particular model was used to analyse market efficiency and self-regulation. The results suggested that the asset prices in their model reflect fundamentals in broad terms but under-or over-valuation phenomena are sustained for prolonged periods of time. The authors found weak self-regulation ability of the market, indicating that institutional settings alone, such as the centralised exchange based on double auction trading are not sufficient to guarantee effective market functioning. Moreover, they found a positive relationship between asset returns and liquidity changes, and therefore exogenous shocks to investors' cash holdings are leading to substantial changes in asset prices.

Yeh and Yang (2010) analysed the effectiveness of price limits from the perspectives of volatility, price distortion, volume and welfare in an artificial stock market. The market was populated by boundedly rational and heterogeneous traders whose learning behaviour was based on genetic programming algorithm. The authors observe both positive and negative effects from the imposition of price limits. When compared to the market without price limits, it turned out that properly developed price limits reduce volatility and price distortion, and have positive impact on liquidity and welfare.

Not very different from the zero-intelligence traders in Chapter 6 of this thesis, Ponta *et al.* (2011) created a multi-asset artificial stock market populated by zero-intelligence traders with finite financial resources to study the statistical properties of their market. The market consisted of different types of assets representing companies from different economic sectors.

Companies operating in the business sectors of the economy were not allowed to pay dividends as a consequence of random restrictions on the allocation universe of zero-intelligence traders. When dividend-paying assets were introduced to the market, the artificial stock market returned the same structural results observed in the experiments without dividends. These findings suggest a significant structural influence on statistical properties of the artificial stock market.

In another study Ponta *et al.* (2011) examined the statistical properties of the univariate and multivariate processes of prices and returns in an artificial stock market characterised by heterogeneous informed traders. Traders disseminate information and sentiments through interactions determined by sparsely connected graphs. There was a central market maker that governed the price processes for each asset at the intersection of the demand and supply curves (in comparison, there is no market maker in all my experiments). In terms of the univariate price processes their artificial stock market was able to generate stylised facts such as unit roots, volatility clustering and fat tails. The proposed simulation model endogeneously reproduced the static and dynamic stylised facts such as cross-correlation between returns of different securities.

LeBaron's (2012) model of an artificial stock market not only replicated most of the key features of financial time series, but highlighted some genetic stabilizing and destabilizing properties that may appear in real stock markets. The author went further and tested the Minksy's (1986) macroeconomic effects (Minksy argues that financial markets exhibit bubbles as investors become increasingly confident about markets) in a laboratory environment. Remarkably, LeBaron observed that market dynamics are predominated by irregular swings around fundamental values and asset prices slowly increase, and then crash fast and dramatically with high volatility and high trading volume. During the price rise, traders with similar volatility forecast models lowered their assessment of market risk encouraging more aggressive trading behaviour leading to a crash. This process fully matches the Manksy market instability dynamic, and other contemporary approaches to financial instability.

Cooke (2012) proposed a new class of artificial stock markets that use general equilibrium price clearing with built in exogenously dividends from a geometric vector autoregressive model. There was no risk free asset and traders were enabled to form their demands for an arbitrary number of stocks with arbitrary covariance structure. Cooke (2012) derived an optimisation problem for estimating equilibrium prices and demonstrated that his simulated market successfully generated many real stock market dynamics and phenomenon.

Kimura *et al.* (2012) proposed a different approach for the analysis of stock prices by investigating the simulation of traders' behaviour in an artificial stock market where traders interact, by demanding and supplying stock, driving asset prices to an equilibrium value. Their empirical results of the dynamics of asset prices indicated that, under the assumption of utility maximizing traders with different expectations about future dividends, stock prices may under-react. Moreover, the gradual change of stock prices in the sub-reaction confronted the principles of the EMH.

Huang *et al.* (2012) constructed an order-driven artificial stock market to analyse the transaction costs of ten securities in the Taiwan stock market. Their empirical findings showed that the liquidity costs of market order of the ten financial instruments ranged between -1 per cent and 1 per cent. However, the simulation costs of market order in their artificial stock market are larger than those of real data ranging between 0 per cent and 10 per cent. The authors assumed that the reason for this difference is that investors in real-life stock markets do not execute their orders blindly. Regardless of the difference, this models represent an appropriate simulation tool for transaction cost assessment when investors would like to liquidate their stocks in a short span of time.

Panayi *et al.* (2012) introduced a semi-synthetic artificial stock market based on a new concept in the area of agent-based modelling of financial markets. Essentially the model was partially based on real order book data and partially based on artificial agents, making it closer to the real-life market rather than a pure agent-based model. The authors performed tests of realistic daily trading runs believing that traders from real markets execute their buy or sell orders not as suggested in the dataset, but based on the difference between the price they observe in the artificial stock market and the price they has seen on the real stock market.

The experiment examined the effect of using homogeneous, limited intelligence traders compared to realistic traders to evaluate a number of financial metrics for intra- and inter-day variability. The homogeneity of limited intelligence traders resulted in constant upward or downward trends, as well as atypical volatility indicating that results using realistic traders and relative pricing of real order tend to outperform other modelling approaches.

The main criticism even in advanced artificial stock markets is that most simulations assume trading of only one risky asset and one risk-free asset alone. It is true that, with all of the latest technological innovations in the latest simulations, it was important to begin with the simplifying case of one risky and one risk-free asset. LeBaron (2006) suggest that the process of such simplification could prevent the observation of many interesting features of artificial stock markets. Trading of these two assets usually take place in one stock market representing another common issue in contemporary artificial stock markets. Often the risk-free rate within all simulations is fixed, and therefore the market is not a general equilibrium model. This can be problematic in that examining the level and volatility of the risk-free rate of the two assets itself represents another asset pricing dilemma. Setting the risk-free rate to be as low and stable as it is in real-life macro datasets proved very difficult, and most researchers ignore this important issue. The very recent artificial stock markets discussed in the next section aims to avoid these criticisms.

2.1.5 Recent artificial stock market modelling trends

Mathieu and Brandouy (2013) developed the Artificial Trading Open Market (ATOM) with the aim of simulating the main features of the Euronext-NYSE stock exchange microstructure. Consistent with the order generation process described in Section 3.2.4, ATOM successfully matched bid and ask orders submitted by artificial traders to determine quotations and prices. These market evaluations were governed by a negotiation system between sellers and buyers based on an asynchronous, double action mechanism built in the order book. Hence, ATOM was able to generate, play or replay order flows from real-life or artificially generated stock exchanges in a very quick manner. Remarkably, the model enabled the combination of human-beings and artificial traders on a single market using its network capabilities. ATOM has been involved in various investment practices, portfolio management, risk management, stock market regulation, and algorithmic trading.

Yang and Sun (2013) proposed a novel model of a European option market which better replicated some features of real option markets. Their artificial market consisted of three types of option traders and multi-agent matchmaking tradeoff that closely mimicked real option markets. The experimental results revealed interesting findings. The average option price increased with an increase of information inflow, but the variance decreased. This proved the significance of information disseminated by the option market in the model. By using different proportions of hedging and speculating option traders, the results suggested that an increase of hedging option traders can stabilize the market. On the other hand, an increase in the proportion of speculative option traders can make the market more volatile and this finding is consistent with the results from real financial markets.

Brandouy *et al.* (2013) constructed an agent-based artificial stock market populated by heterogeneous mean-variance traders with quadratic utility functions to study the effect of individual investor's preferences on their portfolio dynamics from the wealth and risk adjusted return point of view. The authors compared the relative performance of investment strategies using ecological competitions, where populations of artificial traders co-evolve. The empirical findings indicated that a higher relative risk aversion process helped the population of traders to survive in the long-run measured by a higher wealth indicator or the Sharpe ratio of constrained portfolios.

When short-selling was introduced in the model, the highest risk aversion did not secure the highest profits, eliminating risk takers and traders with high risk aversion from the market and leaving only traders with moderate level of risk aversion to survive in the long-run.

Utility-maximizing consumption and investment strategies in closed form are unknown for realistic settings which include portfolio constraints, incomplete markets and very high number of state variables. Conventional numerical techniques experience difficulties in solving those issues. Kraft and Munk (2013) developed artificial market strategies to tackle the constrained consumptions-investment issues. The authors applied their artificial market strategies to the life-cycle problem of an individual who gets unspanned labour income and is prohibited from borrowing and short selling.

Youki *et al.* (2013) investigated the existence of and the possible origin of the disposition effect in an experimental environment that closely mimics real stock markets. The authors accurately depicted actual individual investor trading behaviour and found the presence of the disposition effect in their artificial stock market. The loss aversion process has been pointed out as being one of the possible sources of the disposition effect.

Another often ignored and unsolved issue even by the latest technological artificial stock market innovations is timing. In an environment of constantly evolving strategies timing can play a significant role in process of adaptation of the strategies to the specific timing and trading structures. Moreover, developing multiple artificial stock markets for trading is still a difficult problem for all agent-based modellers. This issue has not been solved yet even after the introduction of powerful and sophisticated modelling tools such as Altreva Adaptive Modeler (Section 3.2). Once agent-based modellers are more confident in advanced mastering of Genetic Programming technique, they will probably develop multi-asset markets. It will be interesting to see whether researchers will be able to develop asynchronous trading actions performed on multi-asset artificial stock markets as well. This remains to be seen in the near future.

Chapter 3.

Artificial Intelligence Tools and Main Software Platform

3.1 Artificial Intelligence Tools in Finance

3.1.1 Artificial Neural Network (ANN) in Finance

The aim of artificial intelligence is to develop a system that could compute, learn, remember, and optimize in the same way as a human brain (Cheng and Titterington, 1994). The history of neural networks begins when scientists were trying to model the neuron. McCulloch and Pitts (1943) created the first model of a neuron with two inputs and one output. The authors pointed out that a neuron won't be active if only one of the inputs is active. The weights for each input were equal, and the output was binary. McCulloch and Pitts' neuron is known as *logic circuit* in science. In general, neural networks represent a collection of simple computational units interlinked by a system of connections. The number of units can vary widely and the connections be intricate (Cheng and Titterington, 1994).

Artificial Neural Network (ANN) is in fact a functional simulation of a simplified model of the biological neurons aiming at the simulation of intelligent data combined with evaluation methods such as pattern recognition, classification and generalization by using simple processing units called neurons (Malik, 2005). The rationale behind the development of the ANN models was the idea of detailed study on how the brain works and to create a mechanism that would function in the same way. While powerful computers process elementary operations in nanoseconds, the human brain requires milliseconds. However, the critical difference is not in the speed of processing but in the organization of processing. A key notion in neural networks is the connectionism. All processing tasks in the human brain are spread over 10^{12} neurons connected each other and the brain achieves complex tasks based on the massively parallel way in which many simplistic operations are performed simultaneously (Cheng and Titterington, 1994). The purpose of the artificial neuron is to mimic the first-order characteristics of actual biological neurons. The usefulness of ANN is based on its capability to solve difficult problems through the high degree of connectivity that provides the neurons its high processing ability.

Each neuron performs only very limited operation, but the parallel- distributed architecture is massive and enable quick solutions by working in parallel. The ANN's are capable of developing a generalised solution to any particular problem other than that used for training and to generate valid solutions, even when there are errors in the training dataset (Malik, 2005). They have an excellent ability of approximating any nonlinear mapping to any degree of accuracy (Hornik, 1989) and don't require a priori model to be assumed (Bishop, 1995). The major advantage of ANN's is that the domain knowledge base in disseminated in the neurons and information processing is performed in parallel-distributed manner (Udapa, 1997). Moreover, ANN's are highly parallel data processing tools capable of learning functional data dependencies (Malik, 2005). These advantages make the ANN a powerful tool for modelling issues in which functional relationships are uncertain or vary through time.

There are different variations of ANN's but all of them have three important key elements such as the individual neuron, the connections between the neurons, and the learning algorithm. Every single different variation is characterised by the kind of possible connections between neurons. For instance, one neuron could be connected to another, but the second neuron cannot have another connection towards the first. ANN's are composed of one or more layers of neurons. In massive neural network entities such as Perceptron, Multi-layer feed-forward network with Back-Propagation (BP) learning, the Boltzmann Machine, Linear Associator and the Grossberg model, the output from the units from a single layer is allowed to activate neurons in the next level only (Dillon, 1991).

However, neural networks suffer from two major limitations, despite their popularity. The unsatisfactory low performance under uncertainty in the data is the first issue. For instance, neural networks are powerless in point forecasts where uncertainty in operation of the system exists. Unexpected passenger demand in public transportation systems, machine breakdowns on the shop floor and abrupt changes in weather conditions in the national energy market are areas where neural network performance is inefficient. This issue is based on the fact that the neural networks generate averaged numbers of solutions dependent on inputs. Unfortunately, such reduction cannot be eliminated through changing the structure of the model or re-execution of the training process (Khosravi, 2010).

The second pitfall of the neural network is associated again with point directional predictions. Neural network models generate point predictions without any accuracy indications. Neural networks are unreliable if training data is sparse and targets are multi-valued or potentially affected by probabilistic events (Khosravi, 2010). Sometimes neural networks are criticised by financial practitioners as being black boxes (Benitez *et al.,* 1997) turning the attention of researchers to more sophisticated evolutionary computation.

All the neural network drawbacks can be avoided by the introduction of more sophisticated artificial intelligence programming tools such as Genetic Algorithms, Genetic Programming, and especially Strongly Typed Genetic Programming (Zhang and Ciesielski, 1999).

3.1.2 Genetic Algorithm in Finance

The Genetic Algorithm (GA) is a simple optimization technique and a powerful learning tool in many agent-based models. GA is based on a simulation of the natural selection process or survival-of-the fittest principle (Darwinian process). In essence, GA computationally mimics the process of natural selection and evolution. Holland (1960) was the pioneer creator of the first computational technique which successfully mimicked natural selection through crossover and mutation operators. The author applied fixed-length simple binary coding to solve difficult problems and real-life issues. The typical genetic algorithm consisted of three main stages on its way to determining optimal solutions.

Initialization was the first step which encompassed the random generation and development of various chromosomes or a pool of different solutions. In most cases the pool included up to 200 chromosomes. Larger pools offer greater diversity and better solutions at the expense of more computer power. Later on in the process, the initially diverse pool of chromosomes experience transformation of similarity and identity through specific genetic recombinations.

Fitness evaluation was the next step where the fittest chromosomes produce more offspring leading to optimisation of GA search for the individual representing the optimal solution to a particular issue. As a result, the objective function could be further maximized.

Selection is the final selection phase where the GA commences reproduction activites and selection of future generation parents. The selective mechanism identifies which individuals have the best genetic material for parental reproduction. The most fitted chromosomes are more likely to reproduce in the next generation of individual artificial agents. The most popular selection method is called 'tournament selection', where randomly picked pairs of strings are compared in fitness value terms and the fittest agent is forwarded to the mating population. The tournament process continues until the pool is fully populated (Goldberg, 1994). The Fundamental Theorem of Genetic Algorithms suggests that once an individual has been processed, the GA reaches many different points in the search space. Hence, a preliminary identified limited set of individual agents effectively explore huge search spaces (Goldberg, 1989).

Drake and Marks (2002) argues that genetic algorithms are useful financial market tools. For example, Hiemstra (1996) applied GA to the tactical asset allocation to observe that the established system outperformed a passive rebalancing policy. In another experiment Lettau (1997) used GA to examine portfolio decisions of boundedly rational traders in a financial market and demonstrated that the GA-based traders learned to carry too much risk as compared to the optimal portfolio of rational investors. Arifovic (1996) implemented GA to analyse the behaviour of the exchange rate, portfolio decisions, and composition decisions. The author reported similar GA performances to those observed in experiments with human subjects. GA can be applied in the development of optimal parameters threshold values for technical trading models, portfolio optimisation, pricing of options and futures, and automatic induction of foreign-exchange trading rules. Due to their transparency of the achieved outcomes, GA, provide wide practical applications in credit scoring, financial spreadsheets, processing insurance applications, portfolio management, and trading stocks. Moreover, GA can effectively detect financial market and exchange rate volatility (Chen, 2002).

Genetic algorithms are well-suited to for financial modelling financial markets because of the following reasons. Simplicity and flexibility are the main advantages of GA when simulating financial markets. GA are pay-off driven (pay-offs means improvements in predictive power or return over a benchmark). GA ability in finding approximate solutions to combinatorially explosive issues means that a very good match between the problem solving model and the problem exists. Furthermore, GA can be implemented in most parameter optimization processes. Unlike conventional models, GA simulates a wide range of extensions, flexible modelling and constraints. For instance, Multiple Discriminant Analysis (MDA) was for many decades the prominent bankruptcy prediction technique based on balance-sheet figures. GA is capable of formulating financial ratios which successfully determine future bankruptcy signals which significantly outperform Multiple Discriminant Analysis by more than 10 per cent (Kingdon and Feldman, 1995).

3.1.3 Genetic Programming in Finance

Genetic Programming (GP) was introduced by Koza in 1992 and represents a programming technique of automatically generating computer programs to perform specified tasks. It uses a GA to search through a space of possible computer programs for a program which is optimal in its ability to perform a particular task. GP is a branch of GA but much more powerful than the later. The major difference between GP and GA is the representation of the problem solution. While the outcome of the GA is a quantity, the result of GP operations is another computer program. GP is best used in situations where there is no ideal solution or in cases when variables are constantly changing. While in GA, a solution candidate is labelled a chromosome, in GP a solution candidate is organised as a parse tree whose nodes are procedures, functions, variables and constants in hierarchically structured programs. The tree consists of different building blocks (functions) which processes its branch nodes as inputs and provide an output at the end. The subtrees of a node in any parse tree describe the arguments to the procedure or function in that node. In a GP process, the researcher identifies all the possible variables, functions and constants implemented as nodes in a parse tree (Montana, 2002). The content as well as the size and shape of the tree might experience dramatic change during the evolution process. Similarly to GA, Genetic Programming replicates the stochastic process where the fittest survive and transfer their genetic material to the next generation. Then the fittest solutions multiply and create completely new generations until the very best solution is found. GP models can flexibly search among complex patterns and evolve the best solutions in large search spaces. Figure 1.0 illustrates the main four execution steps in a genetic programming run.

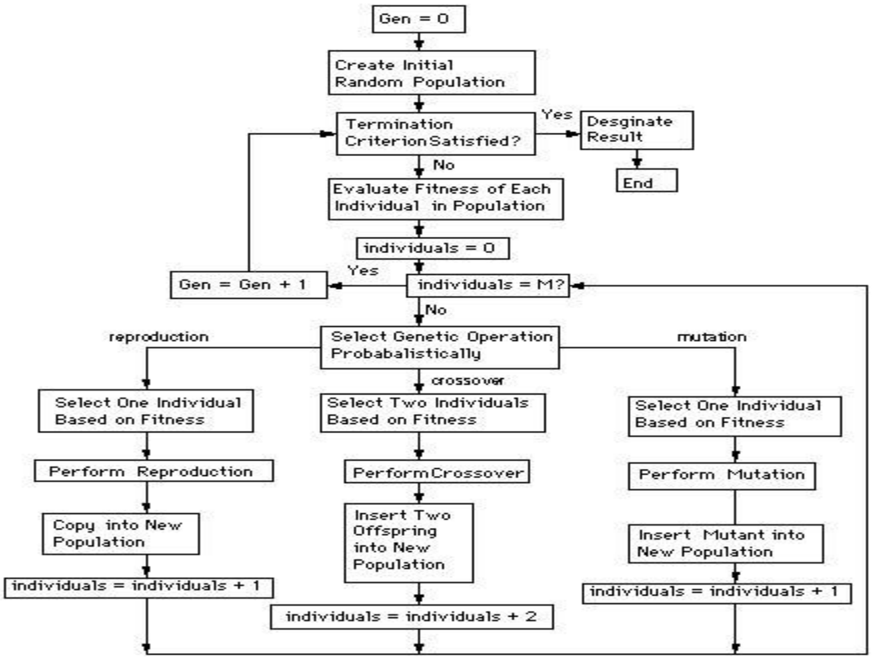

Figure 1.0 Main GP problem solving steps. Copied from http://www.geneticprogramming.com/Tutorial/index.html

The GP process begins with the random generation of a wide variety of different computer programs designed to closely match the description of the problem. Then the researcher establishes the maximum size (number of functions and terminals) of the initial individual programs grouped in a rooted point-labelled program tree. In the second step, each program in the computational population is executed and then assigned a fitness value measuring how well it solves the issue. The fitness value could potentially take many different forms such as the amount of error between an output and the desired output, program adequacy in capturing patterns, or the process of categorizing objects into classes. The third phase encompasses the creation of new population of offspring programs through crossover, mutation or reproduction. The process of fitness estimation and performing genetic operations is repeated over many generations. The crossover technique is the most important process related to GP modification where two of the fittest parents sexually combine to form two new offspring. This is usually done through the 'tournament selection' where two solutions are chosen randomly and the fittest of them will win the biological mating simulation.

Figures 2.0 and 3.0 illustrate crossover operations with both different and identical parents. The two figures suggest that crossover create offspring by deleting the crossover component from the first parent and then inserting the crossover element of the second selected parent (Koza, 1992). In fact, Figure 1.2 represents one of the main differences between GA and GP. While in GA, identical parents can generate only identical offspring, in GP identical parents are capable of generating different offspring (the main advantage of GP). In other words GP can develop two completely new solutions out of the same solution.

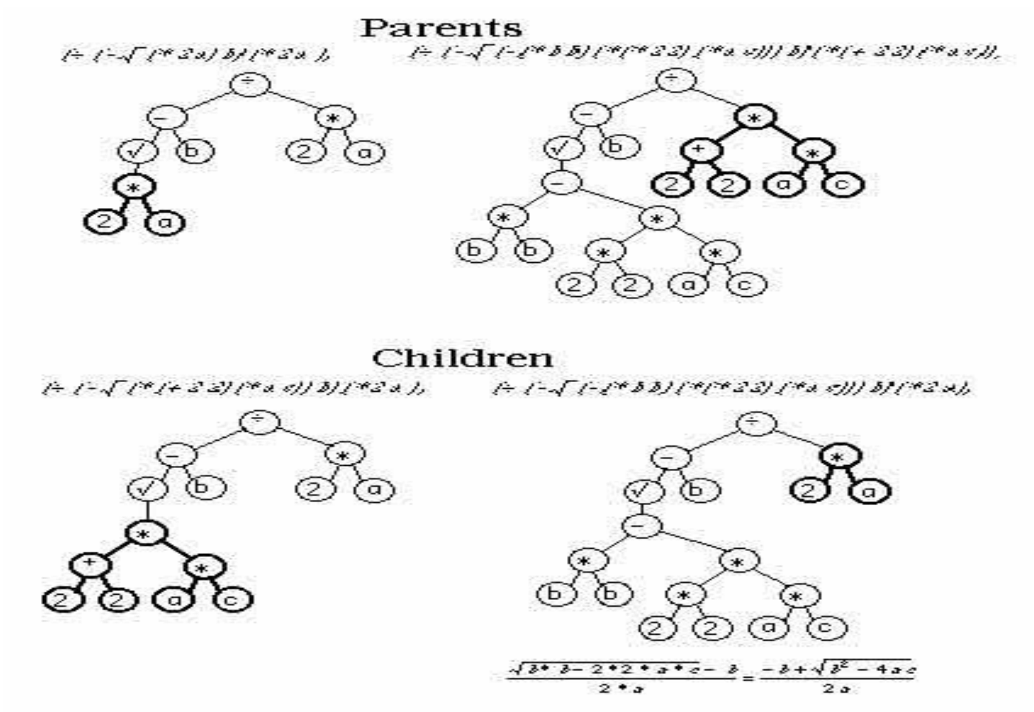

Figure 2.0 Crossover process with different parents. Copied from: http://www.geneticprogramming.com/Tutorial/index.html

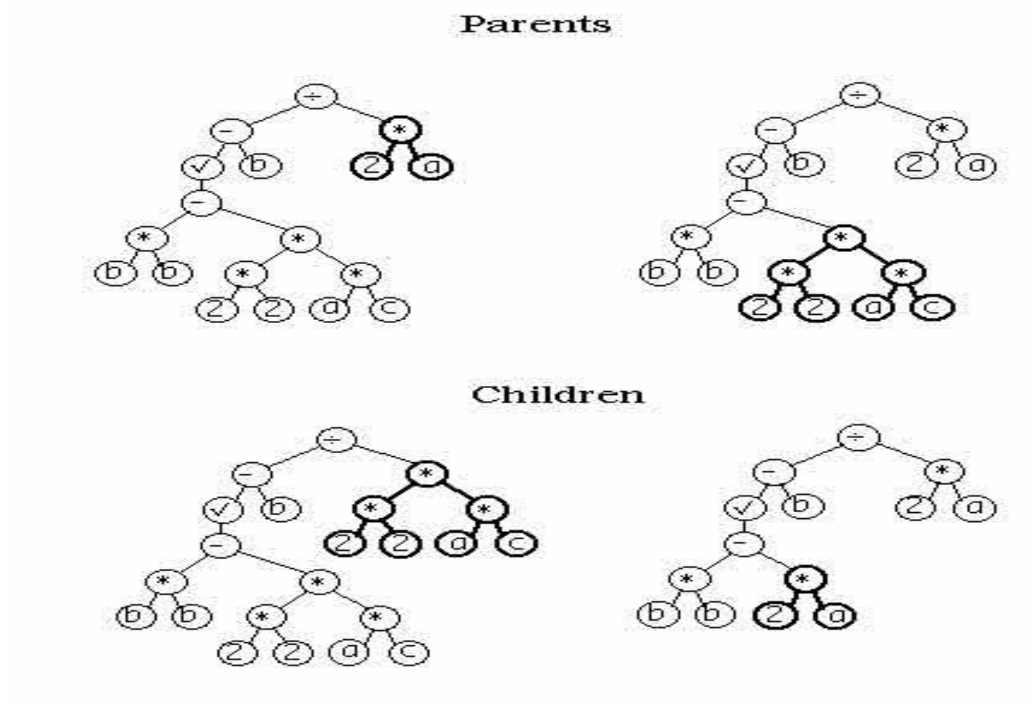

Figure 3.0 Crossover process with identical parents. Copied from: http://www.geneticprogramming.com/Tutorial/index.html

Figure 3.0 illustrate how identical parental groups can create different offspring output. The bold parts of both systems represent the subtree selections to be combined in a crossover process. The crossover process begin by random selection of a node within each tree as crossover points and the subtree rooted at the selected node in the second parent replace the subtree rooted at the selected node in the first parent to generate a child. Then the child could be used further in the process if its maximum depth is less than or equal to the maximum depth of any tree (Montana, 2002).

Mutation is another significant genetic programming process (Figure 4.0). There are two mutation categories. In the first type different single functions or terminals could be easily substituted, while in the second mutation type a whole subtree can replace another subtree. In a similar way to crossover, the mutation process begins by random selection of a node within the parent tree as the mutation point and generates a new tree of maximum mutation tree depth. The process continues by replacing the subtree rooted at the selected node with the generated tree and if the maximum depth of the newly created child is less than or equal to the maximum depth of any tree the whole process is completed.

In cases when the maximum depth of the child is greater than the maximum depth of any tree, then the researcher can use the parent or start the computing process from the beginning (Montana, 2002).

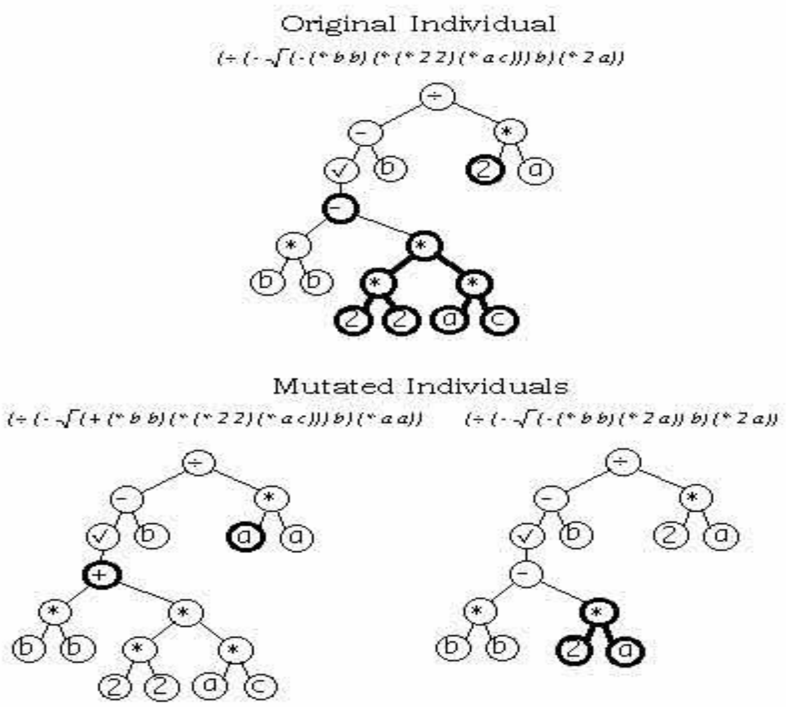

Figure 4.0 The two different mutation types. Copied from: http://www.geneticprogramming.com/Tutorial/index.html

The tree located at bottom right of Figure 4.0 represents successful replacement of two subtrees. The bottom left tree shows a mutation operation of two terminals-terminal (2) and terminal (a). The same configuration illustrates a mutation of a single function (-) for single function (+). The reproduction operation is a relatively straightforward process of copying only a single individual and implementing it into the next generation.

The GP methodology finds wider applicability in financial markets due to its adequate tree structure of the solution candidates which very closely matches the decision behaviour of real human traders. In computer simulated markets traders' parameters and the market mechanism can be precisely modelled and controlled to test a wide range of different hypotheses. The dynamic behaviour of evolved agents provides significant insights into the persistence of various types of trading behaviours.

Neely *et al.* (1997) and Neely and Weller (1999) implemented GP to search for profitable technical trading rules in the foreign exchange market. They captured economically significant out-of-sample excess returns to those rules for the selected six exchange rates. Bhattacharyya *et al.* (1998) used GP to find trading patterns and demonstrated that the acquired solutions were simpler, easier to interpret and less likely to overfit for the high-frequency data from the foreign exchange market. Chen *et al.* (1998) adopted GP to derive option pricing formulas by using real S&P 500 index options data. They compared the results generated by the GP to the traditional Black-Scholes formula and reported GP outperformance. In a similar experiment Chidambaran *et al.* (1998) applied GP to approximate the relationship between the price of a stock option, the terms of the option contract, and properties of the underlying stock price and demonstrated that the superiority of GP over the Black-Scholes model. Marney and Tarbert (2000) and Marney *et al.* (2001) developed profitable stock market trading rules through the implementation of GP.

3.1.4 Strongly Typed Genetic Programming in Finance

Strongly Typed Genetic Programming (STGP) is an enhanced version of GP which enforces data type constraints and whose use of generic functions and data types makes it a lot more powerful than GA and GP. STGP was introduced by Montana back in 1995, and unlike GP each variable and constant has an assigned type, ensuring that all generated parse trees consist of these data types. Moreover, each function has a specified type for each argument and for the value it returns. This in turn defines that functions and terminals can be implemented as bases for other functions enhancing the development of more meaningful and fitter solutions. While functions are classified as operators, such as add(...) or average(...), created to make arguments in the process of finding solutions, terminals are variables that take no arguments. This design condition represents an important difference with GP because specified data types in STGP greatly decrease the space search time and improve the generalization performance of the solutions found. The typing system in STGP keeps track of the parity of the GP trees leading to a reduced search space from all possible GP trees down to only the subspace of trees that are characterised by proper symmetry (Zumbach *et al.,* 2001).

Whilst in conventional GP, the agent's trading rules are evaluated by the same fitness function in every generation, STGP evaluate the fitness of agents through a dynamic fitness function. The dynamic nature of the fitness function enables the process of finding appropriate solutions to move forward and include the most recent experimental observations. Another important difference between the conventional GP and STGP is that whilst GP replaces the entire genetic population through crossover and mutation techniques, STGP only replaces a small proportion of the entire population at a time. This enables the population to change gradually and is an essential factor in maintaining a certain degree of model stability. (Witkam, 2013).

In GP, Boolean functions (switching functions) are represented by a real number, with $\mathbb{R} = 0$ for false assumptions, and $\mathbb{R} \neq 0$ for true statements, allowing free mixing of the types. However, when the type become more complex in cases such as vectors, matrices or lists, the process of embedding $\mathbb{R}$ is hardly possible to be completed. Hence, SPGP enabled the computation of more complex random programs by imposing specific symmetry of parity to the solution (Zumbach *et al.,* 2001).

In order to construct a parse tree researchers need to bear in mind important additional criteria beyond those required for GP. For instance, the root node of the tree returns a value of the type required by the problem and each non-root node returns a value of the type required by the parent node as an argument (Montana, 2002). While GA and GP can be developed in any programming language, the STGP is normally written in a new programming language, which is a combination of Ada (Barnes, 1982) and Lisp (Steele, 1984) programming languages. The important element taken from Ada programming language is the concept of generics as a method of developing strongly typed data. Lisp provided the concept of having programs represented by their actual parse trees (Montana, 1995).

The application of STGP to financial markets is not well documented in the literature with only a few published studies providing great scope for future research. For instance, Zumbach *et al.* (2002) applied STGP to thirteen years of high frequency data for two foreign exchange time series to discover new types of volatility forecasting models. The out-of-sample forecasting performance was compared to traditional ARCH-types models. The authors observed that STGP consistently outperformed the benchmarks indicating that the cross products of returns at different time horizons significantly improve the predictive performance. The application of STGP technique to financial markets is present in all chapters of my thesis.

3.2 Main Software Platform

3.2.1 Altreva Adaptive Modeler

Altreva Adaptive Modeler is a platform for developing agent-based stock market simulation models. The artificial stock market model is populated by up to 10,000 of traders each with their own technical trading rule. All 10,000 traders and their trading rules evolve through an adaptive genetic programming mechanism. Self-organization through the evolution of traders and the price dynamics drives the model to learn to recognise price patterns while adapting to changing stock market behaviour. The actual evolution of the model never stops and evolves with parallel with the real stock market (Witkam, 2013).

In essence, Altreva Adaptive Modeler comprises two main parts: the agent-based model which receives real historical data and generates price forecasts, and the trading system which decides when a new trading signal should be produced based on the forecasts.

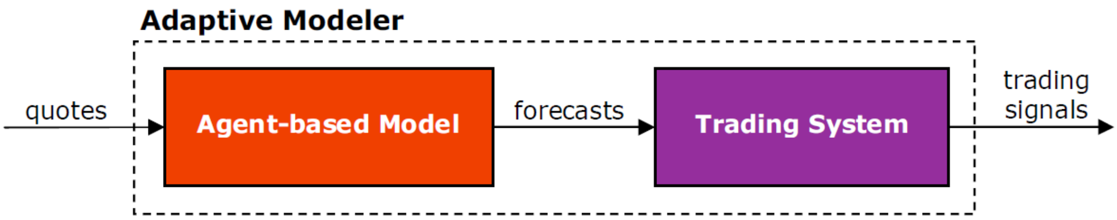

Figure 5.0 Altreva Adaptive Modeler main operational parts. Copied from: http://altreva.com/Adaptive_Modeler_Users_Guide.htm

The software typically requires a sufficient amount of historical data for its evolutionary learning process. Altreva Adaptive Modeler reads historical data from Comma Separated Values (CSV) ASCII text files in formats compatible with charting and technical analysis software packages. The software processes any data intervals ranging from 1 millisecond to multiple days.

The following processes developed within Altreva Adaptive Modeler are present in all experiments presented in Chapters 4, 5, 6, 7, 8 and 9 of my thesis.

3.2.2 Developing initial trading rules

Every agent in my experiments has only one trading rule. The software uses a special adaptive form of Strongly Typed Genetic Programming (STGP), where the genomes are the actual trading rules of agents. The very first generation of trading rules is created randomly. The crossover recombination technique (randomly chosen parts of two trading rules are exchanged in order to create two new trading rules, Figure 6.0) and mutation operation that randomly change a small part of the trading rule are applied to create later generations.

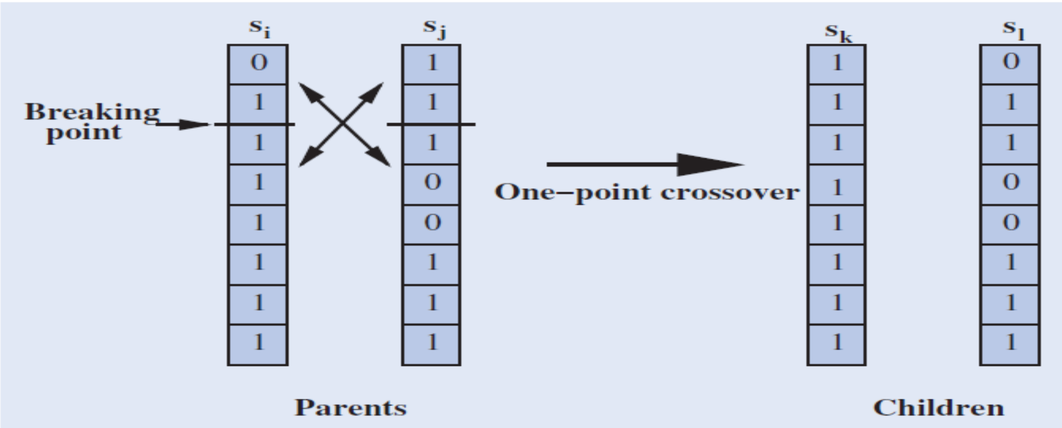

Figure 6.0 The process of genetic crossover for generating new trading strategies. Trading strategies s_i and s_j are the parents. The breaking point has been chosen randomly and then one-point crossover has been used to create children s_k and s_l . Copied from: http://www.tandfonline.com/doi/full/10.1080/14697688.2010.539249

The best performing agents from the initial selection are selected based on the breeding fitness return to be parents in the crossover process. The breeding fitness return process represents a trailing return of a wealth moving average. This is the return over the last n quotes of an exponential moving average of trader's wealth, where n could have the maximum breeding value of 250. Every pair of parents creates two offspring traders, so the number of parents and the number of offspring is equal. The newly created traders replaces poorly performing traders of the initial selection based on the replacement fitness return. The replacement fitness return represents the average return of a wealth moving average per quote since the initial trader creation. In quantitative terms, this is the cumulative return of an exponential moving average of trader's wealth, divided by trader's breeding value.

Hence, the crossover process keeps the population of traders constant. The process of mutation includes randomly changing a small part of a program in the GP tree in order to create new generations. The whole process is repeated until at least one programme in the population achieves a satisfactory fitness level (Witkam, 2013).

There is no stopping condition in my experimental models because they keep evolving as long as new prices get imported. The random nature of the initial rules is to ensure that a large variety of possible trading rules is fully investigated. To avoid the creation of trading rules that cannot be properly evaluated and to reduce the creation of meaningless trading rules, the modelling software uses a form of STGP. The STGP (Montana, 1995) involves the definition of a specific set of types that fit the problem domain. Every function and terminal is than defined to return a specific type and every function argument is defined to be of a specific type. This in turn defines which functions and terminals can be used as arguments for other functions.

The trading rules use historical price and volume data as input, and according to their internal logic, generate advice which consists of a desired position in the security as a percentage of wealth and an order limit price for purchasing or selling the security. The internal logic of the trading rules consists of the following operators:

•Price and volume data access functions;

•Average, min, max functions on historical price or volume data;

•Various logical and comparison operators;

•Some basic Technical Indicators (Witkam, 2013).

For the purpose of my experiments we consider the degree of intelligence to be proportional to the level of complexity of the trading rule. I control the complexity of trading rules by varying the maximum genome depth, which is the maximum number of hierarchical levels a genome is allowed to have.

3.2.3 Artificial stock market structure

I study individual investor behaviour within the context of the artificial stock market populated by 10,000 boundedly rational agents. All of the agents are characterised by adaptive learning behaviour represented by the genetic programming algorithm. The artificial traders all have different trading rules. Hence, the agents in the model are not orientated towards predetermined formation of strategies, and therefore are free to develop and continually evolve new trading rules. Agents' trading rules will improve by a natural selection process because the survival-of-the-fittest principle is in place.

The traders' forecasting rules are represented and evolved by STGP. The actual difference between the levels of intelligence determines the complexity of the forecasting expectations that artificial agents are able to evolve. The complexity of the forecasting expectations is measured by the depth of the STGP tree. The various genome depths affect the memory length of traders. A greater genome depth means that more complex trading rules can be developed that look further back in history. According to LeBaron (2004), traders use different amounts of past information to evaluate trading strategies and, therefore, they possess various memory lengths when evaluating forecasting rules. Additionally, markets composed of different intelligence levels offer the opportunity to analyse market efficiency in depth, rather than examining whether intelligence improves market properties where zero intelligence is present.

Artificial traders generate wealth by investing in two assets that are available in the artificial stock market- the risky stock asset and the risk free asset represented by cash. Because the models continuously evolve, the agents with trading rules that perform well become wealthier, positively influencing the forecasting accuracy of the model.

In each period, an artificial trader has two methods of keeping his wealth:

$$W_{i,t} = M_{i,t} + P_t h_{i,t} \tag{15}$$

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.

All traders get the same Initial wealth, which can be assigned to new traders either by Pareto distribution or Maxwell-Boltzmann distribution. The first method involves an unequal distribution where a large part of the total wealth is possessed by a small fraction of traders (known as ’80-20 rule’). This particular method distributes wealth to very wealthy traders but not necessarily to the remainder of the population. However, this shortcoming is compensated with the Maxwell-Boltzmann distribution, which consider the distribution of a conserved quantity such as money among components of a closed model such as agents in an economy (Dragulescu and Yakovenko, 2000 and 2001). Although money could be considered conserved in a closed economy, wealth when includes non-cash assets is not necessarily conserved. Moreover, artificial stock market models simulated in Altreva Adaptive Modeler in some cases do not represent a closed economy. However, according to Witkam (2013) these inconsistencies are rather insignificant for assigning initial wealth values to traders. The Maxwell-Boltzmann distribution has a parameter a that has a significant impact on the magnitude of values:

$$a = \frac{rms^2}{3} \tag{16}$$

Where rms is the root mean square of the values (Witkam, 2013).

3.2.4 Virtual Market clearing mechanism and order generation process

The Virtual Market is a simulated double auction call (or batch action) market where all buy and sell orders from artificial agents are collected. After having received the real-world market price (prices are imported into the model from a CSV file), agents evaluate their trading rule and place their order (if any). The Virtual Market then calculates the clearing price. The clearing price is the price at which the highest trading volume from limit orders can be matched. If the same highest trading volume can be matched at multiple price levels, then the clearing price will be the average of the lowest and the highest of those prices. The market orders have no influence on the clearing price because the clearing price calculation algorithm is designed based on commonly used mechanisms in real-life call markets.

After the clearing price has been calculated, all of the executable orders are executed for the clearing price. Therefore, the buyers and sellers automatically receive "price improvement". There is no market maker. The number of shares bought by agents is always equal to the number of shares sold by agents. In other words, prices in the artificial market are determined by the traders' orders (Figure 7.0).

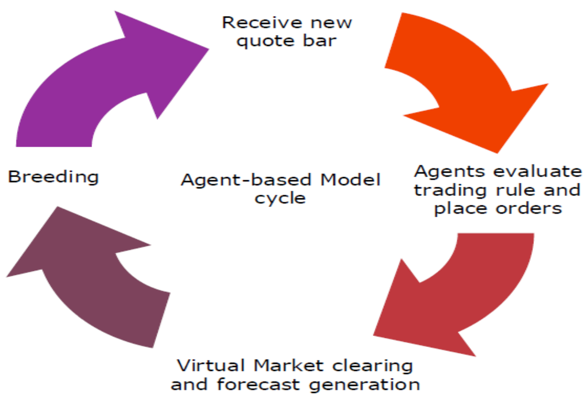

Figure 7.0 Order generation process. Copied from: http://altreva.com/Adaptive_Modeler_Users_Guide.htm

When the total number of shares offered (at or below clearing price) exceeds the total number of shares asked (at or above clearing price) or vice versa, the remaining orders will not be (fully) executed.

In this case, the orders at the clearing price will be selected for execution with priority for market orders over the limit orders and then on a first-in-first-out (FIFO) basis. The FIFO principle only applies to the priority assigned to agent orders (within one round) when not all orders can be executed if supply does not equal demand.

The orders can be partially executed. If there are no matching limit orders at all, no market orders will be executed either. In that case, the published Virtual Market price will be the Virtual Market price of the previous quote (Witkam, 2013).

The order generation process consists of the following components. The software transforms the output of an agent's trading rule into a buy or sell order through comparison of the desired position and the agent's current position. Hence, the agents do not directly determinate the number of shares to buy or sell but determinate their desired position size in the asset as a percentage of their wealth, using their trading rule.

Then the software calculates the number of shares that need to be purchased or sold based on the difference between the desired position size and agent's current position size. In the case that shares need to be purchased or sold, an order is generated to buy or sell the required amount of shares and given by the specified limit price or market order indication. Agents determinate their order limit price using their trading rule (the 'Advice' gene combines all this information into the output returned by the trading rule). Functions such as 'average', 'min' and 'max' calculate the average, minimum and maximum of the closing prices over given number of quotes on the given market. For instance, average (*l,*m) is the average closing price over quotes [*t*-*l*,*t*-1] on market *m*, where *t* is the current quote. The 'min' and 'max' functions return the lowest and respectively the highest of the closing prices over the last *n* quotes, on market *m* (where *n* and *m* are specified function parameters).

I would like to illustrate the order generation process described above with the following example-if an agent holds 1000 shares of the Coca-Cola Company, priced at 38.50 and 80,000 in cash his wealth will be 118,500 and its position is 32.5%. Its trading rule generates an advice of a position of 50% and a limit price of 38.50. Then a limit order will be produced to purchase 539 (=50%*118,500/38.50-1000) additional Coca-Cola shares with a price of 38.50.

Chapter 4

Artificial stock market dynamics and market efficiency: An econometric perspective.

4.1 Introduction

A few decades ago, the Efficient Market Hypothesis (EMH) was widely admired by academic financial economists. The EMH postulates that market prices should reflect all available information. As a consequence, market prices should always be consistent with their fundamental values. The hypothesis was independently developed by Samuelson (1965) and Fama (1963, 1965a, 1965b, 1970). Samuelson (1965) generated a series of non-linear programming solutions to spatial pricing models with no uncertainty, and proposed that in informationally efficient markets price changes are unpredictable if market prices fully incorporate the information disseminated from all market participants. Since then, the concept of efficient markets has been applied to many theoretical models and empirical studies of asset prices, generating several controversial debates.

Advocates of the EMH, such as Jensen (1978), have argued that there is no other proposition in finance which has more solid supporting empirical evidence than the EMH. In the late 1970s, the EMH began its transition from theory to doctrine. Thompson (1978), Galai (1978), Charest (1978a), Davidson and Froyen (1982), and in the early 1990s, Nichols (1993), Conrad (1995), and Shanken and Smith (1996) provided evidence that supported the EMH. Malkiel (2003) suggested that stock markets are more efficient and rather less predictable than many academic studies would have us believe.

By the beginning of the twenty-first century, the academic dominance of the EMH has become less prominent. A group of researchers equipped with anomalous evidence inconsistent with the EMH, suggested that the EMH should be replaced by a behavioural finance approach (Thaler, 1993; Haugen, 1999; Schleifer, 2000; Shiller, 2003). It has been observed that financial markets do not process information instantaneously (Chan *et al.,* 1996), and that markets can overreact as a result of investor optimism or pessimism (Dissanaike, 1997).

Furthermore, many empirical studies support the fact that markets are predictable and technical trading strategies generate significant profits (Brock *et al.,* 1992; Jeegadesh and Titman, 2001; Chiarella and He, 2002).

More recently a number of economists have begun to question the accuracy of the empirical results associated with the EMH by using more sophisticated data sets and greater computing power. The concepts of heterogeneity, bounded rationality, and evolutionary adaptive agents have been explored by Brock and Hommes (1997, 1998), Chiarella and He (2002a, 2002b), Gaunersdorfer and Hommes (2000), and Hommes (2001). A few years later, in an attempt to accommodate most of the complexities of the real world, Lo (2004, 2005) proposed the Adaptive Market Hypothesis (AMH). This hypothesis modifies the EMH paradigm to suggest that the forces that drive prices to their efficient levels are weaker and the processes of learning and competition and evolutionary selection pressures govern these forces.

The controversy intensified and researchers began to question whether the EMH could ever be validated or discredited (Langevoort, 1992). Roll (1977) went even further suggesting that EMH and CAPM are joint hypotheses. Nevertheless, the majority of the studies related to market efficiency and adaptability have one major shortcoming. They failed to investigate the relation between market diversity and market efficiency, and the impact of individual learning and adaptability on the diversity of traders' expectations. Chen and Yeh (2001) expressed the view that the market size could potentially have a dramatic impact on market efficiency. The questions this study is trying to answer are whether market organization influence traders' strategies and in turn market efficiency and whether market structure affect individual learning in the AMH?

In this study, I developed ten stock markets- each populated by different numbers of artificial traders- for each of the FTSE 100, S&P 500 and Russell 3000 indices. I also implemented a special adaptive learning form of Genetic Programming (GP), called Strongly Typed Genetic Programming (STGP), in order to investigate the relationship between market efficiency and adaptability. The reason for using STGP is because Lo (2004, 2005) regarded the market as an evolutionary process where the principles of evolution- such as competition, adaptation, and natural selection- are applicable to financial markets.

Hence, the artificial traders in my experiment can be considered to be agents that adapt, learn, evolve, and try to survive. The random nature of the initial trading rules of the agents allowed me to observe how they learn, adapt, and survive (the worst performing traders were replaced). The scientific advantage of the STGP over the conventional Genetic Programming (GP) used in most studies so far is that STGP evaluates the fitness of agents through a dynamic fitness function which processes the most recent quotes of the three indices in my experiment, rather than a re-execution of the same trading rules.

I then empirically evaluated the price series of these three indices to investigate the relationship between markets populated by different numbers of heterogeneous agents with different dynamics and the validity of the Efficient Market Hypothesis (EMH) and the Adaptive Market Hypothesis (AMH). I also explored the dynamic behaviour of the models when testing for the presence of nonlinearity.

Despite the voluminous literature on the topic, no other study has implemented the STGP technique and 10,000 artificial agents, which has enabled me to develop of a wider variety of trading rules. My financial markets can, therefore, be viewed as co-evolving ecologies of different trading strategies. These strategies are analogous to a biological species, and the amount of funds deployed by traders following a given strategy is analogous to the population of that species (Farmer and Lo, 1999). The presence of 10,000 heterogeneous and interacting adaptive traders, rich in dynamics, provides the opportunity to study the stock market as a complex adaptive system. Artificial traders are, by definition, capable of adapting, learning, and evolving, which makes them extremely suitable for the analysis of market efficiency and adaptability, because adaptation and learning in heterogeneous structures are known as important tools for analysing financial market behaviour (Hommes, 2001). Hommes (2011) argued that heterogeneity is a critical aspect of the theory of expectations, because a model of heterogeneous expectations can explain different aggregate outcomes across different market settings.

To summarise, the contributions of this chapter are as follows. Firstly, I am the first to apply the STGP technique in the analysis of market efficiency and adaptability, whilst taking into account different market structures and individual trader cognitive abilities and heterogeneity. Recent studies, such as Urquhart and Hudson (2013), suggest that the AMH better describes the behaviour of stock returns than the EMH.

However, the authors based their conclusions entirely on econometric tests, and failed to observe the processes of adaptation, learning, competition, and evolutionary selection pressures that govern the AMH. My study aimed to fill this gap by providing significant empirical findings combined with evidence gained from evolutionary dynamic processes. Secondly, since their creation, the EMH and the AMH have not been formalised appropriately. I hope that the solid empirical evidence that I present could shed light on the formation of stock market dynamics and the formalisation of both hypotheses within artificial laboratory stock market settings. Thirdly, I have found that different trader populations behave as an efficient adaptive system. I observed that market efficiency is not necessarily associated with rational assumptions and that nonlinear dependence in index returns evolve over time. Hence, I think that market efficiency is not a static characteristic as assumed in most of the studies published so far. My findings are consistent with the perception of financial markets as adaptive systems subject to evolutionary selection pressures.

4.2 Background

4.2.1. Origins and supporting empirical evidence on the EMH

More than a century ago, Bachelier (1900) analysed the mathematical theory of random processes and expressed the view that stock price movements follow a Brownian motion and that, therefore, stock prices are unpredictable. Several years later, Samuelson (1965) generated a series of non-linear programming solutions to spatial pricing models with no uncertainty, and proposed that the price changes in informationally efficient markets are unpredictable if market prices fully incorporate the information disseminated from all market participants.

In four different seminal papers, Fama (1963, 1965a, 1965b, 1970) measured the statistical properties of market prices and operationalised the EMH by allocating structure on various information sets available to market participants.

Fama (1970) reviewed the empirical evidence gained in the 1960s and proposed three major versions: (a) weak form tests of the efficient market models, (b) tests of martingale models of the semi-strong form, and (c) strong form tests of the efficient markets.

During the early years of development, the EMH gained massive academic attention. Jensen (1978) stated that 'there is no other proposition in economics which has more solid empirical evidence supporting it than the EMH'. Thompson (1978), Galai (1978), Charest (1978a), and Davidson and Froyen (1982) provided early empirical evidence in support of the EMH. Studies by Nichols (1993), Conrad (1995), and Shanken and Smith (1996) generated further support for market efficiency. Malkiel (2003) acknowledged that market participants are less rational and that predictable patterns in stock returns can appear for short periods of time, but that stock markets are more efficient and less predictable than many recent research papers demonstrate.

Chen *et al.* (1997) approached the EMH through the application of Genetic Programming (GP). The authors tested a short-term sample of TAIEX (Taiwan index) and the S&P 500 and concluded that the EMH is sustained, although they also confirmed the existence of short-term nonlinear regularities. However, these nonlinear regularities could not be exploited further to a profitable level due to the very high search costs involved in the process of discovering them.

4.2.2. Challenging empirical evidence on the EMH

Studies in the last three decades suggest a rejection of the EMH (Lehman, 1990; Jegadeesh, 1990; Hsieh, 1991; Richardson and Smith, 1993). Lo and MacKinlay (1988) rejected the random walk model for weekly stock market returns and suggest that this rejection does not necessarily mean the inefficiency of stock-price formation.

Timmermann and Granger (2004) that even if the EMH is correct, stock prices can still be predicted. The authors argue that the EMH does not rule out stock market predictability if the behaviour of investors results in efficient markets by their constant profit-seeking attitude.

Timmermann and Granger (2004) proposed a forecasting procedure that could work even if the EMH is correct. The EMH suggest the absence of arbitrage opportunities but it does not eliminate all forms of predictability in financial returns. Under conditions of no arbitrage, the current price of any financial instrument, P_t is represented by the conditional expectation of the financial instrument's payoffs- consisting of its future price, P_{t+1} and any dividends, D_{t+1} - multiplied by a variable known as 'pricing kernel', Q_{t+1}, that comprises variations in economic risk premia:

$$P_t = E\left[Q_{t+1}\left(P_{t+1} + D_{t+1} \right)\Omega_t \right] \quad (17)$$

Where $E\left[. / \Omega_t \right]$ is the population expectation, dependent on the information set Ω_t. The EMH translates into a simple moment condition under a set of restrictive assumptions (Harrison and Krebs, 1979). As most stock prices are trended, tests for predictability eliminate such trends by considering the excess rate of return, R_{t+1}, defined as the return, $\left(P_{t+1} + D_{t+1} - P_t \right) / P_t$, above the risk-free rate (for instance the return on T-bills used in my experiments), $r_{f,t}$. Dividing Equation (17) by P_t and subtracting $r_{f,t}$, we get:

$$E\left[Q_{t+1} R_{t+1} / \Omega_t \right] = 0 \quad (18)$$

Timmermann and Granger (2004) pointed out that since the process generating risk-premium is model-dependent and is not observable, testing the EMH can only be performed jointly with auxiliary hypotheses about Q_{t+1}. This could be done by rearranging Equation (18) to get:

$$E\left[R_{t+1} / \Omega_t \right] = \frac{-Cov\left(R_{t+1}, Q_{t+1} / \Omega_t \right)}{E\left[Q_{t+1} / \Omega_t \right]} \quad (19)$$

Timmermann and Granger (2004) argue that predictability of returns thus does not violate the EMH. Forecasting models are effective because they predict the conditional variance of returns with the pricing kernel, Q_{t+1}, scaled by its conditional mean. In fact the EMH does not take into account how the information variables in the information set, Ω_t are used to generate actual forecasts.

The authors went even further suggesting that market efficiency should not be associated with the random walk model for stock prices because stock prices plus cumulated dividends discounted at the risk free rate should follow a martingale process under the risk-neutral or equivalent martingale probability measure. In its strictest form the random walk hypothesis states that price increments are IID (identically and independently distributed), then this hypothesis is rejected by the presence of conditional heteroskedasticity in returns regardless any risk premium effects.

Grossman and Stiglitz (1980) argued that if financial markets were efficient, there would not be any profit generated through information gathering, therefore, there would be no reason to trade and the markets would eventually collapse. Thus, there must be profit-making-opportunities to compensate investors for the cost of acquiring information and trading.

The 'behaviour finance' group of researchers equipped with anomalous empirical evidence against the validity of the EMH, challenged the advocates of market efficiency (Thaler, 1993; Haugen, 1999; Schleifer, 2000; Koonce, 2001; Shiller, 2003). Chan *et al.* (1996) investigated whether the predictability of future returns from past returns is based on the market's under reaction to information associated with past earning news. The authors found insignificant evidence of subsequent reversals in the returns of stocks with high price and earnings momentum, suggesting that financial markets respond gradually to new information. Studies by De Bondt and Thaler (1985, 1987); Kahneman and Tversky (1982); Arrow (1982); and Dissanaike (1997) demonstrate that investors do not behave in a rational way because they are overly influenced by current information and pay little attention to past information. Hence, investors tend to overreact, building the foundations of the stock market Overreaction Hypothesis (ORH). The ORH, postulates that if stock prices systematically overshoot as a result of excessive investor optimism or pessimism, price reversals should be predictable from past price performances (Dissanaike, 1997). Brock *et al.* (1992) implemented two of the simplest trading rules- moving average and trading rule break- to analyse the Dow Jones movements from 1897 to 1986, and reported a rejection of the EMH in favour of strong support for technical trading strategies. Jegadeesh and Titman (1993, 2001) provided evidence of substantial momentum trading profits that were not a product of data snooping bias, confirming the assumptions behind the behavioural models.

4.2.3. Ongoing debate of the EMH and the emergence of the AMH

Sargent (1993) used the notion of bounded rationality as opposed to perfect rationality, to describe how traders with limited information about fundamental values develop expectation price models. Traders are not irrational, but -considering the amount of limited information they possess- they tend to adapt to optimal beliefs and act in a rational way. The endogenous nature of uncertainty of the state of the world does not allow traders to develop life-time optimisation strategies in favour of more simple reasoning and rules of thumb (Shefrin, 2000). Brock and Hommes (1997a, 1998) demonstrated that evolutionary adaptive systems with many heterogeneous agents, which implemented various trading strategies, represent a nonlinear system capable of generating a wide variety of stylised facts. The authors observed an evolutionary competition between trading strategies where traders implemented their strategies according to an evolutionary fitness measure, such as accumulated past profits. Hence, Brock and Hommes proposed to model financial markets as Adaptive Belief Systems populated by boundedly rational traders.

Hommes (2000) studied the financial markets through the concept of evolutionary systems with different competing trading strategies. All traders involved in the experiment were boundedly rational, in the sense that they were capable of following well performed strategies according to wealth accumulated in the past. The author showed how simple technical trading rules exist and survive evolutionary competition in an entirely heterogeneous environment, where prices and beliefs co-evolve over time. The evolutionary model successfully replicated and described the formulation of various stylised facts.

Chen and Yeh (2002) investigated the emergent properties of artificial stock markets in the light of the EMH and the Rational Expectations Hypothesis (REH). The authors inquired whether the macro-behaviour depicted by the two hypotheses was consistent with the behaviour of the micro-level. A conjecture based on a sunspot-like signal indicated that macro-behaviour can be very different from micro-behaviour and the aggregate results cannot be regarded as a simple scaling-up of individual behaviour.

Kaizoji *et al.* (2002) implemented a spin model in the context of a stock market with fundamentalists and interacting heterogeneous traders to investigate stock market dynamics. The authors demonstrated that magnetisation in the spin model is associated with the actual trading volume in the stock market, and, most importantly, that the market price is determined by magnetisation under natural assumptions.

The Adaptive Market Hypothesis (AMH) was proposed by Lo (2004) and can be regarded as a new version of the EMH, based on revolutionary principles. Lo argued that market prices reflect as much information as required by the mixture of environmental factors and the number of distinct groups of market participants. The AMH postulates that market efficiency is not an isolated process, but it is a very dynamic and context-dependent process where market participants adapt to a changing environment and the processes of learning and competition, as well as the evolutionary selection pressures that govern the AMH. The agents are not perfectly rational, but rather they are boundedly rational satisfiers that operate in ecological systems competing for scarce resources. The ecological systems exhibit cycles in which competition for resources depletes trading opportunities, but completely new opportunities appear later on in the process.

Lo (2005) extended his work further and highlighted that traders act in their own-self-interest, but they also make mistakes. However, they tend to learn from their mistakes and adapt to changing market conditions. Competition drives adaptation and innovation and evolution determinates market dynamics.

Lim (2007) analysed eleven emerging and two developed markets through the portmanteau bicorrelation test and concluded that market efficiency evolves over time. Their rolling sample framework was able to capture periods of efficiency and inefficiency by comparing the time windows that these markets generate significant nonlinear serial dependence. It appeared that the U.S. market was the most efficient and the Argentine market the most inefficient.

Potters *et al.* (2008) studied in great detail the market prices of options on liquid markets, where the market corrected the inadequate Black-Scholes formula to investigate two statistical aspects of asset fluctuations: volatility clustering and correlations in the scale of fluctuations.

These two aspects were not initially included in the pricing models but later appeared in the price fixed by the market as a whole. Hence, the authors concluded that financial markets behave as rather efficient adaptive systems.

Neely *et al.* (2009) used daily exchange rate data from the Federal Reserve H.10 Statistical Release and concluded that financial markets deviate substantially from the EMH and they are adaptive systems based on evolutionary selection pressures. More complex trading strategies in their experiment survived longer than simple strategies, suggesting that financial markets function as adaptive systems.

Kim (2009) developed a monetary model based on the incomplete knowledge of market participants governed by adaptive learning rules, which allowed agents to learn about the economic environment. Simulation results demonstrated that the model under adaptive learning dominates the market and explained why fundamentals predict exchange-rate returns over long horizons but not over short horizons.

Benink *et al.* (2010) used artificial financial markets to study market efficiency and learning in the context of the Neo-Austrian economic paradigm. The authors demonstrated that markets are more efficient when informational advantages are small and the learning of traders leads to a more informationally efficient market but also a less efficient market in terms of excess returns.

Urquhart and Hudson (2013) empirically investigated the US, the UK and the Japanese markets using long run historical data and concluded that the AMH provides a better description of the behaviour of stock returns than the EMH. Their empirical results suggested that each of the three markets showed evidence of being an adaptive market, with returns going through periods of independence and dependence.

4.3 Artificial stock market structure for this particular experiment

Table 2.0 represents the main technical settings of my artificial stock markets. I developed ten different markets (denoted A-J) populated by different number of traders for each of the three financial instruments. Market A is populated by 1,000 traders; Market B has 2,000 traders; Market C has 3,000 traders; Market D has 4,000 traders; Market E has 5,000 traders; Market F has 6,000 traders; Market G has 7,000 traders; Market H has 8,000 traders; Market I has 9,000 traders and Market J is populated by 10,000 artificial traders.

Artificial stock market parameters	
Total population size (agents) per market	10,0000 in Market J; 9,000 in Market I; 8,000 in Market H; 7,000 in Market G; 6,000 in Market F; 5,000 in Market E; 4,000 in Market D, 3,000 in Market C; 2,000 in Market B; 1,000 in Market A.
'Best Performing Traders' size(percentage of the total population)	5%
Initial wealth(equal for all agents)	100,000
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 (bars)	1
Minimum breeding age (bars)	80
Initial selection type	random
Parent selection (percentage of initial selection that will breed)	5%
Mutation probability (per offspring)	10%
Total number of quotes processed- FTSE 100	7,262
Total number of quotes processed-S&P 500	7,262
Total number of quotes processed-Russell 3000	7,262
Seed generation from clock	Yes
Creation of unique genomes	Yes
Offspring will replace the worst performing agents of the initial selection	Yes

Table 2.0 Artificial Stock Market Parameter Settings

*Maximum genome size measure the total number of nodes in an agent'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 an agent's genome (trading rule). The depth of a trading rule can be an indicator of its complexity.

4.4 Simulation Results

4.4.1. Is a financial market populated by more heterogeneous adaptive traders efficient?

Throughout the years, the EMH was mainly formalised based on the concept of probabilistic independence in probability theory. Malkeil (1987) quantified the notion of efficient markets by considering the rate of return R_t,a random function defined in the L^2 probabilistic Hilbert space, as well as Ω_{t-1},the σ-algebra produced by the history of rate of return $\left\{R_j\right\}_{j=-\infty}^{t-1}$.

Hence, the EMH states that R_t is independent of any random variables in Ω_{t-1}. Moreover, when taking into account the conditional expectation $E\left\langle R_t \middle| \Omega_{t-1}\right\rangle$, the EMH then implies the following:

$$E\left\langle R_t \middle| \Omega_{t-1}\right\rangle = 0 \quad (20)$$

Because Equation (20) is a result of the random walk in a discrete-time stochastic process, the EMH is associated with the random walk process. Recent research based on nonlinear tests indicates the existence of nonlinear dependence in stock market data. For instance, the studies of Brock *et al.* (1987), Frank *et al.* (1988), Savit (1988, 1989), Hsieh (1989), Scheinkman and LeBaron (1989), Peters (1991) and Willey (1992) demonstrate the presence of nonlinear dependence between R_t and Ω_{t-1},or that $\left\{R_i\right\}_{i=-\infty}^{t}$ represents a chaotic time series which seems to be characterised with random behaviour, but it is in fact deterministic. Moreover, Farmer and Lo (1999) argue that to make the EMH operational, researchers have to specify trader cognitive abilities, information structure, risk preferences, etc. But then the EMH will consist of a test of several auxiliary hypotheses and any potential rejection of such a joint hypothesis could provide misleading information which part of the joint hypothesis is inconsistent with the dataset. Is there any other hypothesis that better represent market efficiency taking into account market complexity and trader behaviour?

To analyse the implications of market size on market efficiency and adaptability, I consider experiments associated with ten different market sizes (number of traders). Lo (2004) argue that measuring the level of market efficiency of a particular market should be seen as relative to other markets. In view of this claim, I investigated the outcome of various tests for nonlinearity and compare market efficiency between ten different markets.

The STGP modelling approach represents an appropriate tool for examining the stock market mechanism in isolation from the traders who populate the artificial stock market. An important addition is that I am able to investigate the relationship between the market dynamics and trading activities and therefore to analyse the efficiency of the stock market in terms of the EMH and the AMH. The heterogeneous environment where stock prices and traders' beliefs co-evolve over time provides an appropriate laboratory platform to investigate market efficiency and the emergent behaviour of the stylised facts of financial returns. Also the heterogeneity of expectations among artificial traders provides important nonlinear conditions for the market.

First, I would like to investigate whether my artificial stock markets for the three financial instruments are efficient in the sense that the stock returns are statistically independent. In order to test for statistical independence, I adopted the procedure of Chen *et al.*(2000) which consists of the Rissanen's predictive stochastic complexity (PSC) filtering (Rissanen, 1989), followed by the celebrated BDS testing proposed by Brock *et al.*(1996). In econometric terms, a stock market is efficient when its return series are unpredictable, or in other words series are identically and independently distributed (IID). Similarly to 4.3.3 and 5.4.2, I begin my econometric analysis by applying the Augmented Dickey-Fuller (ADF) test to detect the presence of a unit root. My ADF testing procedure includes running a regression of the first difference of the log price series against series that have been lagged once and then combined with a drift and time trend. The null hypothesis of a unit root presence was rejected in all three index price series in all different market levels (Tables 3.0, 4.0 and 5.0). Hence, the return series generated by all artificial traders in all markets are stationary at the 95% significance level. This finding is consistent with Lee *et al.* (2010) who reported stationary price series in 32 developed and 26 developing countries. I then applied the Rissanen's PSC criterion to the return series generated by the STGP mechanism to identify the linear ARMA model by selecting the model with minimum PSC.

If some of the three financial instruments at any market level satisfies the EMH, both *p* and *q* of ARMA should equal zero. Hence, there will not be any linear dependence and the return series are not linearly predictable. The seventh column of Tables 3.0, 4.0 and 5.0 shows the ARMA process extracted from the return series. All FTSE 100, S&P 500 and Russell 3000 return series in Markets A, B, C, D, E and F are linearly dependent and therefore inefficient.

FTSE 100									
Market	***Std.dev***	***SK***	***KU***	***JB***	***ADF****	***PSC***	***BDS***	***GARCH***	***Kaplan***
A	1656.6	-0.1717	1.6546	575.35	-37.99	(1,0)	0.13	(0,1)	R
B	1656.3	-0.1717	1.6545	575.30	-37.12	(1,0)	0.43	(0,1)	R
C	1637.1	-0.1706	1.6221	570.05	-37.89	(0,1)	0.29	(0,1)	R
D	1600.8	-0.1701	1.5651	569.23	-37.16	(1,0)	0.71	(1,2)	R
E	1489.4	-0.1687	1.4769	568.11	-38.23	(1,0)	1.48	(1,0)	R
F	1204.7	-0.1611	1.4476	567.58	-35.68	(0,1)	1.72	(0,1)	R
G	1201.2	-0.1565	1.4290	566.47	-35.54	(0,0)	0.38**	***	R
H	1011.3	-0.1522	1.3618	565.78	-35.46	(0,0)	0.30**	***	R
I	800.87	-0.1489	1.2767	560.03	-35.17	(0,0)	0.27**	***	A
J	781.89	-0.1119	1.2001	559.32	-34.94	(0,0)	0.20**	***	A

SK-skewness; ***KU***-kurtosis; ***JB***-the Jarque-Bera test; the numbers in brackets (p,q) in the column PSC are the orders of the ARMA (p,q) selected by the PSC criterion.

*The MacKinnon (1996) one-sided critical value for rejection of the Null hypothesis of a unit root at 5% level is -3.410060.

**Failed to reject the Null hypothesis that series are identically and independently distributed (IID).

***No presence of ARCH effect.

Table 3.0 Econometric statistics for FTSE 100 price series generated by a different number of traders in various artificial stock markets.

S&P 500									
Market	***Std.dev***	***SK***	***KU***	***JB***	***ADF****	***PSC***	***BDS***	***GARCH***	***Kaplan***
A	436.61	-0.0732	1.4571	715.57	-64.42	(0,1)	0.42	(0,1)	R
B	436.57	-0.0730	1.4569	715.50	-64.88	(0,1)	0.85	(0,1)	R
C	436.20	-0.0727	1.4560	714.23	-64.35	(1,0)	1.21	(1,0)	R
D	411.01	-0.0699	1.4555	714.11	-64.10	(1,0)	1.38	(1,2)	R
E	390.77	-0.0690	1.4490	713.78	-63.37	(1,0)	0.90	(1,2)	R
F	387.32	-0.0687	1.4474	713.32	-63.19	(1,0)	0.36	(0,1)	R
G	380.93	-0.0680	1.4470	713.27	-63.67	(0,0)	1.37	***	R
H	373.56	-0.0674	1.3329	713.01	-62.46	(0,0)	0.35**	(0,1)	R
I	370.31	-0.0645	1.3012	712.23	-62.33	(0,0)	0.31**	***	A
J	368.89	-0.0631	1.2991	712.01	-62.01	(0,0)	0.24**	***	A

SK-skewness; ***KU***-kurtosis; ***JB***-the Jarque-Bera test; the numbers in brackets (p,q) in the column PSC are the orders of the ARMA (p,q) selected by the PSC criterion.

*The MacKinnon (1996) one-sided critical value for rejection of the Null hypothesis of a unit root at 5% level is -3.410060.

**Failed to reject the Null hypothesis that series are identically and independently distributed (IID).

***No presence of ARCH effect.

Table 4.0 Econometric statistics for S&P 500 price series generated by a different number of traders in various artificial stock markets.

Russell 3000									
Market	***Std.dev***	***SK***	***KU***	***JB***	***ADF****	***PSC***	***BDS***	***GARCH***	***Kaplan***
A	458.59	-0.0453	1.4742	720.21	-90.77	(1,0)	0.20	(0,1)	R
B	458.58	-0.0452	1.4740	720.20	-90.73	(1,0)	0.30	(1,0)	R
C	458.00	-0.0443	1.4738	719.98	-90.12	(1,0)	1.38	(1,0)	R
D	446.21	-0.0411	1.4732	719.34	-90.56	(0,1)	1.36	(0,1)	R
E	433.78	-0.0410	1.4727	719.21	-89.63	(1,0)	0.34	(1,1)	R
F	430.25	-0.0390	1.4001	719.10	-89.39	(1,0)	0.99	(0,1)	R
G	428.11	-0.0378	1.3998	718.88	-89.12	(0,0)	0.12	(1,1)	R
H	420.20	-0.0372	1.3991	718.24	-88.85	(0,0)	0.32**	***	R
I	415.11	-0.0321	1.3983	718.11	-88.50	(0,0)	1.43**	***	A
J	404.80	-0.0299	1.3930	718.01	-88.11	(0,0)	0.56**	***	A

SK-skewness; ***KU***-kurtosis; ***JB***-the Jarque-Bera test; the numbers in brackets (p,q) in the column PSC are the orders of the ARMA (p,q) selected by the PSC criterion.

*The MacKinnon (1996) one-sided critical value for rejection of the Null hypothesis of a unit root at 5% level is -3.410060.

**Failed to reject the Null hypothesis that series are identically and independently distributed (IID).

***No presence of ARCH effect.

Table 5.0 Econometric statistics for Russell 3000 price series generated by a different number of traders in various artificial stock markets.

The return series generated by artificial traders in Markets G, H, I and J are linearly independent (p=0, q=0). Lack of linearity in these four markets suggest an important initial finding that artificial stock markets populated by 7,000 8,000, 9,000 and 10,000 artificial traders are so efficient that there are no linear signals found.

I estimated the most appropriate ARMA (p,q) model and fitted it to the data set in order to discard all linearity from the sample. Any signal left in the residual series must be non-linear. I used the BDS test to investigate for nonlinearity. The BDS test detects significant deviations in the correlation of integral behaviour from that anticipated under the IID of the dataset. The correlation integral is quantified by:

$$C_{\varepsilon,m} = \sum_{1\leq s} \sum_{\leq t\leq n} \frac{\mathrm{I}_e\left(y_t{}^m, y_s{}^m\right)}{\begin{pmatrix} n \\ 2 \end{pmatrix}} \tag{21}$$

Where $y_t^{\ m} = (y_t, y_{t+\tau}, ..., y_{t+(\tau-1)m})$ an 'm-history' computed from the underlying uni-variate dataset and $\mathrm{I}_\varepsilon(\ \ )$ an indicator function: $\mathrm{I}_\varepsilon\left(y_t^{\ m}, y_s^{\ m}\right) = 1 if \left\| y_t^{\ m} - y_s^{\ m} \right\| < \varepsilon$ and zero otherwise. The correlation integral establishes the frequency and connectivity with which different points are within radius ε of each other. Here m represents the embedding dimension within which lag τ has been implemented in the computing of 'm-history' to prevent the formation of a very high correlation between the elements of an m-tuple (Chen *et al.* 2000).

If the returns generated by the artificial traders in my experiments are identically and independently distributed, then the correlation function (Equation 21) suggests that $\lim_{x\to\infty} C_{\varepsilon,m} = \left(\lim_{x\to\infty} C_{e,1}\right)^m$ for sure for all $\varepsilon > 0$ and m =2,3,.....The test statistic, with limiting standard normal distribution under the null hypothesis was proposed by Brock *et al.* (1996):

$$V_{\varepsilon,m} = \sqrt{n}\left(C_{\varepsilon,m} - C_{\varepsilon,1}^{\ \ m}\right) / \sigma_{\varepsilon,m} \qquad (22)$$

Brock *et al.* (1996) offered an estimation process for the standard deviation $\sigma_{\varepsilon,m}$.

I applied the BDS test directly to the data generated by the artificial traders in Markets G, H, I and J because there are no linear signals detected in these series. The empirical results of the test are reported in the eighth column of Tables 3.0, 4.0 and 5.0.

The null hypothesis of IID is significantly rejected in Markets A, B, C, D, E and F indicating nonlinear dependence in the return series. In Markets G, H, I and J the null hypothesis has not been rejected, suggesting that series generated by the artificial agents are identically and independently distributed. In terms of the BDS test, Markets G, H, I and J are nonlinearly dependent, more random and therefore more efficient than markets populated with fewer traders. This finding can be considered to be a match of the classical version of the EMH.

According to the econometric literature, however, a large part of the nonlinearity in data is in their second moment. I performed the Lagrange multiplier (LM) test with up to 14 lags to detect the presence of an ARCH effect of the residual. In case the null hypothesis is rejected, I further identified the GARCH order of the series according to the Schwartz Information Criterion. The ninth column of Tables 3.0, 4.0 and 5.0 represents the results. All markets with reduced numbers of artificial traders reported the presence of an ARCH effect. I proceeded to further identify the GARCH order based on the Swartz Information Criterion. This process can be expressed in quantitative terms by:

$$r_t = \mu + h_t^{1/2}\varepsilon_t, \qquad h_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i x^2_{t-i} + \sum_{i=1}^{q}\beta_i h_{t-i} \tag{23}$$

Where ε_t is the IID normal innovations and the restrictions are $\alpha_0 > 0$, $\alpha_i, \beta_i \geq 0$ and $\sum_i \alpha_i + \sum_i \beta_i < 1$,(Chen *et al.*2000).

While Market H of the S&P 500 and Market G of the Russell 3000 consisted of ARCH effects, the most populated markets- I and J- in all experiments of the three indices did not consist of any ARCH effect. This result is consistent with the BDS test. Barnett *et al.* (1998) highlighted that the BDS and the Kaplan tests are the best performing ones in nonlinear terms. Moreover, there is a possibility that the process in the data might be chaotic, rather than stochastic. To investigate this assumption I adopted the Kaplan test (Kaplan, 1994). Kaplan argues that in deterministic processes, unlike stochastic processes, nearby points are also nearby under their image in the phase space environment. In technical terms, if X_i and Y_j are relatively close to each other, then X_{i+1} and Y_{j+1} are also close to each other.

When $X_i = \left(r_i, r_{i-\tau}, r_{i-2\tau}, ..., r_{i-(m-1)\tau}\right)$ is embedded in m dimensional phase space, I observe the presence of a recursive function given by:

$$X_{i+\tau} = f\left(X_i\right) \tag{24}$$

Where τ is the fixed positive integer time decay. Hence, I can calculate:

$$\delta_{ij} = \left| X_i - X_j \right| \text{ and } \in_{i,j} = \left| X_{i+\tau} - X_{j+\tau} \right| \quad (25)$$

For all time subscripts (i, j) for a specified choice of embedding dimension m. I assume that $E(\zeta) = \sum_{A\zeta} \in_{i,j} / \#\{A_\zeta\}$, where $A_\zeta \equiv \{(i, j) : \delta_{i,j} < \zeta\}$. In a case of a completely deterministic system with continuous f I achieved $\lim_{\zeta \to \infty} E(\zeta) = 0$ and therefore K (the actual value of Kaplan) is the limitation of $E(\zeta)$ as $\zeta \to 0$.

One of the most important moments in this particular test is to estimate a piecewise regression line for $(\delta_{i,j}, \in_{i,j})$ and apply the intercept to calculate the value of K. I based the actual statistical procedure for K on simulated series which have the same histogram and similar autocorrelation functions as the original series. Acceptance of the null of IID is when K is smaller than the test statistic. The opposite is valid for the rejection of the null hypothesis.

In terms of the Kaplan test (the last column of Tables 3.0, 4.0 and 5.0), only Markets I and J of the three financial instruments demonstrate consistency with the test. The two markets populated by the highest number of artificial traders do not reject the null hypothesis of IID based on the BDS, ARCH and Kaplan tests.

de Lima (1998) investigated nonlinearity and nonstationarity in the S&P daily returns from January 2, 1980 to December 31, 1990. The author was unable to reject the null hypothesis of IID series in all subsamples prior to the 1987 crash. Interestingly, when he expanded the sample and included the crash, the outcome was a strong rejection. Chen *et al.* (2000) reported similar experimental findings.

My results are consistent with Chen *et al.*(2000) based on rejection and non-rejection of the null for the entire markets of the three indices, rather than based on subsample tests as performed by de Lima (1998). This supports the doubt raised by Chen *et al.* (2000) of whether de Lima's research interpretations that the presence of nonstationarity rather than dependence might have been the reason for the rejection of the null of IID prior to the 1987 crash.

The EMH postulates that prices should always be consistent with their fundamental values because asset prices reflect all available information. This is sometimes referred to as allocative efficiency which means that stock prices reflect the true fundamental value of the underlying asset. Many academics tend to believe that the prudent behaviour of traders is indispensable for a certain desirable feature of the stock price i.e. that the prices are likely to be more consistent with thier fundamental values (Chen and Yeh, 1999; Binswanger, 1999).

As illustrated in Figures 8.0, 10.0 and 12.0 the FTSE 100, the S&P 500 and the Russell 3000 price series generated by 1,000 traders only (all markets denoted by A) exhibit market inefficiency. The red curve which represents the generated price series by the artificial traders deviates substantially from the fundamental values of the three indices represented by the yellow curve. Figures 9.0, 11.0 and 13.0 illustrate the price series for the three financial instruments generated by 10,000 artificial traders (all markets denoted by J). It is clearly evident that the price series and their fundamental values significantly overlap, suggesting market efficiency.

Figure 8.0 Time series plot of FTSE 100 generated by 1,000 traders. Note: the yellow curve consists of historical FTSE 100 quotes, the red curve represent price series generated by 1,000 traders.

Figure 9.0 Time series plot of FTSE 100 generated by 10,000 traders.

Figure 10.0 Time series plot of S&P 500 generated by 1,000 traders.

Figure 11.0 Time series plot of S&P 500 generated by 10,000 traders.

Figure 12.0 Time series plot of Russell 3000 generated by 1,000 traders.

Figure 13.0 Time series plot of Russell 3000 generated by 10,000 traders.

This claim is confirmed by the empirical tests I performed. Table 6.0 describes the absolute deviations from real prices as a proportion of the real FTSE 100, S&P 500 and Russell 3000 in Market A and Market J. The mean, maximum and standard deviation reported by the markets with reduced numbers of artificial traders are significantly higher than the equivalent statistics for the prices series generated by traders in the most populated markets. The significance of the differences in the mean values has been estimated by *t*-tests on the paired differences between the deviations from intrinsic (fundamental) values for the two markets. All differences recorded in Table 6.0 are significant at the 1% level.

			FTSE 100		
Market	***N***	***Minimum***	***Maximum***	***Mean***	***Std.deviation***
J-10,000 traders	7262	0.0000	0.8456	0.3283*	0.0272
A-1,000 traders	7262	0.0000	0.9041	0.9918*	0.1018
Mean paired difference between Market J and Market A				-0.6635	
			S&P 500		
J-10,000 traders	7262	0.0000	1.0141	0.0236*	0.2387
A-1,000 traders	7262	0.0000	1.0293	0.0890*	0.9310
Mean paired difference between Market J and Market A				-0.0654	
			Russell 3000		
	7262	0.0000	0.7223	0.1736*	0.0113
	7262	0.0000	0.8327	0.7632*	0.1487
Mean paired difference between Market J and Market A				-0.5896	

*Significantly different from 0 at the 1% level.

Table 6.0 Descriptive statistics of FTSE 100, S&P 500 and Russell 3000 price series-absolute deviations from real prices as a proportion of real prices for markets populated by 1,000 and 10,000 traders.

It is important to note, however, that the market prices in inefficient markets periodically and temporarily deviate from their fundamental values. As the figures illustrate, market prices moves back to their intrinsic levels in the long-run. Hence, inefficiency does not cause long-run mispricing of assets, leading to persistent arbitrage opportunities. My empirical results are consistent with the findings of Decamps and Lovo (2000) who pointed out that asset prices converge to their fundamental values in the long-run.

The Hurst exponent proposed by Hurst (1951) provides a measure for long-term memory and fractality of a time series. Hurst (1951) demonstrated that range series scaled by power-law as time increases:

$$(R/S)_t = c * t^H \tag{26}$$

$(R/S)_t$ represents the rescaled range series at time t, c is a constant and H is called the Hurst exponent. $H = 0.5$ indicates a random series, $0 < H < 0.5$ indicates an anti-persistent series and $0.5 < H < 1$ signals a persistent series. A persistent series is trend reinforcing (the direction of the next value is more likely the same as the current value). Figure 14.0 below illustrates the Hurst exponent for FTSE 100 generated by 10,000 artificial traders. The Hurst exponent ranges from 0.390 to 0.620.

Figure 14.0 Hurst exponent for FTSE 100 price series generated by 10,000 traders.

Figure 15.0 shows that the FTSE 100 Hurst exponent peaks around the value $H = 0.5$ which indicates a random series and therefore efficient markets.

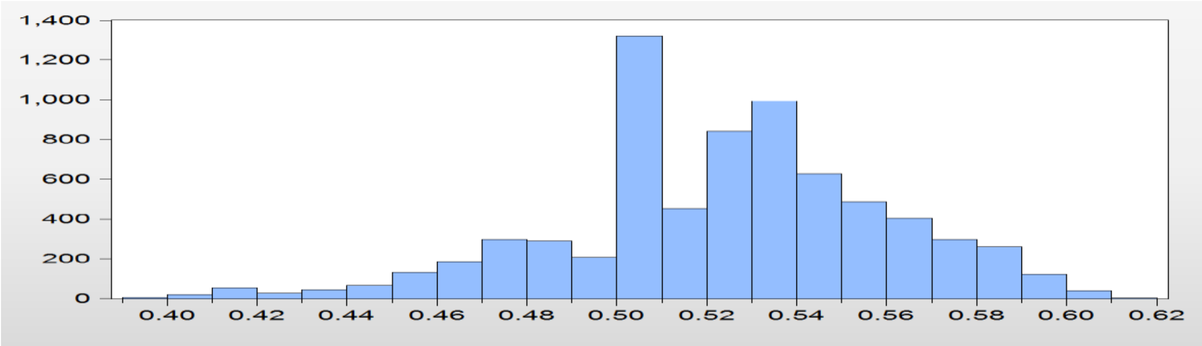

Figure 15.0 Histogram of Hurst exponent for FTSE 100 price series generated by 10,000 traders.

Figure 16.0 illustrates that the Hurst exponent ranges between 0.375 and 0.625 for S&P 500 return series generated by 10,000 traders.

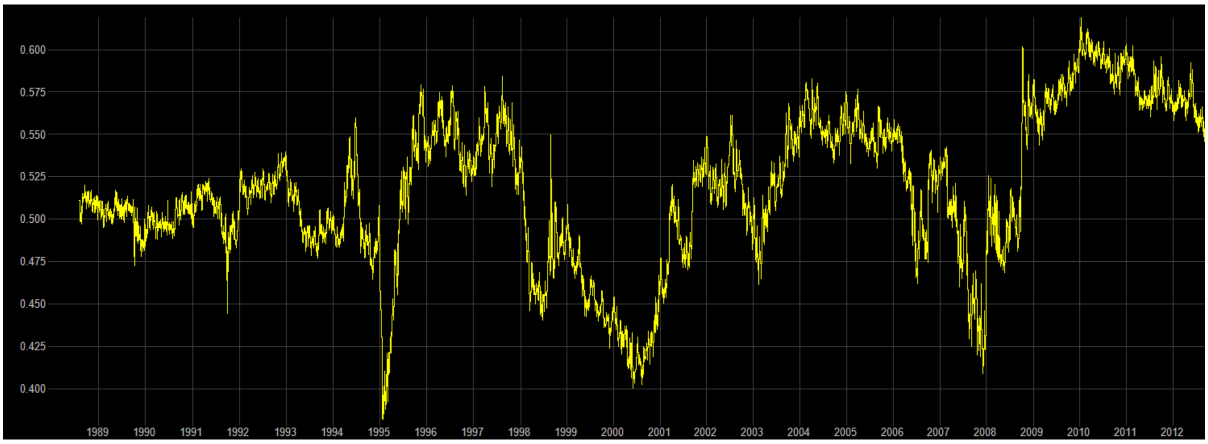

Figure 16.0 Hurst exponent for S&P 500 price series generated by 10,000 traders.

Figure 17.0 clearly indicate that the Hurst exponent has peaked at the value of $H = 0.5$ suggesting randomness and efficient markets.

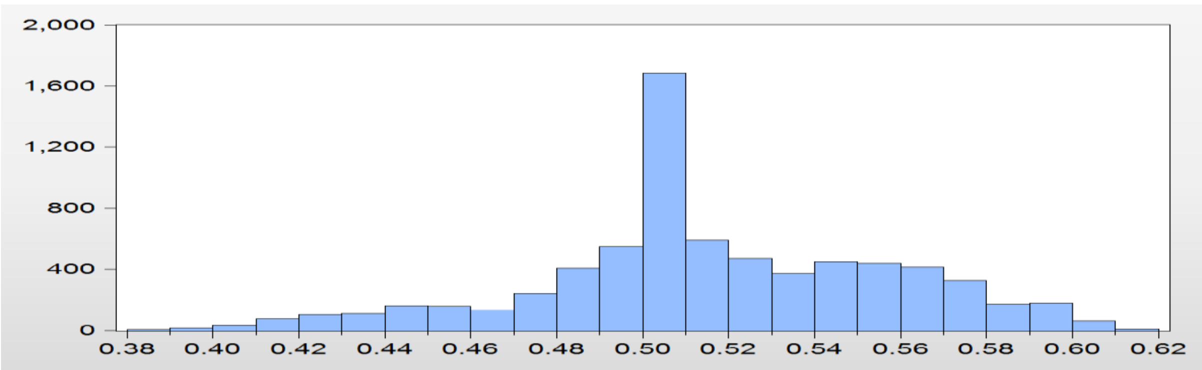

Figure 17.0 Histogram of Hurst exponent for S&P 500 price series generated by 10,000 traders.

Figure 18.0 shows the Hurst exponent for the Russell 3000 ranging between 0.400 and 0.610 and Figure 19.0 shows a peaked value of 0.5. Hence, the return series generated by 10,000 artificial traders for the Russell 3000 are random and the market is efficient.

Figure 18.0 Hurst exponent for Russell 3000 price series generated by 10,000 traders.

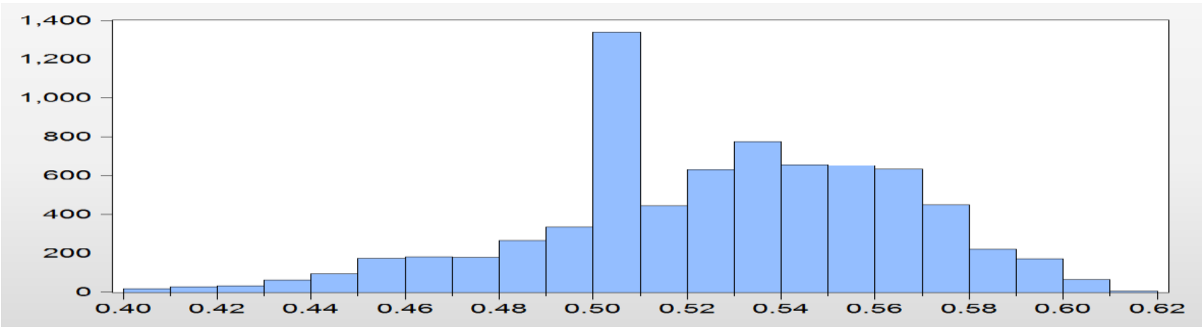

Figure 19.0 Histogram of Hurst exponent for Russell 3000 price series generated by 10,000 traders.

Furthermore, I investigated whether the efficient markets populated by the largest number of traders possess long-memory for FTSE 100, S&P 500 or Russell 3000. A random process characterised by long memory when the autocorrelation function decays asymptotically as a power-law of the form $\tau^{-\alpha}$ with $\alpha < 1$ (Lillo and Farmer, 2004). In other words, values from the past could have significant implications on the present, implying anomalous diffusion under stochastic conditions which emphasises the presence of long-memory.

When the value of the exponent α is smaller than 1, the process might have long-memory (the smaller the value of α , the longer the memory). In order to define the long-memory process I follow Lillo and Farmer (2004):

$$\gamma(k) \sim k^{-\alpha} L(k) \text{ if the limit } k \to \infty \tag{27}$$

Where $0 < \alpha < 1$ and $L(x)$ is a slowly varying function at infinity if $\lim_{x \to \infty} L(tx) / L(x) = 1$ (Embrechts *et al.,* 1997).

In my experiment, I am considering only positively correlated long-memory processes, with a Hurst exponent in the interval (0.5, 1). In terms of the Hurst exponent (H), the long-memory process is characterised by:

$$\alpha = 2 - 2H \tag{28}$$

On the other hand, short-memory processes can be quantified by:

$$H = 1/2 \tag{29}$$

With an autocorrelation function that decays faster than k^{-1}. The relationship between the diffusion properties of the integrated process explains the rationale behind the use of the Hurst exponent. In cases of normal diffusion the increments do not possess long-memory and the standard deviation increases at the rate of $t^{1/2}$. In cases of long-memory increments, the standard deviation increases at the rate of $t^{H} L(t)$, with $1/2 < H < 1$ and a slow-varying $L(t)$ function. As the peaked value of the Hurst exponent is 0.5 in all experiments with markets denoted by J for the three financial instruments, I estimate the corresponding (α) exponent from Equation (36). In all three markets populated by 10,000 artificial traders, I calculated that $\alpha = 1$, suggesting a lack of long-memory processes in place. Consistent with my previous empirical results I can conclude that all markets denoted by J for the FTSE 100, S&P 500 and Russell 3000 are characterised with strong market efficiency- the process possesses short-memory with $H = 0.5$ and the autocorrelation function decay faster than k^{-1}.

Overall my experimental results show that an artificial stock markets populated by a reduced number of traders behaves differently from markets with greater genetic diversity. The price series generated by artificial agents in Markets I and J conform to the EMH. This is clear evidence that enhanced genetic diversity has a beneficial effect on the market. The presence of more artificial traders in Markets I and J corresponds to an enhanced variety of different trading rules, and most importantly, greater market efficiency. Hence, the price formation process produced by a greater number of traders is a better predictor than any small fraction of traders. This is a result of the greater genetic diversity that is presented in the total population. Enhanced diversity means more heterogeneous trading rules and behaviour leading to greater flexibility in the virtual market clearing price mechanism.

Also, with the presence of a greater number of traders, the market is more competitive and more information is reflected in the order flow. It seems that markets populated by a greater number of traders react to price changes in a timely manner, making the entire market more efficient by enhancing the process of adjustment of prices to their fundamental values. Moreover, the BDS, the ARCH and the Kaplan tests revealed that richer dynamic structures, such as stock markets populated by a greater number of traders, helps to describe the findings of complex nonlinear dependence in stock market data.

Despite the EMH view, my empirical results provide evidence that patterns observed in financial markets seems to indicate that markets characterise by internal dynamics of their own. Financial markets dynamics seems influenced by the wider heterogeneity as well as by the microstructure of the market. I think that the key to analyise and understand the rich market dynamics is the mechanism which allows various population of traders to learn and adapt over time. My empirical results better explain market efficiency in terms of the AMH. I observe that market efficiency is not an isolated process, but it is a dynamic and context-dependent process where market participants adapt to their changing environment.

Enhanced genetic diversity provides an appropriate environment where different numbers of artificial traders involved in the evolutionary process adapt to a changing environment. Markets composed of more traders seem to adapt better to the changing environmental conditions leading to increased level of market efficiency.

Moreover, the evolutionary nature of the artificial traders is based on survival of the fittest principle, that is, to better cope with changing circumstances, market dynamics and opportunities. In other words, natural selection operates to select the fittest within an evolutionary framework in which markets and traders interact and evolve dynamically according to the law of economic selection. Under these circumstances, traders compete, learn and evolve. Hence, market efficiency involves reasons and beliefs which have adaptational value because they are changeable in response to changing market circumstances. My findings are consistent with Blume and Easley (1992) who claim that the 'market selection hypothesis' based on the natural selection and survival of economic actors better represent the relationship between market efficiency and market adaptability. The price of FTSE 100, S&P 500 and Russell 3000 in Markets I and J reflect as much information as required by the mixture of environmental factors and the number of distinct groups of artificial traders ranging from 1,000 in Markets A to 10,000 in Markets J. Hence, the AMH seems to measure better market efficiency due to its less theoretically restrictive nature than the EMH. The AMH does not require market participants to uniformly follow the rationality axioms of neo-classical economics.

Overall, the experimental results presented above can be summarized as two implications on financial market efficiency and adaptability, namely, the size effect and the learning effect. On the one hand, the size effect on its own suggests that the market is more efficient when the population size increases. On the other hand, the learning effect indicates that the market is more efficient when traders' adaptive behaviour become more independent. I can conclude that enhanced market size, and greater heterogeneous learning style is leading to improvement in diversity of traders' expectations resulting in more efficient and adaptable financial market structures.

4.5 Conclusions

The controversy surrounding the EMH has stimulated several new directions of research. Many of the well-established financial models rely on two building blocks: market efficiency and homogeneous traders. In this paper, I propose the idea that financial markets should be viewed from a Darwinian 'survival of the fittest' perspective and, specially, within an evolutionary framework in which markets, financial instruments, and traders interact, compete, learn, adapt, and evolve dynamically. My experiment provided such a dynamic environment.

This chapter investigates the relation between market diversity and market efficiency. My empirical findings demonstrate that greater market heterogeneity and adaptability play key roles in market efficiency. The increase of market size significantly and positively contributed to the market efficiency by means of enhancing market diversity. Moreover, individual trader learning, adaptation, and evolution reinforced the motion of efficient markets. My research contributions are twofold. On the one hand, the presence of different market sizes suggests that the market is more efficient when the population size increases. On the other hand, the learning effect indicates that the market is more efficient when traders' adaptive behaviour becomes more independent.

Hence, larger market size and a greater heterogeneous learning style are leading to improvement in the diversity of traders' expectations, resulting in more efficient and adaptable financial market structures.

I generated rich varieties of market dynamics which provided a promising direction to contribute to the current studies on micro-structure and anomalies. My empirical findings suggest that stock markets composed of a reduced number of traders- represented by Markets A, B, C, D, E, and F- behave differently when compared with markets with a greater genetic diversity- represented by Markets G, H, I, and J. I have found that the presence of more artificial traders in the last two markets (I and J) is associated with an enhanced variety of trading rules, leading to greater market efficiency in terms of the EMH. However, I have also found that the EMH on its own cannot explain the internal market dynamics, market micro-structure, and the heterogeneity of market participants.

My experimental results suggest that market efficiency is a dynamic and context-dependent process, where traders adapt to constantly changing market conditions.

The various empirical tests I performed suggest that markets populated by a greater number of traders help to discover the findings of complex nonlinear dependence in stock market data and explain the emergent nature of the EMH and the AMH. I think that the AMH does not require market participants to follow the specific rationality axioms of neo-classical economics and, therefore, market efficiency exists simultaneously with the need for adaptive flexibility.

The evolutionary nature of the STGP technique enabled me to empirically test the idea that different trader populations behave as an efficient adaptive system. My empirical findings revealed that the presence of increased heterogeneity in markets provides ideal conditions for artificial agents to adapt to the changing environmental conditions, leading to higher market efficiency.

One area of future research I would like to explore myself is the degree of traders' prudence- the time horizon which traders look back at while they make investment decisions. I also would like to include social learning in my future experiments and analyse the adaptive switch between social and individual learning.

Chapter 5

Herd behaviour experimental testing in laboratory artificial stock market settings. Behavioural foundations of stylised facts of financial returns.

5.1 Introduction

The majority of investors actively trade stocks instead of buying and holding a market portfolio. Active trading may move asset prices around the intrinsic value of the stock and increase long-run price volatility. Analysing herd behaviour in financial markets is of particular interest because it might offer an explanation of excess volatility and bubbles. Traders experience herd behaviour when the knowledge that others are investing changes their decision from not investing to making the investment. In other words, investors copy the behaviour of other investors leading to changes in their decision-making process after observing others. Investors ignore to a certain degree, their private opinions and follow the market leading to a switch from non-trading to trading. Herding might cause changes in the magnitude of trading activity, the assets traders invest in, or even their valuation. Herding behaviour explains why profit-maximising individuals with similar information react similarly in terms of investing funds (Bikhchandani and Sharma, 2001). The financial crises of the 1980s and 1990s have highlighted herding as a possible reason for excess volatility and financial system fragility. Banerjee (1992), Bickchandani *et al.*, (1992) and Welch (1992) were among the first scholars to write about herd behaviour. They analyse herd behaviour under abstract conditions (in the context of fads, fashions and customs) where privately informed individuals develop their decision-making process in sequence. These early research papers attempt to describe herding when a finite number of individuals have already chosen their actions and all following individuals abandon their own specific private information and herd. Devenow and Welch (1996) suggest that agents disregard their prior beliefs and follow the actions of other agents creating herding. Christie and Huang (1995) assumed that herding is most pronounced when market returns are extreme. Their findings show that when market agents abandon their own stock price forecasts in favour of the aggregate market behaviour, their asset returns are very similar to the overall market return.

A few years later, Avery and Zemsky (1998) investigated herd behaviour in real financial market settings with stock prices determined by a market maker according to the order flow. The authors concluded that the price mechanism prevents the development of informational cascades (a market condition in which traders disregard their own information and imitate previous traders' decisions leading to herd behaviour). They demonstrated that informational cascades are impossible because new information can reach the market at any time; thus, consistent with a steady informational flow, prices do not deviate significantly from fundamental values. Moreover, according to their findings herd behaviour does not cause excess volatility and the mispricing of assets in the long run.

Lakonishok, Shleifer and Vishny (1992) and Bikhchandani and Sharma (2001) claim that there seems to be weaker evidence of herding behaviour in individual stocks than in groups of stocks. They stressed that this does not exclude the possibility of more intensive herding in certain stocks such as stocks of a particular size or with particular performance records.

The Marginal Trader Hypothesis (MTH), proposed by Forsythe *et al* (1992) states that a small fraction of savvy individuals are capable of setting market prices and strive for market efficiency. Marginal traders are described as well-informed and active traders who are more capable of inferring true price and willing to explore those inferences. The authors argue that when one removes those 'perfect' individuals from the pool of traders prediction markets lose their accuracy. Prediction markets are markets established to generate knowledge and forecasts about the likelihood of future events. Forsythe *et al* (1992) analyse data from the Iowa Presidential Stock Market (IPSM), which was successfully created in 1988 and operated as a computerised double-auction market in order to forecast the vote shares of the presidential candidates in elections held in the same year. They combined market design and incentive structures familiar from laboratory experiments to find out how the 1988 US presidentials would finish. The accuracy of prediction they achieved was very impressive.

However it is difficult to test those theoretical assumptions directly. The literature related to herding behaviour in financial markets focuses primarily on statistical measures of clustering. The main difficulty from the empirical point of view comes from the fact that there is no database on the private information available to investors and hence it is not possible to prove whether market agents strictly disregard their own information and imitate. This serious obstacle can be avoided in experimental settings such as an agent-based artificial stock market where the information possessed by traders can be controlled.

Under laboratory settings researchers can observe the private information available to individuals for decision-making purposes, and therefore it is possible to test the presence of herding. In a market simulation model that I created using Altreva Adaptive Modeler, artificial traders receive information about the value of a real security and observe the history of past trades. Based on this information they decide if they want to buy or sell one or more units of the security. By observing how artificial agents deal with the same piece of public information and react to the decisions of the previous agents I can detect the possible presence of herding behaviour. My experiments are extremely suitable environment to investigate the occurrence of 'spurious herding' (Bikchandani and Sharma, 2001) or 'investigative herding' (Froot *et al.,* 1992; Hirshleifer *et al.,* 1994) where a group of traders such as 'Best Agents' and 'All Agents' make similar decisions because they face similar information (historical data of Dow Jones, IBM and GE). Several studies use a statistical measure of herding behaviour proposed by Lakonishok *et al.,* (1992). The authors divide and measure herding behaviour as the average tendency of a group of traders to buy (sell) particular stocks at the same time. This herding measure examines the correlation in trading patterns for a particular group of traders and their tendency to buy and sell the same set of stocks.

The modelling software that I use provides a rich environment to examine herd behaviour as artificial traders make independent decisions creating a heterogeneous market structure (the market is populated by 10,000 boundedly rational artificial agents each with different trading rules and behaviour). Traders adaptive behaviour in my artificial stock market is modelled with an evolutionary computing technique called Strongly Typed Genetic Programming (STGP). The STGP evolves the trading rules at the micro level and co-evolves all agents through trading on the artificial market at the macro level.

In this chapter I evaluate the price series of a group of stocks modelled by the Dow Jones and individual stocks represented by General Electric and IBM generated using two main groups of artificial agents- ‘Best Agents’ and ‘All Agents’. I then use econometric evaluation to analyse the following topics:

(i) Do price series generated by artificial stock market agents exhibit herding behaviour in individual stocks as well as in a group of stocks?

(ii) Volatility analysis of price series generated by ‘Best Agents’ and ‘All Agents’. Is the Virtual Market price based on the behaviour of all agents less volatile in comparison with a small subset of agents?

(iii) Artificial stock markets and the Efficient Market Hypothesis. Is the price series generated by ‘Best Agents’ more likely to conform to the Efficient Market Hypothesis, and therefore be more efficient?

Chen and Yeh (1999) developed a genetic programming (GP) based artificial stock market in order to investigate herding behaviour. Chen and Yeh’s paper is a good starting point and I suggest several extensions to their approach based on the following important factors:

-A greater number of artificial agents. While Chen and Yeh’s model consists of only 500 traders, I employ 10,000 agents. A larger population means increased model stability and reduced sensitivity to random issues. The presence of substantially more artificial agents creates a competitive environment where different trading rules compete and evolve in parallel at the same time. Having a greater number of agents open up the opportunity to implement a wider variety of trading strategies programmed in the agents’ trading rules (Witkam, 2013).

-Rather than using a fixed intrinsic value of the stock (fixed at 100 in Chen and Yeh’s model), I feed the software with real historical asset prices. This is done in order to prevent the formation and development of herding behaviour by design. Cipriani and Guarino (2002) argue that when the price is fixed, individuals tend to disregard their own information and strictly follow the decisions made by the previous agents resulting in herding behaviour.

-My model developed within Altreva settings is built incrementally (walk-forward with no overfitting of historical data). It constantly evolves and adapts to market changes instead of being static. I have 10,000 different trading strategies competing simultaneously and evolving on the artificial stock market in real time. Hence, the model is more resilient to changing market conditions and model performance is significantly more consistent and reliable (Witkam, 2013).

No known study examines the heterogeneity, efficiency and behaviour of the stock market by the implementation of Strongly Typed Genetic Programming (STGP) technique. Specifically, the major contributions of this chapter are as follows.

•Initially to provide evidence of the emergent properties of herding behaviour, stock market efficiency and stylised facts of financial returns gained by implementing a special form of Genetic Programming- STGP.

•Secondly to provide unique tests of the MTH within artificial stock market settings.

5.2 Artificial stock market structure for this particular experiment

Table 2.0 illustrates the main parameters of the GP-based artificial stock market model. Caram *et al.* (2010) introduced a self-organised competition scheme related to the agents' sizes. While the authors imposed constraints allowing only agents of similar sizes to compete with each other, I do not implement such size constraints but focus on the wealth-creating attributes of agents. Each of my artificial markets is divided into two main groups- a small fraction of traders called 'Best Agents' and 'All Agents' (the remainder of the market). Four different 'Best Agents' group sizes were designed- 1%, 5%, 10% and 20% of the total population. By agent group size I mean the fraction of a given market that the agent group has. Both 'Best Agents' and 'All Agents' group sizes are part of and operate in the same market. Hence, I have four stock markets encompassing the four different agent group sizes. The markets trade separately each of the three financial instruments- the Dow Jones (DJ), GE and IBM. For instance, the first market is populated by 1% of the best- performing agents (100 traders) and the remaining 9,900 (99%) of the total 10,000 traders represented by the 'All Agents' group.

The most important characteristic of 'Best Agents' groups is that they momentarily perform best in terms of continuous Breeding Fitness Return (a trailing return of a wealth-moving average). This particular type of return is used to measure the fitness criterion for the selection of agents to breed. Breeding in essence is a process of creating new artificial traders to replace poor performing ones. Breeding involve the selection of well performing traders and the production of completely new genomes by recombination of the parent genomes by crossover and mutation operations. Hence, agents' trading rules will improve by a natural selection process as the survival-of-the-fittest principle is in place.Artificial agents produce wealth by investing in two assets available on the virtual market. One of them is the risky stock asset and the other is the risk free asset represented by cash. There are three separate markets each composed of money and different financial instruments such as IBM, General Electric or an index like the Dow Jones.

Artificial stock market parameters	
Total population size (agents)	10,0000
Best performing agents size(percentage of the total population)	1%; 5%; 10; 20%
Initial wealth(equal for all agents)	100,000
Fixed broker fee	10%
Average bid/ask spread	0.01%
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 (bars)	1
Minimum breeding age (bars)	80
Initial selection type	random
Parent selection (percentage of initial selection that will breed)	5%
Mutation probability (per offspring)	10%
Total number of quotes (bars) processed- DJ	19,942 (01/09/1931-17/06/2011)
Total number of quotes (bars) processed- GE	12,353 (02/01/1962-17/06/2011)
Total number of quotes (bars) processed-IBM	12,355 (02/01/1962-21/06/2011)
Trading hours (open/close)	09:30/16:00
Short positions allowed	Yes
Seed generation from clock	Yes
Creation of unique genomes	Yes
Offspring will replace the worst performing agents of the initial selection	Yes

Table 7.0 Artificial Stock Market Parameter Settings

*Maximum genome size measure the total number of nodes in an agent'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 an agent's genome (trading rule). The depth of a trading rule can be an indicator of its complexity.

5.3 Simulation Results

5.3.1. An investigation of whether the price series generated by artificial stock market agents exhibit herding behaviour in individual stocks as well as in a group of stocks. Testing the Marginal Trader Hypothesis.

I use a statistical measure of herding proposed by Lakonishok *et al.* (1992). The authors argue that herding behaviour can only be detected within subsets of traders. Hence, this particular measure of herding behaviour is extremely suitable for my experiments because is based on trades conducted by a subset of market participants such as 'Best Agents' and 'All Agents' over a period of time. I measure herding behaviour as the average tendency of a group of traders ('Best Agents' and 'All Agents') to buy (sell) Dow Jones, IBM and GE at the same time. This herding behaviour measure accesses the correlation in trading patterns for a particular group of traders and their tendency to buy and sell the same set of financial instruments. Hence, herding behaviour leads to correlated trading.

The measure of herding behaviour for a given financial instrument i, in a given trading day t, $H(i,t)$, is defined as

$$H(i,t) = \left| B(i,t) / B(i,t) + S(i,t) - p(t) \right| - AF(i,t) \tag{30}$$

Let $B(i,t)\left[S(i,t)\right]$ be the number of traders in this subset who buy [sell] a financial instrument i in trading day t and $H(i,t)$ be the measure of herding in financial instrument i for trading day t. In other words $B(i,t)$ is the number of traders who increase their holdings in Dow Jones, IBM and GE in the trading day (net buyers), $S(i,t)$ is the number of traders who decrease their holdings (net sellers), $p(t)$ is the expected proportion of cash traders possess in that trading day.

The adjustment factor $AF(i,t)$ is related to the fact that under the null hypothesis of no herding, i.e. when the probability of any trader being a net buyer of any financial instrument is $p(t)$, the absolute value of $\left| B(i,t)\left(B(i,t)+S(i,t)\right)\right| - p(t)$ is greater than zero.

$AF(i,t)$ is, therefore, the expected value of $\left|B(i,t)/\left(B(i,t)+S(i,t)\right)\right|-p(t)$ under the null hypothesis of no herding. Since $B(i,t)$ follows a binomial distribution with probability of $p(t)$ success, $AF(i,t)$ can be estimated given $p(t)$ and the number of agents trading in that financial instrument in that day. For any financial instrument, $AF(i,t)$ declines as the number of agents trading in that financial instrument rises (Lakonishok *et al.,* 1992 and Bikhchandani and Sharma, 2001).

The herding measures in my experiments are computed for each financial instrument – trading day and then averaged across different subgroups such as 'Best Agents' and 'All Agents'. Values of $H(i,t)$ significantly different from zero are interpreted as evidence of herding behaviour.

Tables 8.0, 9.0 and 10.0 represent my main results on herding behaviour. The third and fourth column of the tables reports the mean and median herding measures for the whole sample. The mean herding measure (the key measure) of Dow Jones at 1% 'Best Agents' is 0.048 and it implies that if $p(t)$, the average fraction of changes that are increases, was 0.5, then 54.8% of the traders of 'Best Agents' subgroup were changing their average holdings of Dow Jones in one direction and 45.2% in the opposite direction. The presence of herding is also confirmed by the relatively large median herding measure of 0.011. However, herding behaviour is less prominent when the market is populated by more artificial traders. For instance, the remainder of the market represented by 'All Agents' group indicate that only 53.7% of the traders change their average holdings of the index in one direction and 46.3% in the opposite direction. Table 8.0 illustrate the same statistical trend at 5%, 10% and 20% levels.

IBM herding statistics (Table 9.0) demonstrate the presence of substantially less herding behaviour. For example, the mean herding measure of IBM at 20% 'Best Agents' shows that 50.9% of the traders change their average holdings of the security in one direction and 49.1% in the opposite direction. 'All Agents' at 20% level of the same security indicate insignificant herding behaviour of 50.4% of the traders change their holdings of IBM in one direction and 49.6% in the opposite direction.

The median herding measure is even smaller, only 0.001, which suggest that there is very insignificant herding behaviour. GE herding statistics on Table 10.0 also indicate the presence of less herding comparing to Dow Jones.

1%			
	N	**Mean**	**Median**
Best Agents	17441	0.048*	0.011*
All Agents	17441	0.037*	0.008*
5%			
Best Agents	17441	0.042*	0.010**
All Agents	17441	0.030*	0.006*
10%			
Best Agents	17441	0.039*	0.008*
All Agents	17441	0.024*	0.005**
20%			
Best Agents	17441	0.021*	0.006*
All Agents	17441	0.013*	0.002*

Table 8.0 Dow Jones herding statistic based on 17,441 trading days for 'Best Agents' and 'All Agents'.
*Results are statistically significant at the 1% level. **Results are statistically significant at the 5% level.

The mean and median herding statistics are presented for 'Best Agents' and 'All Agents'. The herding statistic for a given day is defined as $H(i,t) = |B(i,t)/B(i,t)+S(i,t)-p(t)| - AF(i,t)$ where $B(i,t)$ is the number of traders who increase their Dow Jones holdings in the day (net buyers), $S(i,t)$ is the number of traders who decrease their Dow Jones holdings (net sellers), $p(t)$ is the expected proportion of traders buying in that day, and $AF(i,t)$ is the adjustment factor explained in the text. The herding measures are computed for Dow Jones in each day and then averaged across 'Best Agents' and 'All Agents'.

1%			
	N	**Mean**	**Median**
Best Agents	17441	0.031*	0.004*
All Agents	17441	0.025*	0.002**
5%			
Best Agents	17441	0.029**	0.003*
All Agents	17441	0.017*	0.001*
10%			
Best Agents	17441	0.018*	0.002*
All Agents	17441	0.010*	0.001*
20%			
Best Agents	17441	0.009*	0.001*
All Agents	17441	0.004*	0.001*

Table 9.0 IBM herding statistic based on 17,441 trading days for 'Best Agents' and 'All Agents'.
*Results are statistically significant at the 1% level. **Results are statistically significant at the 5% level.

The mean and median herding statistics are presented for 'Best Agents' and 'All Agents'. The herding statistic for a given day is defined as $H(i,t)=|B(i,t)/B(i,t)+S(i,t)-p(t)|-AF(i,t)$ where $B(i,t)$ is the number of traders who increase their IBM holdings in the day (net buyers), $S(i,t)$ is the number of traders who decrease their IBM holdings (net sellers), $p(t)$ is the expected proportion of traders buying in that day, and $AF(i,t)$ is the adjustment factor explained in the text. The herding measures are computed for IBM in each day and then averaged across 'Best Agents' and 'All Agents'.

1%			
	N	**Mean**	**Median**
Best Agents	17441	0.035*	0.005*
All Agents	17441	0.021*	0.002*
5%			
Best Agents	17441	0.027*	0.005*
All Agents	17441	0.018**	0.002*
10%			
Best Agents	17441	0.020*	0.003*
All Agents	17441	0.011*	0.001*
20%			
Best Agents	17441	0.013*	0.002**
All Agents	17441	0.005*	0.001*

Table 10.0 GE herding statistic based on 17,441 trading days for 'Best Agents' and 'All Agents'.
*Results are statistically significant at the 1% level. **Results are statistically significant at the 5% level.

The mean and median herding statistics are presented for 'Best Agents' and 'All Agents'. The herding statistic for a given day is defined as $H(i,t)=|B(i,t)/B(i,t)+S(i,t)-p(t)|-AF(i,t)$ where $B(i,t)$ is the number of traders who increase their GE holdings in the day (net buyers), $S(i,t)$ is the number of traders who decrease their GE holdings (net sellers), $p(t)$ is the expected proportion of traders buying in that day, and $AF(i,t)$ is the adjustment factor explained in the text. The herding measures are computed for GE in each day and then averaged across 'Best Agents' and 'All Agents'.

My observations suggest that the price formation caused by the collective behaviour (competition and co-evolution) of the entire market is more cohesive than that of any small subset of agents. This is due to the greater genetic diversity that is represented in the total population leading to more diverse (heterogeneous) trading rules and behaviour. Moreover, greater genetic diversity means greater flexibility in the virtual market clearing price mechanism. At a broad level my results do not support the Marginal Trader Hypothesis (MTH) which explains market efficiency as a consequence of the actions of a small pool of traders such as 'Best Agents' who are capable of setting prices and acting without bias. In fact, it seems that the virtual market populated by 'All Agents' reacts to price changes in a timely manner.

All my experiments indicate that there is no tendency towards price crashes or bubbles. Hence, herd behaviour causes no long-run mispricing of assets because the market is consistent with the steady flow of information. It is a well-known fact that bubbles and herding behaviour are difficult to be appropriately identified and their magnitude cannot be determined until after the fact. This limits the ability of policymakers to respond to them in efficient manner. Both phenomena have an indirect effect on the economy, because investors and firms alter their behaviour in response to the price changes. For instance, when a particular corporation's asset price rises, the corporation may respond by increasing its physical capital investment spending higher than it otherwise would have. This suggests that herding behaviour and bubbles may cause a misallocation of resources leading to economic inefficiencies. The inflation and growth is likely to rise above a sustainable level, forcing central government intervention such as raising interest rates. On the other hand, high interest rates increase a firm's borrowing costs, leading to reduced profitability. The convergence of action of traders in my experiment provides valuable information to the policymakers about whether they should be concerned about the presence of bubbles and herding and their destabilising effects. I found evidence of herd behaviour over daily time intervals to be much stronger, revealing the short-term nature of the phenomenon.

My experimental results are in line with the theories of Avery and Zemsky (1998) and DeLong *et al.* (1990) that herding causes asset prices to deviate from fundamentals. Decamps and Lovo (2002) argue that herding behaviour prevents agents from learning the market fundamentals.

However, there is no evidence of a price-destabilising process in place, as the market discounts the informativeness of trades during herding (Avery and Zemsky, 1999). Generated Dow Jones prices do in general follow fundamentals in the long run, but they periodically depart from them, which is more evident in the case of the 'Best Agents' price series.

My results confirm the claim of Decamps and Lovo (2002) that asset prices ultimately converge to fundamentals in the long run:

$$\Pr(\lim_{t\to\infty} E\langle V \mid H_t \rangle = V) = 1 \tag{31}$$

Where V represent the true value of the security at time t, E is the expected value of the asset at time t, H_t describes the history of agents actions up until time t.

Tables 8.0, 9.0 and 10.0 empirically show that the presence of herding in individual stocks is broadly consistent with the findings of Lakonishok, Shleifer and Vishny (1992) and Bikhchandani and Sharma (2001), which claim that the possibility of observing intentional herding behaviour at the level of individual stocks is relatively low.

My artificial stock market observations indicate that herding is more likely to occur at the level of investment in a group of stocks, such as Dow Jones than at the level of individual stocks. Lakonishok *et al.* (1992) suggest that this is due to swings in demand for a group of stocks which have a large effect on stock prices than swings in demand for individual stocks. Moreover, the authors argue that companies might also be more apt to herd in industry groups as opposed to individual stocks. Bikhchandani and Sharma (2001) argue that is unlikely that investors observe each other's holdings of individual stock soon enough to change their own portfolios. This is the reason why according to the authors one is more likely to find herding in a group of stocks. The other reason for the presence of more herding in a group of stocks is that different companies within the group migh try to infer information about the quality of investments from each others' trades and herd as a result (Shiller and Pound, 1989; Banerjee, 1992; Bikhchandani *et al.,* 1992). However this does not exclude the possibility of more extensive herding behaviour in particular categories of stocks, such as stocks of a specific size or performance record (Lakonishok *et al.,* (1992).

There has been an argument that herding should be more persistent within particular industry groups of stocks, such as technology ones, due to their uncertain cash flows. From this point of view, one might expect to record significantly more herding in IBM stocks than General Electric stocks. Experimental tests with IBM assets indicate that there is weak presence of herding behaviour observed in price series generated by both ‘Best Agents’ and ‘All Agents’ at all different group sizes.

Artificial traders simulate a dynamic and competitive market based on the survival-of-the-fittest principle. This type of stock market is characterised by large order flow and small price fluctuations. My findings correspond to the well-known empirical fact that large fluctuations in prices are highly likely to emerge in less active markets with small order flow (Cont and Bouchaud, 2000).

Cont and Bouchaud (2000) quantified this process using the statistic:

$$k\left(\Delta_x\right)=\frac{2_c+1}{n_{order}\left(1-c/2\right)A\left(c\right)\left(1-c\right)^3} \tag{32}$$

where k measures the kurtosis, n_{order} is the order flow; $A(c)$ is a normalisation constant with a value close to 1. The equation above explains that kurtosis (k) of the price change is inversely proportional to the order flow n_{order}.

In other words a market with small order flow (illiquid market) is more likely to experience large price fluctuations with higher frequency than a market where there are substantially more orders processed per unit time as is the case in my artificial stock market.

5.3.2 Volatility analysis of price series generated by 'Best Agents' and 'All Agents'. Is the Virtual Market price based on the behaviour of 'All Agents' less volatile in comparison with a price based on the behaviour of a small subset of agents?

Various studies related to the fluctuations in stock prices revealed that distributions of returns and prices series have fat tails, which characterise non-Gaussian distributions (Pagan 1996; Mandelbrot, 1963 and 1997). Bollerslev et al. (1992) argue that leptokurtosis is still present even after empirically testing for heteroskedasticity. To answer this particular research question, I needed to look at return distributions and moments. The return series were estimated by using the following equation:

$$r_t = \ln(p_t) - \ln(p_{t-1}) \tag{33}$$

Tables 11.0, Table 12.0 and Table 13.0 provide the basic econometric statistics of Dow Jones, General electric and IBM return series.

Dow Jones									
Best Agents									
Size	***Std.dev***	***SK***	***KU***	***JB***	***ADF****	***ARMA***	***BDS***	***GARCH***	***DLOG***
1%	4016.13	1.32	3.14	5135.98	-127.00	(1,1)	157.46	(4,1)	0.99
5%	4034.42	1.33	3.23	5222.58	-140.98	(2,1)	163.94	(2,1)	1.00•
10%	3938.21	1.33	3.18	5195.23	-113.23	(0,1)	138.63	(2,1)	1.00•
20%	3937.30	1.32	3.17	5147.76	-115.29	(1,1)	145.70	(4,1)	0.60
All Agents									
99%	4016.32	1.32	3.14	5126.05	-107.57	(2,1)	138.97	(3,1)	0.99
95%	4031.71	1.33	3.20	5187.56	-103.92	(0,1)	160.09	(2,1)	0.99
90%	3938.97	1.33	3.18	5191.12	-101.94	(1,3)	139.84	(2,1)	0.98
80%	3937.12	1.32	3.17	5148.99	-103.75	(1,1)	150.72	(4,2)	0.97

SK-skewness; ***KU***-kurtosis; ***JB***-the Jarque-Bera test.

*The MacKinnon (1996) one-sided critical value for rejection of the Null hypothesis of a unit root at 5% level is -3.410060.

•The IGARCH model has been used to restrict α+β (ARCH term + GARCH term) to one.

Table 11.0 Econometric statistics for Dow Jones price series generated by 'Best Agents' and 'All Agents'.

General Electric									
Best Agents									
Size	***Std.dev***	***SK***	***KU***	***JB***	***ADF****	***ARMA***	***BDS***	***GARCH***	***DLOG***
1%	25.02	0.86	4.31	1921.86	-81.89	(0,1)	53.02	(1,2)	0.90
5%	24.99	0.85	4.27	1857.56	-66.71	(1,1)	8.14	(1,1)	0.98
10%	24.95	0.84	4.25	1815.05	-68.38	(0,1)	-0.05**	(3,1)	1.00•
20%	25.17	0.86	4.31	1933.54	-74.42	(0,1)	45.68	(1,2)	1.00•
All Agents									
99%	25.01	0.86	4.32	1929.79	-68.63	(0,0)	-0.10**	(1,1)	0.88
95%	24.99	0.86	4.29	1890.46	-69.05	(0,0)	48.40	(1,1)	0.74
90%	24.94	0.84	4.25	1810.36	-69.49	(0,0)	-0.01**	***	***
80%	25.21	0.86	4.30	1923.99	-73.92	(1,1)	47.81	(1,1)	1.00•

SK-skewness; ***KU***-kurtosis; ***JB***-the Jarque-Bera test.

*The MacKinnon (1996) one-sided critical value for rejection of the Null hypothesis of a unit root at 5% level is -3.410060.

**Failed to reject the Null hypothesis that series are identically and independently distributed (IID).

***No presence of ARCH effect.

•The IGARCH model has been used to restrict α+β (ARCH term + GARCH term) to one.

Table 12.0 Econometric statistics for General Electric price series generated by 'Best Agents' and 'All Agents'.

IBM									
Best Agents									
Size	***Std.dev***	***SK***	***KU***	***JB***	***ADF****	***ARMA***	***BDS***	***GARCH***	***DLOG***
1%	75.95	1.79	6.14	9324.13	-69.13	(0,0)	55.85	(2,1)	1.00•
5%	75.98	1.78	6.12	9267.37	-67.71	(0,0)	54.61	(2,1)	0.56
10%	75.92	1.79	6.11	9221.44	-68.44	(0,1)	-0.09**	(1,1)	0.99
20%	75.91	1.79	6.12	9245.56	-74.10	(1,0)	-0.02**	***	***
All Agents									
99%	75.93	1.79	6.13	9299.25	-69.10	(0,0)	-0.08**	(1,3)	0.35
95%	75.99	1.79	6.13	9289.91	-67.27	(1,0)	-0.08**	(1,1)	1.00•
90%	75.90	1.79	6.11	9228.27	-69.23	(0,0)	-0.02**	(2,1)	0.58
80%	75.92	1.79	6.12	9253.98	-68.49	(0,0)	56.33	(2,1)	1.00•

SK-skewness; ***KU***-kurtosis; ***JB***-the Jarque-Bera test.

*The MacKinnon (1996) one-sided critical value for rejection of the Null hypothesis of a unit root at 5% level is -3.410060.

**Failed to reject the Null hypothesis that series are identically and independently distributed (IID).

***No presence of ARCH effect.

•The IGARCH model has been used to restrict α+β (ARCH term + GARCH term) to one.

Table 13.0 Econometric statistics for IBM price series generated by 'Best Agents' and 'All Agents'.

The first empirical property to analyse is volatility. When volatility of returns is considered for all three securities, there is little variation in observed volatility as the proportions of 'Best Agents' and 'All Agents vary. My volatility findings thus closely match the research done by Chen and Yeh (1999) who claim that price fluctuations are not affected by different degrees of agents sophistication. The price series generated by 'Best Agents' and 'All Agents' of the Dow Jones, General Electric, and IBM show that the mass of the distribution is skewed to the right, suggesting that the dataset is not normally distributed (positive deviations from the mean). Moreover, all skew values of the Dow Jones and IBM are greater than 1, indicating the presence of substantial skewness and distribution which is far from symmetrical.

Another empirical property is normality. A common econometric tool to test normality is the Jarque-Bera statistics. According to the test results, the null hypothesis that the price series generated by the artificial traders is normally distributed is rejected in all periods and in all different 'Best Agents' group sizes (see Table 11.0, Table 12.0 and Table 13.0). This result confirms an important fact in empirical finance: most financial return series are not normally distributed, which means that the tails are too fat compared to the normal distribution. The fat-tail (excess kurtosis) presence is obvious in all experiments. In other words, the probability of the occurrence of a large return is significantly higher than the normal distribution predicts. The General Electric and IBM return series show greater kurtosis than that of the Dow Jones. Furthermore, the series generated by 'All Agents' exhibit slightly higher values of excess kurtosis than 'Best Agents'.

5.3.3 Artificial stock markets and the Efficient Market Hypothesis. Is the price series generated by Best Agents more likely to converge with the Efficient Market Hypothesis, and therefore more efficient?

The Efficient Market Hypothesis (EMH) states that stock prices should always incorporate and reflect all relevant information. Thus, securities always trade at their fair value. In the econometric literature, a financial market is efficient when its return series $\{r_t\}$ is unpredictable. Return series are unpredictable when they are identically and independently distributed (IID). I applied a number of tests to determinate whether a series is characterised by IID properties.

First, the Augmented Dickey-Fuller (ADF) test has been applied to test for the presence of a unit root. My ADF test settings include running a regression of the first difference of the log price series against the series lagged once (a sufficient condition to eliminate autocorrelation in the residuals) combined with a drift and a time trend:

$$\Delta p_t = \beta_1 p_{t-1} + \sum_{i=1}^{4} \beta_{i+1} \Delta p_{t-i} + \beta_6 + \beta_{7t} \tag{34}$$

The null hypothesis of the ADF test is that $p_t\left(\ln\left(p_t\right)\right)$ contains a unit root $\left(\beta_1 = 0\right)$. The alternative hypothesis of no unit root presence is rejected when $\beta_1 \neq 0$. The null of the presence of a unit root has been rejected for the Dow Jones, General Electric, and IBM price series (Table 11.0, Table 12.0, Table 13.0). Hence, the returns generated by both 'Best Agents' and 'All Agents' are stationary at the 95% significance level, indicating that the process is orientated around a constant long-term mean and has a constant variance independent of time. Stationarity is a significant condition for standard econometric theory, otherwise we cannot achieve consistent estimators.

Then I proceed further to filter the linear process. The seventh columns of Tables 11.0, 12.0 and 13.0, show the *ARMA (p,q)* process obtained from the return series r_t. Nearly all General Electric and IBM series generated by 'All Agents' are linearly independent (*p*=0, *q*=0). Lack of linearity helped me to formulate an important preliminary finding that the agent-based artificial stock market populated by 'All Agents' is so efficient that there are hardly any linear signals left.

Therefore, in terms of ARMA statistics, series generated by ‘All Agents’ are more likely to converge to the Efficient Market Hypothesis than ‘Best Agents’. While the vast majority of other test results are either AR(1) or MA(1), some of the Dow Jones series exhibit higher order of linear dependence such as ARMA(2,1) and ARMA(1,3).

The next econometric property to test is whether price series are identically and independently distributed (IID). First, I estimated the most appropriate ARMA (p,q) model and fitted it to the data to eliminate all linearity from the sample. Once I have identified the linear series, any other series left should be nonlinear. I applied the Brock, Dechert and Scheinkman (BDS) test to the ARMA residuals in order to test for remaining dependence. Under the null hypothesis, the series are identically and independently distributed (IID). The fitted model is the best linear ARMA time series model, and rejection of the null indicates a nonlinear time series process. The test statistic was developed by Brock, *et.al.* (1996):

$$W_{m,n}(\varepsilon)=\sqrt{n}\frac{T_{m,n}(\varepsilon)}{V_{m,n}(\varepsilon)}\sim N(0,1) \tag{35}$$

Where m represents the embedding dimension; ε is the value of the radius (the distance parameter); $W_{m,n}(\varepsilon)$ is the variance of $T_{m,n}(\varepsilon)$, and

$$T_{m,n}(\varepsilon)=C(N,m,\varepsilon)-C(N,1,\varepsilon)^{m} \tag{36}$$

Critical BDS test parameters are the distance parameter $'\varepsilon'$ and the embedding dimension (DIM). My qualitative result proved not to be sensitive to the choice of epsilon and DIM parameters. Therefore, I only report experimental results with the distance parameter ‘ε’ equal to 0.7 standard deviations and embedding dimension of 6. Based on the test results in column 8 of Table 11.0, the null hypothesis that the residuals of the Dow Jones return series are IID is significantly rejected in all subperiods. The result indicates nonlinear dependence in both the ‘Best Agents’ and ‘All Agents’ for the Dow Jones series. The picture is different for the BDS test results for IBM and General Electric. In the case of IBM, the price series generated by ‘All Agents’ at 90%, 95%, and 99% levels are identically and independently distributed (the null cannot be rejected). The ‘All Agents’ series at 90% and 99% in the General Electric case characterise with the IID class too.

The BDS test results suggests that the residuals of 'All Agents' in the IBM and General Electric series seem to be less nonlinearly dependent and therefore more random than those of 'Best Agents'. Overall, in terms of nonlinear dependence, the price series generated by 'All Agents' are more random and hence more efficient. This result can be considered in line with the classical version of the Efficient Market Hypothesis.

Moreover, the BDS test statistic shows that a richer dynamic structure such as an artificial stock market populated by 10,000 agents may help to describe the findings of complex nonlinear dependence in financial market data. However, according to the econometric literature, a large part of the nonlinearity in financial data is located in their second moment. In order to capture the regularities in volatility fluctuations, I implemented the Lagrange multiplier test for the presence of ARCH effects. Evidence of the ARCH effect has been found in 22 series out of 24 in total. The two series without ARCH effects are General Electric, 'All Agents' at 90% level and IBM, 'Best Agents' at 20% level. Logically, those two series failed to reject the null hypothesis of the BDS test. As the ARCH effect is present in 22 series, I further determinated the GARCH structure of the series by using the Bayesian Information Criterion (BIC). The results are exhibited in column 9 of Tables 11.0; 12.0 and 13.0. The existence of ARCH effect suggests that volatility clustering is quite ubiquitous in all three experiments, especially the Dow Jones one. These empirical results only confirm my findings so far.

Furthermore, in order to investigate any presence of volatility clustering and persistence, I fitted GARCH(p,q) models to the first difference of log daily Dow Jones, General Electric and IBM by using backcast value of 0.7 for the initial variances.

The last column of Table 11.0 indicate that the sum of ARCH and GARCH $(\alpha+\beta)$ in Dow Jones series is very close to 1 in almost all size levels, suggesting substantial persistence of volatility clustering. There is slightly less volatility clustering in IBM and General Electric series, confirmed initially by minor fluctuations in the standard deviation coefficients. The series generated by 'All Agents' in both General Electric and IBM are less likely to experience volatility clustering due to the fact that some of their coefficients have a value of less than 1.

If the sum of $\alpha + \beta$ is less than 1, the volatility shock is time decaying and mean-reverting (specific periods of high changeability do not persist indefinitely and continually decrease to its long-run mean level at a rate equal to the sum of $\alpha + \beta$ coefficients). Some of the coefficients in my experiment were larger than 1, indicating that the variance is not stationary. This is more prevalent in the Dow Jones series generated by 'Best Agents', where volatility clustering might have a permanent effect due to the fact that the sum of the ARCH and GARCH coefficients equals unity (the conditional variance does not converge on a constant unconditional variance in the long run). In this case, I applied the IGARCH model to restrict the ARCH + GARCH term to sum to 1. The IGARCH modelling restriction was introduced in order to prevent the parameters violating a significant empirical assumption of the GARCH, model such as:

(the long-run variance constant/(1-(ARCH term+GARCH term)) (37)

If the sum of ARCH and GARCH term is greater than 1, the quantitative result above would have negative dimension.

To sum up, all my empirical results suggests that the price series generated by 'All Agents' are more likely to conform to the Efficient Market Hypothesis. All my evidence suggests that the presence of more artificial agents in my market corresponds to an enhanced variety of trading rules and greater market efficiency.

5.4 Conclusions

This chapter investigates and analyses the behavioural foundations of the stylised facts of empirical data such as leptokurtosis, non-IIDness and volatility clustering that characterise the real-world financial markets. The main contribution of this chapter is the trade-off between reality (real historical data of the three financial instruments) and calibration of the mechanisms and processes (artificial and empirical models developed) and the explanatory power of the stylised facts analysed through STGP techniques. My experimental results show that an artificial stock market populated by a small subset of best-performing agents behaves differently from a market with greater genetic diversity. Although there is no discernible difference in terms of volatility, the market based on the behaviour of 'All Agents' exhibits less herding and is more efficient than the segmented market populated by 'Best Agents'. Hence, the price formation process caused by the collective behaviour (competition and co-evolution) of the entire market is a better predictor than any small fraction of agents. This is a result of the greater genetic diversity that is presented in the total population. Enhanced diversity means more heterogeneous trading rules and behaviour leading to greater flexibility in the virtual market clearing price mechanism. In simple economic terms this refers to an improved manner in submitting market orders, balancing supply and demand, and setting the stock prices. Also, with the presence of more traders, the market is more competitive, and more of the information is reflected in the order flow. In considering my results one might draw an analogy with the situation in many physical systems where the state of the system at given time is dependent on previous states and on random noise (for example, an AR(1) or red noise system where $x(t+1) = r_x(1)x(t) + \varepsilon(t)$ with $r_x(1)$ being the autocorrelation function at lag 1 and $\varepsilon(t)$ being random noise). A high level of random noise can act to dampen the situation where the system goes out of control with high values of x being amplified through time.

In my case, having a large number of agents with different attributes acts to dampen excessive herding.In this particular case, I found no support for the Marginal Trader Hypothesis which holds that a small group of traders such as 'Best Agents' keep an asset's market price equal to its fundamental value and steer markets to efficient levels.

Moreover, in line with previous research, there is some evidence of more herding in a group of stocks than in individual stocks, but even there the magnitude of herding is far from dramatic and does not exhibit the long-run mispricing of assets and bubble formation. There is no consensus about the presence of asset price bubbles. Most academics argue that all historical bubbles can be described by fundamentally justified expectations related to the future returns on the respective underlying asset. I found evidence to support this claim, because herding behaviour detected in my experiments cannot be classified as being excessive.

Greater genetic diversity ('All Agents' groups) also means less nonlinear dependence, more unpredictability and therefore an enhanced level of randomness in the return series. Hence, these series can be considered more efficient. Unlike small groups of artificial agents where substantial volatility clustering persists, the presence of more agents has led the market to lower levels of localised bursts in the amplitude of price fluctuations. My results are consistent with the findings of Caram *et al.*(2010), who demonstrated the complex nature of the markets. The authors showed that agents of equal size are not only in market competition with those of bigger or smaller sizes, but are also in strong competition with each other at their own level.

The existence of herding behaviour in financial markets represents a classic example of the need for regulatory intervention. Herding can lead to systemic risks in financial markets. For instance, investors are likely to copy what others are doing and buy or sell what others are selling, and own what others own. Regulatory troubles caused by systemic risks are likely to occur, due to the fact that investors are rewarded by relative performance, and therefore risk-averse individuals follow the pack. From another point of view, investors are more vulnerable to be dismissed for being wrong and alone than being wrong and in company (Persaud, 2000). The growth of investment institutions over the years has increased the possibility of herding. For example, the percentage of the UK stock market held by individuals alone dramatically decreased from 54% in 1963 to 12.8 % in 2006 (Hudson and Atanasova, 2009). Herding behaviour is more likely to occur in markets dominated by institutions because managers employed by institutions operate in the market to make money and retain their jobs. Their performance is often based on large compensation packages.

The intuition behind this claim is that the profit condition-particularly a mandate to achieve a minimum benchmark return-could lead to weaker incentives for individuals to deviate from the benchmark, and hence it effectively reduces the competition among them. The lack of competition may lead to the convergence of opinions and the adoption of similar investment strategies. Hence, herding behaviour is encouraged causing potential long-term market reverses and relaxed risk-management controls (Gompers and Metrick, 2001; Wermers, 1999; Scharfstein and Stein, 1990).

5.4.1 Limitations of the use of artificial markets in measuring herding behaviour.

As Lakonishok *et al.* (1992) caution, the exact occurrence and impact of herding is difficult to measure and evaluate without precise knowledge of the demand elasticities for Dow Jones, IBM and GE. This is based on the fact that even mild herding behaviour could have large price effects.

Chapter 6

The implications of trader cognitive abilities for stock market properties

6.1 Introduction

The empirical results regarding the extent to which individuals' intelligence or market structure influence market performance and market efficiency have been controversial. There is no consensus as to whether market performance and efficiency can be improved through the structural (institutional rules and regulatory) aspects of the market or whether it is the intelligence of the traders that matters. Some research emphasises the importance of the intelligence of traders in market performance (Yeh, 2007; 2008). This perspective is contradictory to other research that claims that market structure is the main driving force of efficiency, in that Zero Intelligence (ZI) agents have the same impact as intelligent agents on market performance and, thus, the intelligence of the agents has no significant impact on efficiency relative to market structure (Gode and Sunder, 1993; Chen and Tai, 2003).

The reason behind this divergence in the literature is that the essence of individual cognition and its interaction with the market has not received the attention of financial scholars. Because environments are constructed and shaped by the cognitions of the decision makers who act in them, one can conclude that the emergent structure of the market corresponds to the structure of the individuals' cognitions. Subsequently, the emergent structure of the market interacts with the cognitions of the individuals. In her heterogeneous learning experiment with agents' limited rationality, Giannitsarou (2003) argues that the representative agent is often a good approximation of the agents in an economy. Therefore, both intelligence and market structure must be included when examining market performance. Todd and Gigerenzer (2003), in their challenge to the perfect rationality of individuals, stress the importance of including both the environment and the agents' cognition. In their discussion, they introduce ecological rationality which builds on the perspective of Simon's bounded rationality.

Applying the essence of ecological rationality can provide me with a better understanding of how both the environment and the agents' cognition are determinants and driving forces of market performance and can help me to understand the reasons for the divergence in the results when I include only one aspect, intelligence or environmental structure.

Widiputra *et al.* (2009) describe the behaviour of multiple stock markets within the framework of a dynamic interaction network (DIN) capable of developing various dynamic interactions between genes and predicting their future expressions. The DIN model revealed complex dynamic relationships between stock markets that went beyond the scope of traditional econometric models. In this chapter, I follow Yeh (2007), who supports the importance of intelligence as a driving force in market performance albeit he recommends the interpretation of the results with certain caveats because he applied the artificial data framework of the Santa Fe Artificial Market (SF-ASM).

Using real historical data of the S&P 500 and stocks in the Coca Cola Company and a different computational technique, this chapter provides a new perspective on the research in this field by explaining the reason for the divergence of the results in the literature, as neither intelligence nor market structure individually dominates in driving market performance. As expected, the results for my tests, using real data, are different from the results of Yeh (2007, 2008), who uses simulated market SF-ASM and fixed artificial data. I obtain a mixture of positive and negative impacts from individual intelligence on market performance. My empirical results indicate that using only individual intelligence provides us with an incomplete picture. Consistent with Todd and Gigerenzer (2003), I suggest that both intelligence and market structure are equally important and consistent with Yeh (2007, 2008), I suggest that further research should include both intelligence and market structure.

6.2 Background

The topics of market structure and individual rationality have been at the centre of research into markets for many years. Academics have been divided into two main camps. One camp believes that market structure is the main driving force behind market performance and argues that adequate market rules are of prime importance and individual rationality plays an insignificant role in the formation of market properties. Some argue that markets are everywhere and claims that markets emerge spontaneously when a set of required conditions, such as well-defined property rights, is fulfilled (Aslund, 1995; EBRD,1996). Others believe that even the formation of intimate social relations such as marriage is based on markets (cf. Becker, 1996). The other camp believes that individual rationality is the most important factor in market performance. These authors emphasise the role of intelligence and contend that market structures are less dominant.

In a seminal paper, Gode and Sunder (1993) investigated the relationship between individual motivations and market efficiency. The authors implemented zero intelligence (ZI) traders of a type that 'does not seek to maximise profits, and does not observe, remember, or learn'. Then, they designed two types of market structure-the first type imposes budget constraints on traders that forbid them from making a trade that will run them a loss. The budget-constrained artificial traders never engage in selling below their costs or purchasing above their values. The second type of structure does not have these financial limitations on traders. Their results suggest that the financially constrained ZI traders are able to effectively allocate market resources. It is assumed that ZI traders do not bargain in an intellectual manner, but their interactions through the market leads to a high level of allocative efficiency. Hence, the authors concluded that allocative efficiency is not dependent on individual rationality or trader strategy but that market structure is the main driving force.

Sunder (2006a, 2006b) suggested a new direction of research that is entirely concentrated on institutions and structures instead of individual human behaviour. Sunder (2007) highlighted the Sonnenschein-Mantel-Debreu theorem, which states that individual rationality is unnecessary to obtain regularity properties at the macro level.

In the early 1960's, Becker adopted the ZI approach to study a simulated market populated by irrational traders. He concluded that individual rationality does not matter as the movements of the upward slope of the market supply curve and the downward slope of the market demand curve are not affected by individual motivations. All in all, the market mechanism governed the trader's interactions (Becker, 1962). In a paper produced eleven years later, Simon (1973), emphasised the prime importance of market structure rather than the rationality of individuals. In a more recent paper, Chen and Tai (2003) investigated the effect of learning and intelligence on price dynamics and allocative efficiency. The empirical results suggested that 'mediocre' traders achieved a higher allocative efficiency (96%) for the potential social surplus compared to 88% for the 'smart' traders. Hence, intelligence is not vital for market performance.

In contrast, the proponents of the individual rationality doctrine criticised the work of Gode and Sunder (1993) on the basis of inadequate market structure. Cliff and Bruten (1997) argued that Gode and Sunder's experiment is biased on a very specific condition of symmetric supply and demand. If the condition of symmetry is violated, then the zero intelligence market not longer allocates resources in an efficient manner. Brewer *et al.* (2002) went even further, suggesting that the experiment achieved high allocative efficiency only because the process that they implemented follows a Marshallian path. This type of market dynamics was developed by Alfred Marshall, who argued that dynamics tend to follow a path that leads trading prices to the competitive equilibrium price and that the last trade is necessarily at the equilibrium. When this particular path is present, convergence to the competitive equilibrium and the forecasts of the law of supply and demand certainly follows.

Brewer *et al.* (2002) eliminated the Marshallian path in their experiment to observe at a low level of allocative efficiency in a zero intelligence market and a relatively high level in a market consisting of humans. The empirical results of Cliff and Bruten (1997) and Brewer *et al.* (2002) demonstrated the role of intelligence and its positive effect.

Hayek (1945; 1968) argued that markets can operate in an efficient manner even when the participants have a limited knowledge of the surrounding environment or of the other participants (similar to ZI). Hayek's hypothesis contains three main propositions related to market properties: 1-Real competitive markets lead to allocations that fully exhaust the available gains from trade; 2-Competition as a coordination mechanism is superior to the other mechanisms (according to Hayek competition is the main coordination driver that dominates individual activities and determinates the trader's access to societal information); 3- Competition encourages individuals to search, discover and even create new information sets (Beckmann and Werding, 1994).

It is very difficult to investigate the role of the traders' intelligence in a general stock market environment where thousands of individuals with different risk attitudes, expectations, wealth preferences and even pleasure in trading operate. Traders may adopt a fundamentalist strategy, a technical strategy or a combination of both to predict the future behaviour of stocks. The nature of these trading strategies changes over time, and individuals often implement a strategy which is influenced by their own emotions or gut feelings. Hence, it is difficult to control the behaviour of traders and analyse the individual decision-making process. Furthermore, researchers experience great difficulties in interpreting how groups of these strategies behave in a general market environment (Oberlechner, 2001). The agent-based modelling approach and zero intelligence markets, in particular, appear to be an appropriate tool for examining the market mechanism in isolation from the traders who populate the market. An important addition is that researchers are able to investigate the relationship between the market mechanism and trading activities and therefore analyse the efficiency of a trader's strategy (Ladley, 2004). In a market simulation model that I created using Altreva Adaptive Modeler, artificial traders receive information about the value of a real security and observe the history of past trades. Based on this information, they decide whether they want to buy or sell one or more units of the security. The modelling software that I use provides a rich environment to examine the implications of zero intelligence on stock market properties. Artificial traders make independent decisions creating a heterogeneous market structure (the market is populated by 10,000 boundedly rational artificial agents, each with different trading rules and behaviour).

Bounded rationality is a realistic assumption to apply to traders' behaviour. Traders have neither perfect information nor perfect information processing cognitions. In computational terms, intelligence represents an artificial agent's learning ability and adaptability to simulated market conditions. Todd and Gigerenzer (2003) provides a new and inclusive perspective on bounded rationality that places emphasis on both internal (individual cognition) and external (environment) aspects and their interactions from which emerges a definition of rationality. In this definition of rationality, an adaptive learning style contributes to individual rationality.

In this paper, I develop seven stock markets populated by artificial traders with different levels of intelligence. I empirically evaluate the price series of a group of stocks represented by the S&P 500 and individual stocks represented by the Coca-Cola Company to investigate the following topics:

(i) Whether the efficiency of markets is primarily a function of their rules or whether the effect of human motivations and cognitive abilities dominates. How much of the market efficiency is attributable to individual rationality, and how much is attributable to market discipline? That is testing the Hayek's hypothesis.

(ii) Whether the price series generated by the most intelligence agents is more likely to conform to the Efficient Market Hypothesis, and therefore be more efficient?

Yeh (2007) developed a model allowing greater flexibility than previous artificial stock market models for the traders in choosing between buying and selling assets. Most of the empirical findings of the author were consistent with the theories of Cliff and Bruten (1997) and Brewer *et al.* (2002). The introduction of intelligence resulted in reduced levels of price volatility and supported the process of discovery of the intrinsic value of the stock. However, enhanced intelligence caused an increase in the price and the return volatility of the assets.

In another paper, Yeh (2008) investigates the effect of speculation, under various intelligence levels, on price discovery and market efficiency. For the sake of analysis, the author enabled each trader to individually determine the intrinsic value of the assets and prohibited communication between agents. The paper concluded that the impact of speculation on the market properties depends on intelligence levels. The introduction of intelligence improved price discovery and enhanced market efficiency. Hence, intelligence plays a prime role in the influence of market performance. Both papers produced by Yeh are a good starting point for further research. However, I suggest several extensions to Yeh's approach based on the following important factors:

-The modelling software I use for my experimental tests uses a special adaptive form of STGP. In this form, the process of estimating the agent's fitness does not include any re-execution of their trading rules based on historical data. This is due to the fact that the artificial traders have already executed their trading rules on the same historical data set once and the software is looking only at the realistic returns that they have already made, rather than any hypothetical returns that agents could have made if there were sent back in time. Also, while in conventional GP, the agent's trading rules are evaluated by the same fitness function in every generation, Adaptive Modeler evaluates the fitness of agents through a dynamic fitness function. The dynamic nature of the fitness function enables the return estimation period to move forward and include the most recent historical quotes. Another important difference between conventional GP and Adaptive Modeler is that conventional GP replaces the entire genetic population through crossover and mutation techniques in every generation. Adaptive Modeler replaces only a small proportion of the entire population at a time in order to enable the population to change gradually, which is essential for maintaining a certain degree of model stability (Witkam, 2011).

-I use a greater number of artificial agents. While both models produced by Yeh in 2007 and 2008 consist of only 100 traders, I employ 10,000 traders. A larger population means increased model stability and reduced sensitivity to random factors. Various experiments using the software demonstrated that bigger populations reduce sensitivity to random factors (results of multiple runs will show less variation).

The presence of substantially more artificial agents creates a competitive environment where more different trading rules compete and evolve in parallel simultaneously. A greater number of agents open up the opportunity to implement a wider variety of trading strategies programmed within the agents' trading rules (Witkam, 2011).

-Rather than using a fixed intrinsic value for the stock, I feed the software with real historical asset prices. I use real prices to prevent the preliminary development of herding conditions (herding behaviour by design). Cipriani and Guarino (2002) argue that when the price is fixed, individuals tend to disregard their own information and strictly follow the decisions made by previous agents, resulting in herding behaviour.

-I have developed seven different markets instead of six (Yeh's case). The availability of an extra market allows me to experiment with a wide variety of tree depths. The genome depth represents the highest number of hierarchical levels in an agent's genome (trading rule) indicating its complexity. The depth enables the artificial agents to look back further in history and develop more complex and sophisticated trading rules. I gradually increase the depth of the tree from zero intelligence to a maximum of 20. While Yeh (2007, 2008) created a market with maximum depth of 15 in terms of the functions that a trader can perform, my Market G has been set with a maximum genome depth of 20 based on maximum software availability.

- Another important difference is that the orders of each agent and also their bid/ask price in my experiment are determined by their own specific trading rule, not according to a predefined general model described by Yeh (2007), Yeh (2008) and Yen and Young (2010).

6.3 Artificial stock market structure for this particular experiment

For the purpose of my experiments I consider the degree of intelligence to be proportional to the level of complexity of the trading rule. I control the complexity of trading rules by varying the maximum genome depth. This framework is similar to Yeh's (2007) model, where the traders are not able to observe each other during the process of developing forecasting rules. Individuality ensures that the intelligence of each artificial agent is not manipulated by any other agent's intelligence via imitation. Table 9.0 illustrates the main parameters of the GP-based artificial stock market model.

Artificial stock market parameters	
Total population size (agents)	10,0000
Initial wealth(equal for all agents)	100,000
Average bid/ask spread	0.01%
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**	0 (Market A); 5 (Market B); 6 (Market C); 8 (Market D); 10 (Market E); 15 (Market F); 20 (Market G)
Minimum initial genome depth	2
Maximum initial genome depth	5
Breeding cycle frequency (quote)	1
Minimum breeding age (quote)	80
Initial selection type	random
Parent selection (percentage of initial selection that will breed)	5%
Mutation probability (per offspring)	10%
Total number of quotes processed- S&P 500	17,353
Total number of quotes processed-CC	17,355
Trading hours (open/close)	09:30/16:00
Seed generation from clock	Yes
Creation of unique genomes	Yes
Offspring will replace the worst performing agents of the initial selection	Yes

Table 14.0 Artificial Stock Market Parameter Settings.

*Maximum genome size measure the total number of nodes in an agent'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 an agent's genome (trading rule). The depth of a trading rule can be an indicator of its complexity.

I have run 1,250 simulations and report the average values in all tables. I study traders' behaviour within the context of the artificial stock market populated by 10,000 boundedly rational agents. All of the agents are characterised by adaptive learning behaviour represented by the genetic programming algorithm. The artificial traders all have different trading rules. Hence, the agents in the model are not orientated towards predetermined formation of strategies, and therefore are free to develop and continually evolve new trading rules.

I develop seven different markets, which are denoted as Market A to Market G. Market A is populated by zero intelligence traders who randomly form bid and ask prices if they are buyers or sellers. Zero intelligence agents are implemented by enabling only the genes 'RndPos' and 'RndLim'. With these genes the position advice and order limit price are established randomly. 'RndPos' is a function which returns a random position value ranging from -100% to 100% sampled from a uniform distribution. 'RndLim' represents a function which returns a random limit price generated by a method partly based on Raberto *et al.* (2001) with the difference that m=1 instead of 1.01, so that no spread is added/subtracted for increasing the likelihood of an order being executed. The rationale behind this logic is that at the time the 'RndLim' gene is evaluated, it is unknown whether the agent will place a buy or sell order and thus whether I should add or subtract the spread. Raberto's *et al.* (2001) method works by taking the last closing price from a market which has been initially chosen at random and then multiplying it by a normally distributed random value with $\mu = 1$ and $\sigma = 3.5 \times \sigma_m$, where σ_m is the standard deviation of the log returns of the last 20 quotes of the chosen market (Witkam, 2011). The artificial traders in Market A do not learn strategies or even examine the market-their behaviour is completely random. These are agents with diffuse beliefs who do not remember or learn.

The zero intelligence market operates under closed economy conditions-there is no breeding process in place and no broker commissions so that, no money is going in or out of the population during the process of evolution. Hence, the total amount of cash in the zero intelligence market stays constant and thus the average amount of cash per agent.

The total net number of shares in the model is zero and will stay zero because no shares will be added or removed through agent replacement. This is because agents initially don't get assigned any shares but only cash but can then either go long or short in the security causing the total net number of shares to remain zero. This means that changes in the share price have no implications on the total wealth of the population in Market A, because any potential profits from long positions are offset by losses from short positions of other agents. Hence, under conditions of zero intelligence, the total wealth of the population in Market A (and thus the average wealth per agent) will stay constant. The traders' forecasting rules in markets B to G are represented and evolved by GP. The traders in markets B to G possess different levels of intelligence. The insight from zero intelligence traders comes not directly from the results of their simulation, but rather from the difference between those results and the results of more intelligent traders populating the other six markets. The actual difference between the levels of intelligence is modelling the complexity of the forecasting expectations that the artificial agents are able to evolve. The complexity of the forecasting expectations is measured by the depth of the GP tree. The maximum tree depth is 5 in Market B; 6 in Market C; 8 in Market D; 10 in Market E; 15 in Market F and 20 in Market G (Table 14.0).

Artificial agents in each market are heterogeneous in their genome depth as only the maximum genome depth is specified per market. Hence, in each market, the genome depth of particular agents varies between minimum and maximum genome depth. Moreover, even agents with equal genome depth are heterogeneous because their trading rules are different and can cause very different trading behaviour. The various genome depths affect the memory length of traders. Greater genome depth means that more complex trading rules can be developed that look back further in history. Hence, the artificial traders in Market G, use the widest range of information available. According to LeBaron (2004), traders use different amounts of past information to evaluate trading strategies, and therefore they possess various memory lengths when evaluating forecasting rules. Additionally, markets composed of agents with different intelligence levels offer the opportunity to analyse market efficiency in depth, rather than examining whether intelligence improves market properties where zero intelligence is present.

The agents' trading rules evolve and adapt through a breeding process. Breeding, in essence, is a process of creating new artificial traders to replace the poorly performing ones. It involves the selection of well performing traders and the production of new trading rules by a recombination of the parent genomes trough crossover and mutation operations (Witkam, 2011).

6.4 Simulation results

The mean absolute error (MAE), the return series, the average stock price and the trading volume of the S&P 500 and stocks in the Coca-Cola Company are examined using standard statistical methods. The empirical differences between the variables have been investigated to determine market efficiency at different levels of intelligence. Interestingly, there is a substantial difference between the econometric properties of the groups of stocks represented by the S&P 500 and an individual stock as represented by those in the Coca-Cola Company.

5.4.1 Whether the efficiency of markets is primarily a function of their rules or whether the effect of human motivations and cognitive abilities dominates. How much of the market efficiency is attributable to individual rationality, and how much is attributable to market discipline? That is testing the Hayek's hypothesis.

(i) Mean Absolute Error (MAE) analysis.

Investigating the mean absolute error (MAE) is the most efficient way to measure the effectiveness of intelligence. Willmott and Matsuura (2005) argue that evaluations and inter-comparisons of average model-performance error should be based on MAE. The MAE is an average of the absolute errors $e_i = f_i - y_i$, where f_i is the prediction and y_i is the true value in the mean absolute error equation (Coyle, 1988):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i| \quad (38)$$

In terms of the S&P 500 analysis, Table 15.0 clearly indicates that the MAE significantly decreases when the intelligence level is raised.

S&P 500**							
Type	***Market A***	***Market B***	***Market C***	***Market D***	***Market E***	***Market F***	***Market G***
MAE	7.54*	5.98*	5.43*	5.37*	5.24*	5.12*	5.06*
SD	7.76	5.99	5.92	5.91	5.82	5.77	5.58
SK	1.21	1.39	1.44	1.44	1.45	1.47	1.48
K	3.77	4.73	4.75	4.77	4.79	4.81	4.91
Coca Cola*							
MAE	1.08*	0.89*	0.85*	0.79*	0.77*	0.72*	0.69*
SD	0.60	0.58	0.50	0.49	0.42	0.42	0.37
SK	1.44	1.54	1.59	1.66	1.73	1.83	1.94
K	5.43	5.47	5.49	5.82	5.89	6.01	6.03

MAE is the value of the Mean Average Error. The * means that, based on the *t* statistic, the value is statistically different from the previous value at the 5% significance level. For instance, 7.54 in Market A is statistically different from 5.98 in Market B at the 5% significance level. **measures the average values based on 1,250 simulations.

Table 15.0 Econometric statistics of S&P 500 and Coca Cola Company Mean Absolute Error.

The empirical results suggest that the MAE decreases from 7.54 in Market A to 5.06 in Market G, which is populated by the most intelligent traders in my experiment. The important role of intelligence is confirmed by a substantial decrease in the standard deviation value- from 7.76 in Market A to 5.58 in Market G. The Coca Cola empirical results in Table 15.0 experience the same pattern with- the MAE decreasing from 1.08 in Market A to 0.69 in Market G. Again the standard deviation of the MAE decreases when the level of intelligence is raised. All results are statistically significant at the 5% level confirming my findings that more intelligence reduces the mean absolute error and improves the forecasting function. Thus my results indicate that increased intelligence has a positive effect on the forecasting function. My findings are consistent with Good *et al.* (1999), which argue that to take advantage of learning agents, those agents must be designed to accommodate dynamic learning habits rather than random (zero intelligence) strategies. In their experiment, the model with the highest accuracy (lowest MAE) avoided making large errors and performed better than random.

(ii)Stock returns analysis.

I use logarithmic returns to estimate the returns of the two financial instruments:

$$R_t = \ln(P_t) - \ln(P_{t-1}) \quad (39)$$

Where P_t is the price of the stock in period t.

For the S&P 500 the standard deviation increases from 0.004186 in the zero intelligence market to 0.010715 in Market G, which is populated by the most intelligent artificial traders (Table 16.0). The Coca Cola Company return series experiences a similar increase in volatility as the intelligence level increases in the market (from 0.011624 in Market A to 0.020120 in Market G, Table 16.0). My findings show that in terms of return volatility, intelligence plays a more negative role in the market. Furthermore, the volatility of Market G is much closer to the real S&P and Coca Cola volatility reported in Table 17.0. My findings are consistent with LeBaron (2004), who constructed an artificial stock market populated with traders with different memory lengths between 6 months and 30 years and compared the simulations with data from the S&P 500. The author demonstrated that the heterogeneous memory framework amplified volatility.

S&P 500**							
Type	***Market A***	***Market B***	***Market C***	***Market D***	***Market E***	***Market F***	***Market G***
R	0.000047*	0.000361*	0.000366*	0.000377*	0.000498*	0.000514*	0.000602*
SD	0.004186	0.005611	0.007035	0.007534	0.008311	0.009932	0.010715
SK	-2.94	0.02	0.34	0.47	0.55	0.89	1.09
K	12.64	23.57	31.04	39.38	47.37	49.28	51.22
ADF•	-87.19	-112.04	-129.22	-129.58	-121.77	-130.36	-126.88
ARMA	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)
BDS	48.34	35.56	27.82	30.32	30.54	31.61	31.76
GARCH	(1,1)	(5,2)	(2,1)	(2,1)	(4,1)	(3,3)	(2,1)
Coca Cola**							
R	3.41e-05*	0.000112*	0.000134*	0.000168*	0.000172*	0.000191*	0.000208*
SD	0.011624	0.011883	0.011990	0.012598	0.017348	0.019983	0.020120
SK	-21.23	-18.29	-7.18	-9.08	-8.68	-7.20	-9.55
K	1.69	5.15	5.77	6.31	6.96	7.38	7.44
ADF•	-99.16	-116.14	-125.01	-125.07	-127.05	-126.93	-126.91
ARMA	(1,1)	(3,1)	(1,1)	(1,2)	(2,1)	(1,1)	(1,1)
BDS	67.54	35.81	31.20	33.15	34.32	40.07	35.51
GARCH	(3,2)	(1,2)	(3,1)	(6,6)	(3,2)	(3,3)	(5,2)

• The MacKinnon (1996) one-sided critical value for rejection of the null hypothesis of a unit root at 5% level is -3.410060. *based on the *t* statistic, the value is statistically different from the previous value at the 5% significance level. **measures the average values based on 1,250 simulations.

Table 16.0 Econometric statistics of S&P 500 and Coca Cola return series.

Financial Instrument	*Mean*	*R*	*SD*	*SK*	*K*
S&P 500	504.74	0.000391	0.010007	0.95	50.04
Coca Cola	65.69	0.000201	0.020114	-8.14	7.40

SD measures the Standard Deviation of S&P 500 stock prices, ***SK*** represents the Skewness, and ***K*** value is the Kurtosis.

Table 17.0 Econometric statistics of S&P 500 and Coca Cola Company real return series.

Various studies related to the fluctuations in stock prices reveal that the distributions of returns and price series have fat tails, which characterise non-Gaussian distributions (Pagan, (1996); Mandelbrot (1963, 1997). Bollerslev *et al.* (1992) argue that leptokurtosis is still present even after empirically allowing for heteroskedasticity. The conventional coefficient for kurtosis is given by:

$$KR = E\left(\frac{y_t - \mu}{\sigma}\right)^4 - 3 \tag{40}$$

Where $\mu = E(y_t)$ and $\sigma^2 = E(y_t - \mu)^2$, and expectation E is taken with respect to the cumulative distribution function F (Kim and White, 2004). If $KR = 0$ returns follow Gaussian distribution and if $KR > 0$ returns are characterised by fat tails.

The fat-tail property (excess kurtosis) is obvious in both the S&P 500 and the Coca-Cola Company return series. I compare the excess kurtosis of the zero intelligence market and the intelligent traders markets (B-G) with the real return series of the two financial instruments. The comparison is made in order to determinate which market most resembles the real return series. The real return series (Table 17.0) reported excess kurtosis of 50.04 for the S&P 500 and 7.40 for the Coca Cola Company. A comparison with excess kurtosis values listed in Table 16.0 suggest that markets composed of more intelligent agents resembles the real return series the most. Moreover, Table 16.0 shows that a zero intelligence market is characterised by substantially lower excess kurtosis than a market populated by intelligent traders for both the S&P 500 and the Coca Cola Company return series. The excess kurtosis of the S&P 500 significantly increases from 12.64 in the zero intelligence market to 51.22 in Market G populated by the most intelligent agents.

In terms of the Coca Cola Company return series, the excess kurtosis increases from 1.69 in Market A to 7.44 in Market G.

Overall, all of the markets populated by the most intelligent agents experience more pronounced excess kurtosis in comparison with the zero intelligence markets for both the index (the S&P 500) and the security (the Coca Cola Company). This result suggests that the markets operating with more intelligence agents under a genetic based mechanism better replicate the stylised facts of financial returns that are observed in the real financial markets.

My findings are consistent with LeBaron *et al.* (1999) who performed an experiment with fast and slow learning agents and demonstrated that fast learning agents returned higher excess kurtosis which is in line with real asset returns.

(iii)Stock price analysis.

Table 18.0 indicates that the average stock price $\left(\overline{P}\right)$ of the S&P 500 increases monotonically with increasing agent intelligence from 503.11 in Market A, populated by zero intelligence traders, to 504.72 in Market G, which is composed of the most intelligent traders. The $\overline{P}$ of the Coca Cola Company demonstrates similar trend of an monotonic increase- from 64.99 in Market A to 65.68 in Market G. For both securities- all the differences in means are significant at the 5% level.

S&P 500**							
Type	***Market A***	***Market B***	***Market C***	***Market D***	***Market E***	***Market F***	***Market G***
$\overline{P}$	503.11*	504.57*	504.60*	504.65*	504.68*	504.70*	504.72*
SD	79.87	80.63	80.98	81.50	81.87	82.33	82.73
SK	0.51	0.52	0.52	0.53	0.53	0.54	0.54
K	3.66	3.67	3.68	3.69	3.70	3.71	3.78
Coca Cola**							
$\overline{P}$	64.99*	65.51*	65.59*	65.62*	65.63*	65.66*	65.68*
SD	24.13	24.14	24.17	24.17	24.19	24.20	24.23
SK	1.91	1.92	1.93	1.93	1.94	1.94	1.95
K	6.79	6.81	6.88	6.92	6.97	7.00	7.05

*based on the *t* statistic, the value is statistically different from the previous value at the 5% significance level. **measures the average values based on 1,250 simulations.

Table 18.0 Econometric statistics of S&P 500 and Coca Cola Company stock prices.

Yeh (2007) demonstrated that the properties of price dynamics differ substantially when examined at different levels of intelligence. In his experiments the price dynamics in a zero intelligence market were far from the homogeneous rational expectations equilibrium (REE) whereas the average price in the markets populated by intelligent traders was closer to the homogeneous REE. In traditional economic terms, the REE could be achieved if traders fully exploit all properties of the price dynamics. Given the validity of this claim, the intelligent trader should always play a stabilising role on the market by supporting the process of price discovery. A comparison between prices generated by zero intelligent traders of the two financial instruments and the real-life historical data based on 17,353 observations, reveal that both Markets A are characterised by underpriced securities. The average price of the S&P 500 in Market A is 503.11 (Table 18.0) and the real historical average price reported in Table 2.0 is 504.74. The zero intelligence market of the Coca Cola Company shows the same phenomenon of an- underpriced average stock price of 64.99 in comparison with the average real historical price of 65.69. The steady upward price trend in the average price of the S&P 500 and Coca Cola Company stocks when traders are endowed with more intelligence suggest a price process where prices move closer to more realistic levels. The price generated by the most intelligent traders of the S&P 500 in Market G is 504.72, which is a close match to the real historical price of 504.74. I observe the same steady upward trend in Market G of the Coca Cola Company- the generated price is 65.68 and the real historical price in Table 18.0 is 65.69. Although the price rises are characterised by a fairly minor increases, the empirical data indicate the positive role of intelligence in price discovery in individual as well as in a group of stocks.

However, Table 18.0 illustrates that the standard deviation of the price of the S&P 500 and Coca Cola Company also increases at a 5% level of significance in all markets. In this particular case, I observe the negative effect of intelligence because the price volatility worsens when the intelligence level is raised. This finding is consistent with the study of Yeh (2007) which also documents a significant increase in volatility.

These results paint a mixed picture of the implications of trader cognitive abilities on stock market properties. In terms of the stock prices, the empirical results presented so far suggests that intelligence plays a significant role in price discovery and a rather negative one on stability of asset prices. This finding is consistent with Othman (2008), who argues that human qualities do not always play a significant role in price formation.

My findings do not support the Hayek Hypothesis. Market A which is populated by zero-intelligence agents with diffuse beliefs does not generate correct market prices. It is been demonstrated that Market A cannot function properly when the participants know very little about the environment (they receive limited amount of information about the historic prices based on shorter genome depth) resulting in undervalued financial instruments.

(iv) Trading volume analysis.

Table 19.0 illustrates that the average trading volume ($\bar{V}$) decreases monotonically from 1,254,759 in Market A to 999,373 in Market G of the S&P 500 as the level of intelligence is raised. This finding suggests that trading opportunities are substantially reduced if the traders are endowed with more intelligence. The average trading volume of the Coca-Cola Company also shows a monotonically decreasing trend when the level of intelligence is raised- from 1,590,813 in the zero intelligence market to 1,553,273 in the market composed of the most intelligent agents.

S&P 500**							
Type	***Market A***	***Market B***	***Market C***	***Market D***	***Market E***	***Market F***	***Market G***
$\bar{V}$	1,254,759*	1,230,086*	1,194,022*	1,181,622*	1,170,452*	1,004,878*	999,373*
SD	111.36	110.03	108.28	107.21	107.00	106.62	105.23
SK	0.81	0.87	0.89	0.89	0.90	0.91	0.95
K	2.99	3.17	4.72	5.01	6.22	7.05	8.14
Coca Cola**							
$\bar{V}$	1,590,813*	1,587,933*	1,580,012*	1,572,390*	1,563,351*	1,560,120*	1,553,273*
SD	170.21	168.30	167.25	167.00	165.54	164.89	164.00
SK	0.12	1.66	1.66	1.68	1.69	1.70	1.72
K	2.54	5.73	6.55	6.88	7.00	7.33	7.92

*based on the *t* statistic, the value is statistically different from the previous value at the 5% significance level. **measures the average values based on 1,250 simulations.

Table 19.0 *Econometric statistics of S&P 500 and Coca Cola Company trading volume.*

As outlined in section (ii) the standard deviation of the return series (Table 16.0) increases when traders are equipped with more intelligence which is also similar to Yeh (2007). Thus as the intelligence level raises the average trading volume decreases and the standard deviation of returns increases. This seems an unexpected relationship given the well-reported stylised fact of a positive relationship between volume and volatility reported in, for example, Cont (2009).

There are a number of potential explanations for my findings and these are worthy of future investigation. Initially, for a given level of intelligence, my results do not preclude a positive relationship between volume and volatility and so are not necessarily inconsistent with those reported in Cont (2009).

Another reason for the observed volume- volatility relation as intelligence increases may be that, in a similar way to financial futures markets, I do not have a short-selling restriction in my experiments and traders can short-sell up to 100 per cent of their wealth. Fung and Patterson (1999) argue that the short-selling restriction in equity markets is the key factor that drives the positive relationship between volume and return. Their analysis suggests that trading volume is not necessarily positively correlated to asset returns indicating that short-selling generates different impacts on price movements. Jennings *et al.* (1981) argue that under short-selling restrictions placed in equity markets, trading volume may not be closely related to return volatility. The increase in return volatility and the decrease in trading volume in the presence of short-selling is well documented in the literature. Aitken *et al.* (1998) demonstrated that transparent short sales increased return volatility. Henry and McKenzie (2006) reported that the market displays greater volatility following a period of short-selling activity. Hong *et al.* (2012) argue that heavily shorted stocks have increased price sensitivity to news flow and thus volatility. Schwartz and Norris (2012) pointed out that small firms experience significantly higher volatility when short-selling intensifies.

A third potential explanation for the volume-volatility relationship is related to the different type of traders in my experiments. This is consistent with the findings of Bessembinder and Seguin (1993), who suggest that the volatility-volume relationship in financial markets depends on the type of trader.

The authors investigated the relationship between volume, volatility, and market depth in eight physical and futures financial markets and concluded that linking volatility to total volume does not represent all market information. In a study similar to mine, Daigler and Wiley (1999) examined whether specific types of traders distinguished by the information they possess have any effect on the volume-volatility relation. Their study differentiated traders by either the amount of information they hold or the dispersion of expectations they form based on that information.

Interestingly, the authors demonstrated that the positive correlation between trading volume and return volatility found in some studies such as Tauchen and Pitts (1983), Gallant *et al.* (1992), Lee and Rui (2002), Cont (2007) and LeBaron and Yamamoto (2007) is driven by the general public (off-the-floor traders) rather than floor traders.

Daigler and Wiley argue that the general public cannot distinguish the volume related to liquidity demand from the volume due to a change in fundamental value. Thus volume gives an inaccurate information signal combined with a greater dispersion of expectations leading to an increase in volatility. The most important finding of their study is that floor traders often exhibit an inverse relation between volatility and volume. Hence, the authors highlighted that using different trader categories is a better way to understand the relation between volatility and volume. My empirical results are also broadly in line with Shalen's (1993) model, which explains the volume-volatility relation as dependent on the dispersion of different traders' expectations and the negative volume-volatility correlation reported by Wang (2004) , who found that trading volume contributes negatively to the subsequent volatility. The presence of substantial trading volume in my zero intelligence markets could also be explained by the models developed by Harris and Raviv (1993) and Shalen (1993) where uninformed traders' dispersion of beliefs creates excess volume.

6.4.2 Are the price series generated by the most intelligence agents more likely to conform to the Efficient Market Hypothesis, and therefore be more efficient?

According to the EMH, financial markets are efficient when the return series $\{r_t\}$ are unpredictable. On the other hand, the return series experience the presence of unpredictability when the returns are identically and independently distributed (IID). To determinate whether the series are characterised by IIDness, I applied the BDS (Brock, Deche, and Scheinkman, 1996) test to the residual series.

I first applied the Augmented Dickey-Fuller (ADF) test to investigate the presence of a unit root. My ADF test settings include running a regression of the first difference of the log price series against the series lagged once (a sufficient condition to eliminate autocorrelation in the residuals) combined with a drift and a time trend.

As the econometric results of the ADF test suggests, the null of the presence of a unit root has been rejected in both the S&P 500 and the Coca-Cola return series (Table 20.0). Hence, the financial time series generated by artificial agents are stationary at a 95% significance level, indicating that the process is orientated around a constant long-term mean and has a constant variance independent of time. Lee *et al.*(2010) discovered stationary price series in 32 developed and 26 developing countries. Stationarity can be explained by microstructure biases in low-priced stocks (Conrad and Kaul, 1993 and Ball *et al.*,1995), the role of leverage (Chan, 1988 and Ball and Kothari, 1989), and the importance of stock market size and associated risk factors (Zarowin, 1990 and Richards, 1997).

Then, I proceeded further to filter the linear process. Table 20.0 reports the *ARMA (p,q)* process gained from the return series r_t . In terms of the ARMA statistics, there is no difference between the return series generated by the zero intelligence traders and those generated by the agents equipped with different levels of intelligence. Hence, the zero intelligence and the more intelligent traders are equally likely to converge to the Efficient Market Hypothesis. While the vast majority of other test results are either AR(1) or MA(1), some of the Coca Cola return series exhibit a higher order of linear dependence such as ARMA(2,1) ARMA (1,2) and ARMA(3,1).

S&P 500*							
Type	***Market A***	***Market B***	***Market C***	***Market D***	***Market E***	***Market F***	***Market G***
ADF•	-87.19	-112.04	-129.22	-129.58	-121.77	-130.36	-126.88
ARMA	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)
BDS	48.34	35.56	27.82	30.32	30.54	31.61	31.76
GARCH	(1,1)	(5,2)	(2,1)	(2,1)	(4,1)	(3,3)	(2,1)
Coca Cola*							
ADF•	-99.16	-116.14	-125.01	-125.07	-127.05	-126.93	-126.91
ARMA	(1,1)	(3,1)	(1,1)	(1,2)	(2,1)	(1,1)	(1,1)
BDS	67.54	35.81	31.20	33.15	34.32	40.07	35.51
GARCH	(3,2)	(1,2)	(3,1)	(6,6)	(3,2)	(3,3)	(5,2)

• The MacKinnon (1996) one-sided critical value for rejection of the null hypothesis of a unit root at 5% level is -3.410060.
*measures the average values based on 1,250 simulations.

Table 20.0 Econometric properties of S&P 500 and Coca Cola return series.

I then investigated whether the price series are identically and independently distributed (IID) over time in order to capture any presence of unpredictability, and therefore efficiency. I have started the IID determination process by estimating the most appropriate ARMA (p,q) model and fit it to the data to eliminate all linearity from the sample. Once I have identified the linear element of the series, any other elements of the series should be non-linear. I applied the BDS test to the ARMA residuals to test for remaining dependence. Under the null hypothesis, the series are identically and independently distributed (IID). The fitted model is the best linear ARMA time series model and the rejection of the null indicates a non-linear time series process.

The most important BDS test parameters are the distance parameter 'ε' and the embedding dimension (DIM). In qualitative terms my results prove not to be sensitive to the choice of epsilon and DIM parameters. Therefore, I only report the experimental results with the distance parameter 'ε' equal to 0.7 standard deviations and the embedding dimension of 6. Based on the test results in Table 20.0, the null hypothesis that the residuals of the S&P 500 return series are IID is significantly rejected in all sub-periods. The result indicates non-linear dependence in the S&P 500 and Coca Cola return series. In terms of non-linear dependence, there is no difference between the zero intelligence traders and the intelligent ones. Hence, I once again observe that intelligence plays an insignificant role which is consistent with the findings of Othman (2008).

Moreover, the BDS test statistics shows that a richer dynamic structure such as an artificial stock market populated by 10,000 agents can help to describe the findings of complex non-linear dependence in financial market data.

To capture the regularities in volatility fluctuations, I implemented the Lagrange multiplier test for the presence of ARCH effects. Evidence of the ARCH effect is found in all the S&P 500 and Coca Cola series. Because the ARCH effect is present in all series of both financial instruments, I further determinate the GARCH structure of the series by using the Bayesian Information Criterion (BIC). The existence of the ARCH effect suggests that volatility clustering is quite ubiquitous in all markets.

6.5 Conclusions

Studies that support market structure as the driving force behind market performance have been significantly challenged in the literature. The studies of Yeh (2007, 2008) support the significance of individual intelligence in market performance with the caveat that he uses SF-ASM rather than market prices. I apply the approach of Yeh (2007) by using Adaptive Modeler's special adaptive form of STGP to create artificial traders with different learning styles. I apply this approach using real data (S&P 500 and Coca-Cola). I investigate market properties such as mean absolute error (MAE), stock returns, stock price and trading volume to determine the role of intelligence.

The MAE and its standard deviation values for both the S&P 500 and the Coca-Cola Company indicate the positive implications of intelligence because the markets populated with more intelligent artificial traders show reduced levels of MAE and improved forecasting functions, respectively. In terms of stock prices, traders endowed with more intelligence do help price discovery but the relationship of intelligence with price stability seems to be ambiguous. The price stability of the S&P 500 and the Coca-Cola Company deteriorates if the traders are more intelligent. This finding is consistent with the findings of Chen and Tai (2003), who argue that smart traders do not necessarily bring improvements to the market. Trading volume comparison reveals that trading opportunities are reduced if traders are equipped with more intelligence. This result clearly demonstrates that intelligence does not necessarily improve market properties. Trading volume comparison reveals that trading opportunities are reduced if traders are equipped with more intelligence. Moreover, I found as the intelligence level raises the average trading volume decreases and the standard deviation of returns increases.

In terms of the stock returns, all markets of the index and the security that were populated by the most intelligent agents experience a significant increase of excess kurtosis in comparison with the zero intelligence markets. This empirical result suggests that all of the markets operating under a genetic based mechanism better replicate the stylised facts of financial returns than markets composed of traders with random behaviour.

In terms of the econometric tests such as the ARMA statistics and non-linear dependence (the BDS test), there is no difference between the zero intelligence traders and the more intelligent ones. Hence, more research needs to be performed to better understand the formation and the actual role of individual decision rules, the structure of the decision task, and the surrounding decision-making environment in determining the process of market efficiency. If human qualities such as learning and reasoning are driving prices towards efficiency, then, Othman (2008) posed the question of 'why should they stop at an inefficient result'? It appears to be very likely that no single factor such as individual motivation, market structure or the level of familiarity with a particular task will suffice to explain in great detail the observed variation in the market performance. While market structure appears to have more observable properties, individual behaviour appears to be more complex and less predictable.

The conclusions as to whether intelligence plays a significant role in market performance are somewhat blurred. There is no definite answer as to whether the market mechanism or the trading strategies (individual rationality) matters the most. This conclusion is consistent with Todd and Gigerenzer's (2003) findings that both intelligence and market structure are highly important. I suggest a new perspective for future research that includes investigating the sensitivity of models to parameters determining intelligence and market structure.

6.5.1 Limitations of the use of artificial tradesr in evaluating cognitive abilities.

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 major drawback of this particular experiment in comparison to experiments with real-life humans is the use of specific genome length to manipulate memory length as proxy for cognitive ability and intelligence.

The main limitation comes from the fact is that computational memory is rather short-term in nature. Short-term memory is the type of memory that characterise with a sort of workspace that is used to manipulate information (Seamon and Kenrick, 1994). However, Engle *et al.* (1999) demonstrated a weak relationship between short-term memory and human intelligence indicating the limitations of cognitive abilities of artificial traders.

Chapter 7

Using Strongly Typed Genetic Programming for profitable stock market forecasts.

7.1. Introduction

Forecasting stock market returns is not an easy task because many factors, including political events, economic conditions, and investor's trading expectations, influence securities. Stock market predictability remains a topic of continuous controversy. Some academics argue that predictability is due to some irrational phenomenon, such as noise trading or speculative bubbles, which makes stock prices deviate from tier fundamental values. This deviation is associated with negatively autocorrelated and therefore predictable returns. Both researchers and investors have made incredible efforts to predict the future movements of the stock market and develop trading strategies to transfer the forecasts into profits. The application of Genetic Programming (GP) to stock market forecasts has been rather scarce with fewer than thirty papers found in the literature. At the same time, the application of Strongly Typed Genetic Programming (STGP), developed by Montana in 1995, to financial time series forecasts has not been published yet.

I developed forecasting models on STGP basis to present completely new evidence suggesting that there is limited evidence of random walk in financial time series and then describe a concrete measure of predictability of stock prices based on one-day-ahead forecasts. In the market simulation models that I created, artificial traders receive information about the value of a real security and observe the history of past trades. Based on this information, they decide whether they want to buy or sell one or more units of the security. The STGP technique provides a rich environment to examine the heterogeneity and complexity of the trading rules as well as the implications of trader cognitive abilities on forecasting accuracy.

I developed two different stock markets each populated by 10,000 artificial traders and one stock market populated by 'Low Intelligence Traders'. In the first stock market similarly to the vast majority of the previous forecasting literature I did not take into account transaction costs. In the second stock market, I included 1.25 per cent or 125 basis points as transaction costs. Finally, in the third stock market, I created a market composed of 'Low intelligence Traders' with a half genome depth of 10 (out of 20 maximum) in order to investigate their forecasting abilities.

I then compared the four different in-sample and out-of-sample STGP models to traditional econometric forecasting models, such as Box-Jenkins (ARIMA) and the Holt-Winters exponential smoothing, based on five performance measures to investigate the following:

I.Whether, to what extent and in which form the stock returns in excess of the risk free rate were indeed predictable and profitable. This was done by measuring and quantifying the exact level of generated profit, taking into account the transaction costs.

II. Experimentally testing the Optimism Principle developed by Picard and Cook (1984).

III. Whether intelligence and trader cognitive abilities matter in the formation of more sophisticated trading rules. This was done by testing the Adaptive Expectations Hypothesis in the laboratory artificial stock market settings populated with 'Low Intelligence Traders' in the fourth stock market.

No known study accesses the heterogeneity of trading rules and trader cognitive abilities when transaction costs are taken into account. Specifically, the major contributions of this chapter are:

•To demonstrate and verify the predictability and profitability of stock returns when transaction costs are included.

•I have performed very rare laboratory experimental tests of the Adaptive Expectations Hypotheses as well as the Optimism Principle.

•To provide evidence that the level of traders' intelligence plays a significant role in the process of formation of more sophisticated trading rules and improvements in the quality of forecasts.

My research confirms that stock returns in excess of the risk free rate are indeed predictable, as other scholars have also concluded (e.g., Stambaugh,1986; Campbell,1987; Breen *et al.,*1990; Fama, 1991; Brock *et al.,* 1992; Sullivan *et al.,* 1997). Furthermore, I have empirically proven that traders' intelligence plays a significant role in the formation of more sophisticated forecasting rules.

7.2. Background

For the last twenty years, econometric models have expressed the common assumption that the appropriate modelling of financial returns should allow for non-linearity. The failure of the traditional linear models promoted the development of non-linear models, such as autoregressive conditional heteroskedasticity, general autoregressive conditional heteroskedasticity and self-exciting threshold autoregression (Chan and Ng, 2004; Clements and Smith, 1997) which were used in the forecasting of stock returns. More recently combined forecasts from different models have been studied for forecasting purposes (Fang, 2005; Pai and Lin, 2005; Qian and Rasheed, 2006). The reason for combining forecasts from different models is the important assumption that forecasters are not able to identify the true process in full capacity, but different prediction models might complement the approximation of the data generation process.

It is a well-known fact, however, that both linear and non-linear forecasting models perform poorly out-of-sample (Diebold and Nason, 1990; Meese and Rose, 1991). Another option would be to implement a non-parametric model, such as Genetic Programming (GP) or Strongly Typed Genetic Programming (STGP) in particular.

Koza (1995) and Iba and Nikolaev (2000) proved that GP is a valid approach to finding suitable models to describe financial time series. The advantage of the GP approach, in comparison with linear and non-linear models, is that it enables the researcher to be relatively agnostic about the general form of the optimal trading rule and to fully explore the non-differentiable space of trading rules.

Allen and Karjalainen (1995) were the first to apply GP in order to establish profitable trading rules in the stock market. The authors argue that profitable trading rules exist even after taking into account transaction costs. They found significant evidence that there are trading rules that are able to generate significant excess returns over a buy-and-hold strategy during the period 1970-1989. By using monthly data as opposed to daily data, Becker and Seshadri (2003) developed trading rules that were able to outperform a buy-and-hold strategy when dividends are excluded from stock returns.

Neely *et al.* (1997) applied GP techniques to the foreign exchange market to investigate profitable trading rules. They found strong evidence of economically significant *ex-ante* excess returns to technical trading rules for six different exchange rates over the period 1981-1995.

Chen *et al.* (1998) even used GP to derive option pricing formulas. They obtained real data of the S&P 500 index options for the training and testing of their model. The comparison with the traditional model (the Black-Scholes formula) indicated the superiority of the GP pricing mechanism. Chidambaran *et al.* (1998) also implemented GP to investigate the relationship of the price of CBOE index and equity options, the terms of the option contract and various properties of the underlying security price. Again in comparison with the Black-Scholes formula, they demonstrated the advantage of GP in pricing the CBOE index and equity options.

Iba and Sasaki (1999) applied GP to the forecasting of the Nikkei225. They used 33,177 real data observations to compare the forecasting abilities of GP and neural networks. The experimental results showed the superiority of the GP measured by the low values of the Mean Squared Errors (MSE). However, the authors did not include any transaction costs in their models to offset the profit gain. Also, their prediction accuracy needed improvement because they focused on long-term forecasts rather than short-term or real-time forecasts.

Kaboudan (1999) measured time series predictability by applying GP to eight Dow Jones stock return series. The author introduced a new measure of time series predictability designed to reduce model search space and generate more accurate forecasts.

The new measure, which quantified the probability of predictability, consisted of two fitness values-one taken from the GP subject series and the other one from the same GP series after it has been shuffled randomly. The boundaries of the new measure were between zero and one hundred (zero suggested a random walk process and 100 suggested predictability).

Kaboudan (2000) used GP to produce reasonable one-day-ahead forecasts and develop a single day-trading strategy (SDTS) in which trading decisions were based on GP forecasts of daily highest and lowest prices. SDTS returned high profits when executed for fifty trading days.

However, there are some SDTS restrictions: first, the predicted spread (the difference between predicted daily high and low prices) has to be large, otherwise there is no profit to be made; Second a large volume of shares must be traded in order to reduce transaction costs and increase profits; and third the proposed strategy is profitable only for heavily traded securities to increase the probability of occurrence of expected high or low prices.

For more realistic forecasting purposes I implemented an innovative evolutionary STGP technique with no over-fitting of the S&P 500, IBM and GE historical data and the additional important factors:

-I developed more realistic real-life models taking into account transaction costs to offset the profit gained. Most forecasting studies have not included transaction costs, which I believe are important obstacles in applying forecasting models to real stock exchange trading.

-I performed my experimental tests under a special adaptive form of Strongly Typed Genetic Programming (STGP). In STGP the process of estimating the agent's fitness does not include any re-execution of their trading rules based on historical data (over-fitting). This is due to the fact that the artificial traders have already executed their trading rules on the same historical data set and the software is looking only at the realistic returns that they have already made, rather than any hypothetical returns that agents could have made if they were sent back in time again. Therefore I avoided over-fitting of the data which seems to be one of the biggest forecasting pitfalls.

-I used a greater number of artificial agents. While both models produced by Kaboudan (1999; 2000) consisted of 2,000 traders, the experiment conducted by Iba and Sasaki (1999) employed only 1,000 traders. I developed models with 10,000 traders.

-My model, developed within the Altreva Adaptive Modeler settings, was built incrementally (walk-forward with no over-fitting of historical data). The model constantly evolved and adapted to market changes, instead of being static. Ten thousand different trading strategies competed simultaneously and evolved in the artificial stock market in real time. Hence, the model was more resilient to changing market conditions and the model's performance was significantly more consistent and reliable (Witkam, 2011).

-I have developed three different markets to determine the forecasting accuracy of 'Low Intelligence Traders' (genome depth of 10). The maximum possible genome depth in the other three experiments was 20. Changing the genome depth level enabled me to observe how further the artificial traders look back further in history in order to develop more complex and sophisticated trading rules.

The in-sample period in my experiment consisted of 11,405 daily quotes (24/05/1962-14/09/2007) for the S&P 500 index, IBM and General Electric. The out-of-sample included 1261 real data observations (17/09/2007-14/09/2012). The study period was chosen on the basis of rich data availability, and because the software I used requires a long time series in orders to develop more stable and reliable models. There are various other reasons for the inclusion of these three particular financial instruments in my study.

Harris and Gurel (1986) highlighted that the composition of the S&P 500 list of companies does not depend on forecast security returns. Moreover, many large index funds mimic the performance of the S&P 500 by creating a portfolio of the 500 stocks using the same weights as in the real index. IBM is included because tech stocks tend to be more volatile in nature. I also wanted to investigate and distinguish the level of profitability generated by trading a group of stocks and two well established individual stocks.

7.3. Artificial stock market structure for this particular experiment

Table 21.0 illustrates the main parameters of the STGP-based artificial stock market model for this particular experiment.

Artificial stock market parameters	
Total population size (agents)	10,0000
Initial wealth(equal for all agents)	100,000
Transaction costs	1.25% (125 bsp)
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**; 10 in 'Low Intelligence Traders'
Minimum initial genome depth	2
Maximum initial genome depth	5
Breeding cycle frequency (bars)	1
Minimum breeding age (bars)	80
Initial selection type	random
Parent selection (percentage of initial selection that will breed)	5%
Mutation probability (per offspring)	10%
Total number of quotes processed- S&P500	12,666 (11,405 in-sample, 1261 out-of-sample) 24/05/1962-13/09/2007; 17/09/2007-14/09/2012
Total number of quotes processed-IBM	12,666 (11,405 in-sample, 1261 out-of-sample) 24/05/1962-13/09/2007; 17/09/2007-14/09/2012
Total number of quotes processed-GE	12,666 (11,405 in-sample, 1261 out-of-sample) 24/05/1962-13/09/2007; 17/09/2007-14/09/2012
Seed generation from clock	Yes
Creation of unique genomes	Yes
Offspring will replace the worst performing agents of the initial selection	Yes

Table 21.0 Artificial Stock Market Parameter Settings

*Maximum genome size measure the total number of nodes in an agent'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 an agent's genome (trading rule). The depth of a trading rule can be an indicator of its complexity.

7.3.1. Benchmark models and measures of forecasting accuracy.

I chose the Box-Jenkins and Holt-Winters models as benchmarks for comparison to the STGP technique. The Box-Jenkins or ARIMA approach was introduced in the 1970s, and since then has become one of the most popular tools for time series forecasting. The technique has been successfully applied to some of the most difficult forecasting problems. Reid (1969, 1972) argues that the Box-Jenkins technique generates the most accurate one-step-ahead forecasts for most time series when more than 50 observations are available. Granger and Newbold (1972) provided further evidence by examining 50 macro-economic time series. They concluded that Box-Jenkins produce better forecasts than both Holt-Winters and step-wise autoregression.

Lilien and Kotler (1983) suggest that 13 per cent of the financial services industry use exponential smoothing models. The Holt-Winters multiplicative (additive trend, multiplicative seasonality) method represents a weighted average of past data, in which the actual weights decline geometrically over a time horizon in order to capture short-term fluctuations in the values. Granger (1980), and Wilson and Keating (1990) argue that the accuracy of the forecasts generated from exponential smoothing models reflect the conformity of reality with the assumptions of historical patterns in the data set. The most recent observations have the most relevance for forecasting the future values. In view of this evidence I implemented the Box-Jenkins and Holt-Winters techniques in my experiments. Time series forecasting analyses historical data and projects estimates of future data values. This technique develops models of nonlinear functions by a recurrence relation obtained from past data. The same recurrence relation is applied to predict new values in the time series that are good approximation of the actual values.

As is standard in the forecasting literature, I estimate the Mean Absolute Percent Error (MAPE), Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). The lower the output of the error statistics obtained, the better the forecasting accuracy of the model concerned. Mean Absolute Percent Error and Maximum Absolute Percentage Error (MaxAPE) measure how much a dependent series varies from its model-forecasted level. By examining the mean and maximum across all possible scale-independent models, the researcher can get an indication of the uncertainty of the forecast values.

Examining the percentage errors is advisable because the dependent series represents subscriber numbers for stock markets of different sizes. MaxAPE quantifies the largest possible percentage error indicating the worst-case scenario in forecasting terms.

Mean Absolute Error and Maximum Absolute Error (MaxAE) measures how close the forecasts are to the eventual outcomes. Ther Root Mean Squared Error is the square root of the second moment related to the frequency function of a given random variable. RMSE is positioned on the same scale as the data and measures the differences between forecast values and the actual observed values. I use RMSE to compare the forecasting errors within S&P 500, IBM and GE individual datasets, but not between them. This is because this measure of accuracy is scale-dependent (Hyndman and Koehler, 2006). The measures of forecasting accuracy are quantified as follows:

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{d_t - z_t}{d_t}\right| \tag{41}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|d_t - z_t\right| \tag{42}$$

$$RMSE = \left\{\frac{1}{N}\sum_{t=1}^{N}\left(d_t - z_t\right)^2\right\}^{0.5} \tag{43}$$

where N is the number of forecasting periods, d_i is the actual price of the S&P 500, IBM and GE at period t, and z_t is the forecasting stock price at period t (Pai and Lin, 2005).

7.4. Simulation results

7.4.1.An investigation into whether, to what extent and in which form the stock returns in excess of the risk free rate are indeed predictable and profitable. Evidence of in-sample and out-of-sample predictability with and without transaction costs taken into account.

According to the test results, the null hypothesis that the price series generated by the artificial traders is normally distributed is rejected in all in-sample and out-of-sample periods (Table 22.0).

S&P 500		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
Daily mean	0.0375	0.9481
Daily standard deviation	0.0390	0.1484
Skewness	1.116063	-0.5066127
Kurtosis	2.731443	2.551086
J-B	2401.945 (0.0000) ^{p}	64.42563 (0.0000) ^{p}
ADF•	-29.27921 (0.0000) ^{p}	-40.18580 (0.0000) ^{p}
IBM		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
Daily mean	0.0170	0.1099
Daily standard deviation	0.0122	0.0275
Skewness	1.103778	0.412542
Kurtosis	3.159781	2.064995
J-B	2327.967 (0.0000) ^{p}	81.70225 (0.0000) ^{p}
ADF•	-107.4095 (0.0001) ^{p}	-36.25800 (0.0000) ^{p}
GE		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
Daily mean	0.0058	0.0160
Daily standard deviation	0.0021	0.0060
Skewness	0.627540	1.288965
Kurtosis	3.151968	3.804897
J-B	759.5347 (0.0000) ^{p}	383.2167 (0.0000) ^{p}
ADF•	-107.5242 (0.0001) ^{p}	-38.44100 (0.0000) ^{p}

• The MacKinnon (1996) one-sided critical value for rejection of the null hypothesis of a unit root at 5% level is -3.410060; ***J-B:*** the Jarqu-Bera test; ***ADF:*** Augmented Dickey-Fuller Unit Root Test; ***p***: the *p* value

Table 22.0 Descriptive statistics for the S&P 500, IBM, and GE in-sample and out-of-sample daily returns.

This result confirms an important fact in empirical finance: most financial return series are not normally distributed, which means that the tails are too fat compared to the normal distribution. Tables 23.0-28.0 report the in-sample and out-of-sample performance of Box-Jenkins and Holt-Winters forecasting models under conditions of no transaction costs and with transaction costs added on for the three financial instruments.

Box-Jenkins (without transaction costs) S&P500 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
Box-Jenkins model type	ARIMA(1,1,2)	ARIMA(0,1,18)
RMSE	6.541	17.703
MAE	3.123	12.464
MAPE	0.660	1.105
MaxAPE	25.811	9.621
MaxAE	84.728	106.414
Normalized BIC	3.759	5.764
Holt-Winters (without transaction costs) S&P500 statistics		
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	6.543	17.768
MAE	3.126	12.463
MAPE	0.663	1.106
MaxAPE	25.916	9.60
MaxAE	84.550	106.062
Normalized BIC	3.757	5.766

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 23.0 In-sample and out-of-sample performance of Box-Jenkins and Holt-Winters (without transaction costs) forecasting models on S&P 500 daily stock return series.

Box-Jenkins (without transaction costs) IBM statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
Box-Jenkins model type	ARIMA(0,1,1)	ARIMA(1,1,1)
RMSE	5.384	2.028
MAE	2.144	1.468
MAPE	1.208	1.140
MaxAPE	300.123	9.729
MaxAE	308.193	9.904
Normalized BIC	3.368	1.425
Holt-Winters (without transaction costs) IBM statistics		
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	5.385	2.028
MAE	2.144	1.469
MAPE	1.208	1.140
MaxAPE	300.125	9.749
MaxAE	308.200	9.960
Normalized BIC	3.369	1.426

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 24.0 In-sample and out-of-sample performance of Box-Jenkins and Holt-Winters (without transaction costs) forecasting models on IBM daily stock return series.

Box-Jenkins (without transaction costs) GE statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
Box-Jenkins model type	ARIMA(0,1,2)	ARIMA(0,1,4)
RMSE	1.862	0.594
MAE	0.756	0.365
MAPE	1.167	1.963
MaxAPE	203.670	22.900
MaxAE	105.909	4.722
Normalized BIC	1.244	1.030
Holt-Winters (without transaction costs) GE statistics		
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	1.862	0.595
MAE	0.756	0.367
MAPE	1.167	1.964
MaxAPE	203.671	22.901
MaxAE	105.910	4.805
Normalized BIC	1.244	1.031

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 25.0 In-sample and out-of-sample performance of Box-Jenkins and Holt-Winters (without transaction costs) forecasting models on GE daily stock return series.

Box-Jenkins (transaction costs included) S&P500 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
Box-Jenkins model type	ARIMA(0,1,12)	ARIMA(0,1,18)
RMSE	7.192	19.464
MAE	3.437	13.697
MAPE	0.661	1.104
MaxAPE	25.933	9.639
MaxAE	93.808	117.308
Normalized BIC	3.949	5.954
Holt-Winters (transaction costs included) S&P500 statistics		
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	7.198	19.535
MAE	3.430	13.700
MAPE	0.665	1.105
MaxAPE	26.052	10.133
MaxAE	93.605	116.880
Normalized BIC	3.949	5.950

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 26.0 In-sample and out-of-sample performance of Box-Jenkins and Holt-Winters (transaction costs included) forecasting models on S&P 500 daily stock return series.

Box-Jenkins (transaction costs included) IBM statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
Box-Jenkins model type	ARIMA(0,1,1)	ARIMA(0,1,5)
RMSE	5.922	2.199
MAE	2.316	1.603
MAPE	1.199	1.128
MaxAPE	200.739	10.904
MaxAE	340.208	11.013
Normalized BIC	3.558	1.582
Holt-Winters (transaction costs included) IBM statistics		
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	5.920	2.206
MAE	2.307	1.594
MAPE	1.193	1.123
MaxAPE	200.619	10.817
MaxAE	340.000	11.001
Normalized BIC	3.558	1.588

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 27.0 In-sample and out-of-sample performance of Box-Jenkins and Holt-Winters (transaction costs included) forecasting models on IBM daily stock return series.

Box-Jenkins (transaction costs included) GE statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
Box-Jenkins model type	ARIMA(0,1,1)	ARIMA(0,1,1)
RMSE	2.040	0.627
MAE	0.830	1.963
MAPE	1.173	1.765
MaxAPE	129.751	24.490
MaxAE	75.855	5.0520
Normalized BIC	1.427	0.927
Holt-Winters (transaction costs included) GE statistics		
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	2.040	0.627
MAE	0.834	0.363
MAPE	1.183	1.624
MaxAPE	119.770	20.491
MaxAE	75.866	5.0530
Normalized BIC	1.427	0.928

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 28.0 In-sample and out-of-sample performance of Box-Jenkins and Holt-Winters (transaction costs included) forecasting models on GE daily stock return series.

Tables 29.0-31.0 represents the in-sample and out-of-sample performance of STGP forecasting models on S&P 500, IBM and GE daily stock return series.

Strongly Typed Genetic Programming (without transaction costs) S&P500 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
RMSE	6.225	13.387
MAE	2.817	9.507
MAPE	0.581	0.823
MaxAPE	15.138	6.260
MaxAE	77.907	70.929
Strongly Typed Genetic Programming (transaction costs included) S&P500 statistics		
RMSE	7.760	14.000
MAE	3.439	9.787
MAPE	0.797	0.837
MaxAPE	100.985	7.093
MaxAE	174.289	79.367
Strongly Typed Genetic Programming ('Low Intelligence Traders') S&P500 statistics		
RMSE	6.830	11.862
MAE	3.898	8.291
MAPE	0.696	0.702
MaxAPE	36.979	6.154
MaxAE	107.539	86.466

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 29.0 In-sample and out-of-sample performance of Strongly Typed Genetic Programming forecasting models on S&P 500 daily stock return series.

Strongly Typed Genetic Programming (without transaction costs) IBM statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
RMSE	5.378	1.978
MAE	2.084	1.395
MAPE	1.164	1.087
MaxAPE	207.120	9.678
MaxAE	308.103	9.843
Strongly Typed Genetic Programming (transaction costs included) IBM statistics		
RMSE	8.756	1.897
MAE	2.423	1.370
MAPE	1.310	1.062
MaxAPE	207.572	8.712
MaxAE	595.706	8.904
Strongly Typed Genetic Programming ('Low Intelligence Traders') IBM statistics		
RMSE	6.465	2.016
MAE	2.284	1.458
MAPE	1.220	1.136
MaxAPE	307.323	9.709
MaxAE	378.362	8.926

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 30.0 In-sample and out-of-sample performance of Strongly Typed Genetic Programming forecasting models on IBM daily stock return series.

Strongly Typed Genetic Programming (without transaction costs) GE statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	24/05/1962-14/09/2007	17/09/2007-14/09/2012
Number of observations	11405	1261
RMSE	1.849	0.548
MAE	0.741	0.320
MAPE	1.163	1.744
MaxAPE	133.806	16.979
MaxAE	86.554	3.929
Strongly Typed Genetic Programming (transaction costs included) GE statistics		
RMSE	2.134	0.560
MAE	0.836	0.333
MAPE	1.252	1.794
MaxAPE	253.344	22.205
MaxAE	116.607	4.005
Strongly Typed Genetic Programming ('Low Intelligence Traders') GE statistics		
RMSE	1.914	0.548
MAE	0.785	0.311
MAPE	1.213	1.667
MaxAPE	235.348	22.217
MaxAE	117.255	4.002

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 31.0 In-sample and out-of-sample performance of Strongly Typed Genetic Programming forecasting models on GE daily stock return series.

To test whether the forecasts from two competing models are equally accurate, I implemented Wilcoxon signed-rank (WSR) tests, because the time series in my experiment are non-normally distributed. The null hypothesis of the WSR test is that the two populations represented by the respective members of the matched pairs are identical. When the null hypothesis is true, then each of the 2^N possible sets of signed ranks estimated by arbitrarily assigning plus or minus signs to be ranks 1 through N is equally likely (DeFusco *et al.,* 1990). The test statistic is:

$$WSR = \sum_{t=1}^{N} I + (d_t) rank(|d_t|) \tag{44}$$

Where $I + (d_t) = \begin{cases} 1 \\ 0 \end{cases}$ if $D_t > 0$, otherwise

Where $rank(|d_t|)$ denotes the rank of the absolute value of d_t (Alon *et al.,*2001).

First, I compare the in-sample and out-of-sample one-step-ahead forecasting accuracy of STGP and the traditional econometric models for the S&P 500, IBM and GE without taking into account any transaction costs. All five performance measures (MAPE, MAE, RMSE, MaxAPE and MaxAE) suggested that the STGP technique outperforms Box-Jenkins and Holt-Winters in absolutely all in-sample and out-of-sample experiments for the three financial instruments (first row of Tables 32.0-37.0).

STGP vs. B-J in-sample forecasting models based on S&P500 daily stock returns							
GP model	***WSR [a]***	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP without transaction costs	65.398 (0.0000) [p]	***Difference [b]***	-0.316*	-0.306*	-0.079*	-10.673*	-16.821*
STGP with transaction costs	92.430 (0.0000) [p]	***Difference [b]***	0.568*	0.002*	0.136*	75.052*	80.481*
STGP for 'Low Intelligence Traders'	71.407 (0.0000) [p]	***Difference [b]***	0.289*	0.775*	0.036*	11.168*	22.811*
STGP vs. B-J out-of-sample forecasting models based on S&P500 daily stock returns							
STGP without transaction costs	25.800 (0.0000) [p]	***Difference [b]***	-4.316*	-2.957*	-0.282*	-3.361*	-35.485*
STGP with transaction costs	30.747 (0.0000) [p]	***Difference [b]***	-5.464*	-3.892*	-0.267*	-2.546*	-37.941*
STGP for 'Low Intelligence Traders'	25.772 (0.0000) [p]	***Difference [b]***	-5.841*	-4.173*	-0.403*	-3.467*	-19.948*

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.

Difference [b]= STGP-BJ models; Negative values indicate preference to STGP models.

****Indicates significance at the 99% level.***

Table 32.0 WSR test and paired comparisons of forecasting performance: STGP vs. B-J forecasting models of the S&P 500 daily stock returns.

STGP vs. H-W in-sample forecasting models based on S&P500 daily stock returns							
GP model	***WSR*** [a]	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP without transaction costs	76.398 (0.0000) [p]	***Difference*** [b]	-0.318*	-0.309*	-0.082*	-10.778*	-6.643*
STGP with transaction costs	92.435 (0.0000) [p]	***Difference*** [b]	0.562*	0.009*	0.132*	74.933*	80.684*
STGP for 'Low Intelligence Traders'	79.621 (0.0000) [p]	***Difference*** [b]	0.287*	0.772*	0.033*	11.063*	22.989*
STGP vs. H-W out-of-sample forecasting models based on S&P500 daily stock returns							
STGP without transaction costs	26.639 (0.0000) [p]	***Difference*** [b]	-4.381*	-2.956*	-0.283*	-3.340*	-35.133*
STGP with transaction costs	30.757 (0.0000) [p]	***Difference*** [b]	-5.535*	-3.913*	-0.268*	-3.040*	-37.513*
STGP for 'Low Intelligence Traders'	25.995 (0.0000) [p]	***Difference*** [b]	-5.906*	-4.172*	-0.404*	-3.446*	-19.596*

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.

Difference [b]***= STGP-BJ models; Negative values indicate preference to STGP models.***

****Indicates significance at the 99% level.***

Table 33.0 WSR test and paired comparisons of forecasting performance: STGP vs. H-W forecasting models of the S&P 500 daily stock returns.

STGP vs. B-J in-sample forecasting models based on IBM daily stock returns							
GP model	***WSR*** [a]	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP without transaction costs	8.057 (0.0000) [p]	***Difference*** [b]	-0.006*	-0.060*	-0.980*	-93.003*	-0.090*
STGP with transaction costs	92.230 (0.0000) [p]	***Difference*** [b]	2.834*	0.107*	0.111*	6.833*	255.498*
STGP for 'Low Intelligence Traders'	16.346 (0.0000) [p]	***Difference*** [b]	1.081*	0.140*	0.012*	7.200*	70.169*
STGP vs. B-J out-of-sample forecasting models based on IBM daily stock returns							
STGP without transaction costs	5.216 (0.0000) [p]	***Difference*** [b]	-0.050*	-0.073*	-0.053*	-0.060*	-0.061*
STGP with transaction costs	30.747 (0.0000) [p]	***Difference*** [b]	-0.302*	-0.233*	-0.066*	-2.192*	-1.109*
STGP for 'Low Intelligence Traders'	19.224 (0.0000) [p]	***Difference*** [b]	-0.012*	-0.010*	-0.004*	-0.020*	-0.978*

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.

Difference [b]***= STGP-BJ models; Negative values indicate preference to STGP models.***

****Indicates significance at the 99% level.***

Table 34.0 WSR test and paired comparisons of forecasting performance: STGP vs. B-J forecasting models of the IBM daily stock returns.

STGP vs. H-W in-sample forecasting models based on IBM daily stock returns							
GP model	***WSR*** [a]	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP without transaction costs	8.057 (0.0000) [p]	***Difference*** [b]	-0.007*	-0.060*	-0.044*	-93.005*	-0.097*
STGP with transaction costs	92.227 (0.0000) [p]	***Difference*** [b]	2.836*	0.116*	0.117*	6.953*	255.706*
STGP for 'Low Intelligence Traders'	41.199 (0.0000) [p]	***Difference*** [b]	1.080*	0.140*	0.012*	7.198*	70.162*
STGP vs. H-W out-of-sample forecasting models based on IBM daily stock returns							
STGP without transaction costs	5.216 (0.0000) [p]	***Difference*** [b]	-0.050*	-0.074*	-0.053*	-0.071*	-0.117*
STGP with transaction costs	30.741 (0.0010) [p]	***Difference*** [b]	-0.309*	-0.224*	-0.061*	-2.105*	-1.097*
STGP for 'Low Intelligence Traders'	24.372 (0.0000) [p]	***Difference*** [b]	-0.012*	-0.011*	-0.004*	-0.040*	-1.034*

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.

Difference [b]***= STGP-BJ models; Negative values indicate preference to STGP models.***

****Indicates significance at the 99% level.***

Table 35.0 WSR test and paired comparisons of forecasting performance: STGP vs. H-W forecasting models of the IBM daily stock returns

STGP vs. B-J in-sample forecasting models based on GE daily stock returns							
GP model	***WSR [a]***	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP without transaction costs	22.455 (0.0000) [p]	***Difference [b]***	-0.013*	-0.015*	-0.004*	-69.864*	-9.355*
STGP with transaction costs	91.898 (0.0000) [p]	***Difference [b]***	0.094*	0.006*	0.079*	23.593*	20.752*
STGP for 'Low Intelligence Traders'	33.017 (0.0000) [p]	***Difference [b]***	0.052*	0.029*	0.046*	32.000*	11.346*
STGP vs. B-J out-of-sample forecasting models based on GE daily stock returns							
STGP without transaction costs	8.280 (0.0000) [p]	***Difference [b]***	-0.046*	-0.045*	-0.219*	-5.921*	-0.793*
STGP with transaction costs	30.747 (0.0000) [p]	***Difference [b]***	-0.067*	-0.030*	-0.169*	-2.285*	-1.047*
STGP for 'Low Intelligence Traders'	9.638 (0.0000) [p]	***Difference [b]***	-0.046*	-0.054*	-0.296*	-0.683*	-0.720*

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.

Difference [b]= STGP-BJ models; Negative values indicate preference to STGP models.

****Indicates significance at the 99% level.***

Table 36.0 WSR test and paired comparisons of forecasting performance: STGP vs. B-J forecasting models of the GE daily stock returns.

STGP vs. H-W in-sample forecasting models based on GE daily stock returns							
GP model	***WSR*** [a]	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP without transaction costs	40.128 (0.0000) [p]	***Difference*** [b]	-0.013*	-0.015*	-0.004*	-69.865*	-19.366*
STGP with transaction costs	91.895 (0.0000) [p]	***Difference*** [b]	0.094*	0.002*	0.069*	33.574*	20.741*
STGP for 'Low Intelligence Traders'	37.443 (0.0000) [p]	***Difference*** [b]	0.052*	0.029*	0.046*	31.677*	11.345*
STGP vs. H-W out-of-sample forecasting models based on GE daily stock returns							
STGP without transaction costs	9.134 (0.0000) [p]	***Difference*** [b]	-0.047*	-0.047*	-0.220*	-5.922*	-0.876*
STGP with transaction costs	30.740 (0.0000) [p]	***Difference*** [b]	-0.067*	-0.030*	-0.170*	-1.714*	-1.048*
STGP for 'Low Intelligence Traders'	11.337 (0.0000) [p]	***Difference*** [b]	-0.047*	-0.056*	-0.797*	-0.684*	-0.803*

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.

Difference [b]***= STGP-BJ models; Negative values indicate preference to STGP models.***

****Indicates significance at the 99% level.***

Table 37.0 WSR test and paired comparisons of forecasting performance: STGP vs. H-W forecasting models of the GE daily stock returns.

I observed even bigger out-of-sample differences between the competing forecasting models, indicating higher *ex-ante* STGP accuracy. Statistically speaking, the STGP outperformed both econometric models at the 99% significance level based on the WSR *p*-values (in parenthesis) reported in Tables 32.0-37.0. Moreover, all test results indicate a statistical significance, thereby rejecting the null hypothesis of equality in the forecasting difference between the competing models.   I included transaction costs in my next experiment in order to develop a more realistic trading scenario. If transaction costs are very high, predictability is no longer ruled out by arbitrage because it would be too expensive to take advantage of even a substantially large and predictable component in returns. Hence, I suggest that forecasting stock markets approaches have to be seen in relation to the transaction costs of the stocks.

Each of the three financial instruments was estimated and validated by the in-sample data. This model estimation process was then followed by an empirical evaluation based on the out-of-sample data covering 1,261 observations from the 17th of September 2007 to the 14th of September 2012. In-sample empirical results (the paired differences) expressed in Tables 32.0-37.0, suggests that both the Box-Jenkins and Holt-Winters models slightly outperform STGP forecasting models in terms of all performance measures for the S&P 500, IBM, and GE.

Figures 20.0, 21.0 and 22.0 graphically compare the in-sample forecasting performance of the STGP, Box-Jenkins and Holt-Winters modes for the three financial instruments. The graphs illustrate the difference of forecasting direction suggested by the WSR test.

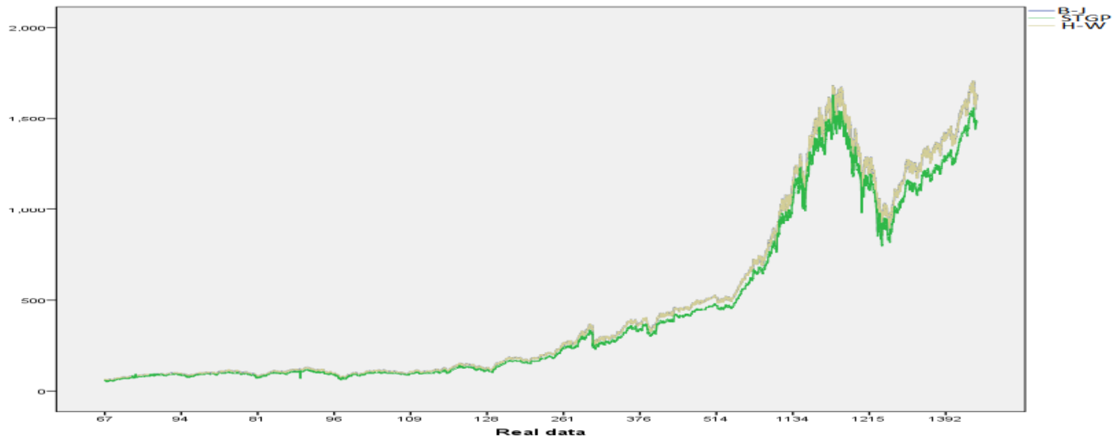

Figure 19.0 In-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the S&P 500 daily stock returns.

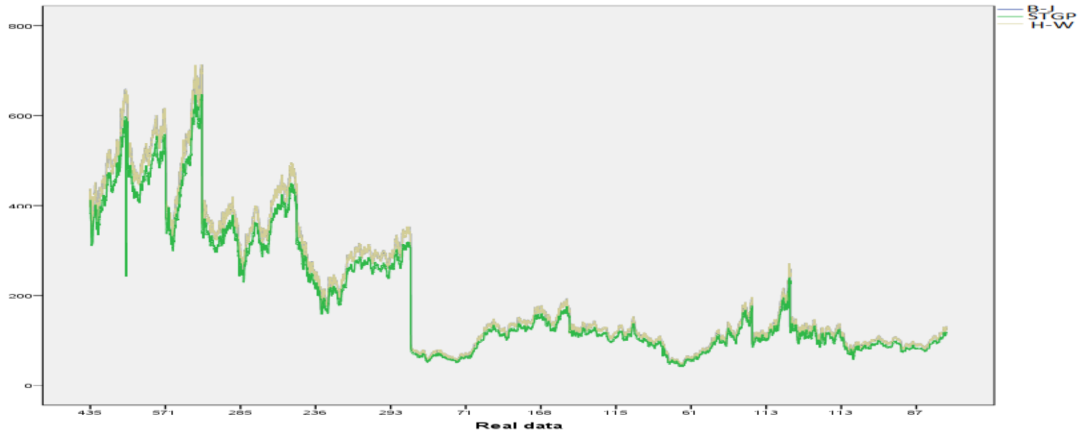

Figure 21.0 In-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the IBM daily stock returns.

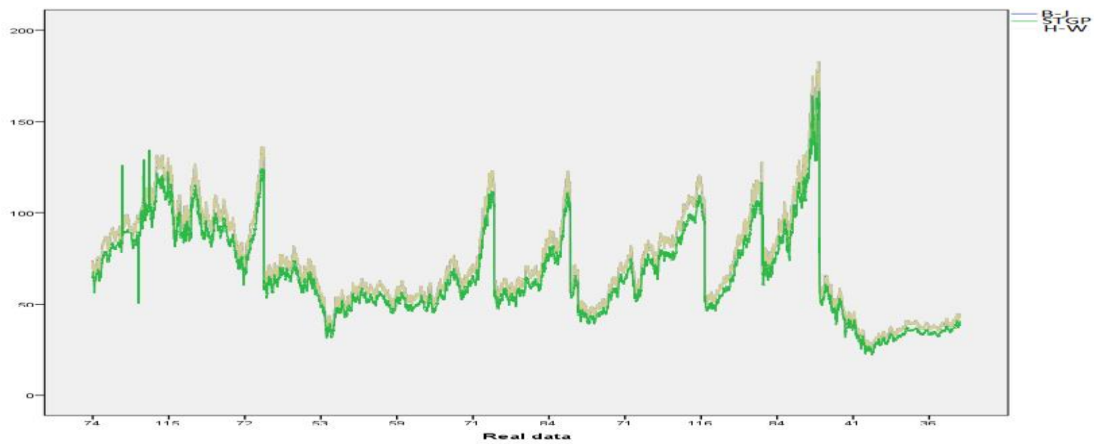

Figure 22.0 In-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the GE daily stock returns.

When I plotted 11,405 in-sample observations of real S&P 500, IBM, and GE data (Figures 23.0, 24.0 and 25.0), I observed that the Box-Jenkins model seems to perform better than both STGP and Holt-Winters (B-J curve does not deviate significantly from the real data curve). This is in line with Newbold and Granger (1974), who claim that the Box-Jenkins procedure markedly outperforms Holt-Winters. However, the out-of-sample validation illustrates the superiority of STGP in absolutely all experiments. The out-of-sample paired comparisons indicate that the STGP substantially outperforms the traditional econometric models. The results are significant at the 99% level, including the WSR tests.

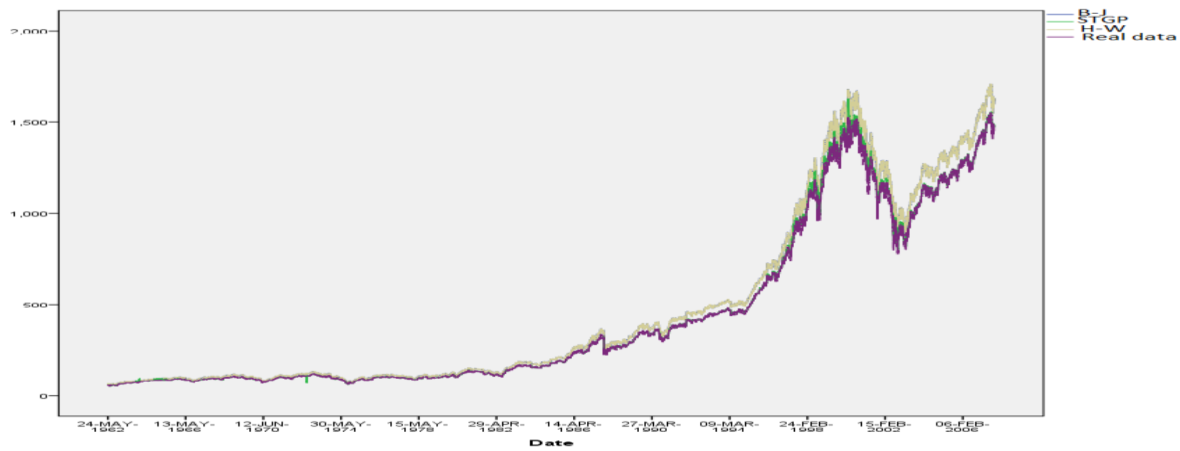

Figure 23.0 In-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the S&P 500 daily stock returns. A plot of real data has been added on

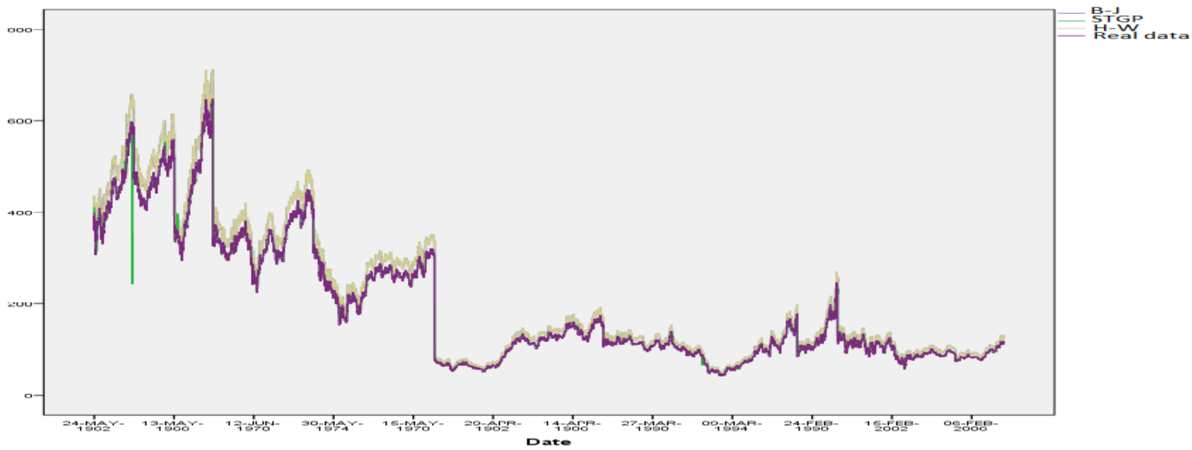

Figure 24.0 In-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the IBM daily stock returns. A plot of real data has been added on for comparison purposes.

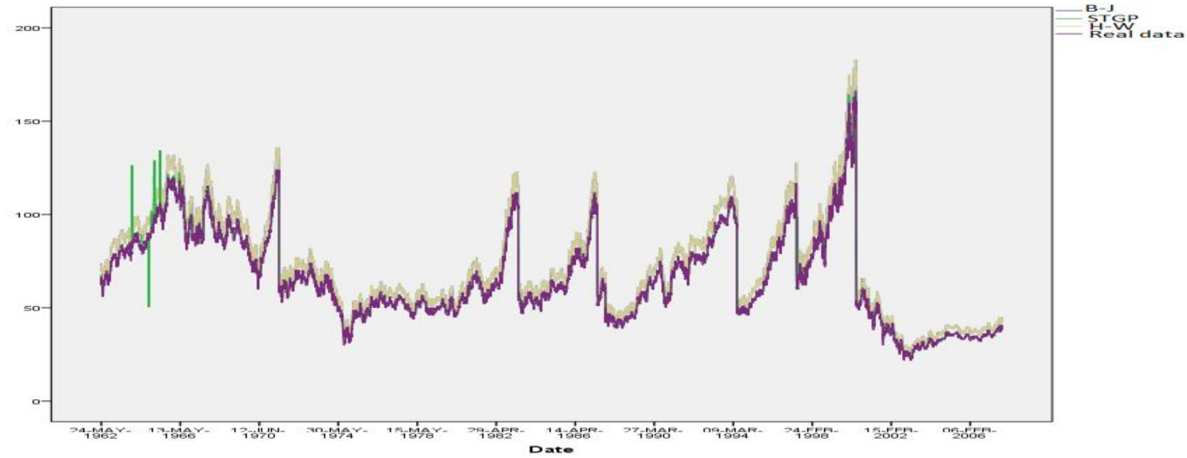

Figure 25.0 In-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the GE daily stock returns. A plot of real data has been added on for comparison purposes.

Figures 26.0, 27.0, and 28.0 clearly illustrate the difference in forecasting direction of the competing models.

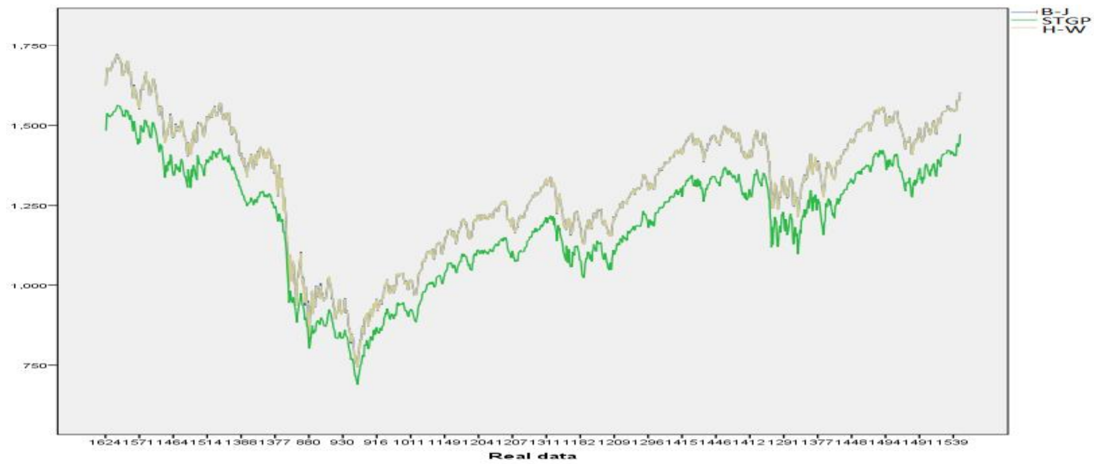

Figure 26.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the S&P 500 daily stock returns.

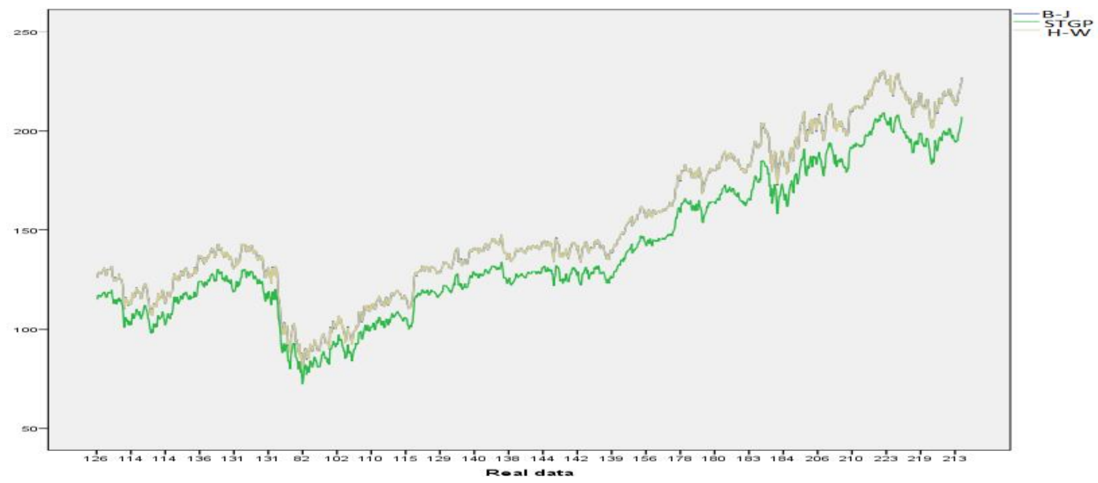

Figure 27.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the IBM daily stock returns.

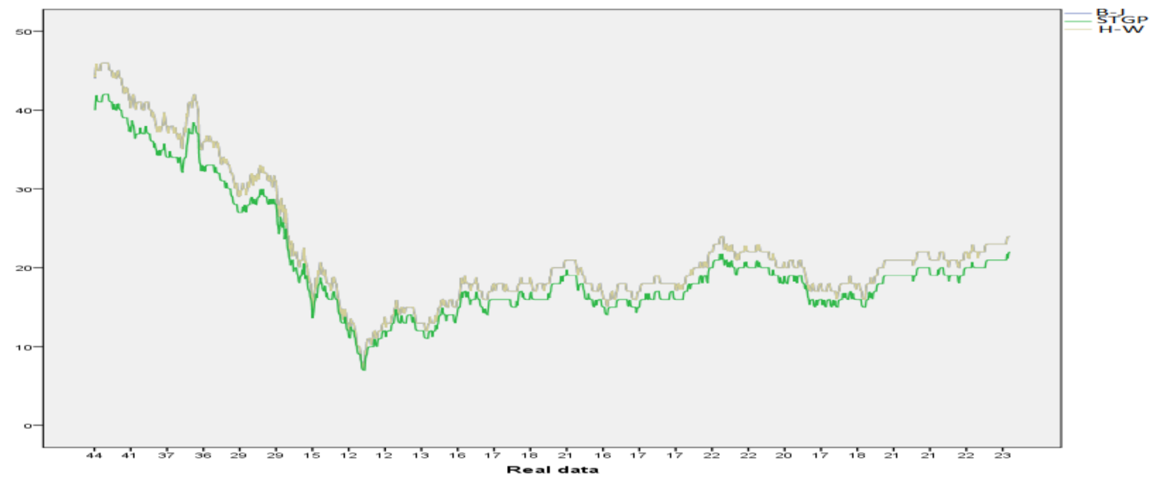

Figure 28.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the GE daily stock returns.

To obtain a more realistic forecasting comparison picture, I plot 1,261 out-of-sample real data quotes (Figures 29.0, 30.0, and 31.0) and observe that the STGP curve overleaps with the real data curve suggesting a superior out-of-sample predictability of the STGP technique. It is clearly visible that there is a substantial difference in the out-of-sample forecasting performance of STGP and the more traditional econometric models.

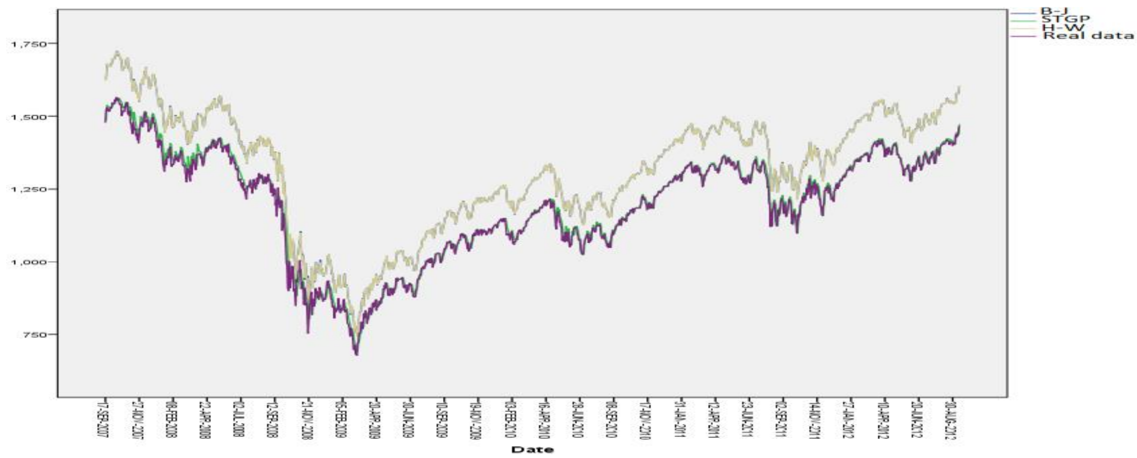

Figure 29.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the S&P 500 daily stock returns. A plot of real data has been added on for comparison purposes.

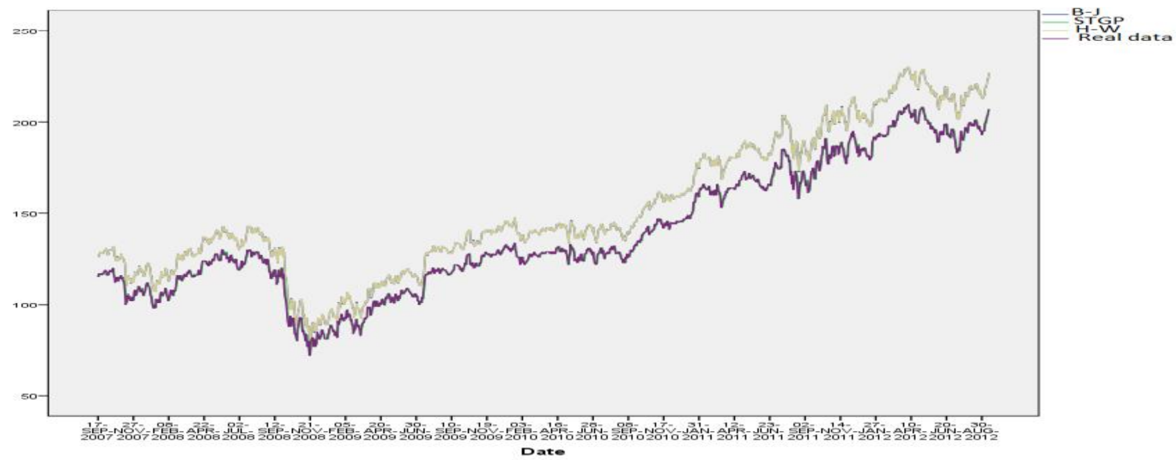

Figure 30.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the IBM daily stock returns. A plot of real data has been added on for comparison purposes.

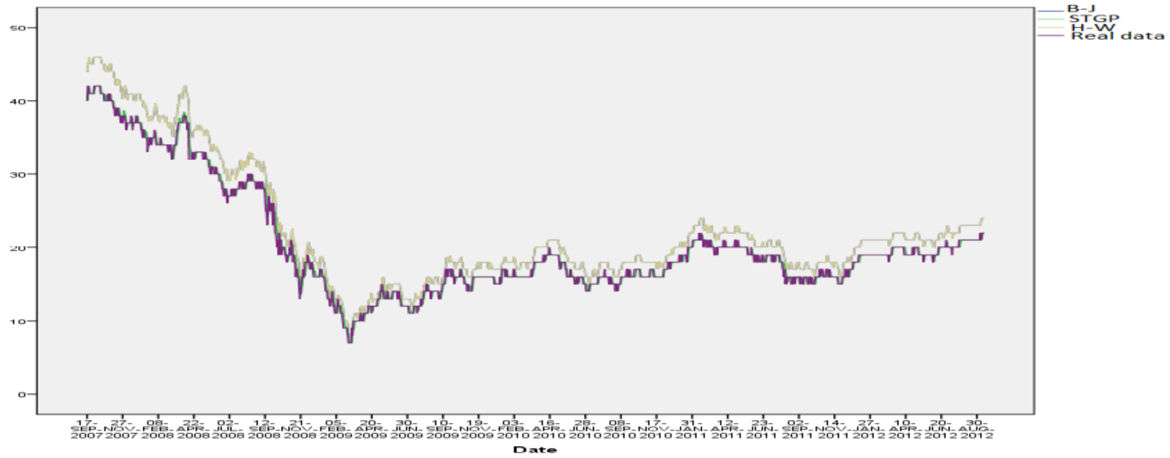

Figure 31.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the GE daily stock returns. A plot of real data has been added on for comparison purposes.

My results are consistent with the findings of Bossaerts and Hillion (1999) and Chatfield (1996), who argue that forecasting models with high in-sample explanatory power do not usually have high out-of-sample fit (external validity) due to model over-fitting or data snooping. The out-of-sample tests in my experiment controlled the over-fitting of data and represented a more powerful framework to evaluate the performance of the competing forecasting models.

The Optimism Principle developed by Picard and Cook (1984) provides an explanation as to why traditional econometric models generate a better in-sample fit, but worse *ex-ante* predictions. The authors developed the Optimism Principe in the light of the standard linear model denoted as:

$$Y = X\beta + \in \qquad (45)$$

where X is an $n \times p$ full rank matrix of known constants, β is a p vector of unknown parameters and the error $\in = (\in_i)$ is independently and identically distributed (IID class) with mean 0 and variance σ^2. The model described above is called the full model with some of the elements of β equalling 0. The residual mean square (RMS), σ_{full}^2 from the least squares fit to the null model is in fact an unbiased estimator of σ^2.

However, the same assumption is not valid for the RMS from fitted forecasting models selected on the basis of in-sample and out-of-sample comparisons. For instance, if all $2^p - 1$ least squares in-sample and out-of-sample are taken into account and the subset associated with the smallest possible RMS, ${\sigma_{\min}}^2$ has been chosen.

Due to the fact that ${\sigma_{\min}}^2 \leq {\sigma_{full}}^2$ for all values of Y and ${\sigma_{full}}^2$ has expectation σ^2, it is logical that ${\sigma_{\min}}^2$ does not. Obviously, ${\sigma_{\min}}^2$ is optimistic with significantly high bias in some cases. As a result, some forecasters believe that their predictions are more accurate than they are in reality (Picard and Cook, 1984). The consequences of possible data over-fitting of Box-Jenkins and Holt-Winters forecasting models in my experiment could potentially lead to underestimated MAPE, MAE and RMSE. The result of this underestimation is narrow prediction intervals and model uncertainty. Hjorth (1989) argues that model selection based on minimising various criterions will lead to underestimation of the same criterions. One of the negative implications of the Optimism Principle is that forecasters often assume that the fit is better than it really is. According to Chatfield (1996), empirical diagnostic tests rarely reject the best-fitting model exactly because it is the best fit, thereby leading to inaccurate forecasts.

In this chapter I confirmed the empirical findings of Kaboudan (2000), who presented scientific evidence that GP models are in fact able to forecast out-of-sample price levels better. The author compared GP forecasts of daily closing prices with simple baseline price forecast strategy where tomorrow's price is simply equated to today's price $(P_t = P_{t-1})$. The most important consequence of my analysis so far is that forecasting comparisons of different methods and models should predominantly be made on the basis of genuine out-of-sample (*ex-ante*) predictions. This result is consistent with the findings of Armstrong (1995) and Chatfield (1996), suggesting that the real test of a forecasting model or method is its *ex-ante* forecasting ability.

MAPE, MAE, and RMSE are all important error measures, yet they may not constitute the best criterion in terms of profitability. Satisfactory low forecasting errors of MAPE, MAE and RMSE do not necessarily guarantee that the model is generating a profit.

For instance, RMSE is an artificial measure of performance, because the root mean squared error is just a quantitative expression of the current model and does not measure the actual profitability of the forecasting model. Moreover, gaining profits requires more accurate forecasts for the critical tendency change of the stock when the asset price suddenly falls reversely after the price rises before.

At this stage, I measured the actual profitability of the most realistic model for trading purposes- the STGP with transaction costs included for the S&P 500, IBM, and GE. I measure profitability by two primary criteria-the number of correct hits (forecasts) and the generated excess return from trading the three financial instruments. By implementing the hit ratio I test the percentage of time that the model has good sign of predictability is quantified by:

$$\text{Sign rate (\%)}=\frac{\text{Number of correct forecasts}}{\text{Number of generated buy / sell orders}}\times 100 \tag{46}$$

The excess return represents the amount received from trading in excess of the risk free rate. It is the continuously compounded return on the S&P 500, IBM and GE price minus the value of daily continuously compounded rate converted from the annualised investment yield on a three month US Treasury bill:

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) - r_{t-1} \tag{47}$$

where P_t is the price of the S&P 500, IBM, and GE traded at period t, and r_t is the risk free rate set at the value of daily continuously compounded rate converted from the annualised investment yield on a three month US Treasury bill (data up to 14/09/2012 was downloaded from the Federal Reserve statistical release website at www.federalreserve.gov/releases/h15). Dividends are not included in my study. Moreover, I trade American financial instruments and this is the reason why I adopted the US risk free T bill rate, instead of employing other risk free instruments in deriving the excess returns. On the other hand, the reason for forecasting the excess return is because it provides a measure of how well my models perform relative to the minimum returns gained from depositing the money in a risk free manner. The number of correct out-of-sample forecasts of the sign of return for each of the three financial instruments is reported in Table 38.0.

Financial instrument	S&P 500	IBM	GE
Number of generated buy/sell orders	247	298	271
Number of correct forecasts (hits)	145	162	151
Successful hits ratio	58.7%*	54.4%*	55.7%*
Excess return	8.12%[b]	4.93%[b]	6.01%[b]

[a] The table reports the number of times a STGP forecasting out-of-sample model correctly predicts the direction of S&P 500, IMB and GE returns and profitability of 1,261 observations (17/09/2007-14/09/2012) for each financial instrument. A ratio market with asterisk (*) indicates a 95% significance based on a one-sided test of H_0:p=0.50 against H_a:p>0.50. [b]The risk-free rate is set at the value of daily continuously compounded rate converted from the annualized investment yield on a 3-month US Treasury bill (up to 14/09/2012).

Table 38.0 Out-of-sample comparison of the predictive strength and profitability of the STGP with included transaction costs of 1.25% for the S&P 500, IBM, and GE forecasting models [a].

The corresponding hit ratios are also given. A hit ratio above 50% is a sign of actual profitability from trading. The S&P 500 reports a hit ratio of 58.7%, IBM has 54.4% and GE has 55.7%, with all transaction costs included. Table 39.0 represents the hit ratios for the three financial instruments over the entire sample of 12,666 observations- S&P hit ratio is 57.3%, IBM hit ratio is 53.7% and GE hit ratio is 54.8%.

Financial instrument	S&P 500	IBM	GE
Number of generated buy/sell orders	2,354	3,186	2,944
Number of correct forecasts (hits)	1,349	1,711	1,613
Successful hits ratio	57.3%*	53.7%*	54.8%*
Excess return	7.98%[b]	4.62%[b]	5.81%[b]

[a] The table reports the number of times a STGP forecasting model correctly predicts the direction of S&P 500, IBM and GE returns and profitability for the entire sample of 12,666 observations (24/05/1962-14/09/2012) for each financial instrument. A ratio market with asterisk (*) indicates a 95% significance based on a one-sided test of H_0:p=0.50 against H_a:p>0.50. [b]The risk-free rate is set at the value of daily continuously compounded rate converted from the annualized investment yield on a 3-month US Treasury bill (up to 14/09/2012).

Table 39.0 Comparison of the predictive strength and profitability of the STGP (over the entire sample) with included transaction costs of 1.25% for the S&P 500, IBM, and GE forecasting models [a].

Additionally I performed another test to investigate whether the hit ratio of the S&P 500, IBM and GE is significantly different from the benchmark of 0.5 (a 95% significance level based on a one-sided test). Under the null hypothesis that the test has no predictive effectiveness power $(H_0 : p = 0.50$ against $H_0 : p > 0.50)$.

The statistical test shows that the hit ratios of the S&P 500, IBM and GE are significantly different from 0.50, which confirms the ability of these forecasting models in the prediction of the returns of the three financial instruments.

Tables 38.0 and 39.0 also reports the excess return gained from trading the index and the two securities. The out-of-sample excess return of S&P 500 is 8.12%, followed by GE with recorded profitability of 6.01% and IBM with 4.93%. The excess returns based on the whole sample generated slightly lower values- S&P 500 has excess return of 7.98%, followed by GE with 5.81% profitability, and IBM with 4.62% excess return. The excess returns differences between out-of-sample and the whole sample can be explained by the initial chaotic behaviour of the STGP models. The very first years of historical data serve as training period leading to lower in-sample profitability of the STGP forecasting models.

Jones (2002) reported that when summed together, bid-ask spreads and commissions for trading US equities account for 0.84%. I chose higher transaction costs of 1.25%, or 125 basis points, in order to assess the economic significance of the profits earned. There are two reasons for the selection of this particular level of transaction costs.

First, while the level of 1.25% may be relatively high by current standards, it appears reasonable for the earlier period of the experiment. Second, I used higher transaction costs to guard against over-fitting in-sample, which seems intuitively reasonable. My results are consistent with the findings of Kaboudan (2000), who claims that stocks are not equally profitable. His empirical results suggested that low priced stocks ($50 or less) are more profitable only when their daily spread average is fairly high (>$1.75). One of the reasons for the difference in profitability in my case might be the high price of the index (an out-of-sample average of $1.195) comparing to only $20 of the GE.

Overall, I believe that stock returns in excess of the risk free rate are indeed predictable, as many other studies have also concluded e.g. Keim and Stambaugh (1986); Campbell (1987); Breen, Glosten, and Jagannathan (1990); Fama (1991); Brock, Lakonishok, and LeBaron (1992) Sullivan, Timmermann, and White (1997). Moreover, have I expanded upon these studies by the inclusion of appropriate transaction costs making my models more realistic for stock exchange trading purposes.

7.4.2.An investigation into whether intelligence and trader cognitive abilities matter in the formation of more sophisticated trading rules or if the market mechanism is the main driving force.

Logically, the in-sample results generated by the STGP for 'Low Intelligence Traders' has been outperformed in the three experiments of the S&P 500, IBM and GE time series. Although the 'Low Intelligence Traders' generated decent out-of-sample forecasts, their predictive ability has a bias, judging by the forecasting measurement errors of the other three experiments.

While the 'Best Performing Traders' group investigated the complexity and heterogeneity of the trading rules, the pool of 'Low Intelligence Traders' with half genome depth examined the emergent behaviour of traders. I have achieved relatively stable 'Low Intelligence Traders' out-of-sample forecasting accuracy.

Surprisingly, the market populated by 'Low Intelligence Traders' returned low MAPE, MAE, and RMSE (Tables 29.0-31.0) in comparison with some of the S&P 500, IBM and GE markets composed of traders equipped with more intelligence (genome depth of 20).

The low measuring forecasting errors demonstrated by the 'Low Intelligence Traders' in some of the markets is because artificial traders with reduced levels of intelligence generate forecasts that tend to stay relatively close to the previous real price. This trader cognitive behaviour is consistent with the Adaptive Expectations Hypothesis, where decision-makers rely predominantly on recent trends in forming their future forecasts (Weigand *et al.,*2004). In my experiment traders with lower levels of intelligence tended to develop their forecasts extrapolatively, that is, based mainly on the past history of the data under consideration. The current expectations of 'Low Intelligence Agents' seem primarily based on a geometrically weighted moving average of past observations:

$$y^{*}_{it} = (1-\lambda)\sum_{j=0}^{\infty}\lambda^{j} y_{i,t-1-j} \quad (48)$$

where y_t is the endogenous variable, y^{*}_{it} represents unobservable anticipations, formed in period t-1 and λ represents the smoothing data weight parameter (Wallis, 1980).

The forecasting price generated by the 'Low Intelligence Traders' averages out to give an artificial stock market price close to the previous real market price. This in turn causes the forecasts to also be relatively close to the next real price, compared with traders with enhanced intelligence (the first experiment) that generate forecasts that tend to deviate more from the previous and next real price but have improved forecasting errors. This experiment also demonstrates the emergent consequences of traders trying to make money.

7.5. Conclusions

Developing and selecting a plausible model are the two most important ingredients of the forecasting process. My empirical results suggest that the comparison and selection of forecasting models should be based entirely on the basis of genuine out-of-sample (*ex-ante*) predictions. Generally speaking it is impossible to specify a forecasting evaluation criterion that is universally acceptable (Diebold, 1998). However, I applied formal statistical model selection criteria to establish the most accurate forecasting model to predict the stock returns of an index and two stocks. The STGP forecasting performance was compared with that of traditional econometric approaches, which demonstrated the superiority of the STGP-based algorithm. I found significant evidence that excess returns in my experiment are both statistically and economically significant, even when transaction costs are taken into account. Such evidence does not appear to have been caused by model over-fitting because I implemented an evolutionary STGP technique which does not use the same data set for forecasting purposes. My incremental walk-forward (out-of-sample tested) forecasting approach does not over-fit historical data, thereby leading to a more resilient and stable forecasting model. The possible dangers of data over-fitting emerge in many areas of finance and economics, such as the predictability of stock returns (Foster *et al.*,1997). Although my in-sample forecasts seem less impressive, there are certainly more reliable and more indicative of future predictive performance when market behaviour changes. The round-trip transaction costs level of 125 basis points suggests that significant opportunities for excess returns can be exploited in an efficient manner, regardless of transaction costs.

Bearing in mind the fact that the level of excess returns I achieved is not negligible, investors may want to consider the forecasting model described in this paper. Taking into consideration the fact that transaction costs are more likely to continue to decrease over time, one should expect the actual profitability of my model to grow accordingly. However, I think that no single investment strategy no matter how accurate is able to achieve a hit ratio approaching 100% and therefore constantly beat the stock market. Stock markets never stop their evolution and trading rules need to evolve in order to stay profitable.

The 'Low Intelligence Traders' emergent trading behaviour analysis emphasised the importance of intelligence in the trading and decision-making processes. I found empirical evidence that trader cognitive abilities play an important role in the formation of trading rules. Lack of sufficient levels of intelligence led towards more extrapolative formation of trading rules that rely on more recent data observations. Although my findings are consistent with the Hayek and the Adaptive Expectations hypotheses, I think that the presence of greater intelligence have beneficial forecasting effects and leads to more complex trading rules. Hence, traders equipped with increased intelligence demonstrate enhanced complexity of trading rules and forecasting accuracy. To summarise, this chapter makes several important contributions to the literature- first, I have provided new evidence of significant stock market predictability and profitability, even when substantial transaction costs are taken into account, and second I have provided detailed analysis of the impact of trader cognitive abilities on the accuracy of forecasting rules. Third and finally, this study conducted very rare experimental tests of the Adaptive Expectations, and the Optimism Principle.

I believe further research should include a direct comparison of the forecasting abilities of Genetic Programming and Strongly Typed Genetic Programming. Another possible avenue of future research could be the comparison of STGP with a combination of forecasts. It has been established that a weighted linear combination of different forecasts is more accurate than any individual forecast (Clement, 1989).

The existing literature also provides a number of results suggesting that small cap stocks are characterised with higher predictive ability than large cap stocks. This is clearly worth a detailed empirical analysis and, will be the subject of forthcoming research.

Chapter 8

New evidence of small-cap stocks profitability and price-volume relation

8.1. Introduction

Stock market forecasting remains a topic of continuous controversy. Some academics argue that stock price predictability is due to irrationalities, such as noisy trading or bubbles which make stock prices deviate temporarily from their fundamental values and lead to negatively autocorrelated and therefore predictable returns. In terms of equity forecasting, the question of how much to invest in large-or small-capitalisation companies is often seen as a vitally important decision due to the risk implications associated with returns in the long-term.

It is more than thirty years since the publication of two seminal papers- Banz (1981) and Reinganum (1981) on the performance of small cap companies. Banz (1981) found that market equity (ME) explains the cross-section of average returns provided by market β. The author demonstrated that the average returns on small stocks are too high given their β calculations and the corresponding average returns on large stocks are too low, leading to the misspecification of the capital asset pricing model (CAPM). He divided the stocks on the NYSE into quintiles based on market capitalisation data. Stock returns for the smallest quintile during 1926-1980 dominated the other quintiles and indexes. Reinganum (1981) argued that small-cap stocks systematically indicate average rates of return that are substantially greater than those of large-cap stocks with equivalent beta risk for at least two years. The author highlighted CAPM misspecification due to omitted risk factors from the same model leading to small-cap stock outperformance for at least two years.

My research focuses on the dynamic relation (causality) between daily stock returns and trading volume, taking into account the notion of Granger causality proposed by Wiener (1956) and Granger (1969). Many studies report uni-directional (Smirlock and Starks, 1988; Saatcioglu and Starks, 1998; Antoniewicz, 1992) and bi-directional (Copeland, 1976; Jennings *et al.*, 1981) Granger causality price-volume relationships. The findings of these studies highlight the predictability of stock returns based on trading volume. The dynamic nature of the price-volume relation is an important indicator of whether information about trading volume is useful in improving the predictability of stock returns.

I used a Strongly Typed Genetic Programming (STGP) technique, developed by Montana in 1995, to present a new type of laboratory evidence in the small-cap equity puzzle. No known study has investigated the small-cap phenomenon through the application of STGP modelling. I developed two stock markets populated by 10,000 traders each and one market with 1,000 traders only, where traders receive information about the value of the Russell 1000, the Russell 2000 and the Russell 3000 and generate forecasting orders. The Russell 3000 consists the top 3,000 US small-and large-cap companies representing 98% of the US stock market. The other two indices on my study- the Russell 1000 index and the Russell 2000 index- are subsets of the Russell 3000 index. The Russell 1000 index represents the 1,000 largest stocks in the Russell 3000 index. The Russell 2,000 index consists of 2,000 of the smallest stocks based on a mixture between market capitalisation and current index membership. While the Russell 1000 is reconstructed annually to ensure the inclusion of new and growing equities, the Russell 2000 is reconstructed annually to ensure larger stocks do not distort the characteristics of the pool of small-cap stocks.

The traders then decide whether they want to buy or sell one or more units of the index. The artificial traders in my experiment possess different trading rules and behaviours. Traders make independent forecasts creating a heterogeneous market structure. I uploaded each market with real-life historical data of the Russell indices. Consistently with the vast majority of the forecasting literature published to date, in the first stock market containing 10,000 traders I did not incorporate any transaction costs.

To determinate whether small- or large-cap stocks dominate the market in a more realistic trading scenario, I included transaction costs to my second market containing 10,000 traders of 0.25% or 25 basis points for Russell 1000 and Russell 3000 and 0.60% or 60 basis points for Russell 2000. My last stock market was populated by 1,000 artificial traders and had no transaction costs.

I then compared the in-sample and out-of-sample STGP models to traditional econometric forecasting techniques such as Box-Jenkins and Holt-Winters. The forecasting performance is evaluated through five measures in order to investigate the following:

I.Whether small-cap stocks are more predictable than large-cap stocks. An investigation into whether the three indices returns in excess of the risk free rate are predictable and most importantly profitable (quantifying the precise level of generated profit after taking appropriate transaction costs into account).

II. The dynamic causal relationship between trading volume and index returns. An investigation into whether the level of in-sample trading volume is a good predictor for the out-of-sample stock returns.

III. Whether a market with a reduced population of only 1000 traders is capable of generating accurate and profitable forecasts.

Specifically, the major contributions of this chapter to the field are:

•To show the index returns and the exact level of profitability when appropriate transaction costs are taken into account.

•To shed light on the ‘size effect’ equity puzzle which generated a lively debate on market efficiency and asset pricing.

•To empirically demonstrate the impact of trading volume on the future direction of stock returns.

•To conduct unique experimental tests of the errors-in-expectations hypothesis and the visibility hypothesis within laboratory-based artificial stock market settings.

•To demonstrate that genetic diversity plays a vitally important role in forecasting accuracy and stock market behaviour.

My findings are consistent with the recent wave of research suggesting that small-cap stocks are more predictable and tend to cyclically outperform large-cap stocks. Moreover, I demonstrate that small-cap stocks generate higher returns after transaction costs and therefore they are more profitable in overall terms.

I showed that enhanced genetic diversity and the presence of more artificial traders have beneficial effects on forecasting accuracy. My investigation on the dynamic price-volume causal relationship revealed a very weak presence of bi-directional and a few cases of uni-directional relationships.

8.2. Background

By approximate definitions, small-cap stocks have a smaller market capitalisation within the range of $250 million and $2 billion. Small-cap stocks often outperform large-cap stocks, but their dominance is characterised by a rather cyclical nature. From 1925 to 1964, small-caps and large-caps had identical returns. In the next four years to 1968, the small-cap asset returns doubled compared to the large-cap ones. During the following five years this advantage disappeared. The period from 1973 to 1983 is characterized by the largest small-cap stocks outperformance (Bogle, 2000). However, from an historical point of view, small-cap stocks delivered higher Earnings Per Share (EPS) growth rates than large-cap stocks (about 50% higher on average).

Recent data released by Russell Investments shows that in the last decade small-cap stock returns exceeded those of large-cap stocks seven times (Oharazawa, 2010). The small-cap Russell 2000 index outpaced the large-cap Russell 1000 index by nearly fifty percentage points during the last decade (Opdyke, 2010). For the last fifty years (up to December 31, 2011), $1 invested in large-cap stocks would have grown to $83. The same investment in small-cap stocks would have grown to $263 (Standard and Poor’s, 1957-2011). Moreover, small-cap stocks appear more predictable than large-cap stocks (Avramov, 2001).

Although the dominance of the small-cap stocks has not been constant and steady, recent studies have continued to document the importance of size in determining asset returns.

Johnson *et al.* (1999) argued that the returns to small-cap stocks are over four times higher (about 2,000 basis points) than the returns to large-cap stocks during US Democratic administrations. Siegel (2002) suggested that relatively young markets are more likely to reject the weak form of market efficiency. Ready (2002) and Qi and Wu (2006) demonstrate that more arbitrage opportunities exist in young markets than in mature ones.

Hansen *et al.* (2005) found significant calendar effects in small stock indices. Hsu and Kuan (2005) reported the existence of significantly profitable trading rules in young markets proxied by NASDAQ Composite and Russell 2000 from 1989 to 2002, but not in mature markets such as the DJIA and S&P 500 for the same period.

Switzer and Fan (2007) suggested that the small-cap stock outperformance is country dependent and demonstrated this by adding Canadian small-cap stocks for international investors in enhancing their risk-return performance. In another study, Switzer (2010) investigated the small-cap premium in US and Canada. The author demonstrated that since 2000, economically and statistically significant abnormal performance has been observed in small-cap stocks in both countries. Hsu *et al.* (2010) used S&P SmallCap 600/Citigroup Growth, Russell 2000, and NASDAQ Composite to proxy for the small and growth segments of the US stock markets and found significant predictive power for those stock portfolios in the pre-exchange trading periods. Shynkevich (2012) found profitable technical trading rules for a number of small-cap sector portfolios after adjusting for data-snooping bias, but only before the inclusion of transaction costs.

Allen and Karjalainen (1999) were the first to implement a Genetic Programming (GP) approach to create trading rules for the S&P index from 1928 to 1995. Kaboudan (1999) achieved more reliable forecasting results by the introduction of a new GP probability measure of the time series' predictability. Kaboudan (2001) used a GP approach to generate reasonable one-day-ahead forecasts of stock prices. Lawrenz and Westerhoff (2003) applied GP techniques to design a simple but efficient exchange-rate model to understand the driving forces of the foreign exchange markets. Zhou (2004) used GP to develop an effective emerging markets stock selection model.

Potvin *et al.* (2004) generated positive excess returns (without considering transaction costs) by applying GP on 14 Canadian companies for 1992-2000. Fyfe *et al*. (2005) found that basic GP returns are higher than buy-and-hold returns for three different S&P indices. How *et al.* (2010) used a GP approach to find substantial evidence that technical trading rules have a higher predictive ability for the Russell 2000 index than the Russell 1000 index. However, once transaction costs were included the small-cap (Russell 2000 index) disappeared. The positive excess returns for the Russell 2000 index turned into large negative annual average excess returns of up to 9.9%.

Research on the stock price-volume relationship goes back to the early 1950's. The early studies on price-volume relation examined the contemporaneous relationships between absolute price changes and trading volume and suggested positive relationships between daily price changes and daily trading volume for both stocks and indices (Ying, 1966; Westerfield, 1977; Rutledge, 1984).

According to Karpoff (1987), there are at least three reasons why the price-volume relation is important: 1) it provides a clear view of the financial markets structure; 2) it plays a significant role in the empirical distribution of speculative prices; and 3) it has substantial implications on futures markets and event studies that use a mixture of price and volume data from which to draw inferences.

Several subsequent studies have found a positive price-volume relationship (Courouch, 1970; Epps and Epps, 1976; Haris, 1986; Chen *et al.,* 2001; Khan and Razwan, 2001; Lee and Rui, 2002; Pisedtasalasai and Gunasekarge, 2008).

Chordia and Swaminathan (2000) found that daily returns of stocks with high trading volume lead daily returns of stocks with low trading volume. Wang (1994) investigated the dynamic (causal) relationships between volume and returns and stated that volume may provide information about expected future returns. Hiemstra and Jones (1994) and Fajihara and Mougoue (1997) discovered bi-directional nonlinear causality in the prices and trading volume. Silvapulle and Choi (1999) found strong evidence of nonlinear bi-directional causality between stock returns and volume series.

Chen *et al.* (2001) conducted a comprehensive study examining the dynamic stock price-volume relation using daily data from nine major markets. The authors reported strong evidence for the argument that returns causes volume but no evidence to suggest that volume causes returns.

Chen and Liao (2002) investigated the stock price-volume relationship from the perspective of an agent-based model of stock markets and found the presence of bi-directional causality between stock returns and trading volume in all four artificial stock markets of different design. Chen and Liao (2005) regarded the price-volume relationship as a genetic property of a financial market and its full understanding cannot be accomplished unless the causal relationship between individual behaviour at the bottom and aggregate phenomena at the top is well understood.

The significance of all these findings is that trading volume can help predict stock returns. In other words, the knowledge of current trading volume improves the ability to forecast stock prices, because the trading volume has predictive power on stock returns enabling traders to develop volume-based strategies.

8.3. Artificial stock market settings for this particular experiment

Table 40.0 represents the main settings of the STGP-based artificial stock market model for this particular experiment.

Artificial stock market parameters	
Total population size (agents)	10,0000
Initial wealth(equal for all agents)	100,000
One way transaction costs	0.60% (60 bsp) for small cap index; 0.25% (25 bsp) for large cap index.
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 (bars)	1
Minimum breeding age (bars)	80
Initial selection type	random
Parent selection (percentage of initial selection that will breed)	5%
Mutation probability (per offspring)	10%
Total number of quotes processed-Russell 1000	8,840 (7,306 in-sample, 1534 out-of-sample) 21/05/1979-01/01/2007 ; 02/01/2007-16/11/2012
Total number of quotes processed-Russell 2000	8,840 (7,306 in-sample, 1534 out-of-sample) 21/05/1979-01/01/2007 ; 02/01/2007-16/11/2012
Total number of quotes processed-Russell 3000	8,840 (7,306 in-sample, 1534 out-of-sample) 21/05/1979-01/01/2007 ; 02/01/2007-16/11/2012
Seed generation from clock	Yes
Creation of unique genomes	Yes
Offspring will replace the worst performing agents of the initial selection	Yes

Table 40.0 Artificial Stock Market Parameter Settings

*Maximum genome size measure the total number of nodes in an agent'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 an agent's genome (trading rule). The depth of a trading rule can be an indicator of its complexity.

8.4. One-step-ahead (static) forecasts and benchmark forecasting models.

I adopted one-step-ahead forecasts because it offers the opportunity to estimate a sequence of one-step-ahead predictions, using the actual rather than forecasted values for lagged dependent variables (if available). The dynamic or multi-step forecasts adopt previously forecasted values for the lagged dependent variables that are used in the process of generating forecasts of the current value. One of the most important advantages of the one-step-ahead forecast is the ability to avoid problems associated with cumulative errors from the previous period for out-of-sample forecasting (Makridakis and Winkler, 1989). I chose the Box-Jenkins model and Holt-Winters (three parameters) multiplicative exponential smoothing model as benchmarks for comparison to the STGP technique.

The Box-Jenkins or ARIMA approach was introduced in the 1970, and since then has become one of the most popular tools for time series forecasting. The technique has been successfully applied to some of the most difficult forecasting problems. Some studies suggest that 13 per cent of the financial services industry use exponential smoothing models such as the Holt-Winters model. The Holt-Winters multiplicative (additive trend, multiplicative seasonality) method represents a weighted average of past data, in which the actual weights decline geometrically over a time horizon in order to capture short-term fluctuations in the values. The most recent observations in the Holt-Winters forecasting model have the most relevance in forecasting the future values.

8.5. Measures of forecasting accuracy.

As is standard in the forecasting literature, I estimated the Mean Absolute Percent Error (MAPE), Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). The lower the output of the error statistics achieved, the better the forecasting accuracy of the model concerned. I used the three forecasting accuracy measures to compare the forecasting errors within the Russell 1000, the Russell 2000 and the Russell 3000 individual datasets, but not between them because this measure of accuracy is scale-dependent (Hyndman and Koehler, 2006). The three measures of forecasting accuracy are quantified as follows:

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{d_t - z_t}{d_t}\right| \tag{49}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|d_t - z_t\right| \tag{50}$$

$$RMSE = \left\{\frac{1}{N}\sum_{t=1}^{N}\left(d_t - z_t\right)^2\right\}^{0.5} \tag{51}$$

Where N is the number of forecasting periods, d_i is the actual price of the Russell 1000, 2000 and 3000 at period t, and z_t is the forecasting stock price at period t (Pai and Lin, 2005). Additionally I discuss the values of the Maximum Absolute Error (MaxAE) and the Maximum Absolute Percentage Error (MaxAPE).

8.6. Simulation results

8.6.1. Comprehensive analysis of whether small-cap stocks are more predictable than large-cap stocks. An investigation into whether the three indices' returns, in excess of the risk free rate are predictable, and most importantly profitable.

The in-sample and out-of-sample summaries of the descriptive statistics of the daily returns for the Russell 1000, the Russell 2000 and the Russell 3000 are illustrated in Table 41.0.

Russell 1000		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Daily mean	0.0843	0.8492
Daily standard deviation	0.0617	0.1367
Skewness	0.5471	-0.6755
Kurtosis	1.7686	2.5955
J-B	826.0435 (0.0000) ^{p}	127.1426 (0.0000) ^{p}
ADF*	-52.5321 (0.0000) ^{p}	-52.5351 (0.0000) ^{p}
Russell 2000		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Daily mean	0.0985	1.1382
Daily standard deviation	0.0659	0.1925
Skewness	0.6956	-0.8539
Kurtosis	2.4016	2.8394
J-B	688.6519 (0.0000) ^{p}	188.0888 (0.0000) ^{p}
ADF*	-61.5108 (0.0001) ^{p}	-43.3756 (0.0000) ^{p}
Russell 3000		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Daily mean	0.0851	0.8671
Daily standard deviation	0.0615	0.1397
Skewness	0.5335	0.7080
Kurtosis	1.7536	2.6286
J-B	819.5089 (0.0000) ^{p}	136.9812 (0.0000) ^{p}
ADF*	-62.7357 (0.0001) ^{p}	-43.7368 (0.0000) ^{p}

*The MacKinnon (1996) one-sided critical value for rejection of the null hypothesis of a unit root at 5% level is -3.410060; ***J-B:*** the Jarqu-Bera test; ***ADF:*** Augmented Dickey-Fuller Unit Root Test; ***p***: the *p* value

Table 41.0 Descriptive statistics for the Russell 1000, Russell 2000, and Russell 3000 in-sample and out-of-sample daily index returns.

Following the existing forecasting literature, return is defined as the natural logarithm of value relatives. The return series are asymmetric as illustrated by a non-zero skewness. While there is no presence of excess kurtosis in all of the three indices, the out-of-sample descriptive statistics indicate higher kurtosis than the in-sample descriptive statistics (more of the out-of-sample variance is the result of infrequent extreme deviations). According to the Jarque-Bera test results, the null hypothesis that the price series generated by the artificial traders is normally distributed is rejected in all in-sample and out-of-sample periods (Table 41.0).

Tables 42.0-43.0 represents the in-sample and out-of-sample forecasting performance of Box-Jenkins and Holt-Winters for the three indices without any transaction costs included.

Box-Jenkins (without transaction costs) Russell 1000 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Box-Jenkins model type	ARIMA(2,1,1)	ARIMA(1,1,1)
RMSE	8.003	17.702
MAE	4.389	12.315
MAPE	0.663	1.1012
MaxAPE	22.642	9.838
MaxAE	87.379	107.190
Normalized BIC	4.165	5.762

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 42.0 In-sample and out-of-sample performance of Box-Jenkins (without transaction costs) forecasting models of Russell 1000 daily index return series.

Holt-Winters (without transaction costs) Russell 1000 statistics		
Forecasting sample	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	8.006	17.780
MAE	4.385	12.283
MAPE	0.663	1.010
MaxAPE	22.757	10.218
MaxAE	86.929	109.107
Normalized BIC	4.162	5.761

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 43.0 In-sample and out-of-sample performance of Holt-Winters (without transaction costs) multiplicative exponential smoothing forecasting models of Russell 1000 daily index return series.

Box-Jenkins (without transaction costs) Russell 2000 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Box-Jenkins model type	ARIMA(1,1,5)	ARIMA(0,1,1)
RMSE	9.655	30.008
MAE	5.346	22.064
MAPE	0.648	1.357
MaxAPE	13.596	13.332
MaxAE	86.159	22.064
Normalized BIC	4.540	6.808

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 44.0 In-sample and out-of-sample performance of Box-Jenkins (without transaction costs) forecasting models of Russell 2000 daily index return series.

Holt-Winters (without transaction costs) Russell 2000 statistics		
Forecasting sample	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	9.619	29.999
MAE	5.357	22.050
MAPE	0.655	1.356
MaxAPE	14.039	13.331
MaxAE	87.431	32.469
Normalized BIC	4.531	6.807

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 45.0 In-sample and out-of-sample performance of Holt-Winters (without transaction costs) multiplicative exponential smoothing forecasting models of Russell 2000 daily index return series.

Box-Jenkins (without transaction costs) Russell 3000 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Box-Jenkins model type	ARIMA(0,1,1)	ARIMA(0,1,4)
RMSE	7.986	18.344
MAE	4.402	12.835
MAPE	0.654	1.034
MaxAPE	21.676	10.098
MaxAE	86.119	107.533
Normalized BIC	4.158	5.833

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 46.0 In-sample and out-of-sample performance of Box-Jenkins (without transaction costs) forecasting models of Russell 3000 daily index return series.

Holt-Winters (without transaction costs) Russell 3000 statistics		
Forecasting sample	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	7.976	18.426
MAE	4.390	12.800
MAPE	0.654	1.031
MaxAPE	22.006	10.112
MaxAE	86.960	110.214
Normalized BIC	1.244	5.832

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 47.0 In-sample and out-of-sample performance of Holt-Winters (without transaction costs) multiplicative exponential smoothing forecasting models of Russell 3000 daily index return series.

Tables 48.0-53.0 illustrate the in-sample and out-of-sample forecasting statistics of both traditional econometric models for the three financial instruments with transaction costs added on.

Box-Jenkins (transaction costs included) Russell 1000 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Box-Jenkins model type	ARIMA(2,1,1)	ARIMA(1,1,1)
RMSE	8.804	19.472
MAE	4.828	13.546
MAPE	0.663	1.012
MaxAPE	22.642	9.839
MaxAE	96.118	117.908
Normalized BIC	4.355	5.952

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 48.0 In-sample and out-of-sample performance of Box-Jenkins forecasting models with transaction costs of Russell 1000 daily index return series.

Holt-Winters (transaction costs included) Russell 1000 statistics		
Forecasting sample	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	8.806	19.558
MAE	4.823	13.512
MAPE	0.663	1.010
MaxAPE	22.757	10.218
MaxAE	95.622	120.679
Normalized BIC	4.352	5.952

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 49.0 In-sample and out-of-sample performance of Holt-Winters multiplicative exponential smoothing forecasting models with transaction costs of Russell 1000 daily index return series.

Box-Jenkins (transaction costs included) Russell 2000 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Box-Jenkins model type	ARIMA(1,1,5)	ARIMA(0,1,1)
RMSE	10.620	33.009
MAE	5.881	24.271
MAPE	0.648	1.357
MaxAPE	13.595	13.331
MaxAE	94.775	178.709
Normalized BIC	4.730	6.998

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 50.0 In-sample and out-of-sample performance of Box-Jenkins forecasting models with transaction costs of Russell 2000 daily index return series.

Holt-Winters (transaction costs included) Russell 2000 statistics		
Forecasting sample	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	10.581	32.999
MAE	5.893	24.255
MAPE	0.655	1.356
MaxAPE	14.037	13.331
MaxAE	96.175	178.716
Normalized BIC	4.722	6.998

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 51.0 In-sample and out-of-sample performance of Holt-Winters multiplicative exponential smoothing forecasting models with transaction costs of Russell 2000 daily index return series.

Box-Jenkins (transaction costs included) Russell 3000 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Box-Jenkins model type	ARIMA(0,1,1)	ARIMA(1,1,1)
RMSE	8.785	20.178
MAE	4.842	14.118
MAPE	0.654	1.034
MaxAPE	21.676	10.098
MaxAE	94.731	118.287
Normalized BIC	4.394	6.024

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 52.0 In-sample and out-of-sample performance of Box-Jenkins forecasting models with transaction costs of Russell 3000 daily index return series.

Holt-Winters (transaction costs included) Russell 3000 statistics		
Forecasting sample	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
Holt-Winters model type	Multiplicative smoothing	Multiplicative smoothing
RMSE	8.774	20.268
MAE	4.829	14.080
MAPE	0.654	1.031
MaxAPE	22.006	10.112
MaxAE	95.656	121.235
Normalized BIC	4.345	6.023

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 53.0 In-sample and out-of-sample performance of Holt-Winters multiplicative exponential smoothing forecasting models with transaction costs of Russell 3000 daily index return series.

Tables 54.0-56.0 report the in-sample and out-of-sample forecasting ability of the STGP for Russell 1000, 2000 and 3000.

Strongly Typed Genetic Programming (without transaction costs) Russell 1000 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
RMSE	7.654	15.045
MAE	3.930	10.435
MAPE	0.646	0.848
MaxAPE	19.707	8.183
MaxAE	77.035	100.694
Strongly Typed Genetic Programming (transaction costs included) Russell 1000 statistics		
RMSE	9.023	17.363
MAE	4.965	10.066
MAPE	0.717	0.852
MaxAPE	103.597	6.217
MaxAE	144.407	100.202
Strongly Typed Genetic Programming (1000 traders) Russell 1000 statistics		
RMSE	13.140	17.672
MAE	5.827	11.420
MAPE	1.252	0.936
MaxAPE	234.955	8.882
MaxAE	292.506	99.899

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 54.0 In-sample and out-of-sample performance of Strongly Typed Genetic Programming forecasting models of Russell 1000 daily index return series.

Strongly Typed Genetic Programming (without transaction costs) Russell 2000 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
RMSE	9.618	28.476
MAE	5.222	20.927
MAPE	0.610	1.275
MaxAPE	10.762	11.484
MaxAE	73.904	14.123
Strongly Typed Genetic Programming (transaction costs included) Russell 2000 statistics		
RMSE	12.912	28.476
MAE	6.930	20.927
MAPE	1.145	1.275
MaxAPE	113.639	11.484
MaxAE	173.900	141.123
Strongly Typed Genetic Programming (1000 traders) Russell 2000 statistics		
RMSE	14.902	28.649
MAE	7.678	20.716
MAPE	1.225	1.269
MaxAPE	101.734	12.216
MaxAE	185.431	171.018

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error

Table 55.0 In-sample and out-of-sample performance of Strongly Typed Genetic Programming forecasting models of Russell 2000 daily index return series.

Strongly Typed Genetic Programming (without transaction costs) Russell 3000 statistics		
Forecasting range	***In-sample***	***Out-of-sample***
Dates	21/05/1979-01/01/2007	02/01/2007-16/11/2012
Number of observations	7306	1534
RMSE	6.012	12.600
MAE	3.111	8.950
MAPE	0.548	0.700
MaxAPE	20.007	6.622
MaxAE	80.643	84.236
Strongly Typed Genetic Programming (transaction costs included) Russell 3000 statistics		
RMSE	9.143	12.783
MAE	5.018	8.785
MAPE	0.713	0.689
MaxAPE	63.506	8.208
MaxAE	102.986	97.347
Strongly Typed Genetic Programming (1000 traders) Russell 3000 statistics		
RMSE	18.038	19.356
MAE	6.334	13.370
MAPE	1.284	1.005
MaxAPE	125.980	8.762
MaxAE	504.558	111.276

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

Table 56.0 In-sample and out-of-sample performance of Strongly Typed Genetic Programming forecasting models of Russell 3000 index return series.

Prior to the main analysis of this study, the Russell 1000, 2000 and 3000 returns and trading volume were subjected to stationary tests. A stationary time series is a necessary condition for developing the Box-Jenkins forecasting models. For this purpose, the Augmented Dickey-Fuller (ADF) test was conducted. I performed the ADF test by running a regression of the first difference of the log price series against the series lagged once (a sufficient condition to eliminate autocorrelation in the residuals) combined with a drift and a time trend. The process is quantified by:

$$\Delta\rho_t = \beta_1\rho_{t-1} + \sum_{i=1}^{4}\beta_{i+1}\Delta\rho_{t-1} + \beta_6 + \beta_7 t \qquad (52)$$

Under the null hypothesis of the ADF test $p_t\left(\ln(p_t)\right)$ contains a unit root $(\beta_1 = 0)$. The alternative hypothesis of no unit root presence is rejected when $\beta_1 \neq 0$. The null hypothesis of the presence of a unit root has been rejected for both the in-sample and out-of-sample periods of the Russell 1000, the Russell 2000 and the Russell 3000 return series (Table 41.0). Hence, the returns of the three indices are stationary at the 95% significance level:

$$r_t = \ln(P_t) - \ln(P_{t-1}) \tag{53}$$

where r_t denotes the returns of the three indices and P_t represents the price series at time t. This finding is consistent with Lee *et al.* (2010) who discovered stationary price series in 32 developed and 26 developing countries.

Based on non-normal distribution nature of the time series in my experiment, I apply the Wilcoxon signed-rank (WSR) test to determinate whether the forecasts from two competing models are equally accurate. The null hypothesis of the WSR test is that the two populations represented by the respective members of the matched pairs are identical. When the null hypothesis is true, then each of the 2^N possible sets of signed ranks estimated by arbitrarly assigning plus or minus signs to be ranks 1 through N is equally likely (DeFusco *et al.,* 1990). The test statistic is:

$$WSR = \sum_{t=1}^{N} I + (d_t) rank(|d_t|) \tag{54}$$

Where $I + (d_t) = \begin{cases} 1 \\ 0 \end{cases}$ if $D_t > 0$

Where $rank(|d_t|)$ denotes the rank of the absolute value of d_t (Alon *et al.*, 2001).

I began my econometric analysis with a comparison of the in-sample and out-of-sample one-step-ahead forecasting accuracy of the STGP, Box-Jenkins and Holt-Winters models for the three indices, without taking into account any transaction costs.

All five performance measures (MAPE, MAE, RMSE, MaxAPE and MaxAE) suggest that the STGP technique outperforms Box-Jenkins in all in-sample and out-of-sample experiments for the three financial instruments (Tables 57.0, 58.0 and 59.0). I detect substantially larger out-of-sample difference between the two competing forecasting models, indicating higher *ex-ante* STGP accuracy.

STGP vs. B-J in-sample forecasting models based on Russell 1000 daily index returns							
GP model	***WSR [a]***	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP-no transaction costs	39.044 (0.0000) [p]	***Difference [b]***	-0.349*	-0.459*	-0.017*	-2.935*	-10.344*
STGP with transaction costs	36.397 (0.0000) [p]	***Difference [b]***	0.219*	0.137*	0.054*	80.955*	48.289*
STGP- 1000 traders	44.722 (0.0000) [p]	***Difference [b]***	4.336**	0.999*	0.589*	212.313*	196.388*
STGP vs. B-J out-of-sample forecasting models based on Russell 1000 daily index returns							
STGP-no transaction costs	28.961 (0.0000) [p]	***Difference [b]***	-2.657*	-1.880*	-0.253*	-1.700*	-6.496*
STGP with transaction costs	33.913 (0.0000) [p]	***Difference [b]***	-2.109*	-3.480*	-0.249*	-3.622*	-17.706*
STGP- 1000 traders	33.913 (0.0000) [p]	***Difference [b]***	-1.800**	-2.126*	-0.076*	-0.957*	-18.009*

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.

Difference [b]= STGP-BJ models; Negative values indicate preference to STGP models.

****Indicates significance at the 99% level.***

*****Indicates significance at the 95% level.***

Table 57.0 WSR test and paired comparisons of forecasting performance: STGP vs. B-J forecasting models of Russell 1000 daily index returns.

STGP vs. B-J in-sample forecasting models based on Russell 2000 daily index returns							
GP model	***WSR [a]***	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP-no transaction costs	69.337 (0.0000) [p]	***Difference [b]***	-0.037**	-0.124*	-0.038*	-2.834*	-12.255*
STGP with transaction costs	72.576 (0.0000) [p]	***Difference [b]***	2.292*	1.049*	0.577*	100.041*	79.125*
STGP- 1000 traders	72.623 (0.0000) [p]	***Difference [b]***	4.282*	1.797**	0.577*	88.139*	90.656*
STGP vs. B-J out-of-sample forecasting models based on Russell 2000 daily index returns							
STGP-no transaction costs	6.915 (0.0000) [p]	***Difference [b]***	-1.532*	-1.137*	-0.082*	-1.848*	-7.941*
STGP with transaction costs	33.913 (0.0000) [p]	***Difference [b]***	-4.533*	-3.344*	-0.082*	-1.847*	-37.586*
STGP-1000 traders	33.913 (0.0000) [p]	***Difference [b]***	-4.360*	-3.555*	-0.088*	-1.115*	-7.691*

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.

Difference [b]= STGP-BJ models; Negative values indicate preference to STGP models.

****Indicates significance at the 99% level.***

*****Indicates significance at the 95% level.***

Table 58.0 WSR test and paired comparisons of forecasting performance: STGP vs. B-J forecasting models of Russell 2000 daily index returns

STGP vs. B-J in-sample forecasting models based on Russell 3000 daily index returns							
GP model	***WSR*** [a]	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP-no transaction costs	43.335 (0.0000) [p]	***Difference*** [b]	-1.974*	-1.291**	-0.106*	-1.669*	-5.476*
STGP with transaction costs	44.439 (0.0000) [p]	***Difference*** [b]	0.358*	0.176*	0.059*	41.830*	8.255*
STGP- 1000 traders	46.564 (0.0000) [p]	***Difference*** [b]	9.253*	1.505*	0.630*	104.304*	409.827*
STGP vs. B-J out-of-sample forecasting models based on Russell 3000 daily index returns							
STGP-no transaction costs	32.068 (0.0000) [p]	***Difference*** [b]	-5.744*	-3.885*	-0.344*	-3.476*	-23.297*
STGP with transaction costs	33.913 (0.0000) [p]	***Difference*** [b]	-7.395*	-5.333*	-0.345*	-1.890*	-20.940*
STGP- 1000 traders	13.403 (0.0000) [p]	***Difference*** [b]	-0.822*	-0.748*	-0.029*	-1.350**	-7.02*

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.

Difference [b]***= STGP-BJ models; Negative values indicate preference to STGP models.***

****Indicates significance at the 99% level.***

*****Indicates significance at the 95% level.***

Table 59.0 WSR test and paired comparisons of forecasting performance: STGP vs. B-J forecasting models of Russell 3000 daily index returns.

A comparison between STGP and Holt-Winters reports the same phenomenon (Tables 60.0, 61.0 and 62.0).

STGP vs. H-W in-sample forecasting models based on Russell 1000 daily index returns							
GP model	***WSR*** [a]	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP-no transaction costs	39.329 (0.0000) [p]	***Difference*** [b]	-0.349*	-0.455*	-0.017*	-3.050*	-9.849*
STGP with transaction costs	36.110 (0.0000) [p]	***Difference*** [b]	0.217*	0.142*	0.054*	80.840*	48.785*
STGP- 1000 traders	44.447 (0.0000) [p]	***Difference*** [b]	4.334*	1.004*	0.589*	139.333**	196.844*
STGP vs. H-W out-of-sample forecasting models based on Russell 1000 daily index returns							
STGP-no transaction costs	29.579 (0.0000) [p]	***Difference*** [b]	-2.735*	-1.848*	-0.162*	-2.080*	-8.413*
STGP with transaction costs	33.913 (0.0000) [p]	***Difference*** [b]	-2.195*	-3.446*	-0.158*	-4.001*	-20.477*
STGP- 1000 traders	33.913 (0.0000) [p]	***Difference*** [b]	-1.886*	-2.090**	-0.074*	-1.336*	-20.780*

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.

WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.

Difference [b]***= STGP-BJ models; Negative values indicate preference to STGP models.***

****Indicates significance at the 99% level.***

*****Indicates significance at the 95% level.***

Table 60.0 WSR test and paired comparisons of forecasting performance: STGP vs. H-W forecasting models of Russell 1000 daily index returns.

STGP vs. H-W in-sample forecasting models based on Russell 2000 daily index returns							
GP model	***WSR*** [a]	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP-no transaction costs	70.910 (0.0000) [p]	***Difference*** [b]	-0.001*	-0.135*	-0.045*	-3.277*	-13.527*
STGP with transaction costs	72.570 (0.0000) [p]	***Difference*** [b]	2.331*	1.037*	0.490*	99.602**	77.725*
STGP- 1000 traders	72.622 (0.0000) [p]	***Difference*** [b]	4.321*	1.785*	0.570*	87.697*	89.256*
STGP vs. H-W out-of-sample forecasting models based on Russell 2000 daily index returns							
STGP-no transaction costs	6.915 (0.0000) [p]	***Difference*** [b]	-1.523*	-1.123*	-0.081*	-1.847*	-18.346**
STGP with transaction costs	33.913 (0.0000) [p]	***Difference*** [b]	-4.523*	-3.328*	-0.081*	-1.847*	-37.593*
STGP-1000 traders	33.913 (0.0000) [p]	***Difference*** [b]	-4.350*	-3.539*	-0.087*	-1.115*	-7.698*

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.
WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.
Difference [b]***= STGP-BJ models; Negative values indicate preference to STGP models.***
****Indicates significance at the 99% level.***
*****Indicates significance at the 95% level.***

Table 61.0 WSR test and paired comparisons of forecasting performance: STGP vs. H-W forecasting models of Russell 2000 daily index returns.

STGP vs. H-W in-sample forecasting models based on Russell 3000 daily index returns							
GP model	***WSR*** [a]	***Statistics***	***RMSE***	***MAE***	***MAPE***	***MaxAPE***	***MaxAE***
STGP-no transaction costs	43.614 (0.0000) [p]	***Difference*** [b]	-1.964*	-1.279*	-0.106*	-1.999*	-6.317*
STGP with transaction costs	44.144 (0.0000) [p]	***Difference*** [b]	0.369*	0.189*	0.059*	41.599*	7.330*
STGP- 1000 traders	46.279 (0.0000) [p]	***Difference*** [b]	9.264**	1.505*	0.630*	103.974*	408.902*
STGP vs. H-W out-of-sample forecasting models based on Russell 3000 daily index returns							
STGP-no transaction costs	32.604 (0.0000) [p]	***Difference*** [b]	-5.826*	-3.850*	-0.311*	-3.490*	-25.987*
STGP with transaction costs	33.913 (0.0000) [p]	***Difference*** [b]	-7.485*	-5.295*	-0.342*	-1.904*	-23.888*
STGP-1000 traders	13.403 (0.0000) [p]	***Difference*** [b]	-0.912*	-0.710*	-0.026*	-1.350*	-9.968**

RMSE-Root Mean Squared Error; ***MAE***-Mean Absolute Error; ***MAPE***-Mean Absolute Percentage Error; ***MaxAPE***-Maximum Absolute Percentage Error; ***MaxAE***-Maximum Absolute Error.
WSR [a]-Wilcoxon Signed Ranks Test based on positive ranks. [p] refers to the *p*-value of WSR test.
Difference [b]***= STGP-BJ models; Negative values indicate preference to STGP models.***
****Indicates significance at the 99% level.***
*****Indicates significance at the 95% level.***

Table 62.0 WSR test and paired comparisons of forecasting performance: STGP vs. H-W forecasting models of Russell 3000 daily index returns.

The WSR *p*-values (in parenthesis) reported in Tables 57.0 to 62.0 suggest that STGP outperformed both traditional econometric models at the 99% significance level, thereby rejecting the null hypothesis of equality in the forecasting difference between the competing models.

My next experiment was performed under real-life trading conditions by the inclusion of appropriate transaction costs. Each of the three financial instruments was estimated and validated by the in-sample data. This model estimation process was then followed by an empirical evaluation which was based on the e*x-ante* data of 1,534 observations from 02/01/2007 to 16/11/2012. In-sample empirical results of the Russell 1000, reported in Tables 57.0 and 60.0, indicate that both the Box-Jenkins and Holt-Winters models slightly outperformed the STGP forecasting model with the following paired differences between them: MAE of 0.142, RMSE of 0.219 and MAPE of 0.054 (the highest values reported).

The two econometric forecasting models for in-sample small-cap stocks outperformed STGP by a bigger margin with paired differences of MAE of 1.049, RMSE of 2.331 and MAPE of 0.577 (Tables 58.0 and 61.0). Similar to in-sample of the large-cap stocks, both Box-Jenkins and Holt-Winters of the Russell 3000 slightly outperformed the STGP technique with the following highest values of the paired differences between them: MAE of 0.189, RMSE of 0.369 and MAPE of 0.059 (Tables 59.0 and 62.0).

The out-of-sample validation illustrates, however, the superiority of STGP in absolutely all experiments. The out-of-sample paired comparisons indicate that the STGP substantially outperform the traditional econometric models (statistically significant results at the 99% level). *Ex-ante,* STGP significantly outperformed the Box-Jenkins model for Russell 1000 (MAE of -3.480, RMSE of -2.109 and MAPE of -0.249) and the multiplicative exponential smoothing model (MAE of -3.446, RMSE of -2.195 and MAPE of -0.158). Tables 58.0 and 61.0 demonstrate even bigger Russell 2000 out-of-sample STGP dominance over the two traditional econometric forecasting models. Tables 59.0 and 62.0 report a substantial *ex-ante* STGP outperformance for the Russell 3000.

The paired differences between STGP and the other two competing models are larger: MAE of -5.333, RMSE of -7.395 and MAPE of -0.345 for Box-Jenkins, and MAE of -5.295, RMSE of -7.485 and MAPE of -0.342 for Holt-Winters. Figures 32.0, 33.0 and 34.0 clearly illustrate the difference in forecasting direction of the competing models.

Figure 32.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the Russell 1000 daily index returns.

Figure 33.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the Russell 2000 daily index returns.

Figure 34.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the Russell 3000 daily index returns.

To obtain a more realistic forecasting comparison picture I plotted 1,534 out-of-sample real data quotes (Figures 35.0, 36.0 and 37.0) and observed that the STGP curve (the green curve) overlapped with the real data curve, suggesting the superior out-of-sample predictability of the STGP technique. Figures 35.0, 36.0 and 37.0 clearly illustrate the substantial *ex-ante* forecasting accuracy between STGP and the two traditional econometric models.

Figure 35.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the Russell 1000 daily index returns. A plot of real data has been added on for comparison purposes.

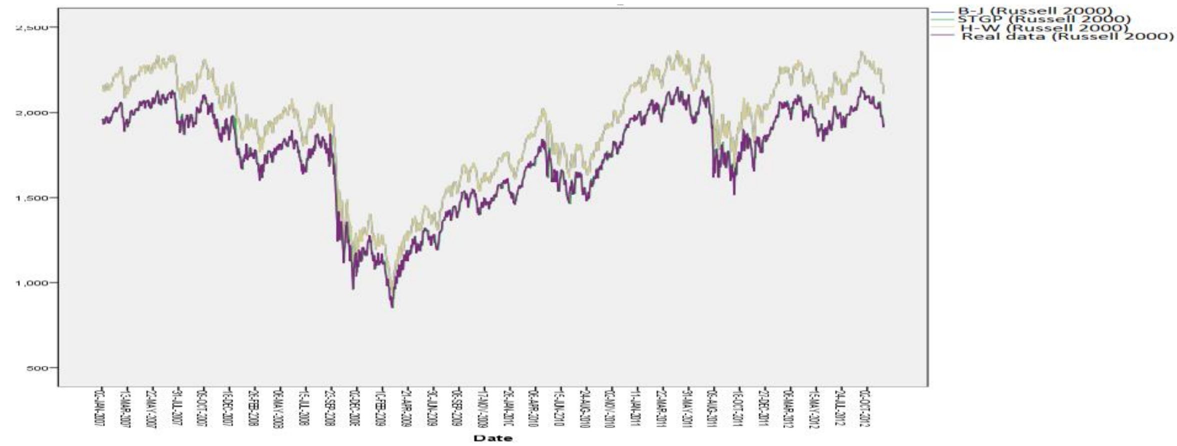

Figure 36.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the Russell 2000 daily index returns. A plot of real data has been added on for comparison purposes.

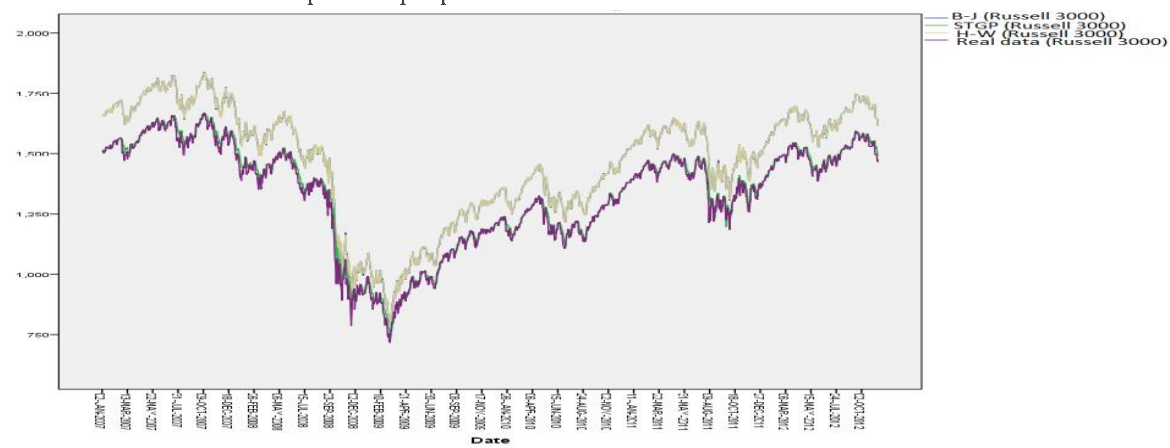

Figure 37.0 Out-of-sample graphical comparisons of forecasting performance: STGP vs. B-J and H-W (transaction costs included) forecasting models of the Russell daily index returns. A plot of real data has been added on for comparison purposes.

My results are consistent with the findings of Chatfield (1996)and Bossaerts and Hillion (1999), who argue that forecasting models with high in-sample explanatory power usually do not have high out-of-sample fit (external validity) due to model over-fitting. I performed out-of-sample tests to guard against data over-fitting and the development of a more powerful framework to evaluate the performance of the three competing forecasting techniques. The above analysis suggests that forecasting comparisons of different methods and models should predominantly be made on the basis of genuine out-of-sample predictions. My results are consistent with the findings of Armstrong (1995) and Chatfield (1996), who argue that the real test of a forecasting model or method is its *ex-ante* forecasting ability.

MAE, RMSE and MAPE are significant forecasting error measures. However, they may not lead to profitable trading. Satisfactory low forecasting errors of MAE, RMSE, and MAPE do not necessarily guarantee that the model is generating a profit. The RMSE is a good example to illustrate this claim because it is just a quantitative expression of the current model and does not measure the actual profitability of the forecasting model. Logically, I investigate and quantify the actual profitability of the most realistic for trading purposes model- the STGP with included transaction costs for the Russell 1000, the Russell 2000 and the Russell 3000.

I adopted two profitability measurement criteria-the number of correct hits (forecasts) and the generated excess return from trading the three indices. The hit ratio detects the percentage of time that the model has good sign predictability:

$$\text{Hit ratio (\%)} = \frac{\text{Number of correct forecasts}}{\text{Number of generated buy / sell orders}} \times 100 \qquad (55)$$

The other profitability criterion- the excess return-represents the amount received from trading in excess of the risk free rate. It is the continuously compounded return on the Russell 1000, the Russell 2000 and the Russell 3000 price minus the value of the daily continuously compounded rate converted from the annualised investment yield on a 3-month US Treasury bill:

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) - r_{t-1} \tag{56}$$

Where P_t is the price of the Russell 1000, the Russell 2000, and the Russell 3000 traded at period t, and r_t is the risk free rate set at the value of the daily continuously compounded rate converted from the annualised investment yield on a 3-month US Treasury bill (data up to 16/11/2012 has been downloaded from the Federal Reserve statistical release website at www.federalreserve.gov/releases/h15).

The number of correct out-of-sample forecasts for the Russell 1000, the Russell 2000, and the Russell 3000 is reported in Table 63.0. The corresponding hit ratios are also given. A hit ratio above 50% is a sign of actual profitability from trading.

Financial instrument	***Russell 1000***	***Russell 2000***	***Russell 3000***
Number of generated buy/sell orders	698	741	766
Number of correct forecasts (hits)	378	433	427
Successful hits ratio	54.1%*	58.4%*	55.7%*
Excess return	3.83%	10.02%	5.11%

[a] The table reports the number of times a STGP out-of-sample forecasting model correctly predicts the direction of the Russell 1000, 2000 and 3000 returns and profitability of 1,534 observations (21/05/1979-16/11/2012) for each financial instrument. A ratio market with asterisk (*) indicates a 95% significance based on a one-sided test of H_0:p=0.50 against H_a:p>0.50. . [b]The risk-free rate is set at the value of daily continuously compounded rate converted from the annualized investment yield on a 3-month US Treasury bill (up to 16/11/2012).

Table 63.0 Out-of-sample comparison of the predictive strength and profitability of STGP with included transaction costs (0.25% for the large cap index and 0.60% for the small cap index) for Russell 1000, Russell 2000 and Russell 3000 [a].

The small-cap stocks reports the highest hit ratio of 58.4% (433 successful hits out of 741), followed by Russell 3000 with 55.7% (427 successful hits out of 766) and the large-cap stocks with 54.1% (378 successful hits out of 698) with all transaction costs included. Table 25.0 reports the hit ratio of the three indices over the entire sample of 8,840 observations. The Russell 2000 has the highest hit ratio of 57.2%, followed by the Russell 3000 with 54.9% successful hits and the Russell 1000 with hit ratio of 53.8%.

Moreover, I conducted a one-sided test to investigate whether the hit ratios of the three indices are significantly different from the benchmark of 0.5 (a 95% significance level). Under the null hypothesis the test has no predictive effectiveness power $(H_0 : p = 0.50$ against $H_0 : p > 0.50)$. The statistical tests rejected the null indicating that the hit ratios of the three indices are significantly different from 0.50. This important finding confirms the ability of my forecasting models in the prediction of the returns of the Russell 1000, the Russell 2000 and the Russell 3000.

Tables 63.0 and 64.0 clearly illustrates that the excess returns gained from trading small-cap stocks are higher than the excess returns of large-cap stocks for out-of-sample and for the whole sample. The Russell 2000 reported out-of-sample excess returns of 10.02% comparing to 3.83% for the Russell 1000 and 5.11% for the Russell 3000. The Russell 2000 reported 9.97% profitability for the entire sample, followed by the Russell 3000 with 5.01% profitability and the Russell 1000 with 3.78% excess return. The difference in hit ratio and excess returns between out-of-sample and the entire sample is due to the very initial chaotic behaviour of the STGP models leading to slightly lower in-sample values.

Financial instrument	*Russell 1000*	*Russell 2000*	*Russell 3000*
Number of generated buy/sell orders	4,596	4,876	4,858
Number of correct forecasts (hits)	2,473	2,789	2,667
Successful hits ratio	53.8%*	57.2%*	54.9%*
Excess return	3.78%[b]	9.97%[b]	5.01%[b]

[a] The table reports the number of times a STGP forecasting model correctly predicts the direction of the Russell 1000, 2000 and 3000 returns and profitability over the entire sample of 8,840 observations (21/05/1979-16/11/2012) for each financial instrument. A ratio market with asterisk (*) indicates a 95% significance based on a one-sided test of H_0:p=0.50 against H_a:p>0.50. . [b]The risk-free rate is set at the value of daily continuously compounded rate converted from the annualized investment yield on a 3-month US Treasury bill (up to 16/11/2012).

Table 64.0 Comparison of the predictive strength and profitability over the entire sample of STGP with included transaction costs (0.25% for the large cap index and 0.60% for the small cap index) for Russell 1000, Russell 2000 and Russell 3000 [a].

Under current stock market conditions, initial transaction costs to invest in small-cap indices exceed transaction costs to invest in large-cap indices worldwide. Keleher (2010) suggested one way transaction costs for the US large-cap indices of 0.17% (or 17 basis points) and 0.52% (or 52 basis points) for the US small-cap indices. I assigned higher transaction costs of 0.25% or 25 basis points for the Russell 1000 and the Russell 3000 and 0.60% or 60 basis points for the Russell 2000. There are two reasons for the selection of these two levels of transaction costs. First, while the levels of 0.25% and 0.60% may be relatively high by current standards, it appears reasonable for the earlier period of the experiment.

Secondly, I applied higher transaction costs in order to guard against data over-fitting, which seems intuitively reasonable.

Studies by Banz (1981), Reinganum (1981), Lakonishok and Sapiro (1986), Lo and MacKinlay (1990), Fama and French (1992), Hackel *et al.*(1994), Avramov (2001), Kaboudan (2001), How *et al.* (2010), Switzer (2010) and Shynkevich (2011) demonstrate small-cap stocks outperformance.

I expanded those studies by the inclusion of transaction costs and the achievement of reliable profits. I found that trading rules generated by STGP have substantially higher forecasting ability for the small-cap Russell 2000 index. This important finding highlights that small-cap stocks are less informationally efficient, which is a necessary condition for excess returns. This is consistent with the findings of Blume *et al.* (1994) and How *et al.* (2010) who argue that small-cap stocks are priced in a less efficient manner than large-cap stocks. As a result small-cap pricing errors can be more readily exploited. Small-cap stocks pricing inefficiencies could be explained by the fact that such assets are less widely purchased by investors and do not receive the same level of attention. On the other hand, the lower level of attention on small-cap stocks indicate that they are relatively more vulnerable to information asymmetry, experiencing gradual price adjustments because the news is absorbed slowly (How *et al.*, 2010).

The errors-in-expectations or extrapolation hypothesis was developed by Lakonishok *et al.* (1994) and LaPorta *et al.* (1997). The errors-in-expectations hypothesis states that investors initially form overly optimistic predictions about the future earnings of growth stocks leading to substantial price declines when these expectations are not met (greater optimistic forecasting bias for growth than for value stocks). Investors make systematic errors in forecasting the future profitability of value stocks and their pessimism about the future performance of value stocks is the actual cause for the value stocks outperformance. Hence, the small-cap outperformance could be based on the assumption that the market slowly realises that earnings growth rates for small-cap stocks are higher than initially expected.

My laboratory experiments can be regarded as direct tests of the extrapolation hypothesis because there is a clear difference in terms of excess returns between the initial forecasts made in-sample and the excess returns recorded out-of-sample. I compared the initial forecasts (in-sample) to the projected predictions (out-of-sample) to find out whether the stock market's initial optimistic forecasting bias is more pronounced for large-cap stocks than small-cap ones. The magnitude of the forecast errors (the difference of the forecast errors) of the Russell 1000 and the Russell 2000 has been estimated in order to determine whether the forecasting expectations for large-cap stocks are significantly more optimistic than small-cap stocks.

If the forecasts of the large-cap stocks are more optimistic than those of small-cap stocks, one would expect the actual difference to be negative (Mian and Teo, 2003).

The magnitude of the forecast errors generated by STGP with transaction costs for the Russell 1000 and the Russell 2000 (the most realistic models) reveal interesting results. The magnitudes of the forecast errors are all negative, indicating more optimistic forecasts for the large-cap stocks. The magnitudes of the forecast errors in-sample are as follows: MAE of -1.965, RMSE of -3.88 and MAPE of -0.428 (these are the differences in forecasting errors between the Russell 1000 and the Russell 2000 listed in Tables 54.0 and 55.0). Out-of-sample MAE is -10.861, RMSE is -11.113 and MAPE is -0.423. The extreme magnitude of the forecast errors demonstrates the same trend- MaxAPE of -10.042 and MaxAE of -29.493 for in-sample and MaxAPE of -5.267 and MaxAE of -40.921 for out-of-sample.

The comparison of the initial forecasts to the projected predictions shows that the stock market's initial optimistic forecast bias is more pronounced for large-cap stocks.

Consistent with the predictions of the errors-in-expectation hypothesis tested within artificial laboratory settings, I found that traders made larger forecast errors in predicting the excess returns for large-cap stocks than small-cap stocks. This finding implies that traders are excessively optimistic about large-cap stocks, rather than small-cap stocks. Although tested within laboratory conditions, the errors-in-expectations hypothesis has some practical implications and proves a possible explanation of the small-cap stocks outperformance.

Individual investors could mistakenly consider the purchase of assets of well-managed companies to be sound investments, even if such assets are bought at a high price. Moreover, the individual investors might favour large-cap stocks because it is easier to justify such kind of investments to their clients.

8.6.2. The dynamic relationship between trading volume and index returns. An investigation into whether the level of in-sample trading volume is a good predictor for the out-of-sample stock returns.

An objective of this study is to investigate whether the finding of a relationship between two variables such as the Russell 1000, 2000 and 3000 returns and their volume is a result of dynamic (causal) factors. I follow the standard econometric procedure and apply the ADF test to examine the stationarities of the in-sample and out-of-sample volume series, V_t of the three indices:

$$v_t = V_t - V_{t-1} \tag{57}$$

I did not accommodate the log difference transformation for volume because trading volume could be zero in some trading periods. The ADF test results listed in Table 65.0 illustrate that all trading volume series are stationary.

		Russell 1000		
Period	***Bid***	***Ask***	***Total***	***ADF****
In-sample	2,836	1,062	3,898	-88.02
Out-of-sample	508	190	698	-45.11
Grand total			***4,596***	
		Russell 2000		
In-sample	3,883	252	4,135	-91.57
Out-of-sample	696	45	741	-51.01
Grand total			***4,876***	
		Russell 3000		
In-sample	3,869	223	4,092	-90.79
Out-of-sample	694	40	766	-50.38
Grand total			***4,858***	

*The MacKinnon (1996) one-sided critical value for rejection of the Null hypothesis of a unit root at 5% level is -3.410060.

Table 65.0 In-sample and out-of-sample trading volume recorded on the artificial stock market and ADF test for Russell 1000, Russell 2000, and Russell 3000.

I then examined the dynamic relation between in-sample V_t and out-of-sample R_t to determinate whether trading volume has any significant predictive power for future returns. To test whether there is any bi-directional or uni-directional causality from one variable to the other, I applied the linear Granger causality test and the modified Baek and Brock (1992) nonlinear causality test.

The following bivariate autoregressions are used to test for linear causality between the two variables:

$$V_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i V_{t-1} + \sum_{j=1}^{n} \beta_j R_{t-j} + \varepsilon_t \qquad (58)$$

$$R_t = \gamma_0 + \sum_{i=1}^{m} \gamma_i R_{t-1} + \sum_{j=1}^{n} \delta_j V_{t-j} + \eta_t \qquad (59)$$

If the β_j coefficients in Equation 58 are statistically significant, the inclusion of past values of return and past history of volume generate a better forecast of future volume. Hence, I conclude that returns cause volume. If a standard *F*-test does not reject the hypothesis that $\beta_j = 0$ for all j ,then returns do not cause volume. If causality runs from volume- to -returns in Equation 59, then the δ_j coefficients will jointly be different from zero. In cases when both β and δ are different from zero, there is a feedback (bi-directional) relation between returns and trading volume (Chen *et al.,* 2001).

The disturbances $\{\varepsilon_t\}$ and $\{\eta_t\}$ are two uncorrelated series following the conventional assumptions of white noises: they are identically and independently distributed (IID) with zero mean and common variance of σ^2 such that:

$$E(\varepsilon_t \varepsilon_s) = E(\eta_t \eta_s) = 0 \quad \forall s \neq t \qquad (60)$$

And

$$E(\varepsilon_t \eta_s) = 0 \quad \forall s, t \qquad (61)$$

Bearing in mind the predictability of indices returns, I am primarily interested in the dynamic relation from in-sample trading volume to out-of-sample returns. Hsiao (1981) argued that the results of the linear Granger causality tests, are sensitive to the choice of lag lengths.

At this stage I need to choose appropriate lag lengths of V_t and R_t, that is, the values of m and n in Equations 58 and 59 through several statistical search criteria such as AIC and the Schwarz criterion (Chen and Liao, 2005). Table 66.0 presents the empirical results for linear and nonlinear causality for the whole sample of 8,840 observations.

Linear Granger causality test	*H_0: Volume changes do not cause index returns*			*H_0: Index returns do not cause volume changes*		
	# of lags	*F-value*	*p-value*	*# of lags*	*F-value*	*p-value*
Russell 1000	12	1.346	0.1192	18	1.044	0.471
Russell 2000	10	1.170	0.1673	17	1.388	0.225
Russell 3000	14	1.899	0.0937	17	1.389	0.226
Nonlinear Granger causality test	***H_0: Volume changes do not cause index returns***			***H_0: Index returns do not cause volume changes***		
	# of lags	***F-value***	***p-value***	***# of lags***	***F-value***	***p-value***
	1	1.901	0.8181	1	1.768	0.0899
	2	1.989	0.0771	2	1.488	0.0801
	3	1.380	0.0699	3	2.900	0.9993
	4	1.237	0.6771	4	1.546	0.3487
	5	2.839	0.7878	5	2.988	0.1001
Russell 1000	6	2.127	0.1101	6	2.011	0.0378
	7	1.892	0.9899	7	2.189	0.0378
	8	1.001	0.0000*	8	1.118	0.0085*
	9	1.120	0.3022	9	1.982	0.3287
	10	1.133	0.0001*	10	2.087	0.3898
	1	2.001	0.9918	1	1.110	0.0012*
	2	1.632	0.1128	2	1.190	0.1126
	3	1.781	0.2788	3	2.091	0.3899
	4	1.235	0.0066*	4	1.347	0.0947
	5	1.110	0.0788	5	1.547	0.3311
Russell 2000	6	1.218	0.0989	6	1.682	0.0022*
	7	1.833	0.9829	7	2.438	0.0963
	8	1.327	0.8129	8	1.349	0.9991
	9	1.237	0.1171	9	1.100	0.1891
	10	1.017	0.1278	10	1.322	0.1178
	1	1.276	0.0011*	1	1.433	0.2221
	2	1.437	0.9237	2	2.383	0.0810
	3	2.878	0.2366	3	1.376	0.0033*
	4	1.375	0.0178*	4	2.001	0.0395
	5	1.115	0.0001*	5	1.378	0.1982
Russell 3000	6	1.985	0.1200	6	1.010	0.5662
	7	1.450	0.9909	7	1.540	0.2878
	8	1.189	0.1117	8	1.578	0.2378
	9	2.225	0.2377	9	1.457	0.8989
	10	2.548	0.9873	10	1.578	0.2178

*Rejection of the null at the 5% significant level.

Table 66.0 Linear and nonlinear Granger causality tests for the whole sample of 8,840 observations.

These empirical tests indicate no presence of linear causality in the three indices. In all experiments, the causal relation is not found to exist in either direction. However, relying only on linear Granger causality test results could lead to inappropriate conclusions, because these tests might overlook significant nonlinear relations.

Baek and Brock (1992) developed a nonparametric test for determining nonlinear causal relations. The authors suggested filtering out linear predictive power in Equations 58 and 59 and then the remaining predictive power between the two residual series of $\{\varepsilon_t\}$ and $\{\eta_t\}$ could be considered to be of a nonlinear nature.

I followed Hiemstra and Jones (1994), who allowed for the residuals to be weakly dependent (short-term temporal dependence) in order to specify the appropriate lagged lengths of m and n in Equations 58 and 59. In terms of the nonlinear Granger causality test, the existence of the causal relation is not definite. The bi-directional nonlinear causality has been found in only one case- the Russell 1000 with eight lags (Table 66.0). While the uni-directional nonlinear causality from volume-to-returns is present in five cases, the returns-to-volume causality has been detected in three other cases. Overall, the returns-to-volume causal relation seems slightly stronger than the volume-to-returns causality. As the trading volume has no predictive power on stock returns under linear conditions and very weak predictive power in nonlinear Granger causality tests, I cannot develop volume-based trading strategies. Furthermore it was interesting to analyse the implications of in-sample trading volume on out-of-sample index returns. The empirical results reported in Table 67.0 clearly indicate that in-sample trading volume does not Granger cause out-of-sample index returns.

Hypothesis	***F-value***	***p-value***
In-sample Russell 1000 trading volume $\xrightarrow{G.C}$ ***Out-of-sample Russell 1000 returns***	1.661	0.5710
In-sample Russell 2000 trading volume $\xrightarrow{G.C}$ ***Out-of-sample Russell 2000 returns***	1.920	0.1018
In-sample Russell 3000 trading volume $\xrightarrow{G.C}$ ***Out-of-sample Russell 3000 returns***	1.883	0.9331

Table 67.0 Granger causality test for 7,306 in-sample and 1,534 out-of-sample observations.

8.6.3. An investigation into whether a market with a reduced population of 1,000 traders is capable of generating accurate and profitable forecasts.

A market populated by 1,000 traders revealed interesting forecasting results. In line with the previous results both Box-Jenkins and Holt-Winters outperformed in-sample STGP for the Russell 1000, the Russell 2000, and the Russell 3000. Out-of-sample STGP for the Russell 1000 outperformed Box-Jenkins with the following forecast errors: MAE of -2.126, RMSE of -1.800 and MAPE of -0.076 (Table 57.0). Logically, the out-of-sample STGP technique performed better than Holt-Winters for the same index: MAE of -2.090, RMSE of -1.886 and MAPE of -0.074 (Table 60.0). Out-of-sample STGP for the Russell 3000 slightly outperformed the two traditional econometric forecasting techniques: MAE of -0.748, RMSE of -0.822 and MAPE of -0.029 for the Box-Jenkins comparison (Table 59.0) and MAE of -0.710, RMSE of -0.912, and MAPE of -0.026 for the Holt-Winters comparison (Table 62.0).

When comparing the empirical results with the other two markets populated by 10,000 traders I found relatively accurate forecasts generated by the market with a reduced number of traders. However, the smaller forecasting errors of the two markets consisting of 10,000 agents (Tables 54.0, 55.0 and 56.0) suggest that greater genetic diversity of the entire population is needed in order to achieve more accurate predictions. Hence, the price formation process caused by the collective behaviour of the market populated by 10,000 traders is a better predictor than any small fraction of traders.

8.7. Conclusions

Researchers have given considerable attention to the ability of firm size to explain stock returns. The evidence of small-cap stocks premiums has changed the investment behaviour of traders and investment institutions. This research contributes to the existing literature by investigating whether small-cap stocks provide evidence of profitability and whether trading volume has any predictive power on stock returns. I also performed unique test of the errors-in-expectations hypothesis within artificial laboratory stock markets settings.

To a large degree, the emphasis of this paper has been on finding predictability that is strong enough to produce reliable trading profits after taking into account appropriate transaction costs. I found evidence of statistically and economically significant profitability in all experiments of the Russell 1000, the Russell 2000 and the Russell 3000 indices. One-step-ahead STGP models demonstrate the superior performance of small-cap stocks compared to large-cap stocks. Although large-cap firms provide certain benefits due to their economies of scale, experimental results gained within laboratory artificial stock settings confirm the claim that small-cap stocks are more predictable and profitable than large-cap stocks.

Some researchers express the view that the dominance of small-cap stocks is period-specific. I argue that small-cap stocks' dominance might not be period-specific, because I achieved significant profits based on tests of the entire 8,840 observations (since the creation of the three indices- 8,840 observations up to 16/11/2012). My findings are consistent with the errors-in-expectations hypothesis which posits that excess returns of growth stocks are driven by more optimistic forecasts compared to those of value stocks. Switzer and Tang (2009) argue that small-cap stocks provide a vehicle for a significant entrepreneurship and innovation in the US, and therefore might be less prone to governance issues than large-cap stocks. According to the authors large board sizes have negative implications on operational performance and pay-for performance compensation for the CEO's, which might be viewed as beneficial for small-cap stocks.

An investigation into the price formation process revealed that the collective behaviour of the market populated by 10,000 traders is a better predictor than any small fraction of traders.

Tables 54.0, 55.0 and 56.0 illustrate that the forecasting errors of the three financial instruments generated by 10,000 traders are smaller than the forecasting errors generated by 1,000 traders indicating the need of greater genetic diversity to achieve more accurate forecasts. Hence, enhanced levels of genetic diversity leads to more heterogeneous trading rules and greater flexibility in the virtual market clearing price mechanism.

I found mixed results associated with the dynamic relation between trading volume and index returns. Table 66.0 illustrate only one case of bi-directional nonlinear causality- the Russell 1000 with eight lags. On the other hand, uni-directional causality has been captured in five cases. Moreover, the returns-to-volume casual relation is stronger than the volume-to-return causality. Overall, the trading volume does not have predictive power on stock returns under linear conditions and very weak forecasting power in nonlinear causality tests. Hence, I am unable to develop any volume-based forecasting strategies. My findings are consistent with the existing literature- while some did not find the existence of linear Granger causality, others reported a uni-directional relation only.

Although small-cap stocks tend to dominate large-cap stocks in investment terms, the importance of the latter should not be disregarded. Investing in large-cap stocks provides many benefits based on their economies of scale. In a global context, large-cap firms have an edge because they easily transcend national boundaries to extend their production abilities in order to generate economic profits. However, it is difficult for any firm- small-cap or large-cap- to maintain its superiority for long periods. Bearing in mind the fast pace of technological innovations and strong market competition, I conclude that it is unlikely to get any easier for a firm to continue to dominate in the future. Firms that are in the forefront of technological innovations might continue to succeed in the future.

Chapter 9

Does high frequency trading affect technical analysis and market efficiency? And if so, how?

9.1. Introduction

The extensive use of technical trading rules by currency market practitioners has long been a puzzle for academics. On the one hand, as Cheung and Chinn, (2001) and Gehrig and Menkhoff (2003) note up to 40 per cent of foreign exchange (FX) traders worldwide rely on technical analysis as their main trading tool. On the other hand, one of the best established paradigms in financial economics- the Efficient Market Hypothesis (Fama, 1970)- suggests that on a market with a vast trading volume and virtually non-existant private information about fundamentals, such as the foreign exchange (FX) market (turnover of 2,000 billion US dollars per day; BIS, 2005), trading rules based on historical price information should not yield excess gross profits to traders.

Most academic studies related to technical trading in the FX market are not consistent with the real-life practice of technical analysis because they largely limited their trading strategies to daily data observations (Brabazon and O'Neil, 2004; Qi and Wu, 2002; Reitz and Taylor, 2006). However, nearly all FX traders who use technical analysis operate at a high frequency (Gomber *et al.,* 2011; Ahlstedt and Villysson, 2012; Guo, 2012). In addition, more than 75 per cent of FX trading has been shown to take place within a single day (BIS, 1996), and that the applicability of technical analysis increases with the frequency of trading (Taylor and Allen, 1992).

While some empirical studies of daily FX data report the existence of significant profits (Martin, 2001; Mathur *et al.,* 2001; Saacke, 2002), some other studies demonstrated the contrary (Levich and Thomas, 1993; Lee and Mathur, 1996b; Lee *et al.,*2001). Studies on the profitability of intra-daily technical analysis also do not convey a clear picture. Some authors report significant net profits (Gencay *et al.,* 2002; Gencay *et al.*, 2003), whereas others find technical trading to be unprofitable even at these high frequencies (Curcio *et al.,*1997; Osler, 2000; Neely, 2003).

Even more intriguingly, Curcio *et al.* (1997) simulated some of the technical analysis rules shown in previous studies to be profitable on daily frequency and demonstrated that they would not be profitable when applied to intra-daily data. This runs contrary to the intuition that the currency markets need time to process information and could be inefficient at very high frequencies but efficient at lower frequency, e.g., daily, horizons.

Moreover, studies on FX technical trading profitability typically fail to account for transaction costs, trading rule optimisation over time, out-of-sample verification, and data snooping issues (Park and Irwin, 2007).

So far, I have focused on the relationship between high frequency trading (henceforth HFT) and technical analysis. However, the question could be reversed and the impact of HFT on the market's quality can also be investigated. The vast majority of empirical evidence to date suggests that HFT improves market liquidity, reduces trading costs in the form of narrower bid-ask spreads, and makes stock prices more efficient (Jones, 2013). On theoretical grounds, however, HFT can also be negative because the speed of trading could put other market participants at a disadvantage, leading to adverse selection and reduced market quality. There is also a possibility for an unproductive arms race developing among HFT institutions competing to be fastest (Jones, 2013).
Empirical research is limited due to the massive and complex nature of the raw datasets which are spread over multiple exchanges and trading platforms with academic studies analysing mainly the 120-stock NASDAQ HFT dataset. The apparent lack of conclusive evidence has enabled HFT to operate with a limited regulatory understanding. Policymakers around the world are still debating whether to introduce limits on HFT or even to completely ban it. The major empirical challenge is to measure the incremental effect of HFT beyond other changes in FX markets. The best published studies for this purpose disregard market structure changes that facilitate HFT (Jones, 2013). The ability to observe all trading in my experiments allowed us to investigate the impact of HFT bid and ask orders on market quality. In this study, I argue that the search for answers to the puzzles described above should be conducted on high frequency rather than daily data, as the trading and resulting price adjustments take place on intraday basis.

This paper also argues that, should the FX efficiency puzzle be reliably addressed, a correct, robust methodology should be employed for this task. For this purpose, I used one-minute currency market data for six currency pairs for time periods shorter than six months.[1][2]

In this study, I also implemented a special adaptive form of Genetic Programming (GP), called Strongly Typed Genetic Programming (STGP), in order to avoid the serious technical analysis issues highlighted by Park and Irwin (2007).

The scientific advantage of the STGP over the conventional Genetic Programming (GP) used in most studies so far is that STGP evaluates the fitness of agents through a dynamic fitness function which processes the most recent quotes of the six currency pairs in my experiment, rather than a re-execution of the same trading rules.

Despite the voluminous literature on the topic, no other study has implemented the STGP technique, one-minute high frequency data, and substantially more artificial agents, which has enabled us to develop of a wider variety of trading rules. The presence of 10,000 artificial agents in my experiment resulted in an increased forecasting model stability and significantly low sensitivity to random factors.

To summarise, the contributions of this study are as follows. Firstly, I have investigated the efficiency of currency markets by analysing the profitability of technical trading rules at the frequency at which this trading actually takes place in the real world. I am not aware of any other study utilising minute-by-minute price data in this context. Secondly, I have taken into account all issues identified in the literature as potentially affecting the reliability of trading results and inference based on them: I controlled for transaction costs, allowed agents to learn from their experience and to switch to more profitable rules, evaluated profitability of rules base on their predictive power rather than in-sample fit, and avoided data-snooping biases by allowing all potential rules and their combinations to be traded on and evaluated by agents. Thirdly, I am the first to apply the STGP technique to analyse the impact of HFT on market quality, taking into account the market structure.

[1] Goodhart and O'Hara (1997) suggest that the availability of high-frequency data enables empirical investigations of many financial market issues. Engle (2000) claim that ultra-high-frequency data contains a complete record of transactions and their related characteristics.

[2] The use of a short time period adds realism to the analysis of technical analysis: as Curcio *et al.* (1997) argue, it is yet to be shown that technical trading profits would remain if the experiments were performed in time periods shorter than a couple of years because investors cannot afford to experience losses during several months even in the long run their strategies are profitable.

9.2. Background

9.2.1. Technical analysis on the FX markets

9.2.1.1. Studies using daily and weekly data

The moving average (MA) trading rules are among the most popular in technical analysis. Martin (2001) achieved statistically significant out-of-sample returns in the spot FX market of Brazil by implementing MA trading rules for daily data. Mathur *et al.,* (2001) applied the MA trading rules to 13 Latin currencies and reported excess returns for four currencies based on daily data. Saacke (2002) reported substantial profits of MA rules in daily USD/DEM exchange rates from January 1979 to July 1995.

The profitability of FX trading based on these traditional techniques has, however, been challenged in the literature. Levich and Thomas (1993) applied MA and filter rules to a number of foreign currencies over the period 1976-1990 and found no presence of profitable trading rules. Lee and Mathur (1996b) applied the MA trading rule to six European currency pairs and reported uncorrelated cross rates which were sufficiently transparent to eliminate MA trading rule profits. Lee *et al.,* (2001) applied MA and channel trading rules on nine Asian FX rates for the period 1988-1995 and reported little evidence of serial correlation in the daily rates combined with lack of significant profits.

Chart patterns are another commonly used tool by technical analysts. Chang and Osler (1999) examine the performance of head-and-shoulders chart patterns using daily spot rates for six FX currencies during 1973 and 1994, and reported statistically significant returns of 13 per cent and 19 per cent, respectively, for the mark and yen only.

However, chart pattern analysts do not conduct parameter optimisation and out-of-sample tests, and do not examine data snooping issues.

These problems were addressed by introducing a model-based bootstrapping technique. Conventional *t*-tests were unable to deal with leptokurtic, auto-correlated, time varying, and conditionally heteroskedastic financial returns. The study by Brock *et al.* (1992) inspired other researchers to also use a model-based bootstrap approach, to avoid the weaknesses of conventional *t*-tests in technical analysis.

Maillet and Michel (2000) applied bootstrapping procedures to six daily FX rates from January 1974 to September 1996 and reported excess returns.

To account for a possible data snooping bias, Qi and Wu (2002) used White's (2000) bootstrap reality check methodology in seven FX rates between 1973-1998 and generated substantial profits (7.2-12.2 per cent) in five FX rates after transaction costs. However, the bootstrapping approach poses difficulties in constructing the full universe of technical trading rules required by the methodology (Park and Irwin, 2007).

The rapid development of non-linear forecasting practices partially addressed the shortcomings of bootstrapping. Sosvilla-Rivero *et al.* (2002) applied technical trading rules based on a nearest neighbour regression approach to the mark and yen during 1982-1996 and generated substantial net returns. Fernandez-Rodriguez *et al.* (2003) also used the nearest neighbour predictive model to achieve annual net returns of 1.5-20.1 per cent for the French franc, Danish krona, Italian lira and Dutch guilder during 1978-1994.

However, Timmermann and Granger (2004) pointed out that it may be inappropriate to use the non-linear methods developed in recent years to capture profitability of technical trading rules during the 1970s or 1980s, highlighting major disadvantage of the nearest neighbour forecasting method.

The application of the Genetic Programming (GP) technique to technical analysis was aimed at addressing the methodological issues affecting the more traditional methods, as mentioned above. Most importantly, while traditional technical analysis investigates a pre-determined parameter space of technical approaches, the GP forecasting approach fully examines a search space which consists of logical combinations of trading rules (Koza, 1992).

Neely *et al.* (1997) used a GP method to analyse profits from technical trading in the daily US dollar bid and ask quotations for the Deutsche mark, yen, pound sterling, and Swiss franc. The authors revealed evidence of economically significant out-of-sample excess returns for each of the six exchange rates during the period 1981-1995.

Thomas and Sycara (1999) implemented a GP model to generate considerable excess returns on the daily USD/JPY rate data. Diaz and Alvarez (2007) combined neural networks and GP methodologies in the weekly USD/JPY and GBP/USD exchange rate returns to find short-term weak predictable structures.

9.2.1.2 Studies using intra-daily data

Since the 'flash crash' of May 6, 2010,- when S&P futures fell almost 10 per cent in 15 minutes before rebounding, high-frequency trading has received massive public attention (Kirilenko *et al.,* 2011). Despite the fact that over the last few years this type of trading has progressively become more important in FX markets, there appears to be no large scale empirical research related to the potential profitability and market impact of HFT (Kearns *et al.,* 2010).

Gencay *et al.,* (2002) applied a real-time-trading (RTT) model to seven years of five minute high frequency data of four currency pairs and showed 3.6- 9.6 per cent annual excess returns. Gencay *et al.,* (2003) implemented a RTT model and exponential moving average (EMA) indicator to high frequency around-the-clock data of four currency pairs and found profitable trading patterns. Curcio et al. (1997) analysed the profitability of technical trading rules based on predefined price ranges and applied to hourly rates for DEM, JPY, and GBP against the USD, and concluded that on average the profits were not significant, even before transaction costs were taken into account. Oussaidene *et al.* (1997) investigated the implications of GP on one-hour high frequency data of seven currency pairs and found the presence of profitable trading rules, but this was without taking into account any transaction costs. Dempster and Jones (2002) developed pattern recognition channels, which are used by technical analysts as trade entry signals for the USD/GBP one-minute high frequency data from 1989 to 1997, and reported rather inconsistent profitability of trading. Osler (2000) investigated the predictive power of support and resistance levels for high frequency one-minute data of three currency pairs and failed to report profitability after taking into account transaction costs. Overall, there does not seem to be a consensus regarding the profitability of intra-day technical trading on currency markets.

The performance of GP-based technical trading in currencies using intraday data could be expected to be superior to other methods, based on the results in Zumbach *et al.* (2001). These authors used 13 years of hourly data of USD/CHF and USD/JPY to examine out-of-sample forecasting performance of STGP with syntactic restrictions and ARCH-types models, and demonstrated the superiority of STGP's predictive ability. Dempster *et al.* (2001) reported significant predictability of genetic algorithms applied to fifteen-minute high frequency GBP/USD data.

Dempster and Jones (2002) used a genetic algorithm to find significant profits in the presence of realistic transaction costs in the USD/GBP tick data, from 1994 to 1997, aggregated to different intraday frequencies (one-minute being the highest frequency used). When traded in an adaptive manner, however, the authors found that the resulting portfolio strategies are ultimately loss making, highlighting the penalty for over-reaction to short-term market behaviour. Bhattacharyya *et al.* (2002) applied GP to one-hour high frequency data for the USD/DEM currency pair and while the authors reported reliable predictive performance of the model, they failed to show profitability. Neely and Weller (2003) examined the performance of GP on half-hourly high frequency data of four different currency pairs and found no profitability.

9.2.2. HFT and its impact on market efficiency

Some academics argue that HFT create risks and, increases volatility and financial fragility, because when a market dislocation arises HFT reacts ahead of other investors due to their advantage of fast access to the market. For instance, Zhang (2010) showed that HFT can have harmful effects for the US capital market such as it is positively correlated with stock price volatility among the top 3,000 stocks in market capitalisation and negatively associated with the stock market's ability to incorporate news about company fundamentals into stock prices. Kirilenko *et al.* (2011) analysed the trading of S&P 500 e-mini futures between August 2010 and August 2012 and concluded that high frequency traders make profits at the expense of retail investors, e.g., small retail investors lost $3.49 for every single contract they traded with high frequency traders.

Gai *et al.* (2012) examined the implications of two recent 2010 NASDAQ technology updates that reduced the minimum time between messages from 950 nanoseconds to 200 nanoseconds, and reported that these technological improvements substantially increased the number of cancelled orders. As a result the bid-ask spreads and trading volume experienced a slight change, suggesting that there may be diminishing liquidity benefits to faster exchanges. Jones (2013) highlighted that HFT can intermediate trades at lower cost, but the speed possessed by HFT could disadvantage other investors and the resulting adverse selection could harm market efficiency.

Bias *et al.* (2013) considered a unit mass continuum model of risk-neutral, profit maximising financial institutions trading in the market for one asset, and reported that fast trading increased adverse selection costs for all market participants creating negative externalities. Foucault *et al.* (2013) demonstrated that an investor's optimal trading strategy is significantly different when he observes news faster than others, and reported that price changes are more correlated with news and trades contribute more to volatility when the investor has fast access to news. A report published by the Australian Securities & Investment Commission (ASIC) in 2013 revealed that some HFT algorithms contribute to excessive order messages, fleeting orders, and market 'noise'. This has a disruptive impact on the market, resulting in damaged investor confidence.

Moreover, there are some recent market glitches associated with HFT. On August 1, 2012, Knight Capital: one of the largest market- makers in the U.S. equities: implemented a new trading algorithm which accumulated large positions in 148 listed stocks over approximately 45 minutes, resulting in trading losses of $440 million. Only a few months ago, NASDAQ experienced substantial issues associated with a software program that was unable to handle the pace of order submissions and cancellations by HFT, causing tens of millions of dollars of losses (Jones, 2013).

Some other scholars, on the other hand, suggest that HFT makes financial markets more efficient by creating more liquidity, narrowing spreads, and removing the commissions that drive up costs. For instance, Chaboud *et al.* (2009) used one-minute data of the EUR/USD, USD/JPY, and EUR/JPY traded in 2006 and 2007, and found that HFT is associated with lower volatility and enhanced liquidity. Brogaard (2011) used HFT data of 120 US stocks that took place in NASDAQ and BATS exchange during 2008, 2009, and 2010 and found a statistical relationship between HFT and volatility in a Granger causality context. The author suggested that after controlling for time series variation, increased HFT caused intraday volatility to decrease. Domowitz (2010) went even further suggesting that events such as the 'flash crash' in 2010 are generic features of equity markets and similar events occurred in manual markets. The author highlighted that a similar 'flash crash' occurred on May 28, 1962, when the Dow Jones fell sharply for a period of 20 minutes. Brogaard *et al.* (2013) used the 120-stock NASDAQ HFT dataset to trade in the direction of permanent price changes and in the opposite direction of transitory pricing errors, and concluded that HFT plays a beneficial role for market efficiency.

Hendershott *et al.* (2011) examined the implementation of an automated quote system at the New York Stock exchange in 2003 and concluded that after implementing the autoquote effective spreads narrowed, adverse selection was reduced, and more price discovery took place. Riordan and Storkenmaier (2012) analysed the effect of a technological upgrade on the market quality of 98 of the most traded German stocks and confirmed the findings of Hendershott *et al.* (2011). Boehmer *et al.* (2012) performed an international investigation of electronic message traffic and market quality across 39 different stock exchanges around the world between 2001 and 2009, and found that co-location improved liquidity and the informational efficiency of prices.

Jovanovic and Mankveld (2012) traded Dutch and German index stocks to analyse the advent of a middleman through an event study around the introduction of a new HFT venue. The authors reported that HFT can update limit orders quickly based on new information that coincided with a 23 per cent drop in adverse selection. The UK Foresight report (2012), a study administrated by the UK government, found no connection between HFT and market volatility, nor did it find any evidence suggesting an increase in market abuse. Martinez and Rosu (2013) developed an equilibrium model in which HFT observed a continuous stream of signals called news, and suggested that HFT makes the markets more efficient by quickly incorporating the news. Carrion (2013) reported similar results demonstrating that days with intense HFT activity are connected to better informational efficiency in terms of returns predictability measured by weaker predictive power of lagged order flow and information about the market index. Hansbrouck and Saar (2013) used NASDAQ data to investigate the market quality in the millisecond low-latency environment and concluded that an increased low-latency activity is leading to lower spreads and reduced short-term volatility. In a similar vein, Hagsrömer and Nordèn (2013) examined individual HFTs in a sample from NASDAQ-OMX Sweden and reported that the market-maker activity of HFTs decreased volatility in the short run. Carrion (2013) highlighted the limitations of the last two studies pointing out that they describe the total volatility only. Lastly, the ASIC report (2013) suggests that HFT does not have a significant effect on price formation and liquidity, and apart from rare one-off cases, there is no systemic issue related to predatory trading practices. To summarise, the question about whether HFT benefits or harms market efficiency still remains unresolved, and this is to my best knowledge the first study to address this question using a Genetic Programming methodology.

9.3. Artificial stock market structure for this particular experiment.

I have developed six different currency markets each populated by 10,000 agents, to trade one-minute high frequency data of EUR/USD, USD/JPY, GBP/USD, USD/CHF, and USD/CAD. Table 68.0 illustrates the main parameters of the STGP-based artificial stock market model for this particular experiment.

Artificial stock market parameters	
Total population size (agents)	10,0000
Initial wealth(equal for all agents)	100,000
Transaction costs	1.5 basis points
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 quotes processed for each of the six pairs	206,413 (176,735 in-sample and 29,678 out-of-sample)
Seed generation from clock	Yes
Creation of unique genomes	Yes
Offspring will replace the worst performing agents of the initial selection	Yes

*When the absolute forecast price change is within the specified range, the forecast is considered significant. The lower limit of the range prevent the generation of trading signals when price changes are being forecasted that may be too small to at upon. A forecast is always declared insignificant when the forecasted price change is zero. The upper limit prevent the generation of trading signals which differ too much from the asset's price (i.e. when the forecasts go to extreme values or remain fixed for a long period of time).

**Maximum genome size measure the total number of nodes in an agent'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 an agent'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 an agent is represented by the number of quotes that have been processed since the agent was created. This measure also specifies the period over which agent performance will be compared. My minimum breeding age is set to 80, which means that the agent's performance over the last 80 quotes will be compared.

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

Table 68.0 Artificial Stock Market Parameter Settings

9.4. Data and parameters

In this paper I follow Neely and Weller (2003) who applied a Genetic Programming (GP) technique to half-hourly data of USD/DEM, USD/JPY, USD/GBP, and USD/CHF during 1996, in order to find profitable trading rules and investigates market efficiency. The authors failed to report profitability after taking into account transaction costs of up to 2.0 basis points, demonstrating market efficiency. For more realistic forecasting purposes I implemented an innovative evolutionary STGP technique with no over-fitting of the FX historical data, as discussed above. I included more currency pairs and a higher frequency premium data in my experiment. I obtained one-minute high frequency FX data of the most traded currency pairs- EUR/USD, USD/JPY, GBP/USD, AUD/USD, USD/CHF, and USD/CAD. The in-sample trading period in my experiment started on August 27, 2012 at 10pm (GMT) and finished on February 12, 2013 at 10.59am (GMT) and consisted of 176,735 real data quotes. The out-of-sample trading begun on February 12, 2013 at 11am (GMT) and finished on March 12, 2013 at 11am (GMT) and consisted of 29,678 real data quotes. The study period was chosen on the basis of maximum data availability downloaded from Bloomberg. Similar to the real-life FX market, my experiment allowed 24 hours of trading, except at weekends with trading taking place from 8.15pm (GMT) on Sunday until 10pm (GMT) on Friday. Throughout this study, I impose one-way transaction costs of 1.5 basis points.

9.5. Forecasting methods

9.5.1. Benchmark models

The clearing price generated within the STGP framework at time *t* based on agents' orders is used as a predicted price for the subsequent period, *t+1*. These STGP-based forecasts are compared to two traditional econometric forecasting models, one in the framework of a parametric model represented by AR-GARCH and one of a non-parametric model represented by the K nearest neighbours model.

Firstly, I fitted linear ARMA-GARCH model to one-minute high frequency data:

$$\left(1-\phi_1 B-\ldots-\phi_p B^p\right)\left(x_t-\theta_0\right)=\left(1-\theta_1 B\ldots-\theta_q B^q\right)\in_t \qquad (62)$$

where $V\langle\in_t|\Theta_{t-1}\rangle=\sigma_t^{\,2};\ \sigma_t^{\,2}=a_0+a_1\in_{t-1}^2+b\sigma_{t-1}^2$, Θ_t is the information available at time t, and $\left(\in_t/\sigma_t\right)$ is a Student's t_v random variable (Meade, 2002). I utilise the GARCH model to capture the well known phenomenon of the time dependent variance in financial time series. An optimal model is selected from initial range of values of $p=q=3$, based on the Schwarz Bayes Information criterion. Furthermore, I implemented the Student's t distribution to compute the residual kurtosis, as it is known to be non-normal and have heavy tails. In addition, one of the properties of high frequency one-minute data was the relatively high presence of no price changes in the currency pairs. To account for this empirical phenomenon, a modified density of the standardised residuals is used, by making it dependable on market activity (Meade, 2002):

$$f\left\langle\frac{\in_t}{\sigma_t}\middle|\delta_t\right\rangle=\left\{\begin{matrix} p_0 \\ \dfrac{g_v\left(\dfrac{\in_t}{\sigma_t}\right)}{1-p_0}\end{matrix}\right\}_{if\,\delta_t=0}^{if\,\delta_t=1} \qquad (63)$$

where $g_v(.)$ is the Student's t density function and δ_t denotes market inactivity:

$$\delta_t\left\{\begin{matrix}1\\0\end{matrix}\right\}^{if\,|x_t|+|x_{t-1}|=0,otherwise} \qquad (64)$$

If $\delta_t=1$, the forecast $\hat{x}_{t+i/t}=0$ for $i=1,2,\ldots..$ (Meade, 2002).

In this way, I aim to account for intervals of market inactivity and to detect the presence of kurtosis and heteroskedasticity during active market times. I estimated the coefficients in my ARMA-GARCH models by adopting the quasi-maximum likelihood.

The non-parametric K nearest neighbours model postulates that the lack of linearity is leading to repetitiveness of trends in the time series. When the previous trend is identical to the current trend of one-minute high frequency data, then the previous trend of the one-minute data can predict behaviour in the imminent and very near future (Meade, 2002). In this particular approach, one-minute data is viewed as a continuous series of d histories, $X_t^{\,d}$, which are in themselves continuous of d successive observations (with d being an integer):

$$X_t^{\,d} = \left(x_t, x_{t-1}, \dots x_{t-d+1} \right) \qquad (65)$$

The next step is to estimate a measure of distance between the nearest neighbour and the latest d historical one-minute data. I implemented the geometrically weighted Euclidean distance which measures the distance between d historical one-minute data at two time dimensions, r and s:

$$D\left(X_r^{\,d}, X_s^{\,d} \right) = \sqrt{\frac{\sum_{i=1}^{d} \alpha^{d-i+1} \left(x_{r-i+1} - x_{s-i+1} \right)^2}{\sum_{i=1}^{d} \alpha^{d-i+1}}} \qquad (66)$$

The geometrically weighted Euclidean distance detects similarities in very recent observations of the dataset and determinates the K nearest neighbours. I proceeded further to establish the forecasting model by weighting the predicting abilities of each neighbour. At time t, the nearest neighbours are represented by $X_{T_k}^{\,d}$ where k =1,..., K. The next step forecast is:

$$x_{t+i/t} = E\left\langle x_{t+i} \middle| x_t, x_{t-1}, \dots \right\rangle = \frac{\sum_{k=1}^{K} w_k x_{T_k+i}}{\sum_{k=1}^{K} w_k} \qquad (67)$$

where $w_k = \left(\zeta + D\left(X_t^{\,d}, X_{T_k}^{\,d} \right) \right)^{-1}$, with $\zeta = 10^{-5}$. I followed Meade (2002) to calculate the values of K, d and α by minimising the root mean squared error, mean absolute error, and per cent better than no change forecast for each of the six currency pairs.

9.5.2. Evaluation of point forecasts

In order to measure *ex-ante* forecasting accuracy, I adopted root mean squared error (RMSE) and mean absolute error (MAE):

$$RMSE = \left\{ \frac{1}{N} \sum_{t=1}^{N} (d_t - z_t)^2 \right\}^{0.5} \tag{68}$$

$$MAE = \frac{1}{N} \sum_{t=1}^{N} |d_t - z_t| \tag{69}$$

where N is the number of forecasting periods, d_i is the actual price of the six currency pairs at period t, and z_t is the forecasted currency pairs price at period t (Pai and Lin, 2005).

In addition to the parametric and non-parametric forecasting models, I applied Diebold and Mariano's (1995) test of predictive accuracy which is based on the difference between a quadratic loss function $g(e_{jt}) = e^2{}_{jt}$ and a linear loss function $g(e_{jt}) = |e_{jt}|$. Both loss functions are suitable for comparing RMSE and MAE.

Under the null hypothesis of the test forecasting model j and forecasting model 0 characterised by identical predictive accuracy:

$$H_0 : E\big(g(e_{jt})\big) = E\big(g(e_{0t})\big), \tag{70}$$

where $g(\ )$ represents the loss function. The test statistic S_s is quantified by:

$$S_s = \frac{n_+ - 0.5(T - n_=)}{\sqrt{0.25(T - n_=)}} \ , \tag{71}$$

where n_+ is a count of the occasions where $|e_{jt}| > |e_{0t}|$ is satisfied; and $n_=$ is a count of the occasions where $|e_{jt}| = |e_{0t}|$ is satisfied. PCB is related to S_s because:

$$PCB = 100 \frac{T - n_+ - n_0}{T} \tag{72}$$

where n_0 is a count of the occasions when $|e_{jt}| = |e_{0t}| \neq 0$ is satisfied (Meade, 2002).

9.5.3. Evaluation of directional forecasts.

In some trading situations, the direction of change in the FX rate rather than the point value is of prime importance. I implemented Pesaran and Timmermann's (1994) non-parametric test of predictive performance to detect the direction of my forecasting models. This particular test is appropriate because one-minute high frequency data exhibit observations where there is no price change since the last observation. The null implies that:

$$H_0 : \sum_{i=1}^{m} \left(\pi_{ii} - \pi_{i0}\pi_{0i} \right) = 0 \qquad (73)$$

where π_{ii} is the probability that the actual and predictive category is i ; π_{i0} is the probability that the actual category is i ; and π_{0i} is the probability that the predictive category is i .

9.6. Simulation results

9.6.1. Net profitability of high frequency trading

In this section I investigate whether technical trading associated with one-minute high frequency data for the six most traded currency pairs is profitable in the presence of appropriate transaction costs.

Table 69.0 reports the summary statistics of one-minute high frequency data for the six currency pairs. The distributions of returns are slightly skewed for most of the in-sample and out-of-sample the currency pairs. Apart from USD/JPY out-of-sample pair, the kurtosis is not significantly large and therefore there is no excess of absolute standardised returns: that are either large or close to zero: than would be expected from a normal distribution of returns.

In-sample						
Currency pair	***EUR/USD***	***USD/JPY***	***GBP/USD***	***AUD/USD***	***USD/CHF***	***USD/CAD***
Observations	176,735	176,735	176,735	176,735	176,735	176,735
Mean (x10^6)	7.37	468.12	9.13	5.89	5.26	5.68
Std.dev (10^6)	0.14	26.92	0.09	0.05	0.07	0.05
Skewness	0.25	0.81	-0.48	-0.28	0.19	-0.18
K	3.64	3.32	4.01	3.88	5.27	3.79
·LB-lag=1(1.2)	367.11	818.10	404.12	529.53	399.97	689.49
·LB-lag=10(16.9)	410.20	999.57	516.19	673.58	448.17	639.18
·LB-lag=50(74.0)	494.73	1001.22	597.34	701.76	594.01	796.27
Out-of-sample						
Observations	29,678	29,678	29,678	29,678	29,678	29,678
Mean (x10^6)	44.41	3153.20	51.40	34.61	31.39	34.67
Std.dev (10^6)	0.53	38.25	0.72	0.18	0.35	0.34
Skewness	0.41	0.87	0.42	-0.17	0.23	-0.71
K	2.88	3.73	4.11	3.91	3.12	4.04
·LB-lag=1(4.7)	500.24	998.87	612.00	681.27	499.80	745.02
·LB-lag=10(18.3)	493.10	1091.93	586.14	698.94	515.49	899.48
·LB-lag=50(71.4)	532.19	1187.10	658.49	749.04	683.27	999.92

K-kurtosis; ***LB***-Ljunk Box.

• The calculation of Ljunk Box statistic was modified to accommodate for ARCH effects (Diebold,1986). The numbers in parentheses are determined at 0.1% (0.001) significance levels for H_0:p_k=0 for $k \geq 1$.

Table 69.0 Summary of statistics of FX one-minute high frequency in-sample and out-of-sample data.

In his work with large data sets (Wallace, 1972) observed that there was a tendency to reject the null hypothesis too often unless the significance level is adjusted downward. The justification for this downward adjustment is that some of the increased sample size should be used to reduce type II errors (Klein and Brown, 1984). Lindley (1957) used different sample sizes and a fixed significance level and reported that the null hypothesis can be rejected even if the posterior odds favour the null when the sample size is increased but the significance level is fixed.

Given the large in-sample size of 206,413 and out-of-sample size of 29,678 observations, one issue that arises in the context of my empirical analysis is Lindley's paradox leading to overstatement of statistical significance and a tendency to reject the null hypothesis at conventional significance level even when posterior odds favour the null. To alleviate this issue Connolly (1989) proposed a formula for estimating sample size adjusted critical values for t statistics:

$$t^* = \left[(T-k)\left(T^{1/T}-1\right) \right]^{1/2} \quad (74)$$

where T is the sample size and k is the number of parameters estimated.

I made large-sample adjustments to critical t-values in order to eliminate the overstatement of statistical significance. When absolute value of a regression t-statistic is greater than the value of equation (14), its absolute value is reduced by the adjustment t^* (Corrado and Truong, 2007). When the estimated standard statistic is greater than the critical value of t^*, the null hypothesis should be rejected. I applied the Ljung Box test to the auto-correlation structure of returns. The null hypothesis of no-autocorrelation is rejected for all currency pairs at a 0.1 per cent (0.001) significance level. Hence, the rejection of the null hypothesis shows that an ARMA model with at least one significant coefficient can be detected in the series. This is also a sign of time dependent variation in the variance of returns of the six currency pairs.

I modelled the properties detected in the six data sets using a linear time series model- and an AR-GARCH model was fitted to all available observations. The reason for selecting the GARCH model was because I wanted to detect the time dependent variance and some of the excess kurtosis observed in the six currency pairs (Table 69.0). Previously, Baillie and Bollerslev (1991) and Cecen and Erkal (1996) applied the GARCH model to intra-day currency rates. Cecen and Erkal (1996) argue that intra-day FX returns are caused by a non-linear generating process.

I implemented the likelihood ratio tests to investigate the significance of the AR process and the GARCH process. First, the AR (p) models were determined by the common identification/estimation/diagnostic test cycle. The selected models in my experiment have a p of at least 3 and are characterised by significant coefficients at a 0.1 per cent significance level.

Table 70.0 reports the AR-GARCH model coefficients estimated by the quasi-maximum likelihood. This test demonstrated the presence of significant GARCH behaviour at a 0.1 per cent significance level in all one-minute data sets for the six currency pairs.

In-sample						
Parameters	***EUR/USD***	***USD/JPY***	***GBP/USD***	***AUD/USD***	***USD/CHF***	***USD/CAD***
p	4	3	5	6	6	5
$\theta_0\left(\times 10^6\right)$	-2.202	-1.999	0.900	0.228	-1.898	-2.001
v	2.00	2.87	2.37	2.56	2.72	2.60
$a_0\left(\times 10^6\right)$	0.006	0.011	0.009	0.005	0.010	0.004
a_1	0.99	0.074	0.101	0.092	0.080	0.079
b	0.619	0.728	0.631	0.803	0.711	0.693
ϕ_1	-0.131 (-6.12)	-0.072 (-4.53)	-0.139 (-6.18)	-0.155 (-5.98)	-0.141 (-6.01)	-0.138 (-5.69)
ϕ_2	-0.021 (-3.75)	-0.011 (-3.39)	-0.035 (-4.68)	-0.039 (-4.11)	-0.031 (-3.31)	-0.023 (-4.17)
ϕ_3	-0.012 (-3.08)	-0.009 (-2.18)	-0.019 (-4.01)	-0.018 (-3.79)	-0.010 (-3.99)	-0.008 (-2.92)
p_0	0.009	0.005	0.008	0.062	0.074	0.055
$p-values$ **	0.000	0.000	0.000	0.000	0.000	0.000
$RMSE\left(\times 10^3\right)$	0.616	0.817	0.700	0.559	0.682	0.511
$MAE\left(\times 10^3\right)$	0.399	0.663	0.491	0.334	0.448	0.299

*Likelihood ratio tests were performed to determine the significance of the AR process and the GARCH process.

** $p-values$ obtained for the hypothesis $H_0:\phi_i=0$ for $i=1,\ldots,p$. Robust *t*-statistics corrected for Lindley's paradox are reported in parentheses below coefficient values. All values are estimated at 0.1% (0.001) significance level.

Table 70.0 In-sample AR-GARCH* coefficient estimates based on the quasi-maximum likelihood method.

Table 71.0 represents the AR-GARCH model coefficients based on the STGP technique and reports similar findings with the real historical data analysis reported in Table 70.0. The in-sample RMSE and MAE for the AR-GARCH models of the historical data sets of six currency pairs are slightly higher than the RMSE and MAE for AR-GARCH models based on the STGP technique. This is evidence of STGP's superiority over parametric forecasting methods. Moreover, when the sum of the GARCH coefficients $\left(a_1+b\right)$ equals unity (the conditional variance does not converge on a constant unconditional variance in the long-run) then the unconditional variance is non-stationary.

In-sample						
Parameters	***EUR/USD***	***USD/JPY***	***GBP/USD***	***AUD/USD***	***USD/CHF***	***USD/CAD***
p	4	2	4	6	5	5
$\theta_0\left(\times 10^6\right)$	-2.783	-2.118	0.99	0.310	-2.761	-2.381
v	2.70	2.83	2.79	2.71	2.63	2.71
$a_0\left(\times 10^6\right)$	0.006	0.010	0.005	0.005	0.010	0.006
a_1	0.091	0.078	0.100	0.097	0.091	0.087
b	0.718	0.779	0.691	0.703	0.700	0.725
ϕ_1	-0.101 (-5.91)	-0.056 (-4.15)	-0.100 (-6.28)	-0.121 (-6.72)	-0.111 (-6.10)	-0.101 (-6.00)
ϕ_2	-0.020 (-4.61)	-0.009 (-2.73)	-0.033 (-3.82)	-0.029 (-4.01)	-0.021 (-3.95)	-0.131 (-2.72)
ϕ_3	-0.011 (-3.14)	-0.008 (-2.68)	-0.030 (-4.19)	-0.015 (-4.00)	-0.014 (-3.99)	-0.009 (-2.97)
p_0	0.004	0.002	0.001	0.043	0.033	0.029
$p-values$ **	0.000	0.000	0.000	0.000	0.000	0.000
$RMSE\left(\times 10^3\right)$	0.599	0.800	0.672	0.521	0.655	0.501
$MAE\left(\times 10^3\right)$	0.385	0.637	0.467	0.317	0.424	0.270

*Likelihood ratio tests were performed to determine the significance of the AR process and the GARCH process.

** $p-values$ obtained for the hypothesis $H_0 : \phi_i = 0$ for $i = 1,..., p$. Robust *t*-statistics corrected for Lindley's paradox are reported in parentheses below coefficient values. All values are estimated at 0.1% (0.001) significance level.

Table 71.0 In-sample AR-GARCH* coefficient estimates based on the STGP method.

The sum of the GARCH coefficients listed in Tables 70.0 and 71.0 are less than one, suggesting that volatility clustering does not have a permanent effect. Hence, the volatility shock is time-decaying and mean reverting (specific periods of high changeability do not persist indefinitely and continually decrease to its long-run mean-level at a rate equal to the sum of $a_1 + b$ coefficients).

The next step is to compare the forecasting abilities of non-parametric models such as K nearest neighbours and the STGP technique. The reason for choosing the K nearest neighbour method is based on its popularity. This particular method is among the most popular methods used in statistical pattern recognition, with over 900 different studies published on the method since 1981 (Holmes and Adams, 2002). The K nearest neighbour procedure forecasts a new point y_{n+1} to be the most common class found among the K nearest neighbours of x_{n+1} in the set $\{x_i\}^n_{i=1}$ (Holmes and Adams, 2002).The K nearest neighbour method highlights the importance of the similarities between recent observations based on the value of α (≥ 1).

I have chosen the following optimal K nearest neighbour model parameters:

$K \in \{10, 50, 100, 150, 200, 250, 300, 350, 400\}$, $d \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $\alpha \in \{1.0, 1.5, 2.0\}$.

The values of the optimal model parameters are estimated by minimising two objective functions, such as RMSE and MAE.

Tables 72.0 and 73.0 suggest that the surfaces described by the objective functions and the values of K , d and α are relatively flat and therefore a significant change in K and d does not correspond to a substantial change in the objective function.

In-sample												
Pair	***EUR/USD***		***USD/JPY***		***GBP/USD***		***AUD/USD***		***USD/CHF***		***USD/CAD***	
Value*	***KK(1)-RMSE***	***KK(2)-MAE***	***KK(1)-RMSE***	***KK(2)-MAE***	***KK(1)-RMSE***	***KK(2)-MAE***	***KK(1)-RMSE***	***KK(2)-MAE***	***KK(1)-RMSE***	***KK(2)-MAE***	***KK(1)-RMSE***	***KK(2)-MAE***
K	300	350	400	300	200	300	250	350	200	250	300	400
d	5	6	5	10	7	7	5	8	5	5	8	8
α	1	1	1	1	1	1	1	1	1	1	1	1
RMSE	0.703	0.721	0.981	0.996	0.812	0.867	0.707	0.723	0.798	0.812	0.636	0.650
MAE	0.510	0.538	0.754	0.787	0.544	0.600	0.430	0.434	0.530	0.544	0.400	0.430

*The values of *K, d* and α are estimated by minimizing the RMSE and MAE for each of the six currency pair series.

Table 72.0 In-sample K nearest neighbour forecasting based on conventional non-parametric predictive model coefficient estimates (RMSE and MAE are shown x10^3)

In-sample												
Pair	***EUR/USD***		***USD/JPY***		***GBP/USD***		***AUD/USD***		***USD/CHF***		***USD/CAD***	
Value	***KK(1)-RMSE***	***KK(2)-MAE***	***KK(1)-RMSE***	***KK(2)-MAE***	***KK(1)-RMSE***	***KK(2)-MAE***	***KK(1)-RMSE***	***KK(2)-MAE***	***KK(1)-RMSE***	***KK(2)-MAE***	***KK(1)-RMSE***	***KK(2)-MAE***
K	250	300	400	350	250	350	400	300	300	350	250	300
d	5	7	8	7	6	7	8	8	7	7	5	7
α	1	1	1	1	1	1	1	1	1	1	1	1
RMSE	0.672	0.691	0.942	0.956	0.800	0.812	0.683	0.699	0.733	0.748	0.612	0.638
MAE	0.490	0.507	0.721	0.733	0.533	0.544	0.428	0.432	0.519	0.532	0.390	0.402

*The values of *K, d* and α are estimated by minimizing the RMSE and MAE for each of the six currency pair series.

Table 73.0 In-sample K nearest neighbor forecasting coefficient estimates based on the STGP technique (RMSE and MAE are shown x10^3)

The STGP technique outperformed the non-parametric predictive model evidenced by lower RMSE and MAE levels. On the other hand, the RMSE and MAE of the non-parametric model are slightly higher than the errors of the AR-GARCH model, indicating better in-sample performance.

For *ex-ante* forecasting evaluation, I used summary accuracy measures and a formal hypothesis test, such as Pesaran and Timmermann (1994), which investigates the directional forecasting accuracy and therefore provides important information on the statistical significance of sign forecasts. I have found this particular test to be appropriate for *ex-ante* forecasting evaluation because it compares the number of observed correct forecasts with the estimated and expected probability of a correct forecast during conditions of independence between actual and forecast changes. Additionally, I investigated the market timing ability by performing the Henriksson and Merton (1981) test, which has a hypergeometric distribution under the null hypothesis of no market timing ability. This test provides information on market timing that is independent of any distributional assumptions about the return on the currency pairs. The test includes the possibility that the forecaster's confidence in his forecasts can vary over time and, if such variations are observable, then the test can be refined to evaluate his predictive ability for each variation. The results of both the above tests can be found in Tables 74.0 and 75.0, and they indicate that the sign predictions of the forecasting models have market timing value across all currency pairs.

Error measure	***Model***	***Out-of-sample***					
		EUR/USD	***USD/JPY***	***GBP/USD***	***AUD/USD***	***USD/CHF***	***USD/CAD***
RMSE•	***AR-GARCH***	0.2366	0.4385	0.3859	0.2438	0.3985	0.3287***
	KK(1)	0.3948***	0.5985***	0.3011***	0.2498***	0.3846***	0.3349***
	KK(2)	0.3865***	0.4530	0.3984***	0.3110***	0.3755	0.3437
	KK(3)	0.3857	0.5348***	0.3895	0.2438***	0.3328***	0.3103***
Formal tests	***P-T test***[a]	3.61***	5.04***	3.10***	2.11***	2.99***	3.00***
	H-M test[b]	0.0027***	0.0073***	0.0057***	0.0081***	0.0099	0.0001***
MAE	***AR-GARCH***	0.2218	0.2191***	0.2003	0.2000	0.2247	0.2524
	KK(1)	0.2537	0.2997***	0.2007	0.1999***	0.2338***	0.2119
	KK(2)	0.2502***	0.2889	0.2745	0.2110**	0.2829	0.2130***
	KK(3)	0.2210	0.2945	0.2735***	0.2098	0.2742	0.2538
Formal tests	***P-T test***	1.97	2.89	2.35***	1.90	1.72	1.68***
	H-M test	0.0138	0.0219	0.0004***	0.0120**	0.0105**	0.0004***

P-T: Pesaran and Timmerman test; ***H-M:***Henriksson and Merton test;

***Significant at 0.1% (0.001) level.

•The significance in RMSE differences is performed by using Diebold-Mariano test with quadratic loss function.

a-The Pesaran and Timmerman test is a Husman type test with limiting distribution of the test *N*(0,1).

b-In the Henriksson and Merton test, the number of forecasts has a hypergeometric distribution under the null hypothesis which postulates no market timing ability. The *p*-values of the test are reported.

Table 74.0 Summary of RMSE (x10^3), MAE (x10^3), Pesaran and Timmerman and Henriksson and Merton tests for one period ahead forecasts based on conventional parametric and non-parametric predictive models

Error measure	*Model*	***Out-of-sample***					
		EUR/USD	***USD/JPY***	***GBP/USD***	***AUD/USD***	***USD/CHF***	***USD/CAD***
	AR-GARCH	0.2305***	0.4300	0.3799	0.2390	0.3974	0.3198***
	KK(1)	0.3898	0.5823***	0.3000***	0.2425***	0.3788***	0.3219
RMSE•	***KK(2)***	0.3790***	0.4497	0.3926	0.3099***	0.3622	0.3388***
	KK(3)	0.3802***	0.5267	0.3843	0.2400	0.3277	0.3001***
Formal tests	***P-T test***[a]	3.08***	4.83***	2.99***	1.86***	2.15***	2.28***
	H-M test[b]	0.0006***	0.0018***	0.0016***	0.0013***	0.0056***	0.0001***
	AR-GARCH	0.2190	0.2100	0.1989***	0.1558	0.2179***	0.2494
	KK(1)	0.2455***	0.2980	0.1909	0.1881***	0.2300	0.2090***
MAE	***KK(2)***	0.2492	0.2812***	0.2677	0.2101	0.2788	0.2005
	KK(3)	0.2099	0.2828	0.2654	0.1766***	0.2733	0.2401
Formal tests	***P-T test***	1.12***	2.04***	1.99***	1.01***	1.08***	1.05***
	H-M test	0.0126***	0.0195***	0.0001***	0.0098***	0.00090***	0.0002***

P-T: Pesaran and Timmerman test; ***H-M:***Henriksson and Merton test;

***Significant at 0.1% (0.001) level.

•The significance in RMSE differences is performed by using Diebold-Mariano test with quadratic loss function.

a-The Pesaran and Timmerman test is a Husman type test with limiting distribution of the test *N*(0,1).

b-In the Henriksson and Merton test, the number of forecasts has a hypergeometric distribution under the null hypothesis which postulates no market timing ability. The *p*-values of the test are reported.

Table 75.0 Out-of-sample summary of RMSE (x10^3), MAE (x10^3), Pesaran and Timmerman and Henriksson and Merton tests for one period ahead forecasts based on the STGP technique.

The Pesaran and Timmermann test of directional accuracy revealed significant predictive performance. Based on the RMSE and MAE values, the K nearest neighbour models- KK(1), KK(2) and KK(3)- are each significantly more accurate than the AR-GARCH forecasting models. The actual significance in the differences in RMSE was achieved by applying the Diebold-Mariano testing procedure with quadratic loss function. However, there is no consistent superiority of either parametric or non-parametric models when I consider both in-sample and out-of-sample periods. This finding is consistent with Meade (2002). A direct out-of-sample comparison between traditional econometric forecasting models and STGP technique indicate the superiority of the latter. The RMSE and MAE generated by STGP are significantly smaller than the out-of-sample errors produced by both parametric and non-parametric models.

I measure profitability by two primary criteria: the number of correct hits (forecasts) and the generated excess return from trading of the six currency pairs. The hit ratio determinate the percentage of time that the model has good sign of predictability:

$$\text{Hit ratio (\%)}=\frac{\text{Number of correct forecasts}}{\text{Number of generated buy / sell orders}}\times 100 \qquad (75)$$

The other profitability criterion- the excess return-represents the amount received from trading in excess of the risk free rate. It is the continuously compounded return on the six currency pairs, minus the value of the daily continuously compounded rate converted from the annualised investment yield on a three-month US Treasury bill:

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) - r_{t-1} \qquad (76)$$

where P_t is the price of EUR/USD, USD/JPY, GBP/USD, AUD/USD, USD/CHF, and USD/CAD traded at period t, and r_t is the risk free rate set at the value of the daily continuously compounded rate converted from the annualised investment yield on a three-month US Treasury bill (data up to March 3, 2013 has been downloaded from the Federal Reserve statistical release website at www.federalreserve.gov/releases/h15). The reason for choosing the US Treasury bill is the fact that the US dollar participates in every currency pair in my experiment. The number of correct out-of-sample forecasts for the six currency pairs is reported in Table 76.0. The corresponding hit ratios are also given. A hit ratio above fifty per cent is a sign of actual profitability from FX trading.

Currency pair	*EUR/USD*	*USD/JPY*	*GBP/USD*	*AUD/USD*	*USD/CHF*	*USD/CAD*
Number of generated buy/sell orders	16,787	16,019	16,982	15,384	15,291	15,325
Number of successful hits	8,931	8,378	8,949	8,446	8,104	8,551
Successful hit ratio	53.2%*	52.3%*	52.7%*	54.9%*	53.0%*	55.8%*
Excess return	3.63%	3.11%	3.25%	5.01%	3.52%	5.76%

The table reports the number of times a STGP out-of-sample forecasting model correctly predicts the direction of the six currency pairs returns and profitability of 29,678 observations (12/02/2012-12/03/2013) for each currency pair. A ratio market with asterisk (*) indicates a 95% significance based on a one-sided test of H_0:p=0.50 against H_a:p>0.50. . [b]The risk-free rate is set at the value of daily continuously compounded rate converted from the annualized investment yield on a 3-month US Treasury bill (up to 12/03/2013).

Table 76.0 Out-of-sample comparison of the predictive strength and profitability of STGP in the presence of transaction costs of 1.5 basis points for the six currency pairs.

The USD/CAD pair reports the highest hit ratio of 55.8% (8,551 successful hits out of 15,325), followed by: the AUD/USD pair with 54.9% (8,446 successful hits out of 15,384), the EUR/USD pair with 53.2% (8,931 successful hits out of 16,787), the USD/CHF pair with 53.0% (8,104 successful hits out of 15,291), the GBP/USD pair with 52.7% (8,949 successful hits out of 16,982), and the USD/JPY pair with 52.3% (8,378 successful hits out of 16,019).

I also conducted a one-sided test to investigate whether the hit ratios of the six currency pairs are significantly different from the benchmark of 0.5 (a 95 per cent significance level).

Under the null hypothesis the test has no predictive effectiveness power $(H_0 : p = 0.50$ against $H_0 : p > 0.50)$. The statistical tests rejected the null indicating that the hit ratios of the six currency pairs are significantly different from 0.50. This important finding confirms the forecasting ability of my models.

While predictability is of considerable theoretical interest, profitability of FX trading is of obvious economic importance. Table 76.0 reports the out-of-sample excess return gained from trading the six currency pairs. The USD/CAD currency pair generated the highest excess return of 5.76%, followed by: the AUD/USD pair with 5.01%, the EUR/USD pair with 3.63%, the USD/CHF pair with 3.52%, the GBP/USD pair with 3.25%, and the USD/JPY pair with 3.11%, in the presence of appropriate transaction costs. My profitability findings are consistent with Carrion (2013), who argued that there is economically significant predictability in intraday prices, but contrary to Menkveld (2013), who claims that HFT is unsuccessful at forecasting price evolutions during the day beyond the very short horizon of five seconds. The difference in results can be explained by the fact that Menkveld used a single specific HFT that pursued a market making strategy rather than an aggregate HFT in the NASDAQ dataset implemented by Carrion, which enabled him to detect forecasting of intraday prices beyond five seconds. However, Menkveld (2013) also reported that net long or short positions can have duration of seconds, minutes, and hours generating statistically significant intraday profits. Chordia *et al.* (2008) suggested that intraday market inefficiencies are present in markets that are otherwise efficient at longer horizons due to the fact that investors need sufficient time to process and react to information.

Neely and Weller (2003) investigated the effect of up to 2.0 basis points of intraday technical trading transaction costs for one-way transactions. Various studies have demonstrated that transaction costs have decreased over time (Chordia *et al.,* 2008, 2011; Angel *et al.,* 2011). I assigned transaction costs of 1.5 basis points for high frequency technical trading.

The choice of 1.5 basis points is based on the real-life exchange fee structure. Euronext charges a fixed fee of €1.20 per trade, which represent 0.48 basis points for a trade of approximately €25,000. Additionally, Euronext imposes a variable fee of 0.05 basis points and another 0.04 basis points if the cancellation-to-trade ratio is over 5. On top of that there are post-trade expenses such as clearing fees and necessary trading margins (Menkveld, 2013). LCH-Clearnet, the clearing house used by Euronext, charge clearing fees of €0.23 (0.09 basis points). To enhance the HFT realism in my experiment, I assigned the remainder of 0.84 basis points to funds required to keep margin accounts with the clearer, exchange and clearing house membership fees, development and acquisition of trading software and hardware. More importantly, I used robust transaction costs that allowed protection from data over-fitting, which represents the biggest pitfall in forecasting. Data over-fitting can exaggerate minor fluctuations in the dataset of the six currency pairs leading to poor predictive performance. Moreover, data over-fitting can produce test sensitivities and specificities hardly reproducible in subsequent experiments. If the experiment is over-fitted forecasting models with high in-sample explanatory power may not have high out-of-sample fit (Bossaerts and Hillion, 1999; Chatfield, 1996).

9.6.2. The impact of HFT on market quality

Is HFT beneficial or harmful to market efficiency? The role of HFT on price discovery and market liquidity processes. To what extend the policymakers should control HFT?

Recent years have witnessed a raise of computer-driven trading which is characterised by high speed and processing power leading to accelerated trading times. Transactions are executed in the blink of an eye, forcing humans out of the trading process. Determining whether HFT is beneficial or harmful to market efficiency is not an easy task and the debate so far includes many contradictory studies and claims.

It is a well known fact that there are no market makers in HFT. As described earlier my experimental settings do not have any market makers providing real-life trading conditions to investigate the implications of HFT on market efficiency. Similar to Brogaard *et al.* (2013), I applied a state space model to decompose FX price movements into permanent and transitory (temporary) components and relate those price movements to one-minute HFT. While the permanent component is interpreted as information, the transitory component represents pricing errors which is also known as temporary volatility or noise.

I implemented the state space models in my experiment because they offer several advantages. First, the state space models describe how information is incorporated into prices and the informativeness of prices. Second, these models allowed us to observe the overall roles of HFT in currency prices and the differential role of active and passive HFT. While Hendershott and Riordan (2011) argue that the state space models explicitly separates short-term transitory effects from long-term permanent effects, Moody and Wu (1997) reported that the state space models are able to explain correlational structures in the high frequency data that the conventional random-walk models of efficient market theory do not explain. Hendershott and Menkveld (2011) provide more evidence for why state space models are a more appropriate tool for the investigation of the impact of HFT on market properties. According to the authors, the maximum likelihood calculation in the state space models is asymptotically unbiased and efficient, and while the conventional vector autoregressive models require truncation of the lag structure, the state space model implies an infinite lag autoregressive structure.

Most importantly in the state space model, the Kalman smoother enables a series decomposition where the efficient currency price and the transitory errors are calculated at any point in time by using all 206,413 one-minute observations (both in-sample and out-of-sample data sets). Price pressures caused by increased liquidity demand are leading to noise in prices (transitory pricing error) of the six currency pairs. To investigate whether HFT plays a beneficial or harmful role on price efficiency, I needed to detect the trading direction towards permanent price changes and transitory pricing errors. HFT will have a beneficial impact on the market if one-minute high frequency trading is orientated in the opposite direction of the transitory pricing error, thereby reducing transaction costs.

When HFT trading is orientated towards the direction of the transitory pricing error, transaction costs increase. Trading in the direction of the pricing error could be a result of manipulative trading strategies or predatory trading, and has a harmful effect on the market.

Menkveld *et al.* (2007) suggest that the state space model of the currency pairs can be decomposed into two dinstinctive parts- permanent and transitory components:

$$p_{i,t} = m_{i,t} + s_{i,t} \tag{77}$$

where $p_{i,t}$ is the (log) midquote (the average of the bid and the ask quote) at time t for currency pair i, $m_{i,t}$ is the permanent component of a martingale type- $m_{i,t} = m_{i,t-1} + w_{i,t}$ ($w_{i,t}$ is the innovative element in permanent price component); and $s_{i,t}$ is the transitory component.

To investigate the overall impact of HFT on market properties, I developed two different state space models. The first model analyses HFT^{all} activity and the second model investigates the two elements of HFT^{all}- HFT^{bid} and HFT^{ask}. To estimate the aggregate model, I follow Menkveld (2011) and Hendershott and Menkveld (2011):

$$w_{i,t} = k_i^{all} HFT_{i,t}^{all} + \mu_{i,t} \tag{78}$$

where $HFT_{i,t}^{all}$ is the surprise innovation factor in HFT^{all}, which is the residual of an autoregressive model to eliminate autocorrelation.

In order to calculate the state space model for each currency pair of the one-minute high frequency data in a trading day, I implemented the maximum likelihood via the Kalman filter where price changes- HFT^{all}, HFT^{bid} and HFT^{ask} - are non-zero. To estimate the statistical inference I adopted the clustering techniques of Petersen (2009) and Thompson (2011). Table 77.0 represents the empirical results of the HFT^{all} space model for each of the six currency pairs, and the overall space model. All space models are positively correlated with efficient price changes (permanent price component) and negatively correlated with pricing errors (transitory price component). This is clear evidence of the positive role that HFT plays in the price discovery process.

Permanent price component								
	Measures	***EUR/USD***	***USD/JPY***	***GBP/USD***	***AUD/USD***	***USD/CHF***	***USD/CAD***	***ALL***
k^{all}	bps/$100,000	0.21	0.48	0.17	0.29	0.44	0.38	0.32
(t-stat)		(5.28)	(12.31)	(4.49)	(8.37)	(7.98)	(8.12)	(7.76)
$\sigma^2\left(HFT^{all}\right)$	$100,000	14.11	18.28	13.01	14.43	16.49	10.35	15.9
$\left(k^{all} * \sigma\left(HFT^{all}\right)\right)^2$	bsp.2^2	10.88	19.99	10.21	12.35	16.19	10.83	11.81
(t-stat)		(20.11)	(31.32)	(19.56)	(22.98)	(28.12)	(20.05)	(22.90)
$\sigma^2\left(w_{i,t}\right)$	bsp.2^2	219.07	328.70	201.38	255.56	290.33	203.47	299.27
Transitory price component								
ϕ	bps/$100,000	0.36	0.45	0.27	0.34	0.41	0.30	0.33
ψ^{all}		-0.03	-0.12	-0.03	-0.10	--0.11	-0.05	-0.07
(t-stat)		(-7.01)	(-13.14)	(-4.49)	(-8.11)	(-12.74)	(-8.62)	(-7.21)
$\sigma^2\left(HFT^{all}\right)$	$100,000	0.22	1.11	0.84	1.01	0.93	0.54	0.66
$\left(\psi^{all} * \sigma\left(HFT^{all}\right)\right)^2$	bsp.2^2	3.19	6.81	3.14	5.13	5.01	3.48	5.25
(t-stat)		(9.39)	(12.27)	(10.48)	(11.39)	(6.90)	(5.58)	(11.17)
$\sigma^2\left(s_{i,t}\right)$	bsp.2^2	99.39	150.06	102.77	148.90	101.36	76.44	114.59

The model is estimated for each currency pair, every minute, using HFT variables to decompose the observable historical price $p_{i,t}$ for currency pair i at time t (in one-minute increments) into two components: the permanent price component $m_{i,t}$ and the transitory component $s_{i,t}$: $p_{i,t} = m_{i,t} + s_{i,t}$; $m_{i,t} = m_{i,t-1} + w_{i,t}$; $w_{i,t} = k^{all} HFT^{all}{}_{i,t} + \mu_{i,t}$;

$$s_{i,t} = \phi s_{i,t-1} + \psi^{all} HFT^{all}{}_{i,t} + \upsilon_{i,t}$$

Where $HFT^{all}{}_{i,t}$ is HFT overall order flow; $HFT^{all}{}_{i,t}$ is the surprise component of the order flow. *T*-statistics are calculated using standard errors double-clustered on currency pair and one-minute data.

Table 77.0 State Space Model for each of the six currency pairs based on the whole sample of 206,413 one-minute high frequency data.

The values of k and ψ are calculated in basis points per $100,000 traded. The value of 0.32 for the overall k coefficient listed in the last column suggests that $100,000 of positive surprise order flow (bid minus ask orders) corresponds to a 0.32 basis points increase in the efficient price. My findings are consistent with Brogaard *et al.* (2013) and O'Hara and Ye (2011), but contrary to Carrion (2013), who found that an aggregate HFT makes money on average when providing liquidity and loses money on average when demanding liquidity. The conflicting findings can be explained by the sensitivity of the manner in which profits are described and estimated (Carrion, 2013).

The aggregate proportion of efficient price variance $\left(k^{all} * \sigma\left(HFT^{all}\right)\right)^2$ correlated with overall HFT order flow is 11.81 basis points squared in the one-minute permanent price variance of 299.27 basis points squared. The negative values of ψ coefficients in the pricing errors suggest that HFT is trading in the opposite direction to the pricing errors, leading to reduced transaction costs. Hence, HFT is having a beneficial effect on price discovery and efficiency.

To investigate the individual impact of bid and ask orders on price efficiency and the formation of HFT strategies, I implemented the disaggregated state space model of HFT.

Table 78.0 illustrates that while HFT bid orders are positively correlated (k^{bid} is positive in all currency pairs and overall) with changes in the permanent price component, HFT ask orders are negatively correlated with price changes in the permanent price component. Positive correlation of bid orders is associated with informed trading. The negative coefficients of k^{ask} indicate that passive trading occurs in the opposite direction of permanent price changes.

Permanent price component								
	Measures	***EUR/USD***	***USD/JPY***	***GBP/USD***	***AUD/USD***	***USD/CHF***	***USD/CAD***	***All***
K^{bid}	bps/$100,000	14.03	19.37	10.95	13.58	12.85	5.90	14.82
t-stat		(21.48)	(36.94)	(19.83)	(22.25)	(20.26)	(18.64)	(24.54)
K^{ask}	bps/$100,000	-19.02	-28.37	-18.28	-17.28	-11.12	-4.48	-19.82
t-stat		(-21.98)	(-37.29)	(-20.21)	(-25.49)	(-22.30)	(-15.19)	(-26.49)
$\sigma^2\left(HFT^{bid}\right)$	$100,000	0.91	1.10	1.03	0.98	1.01	0.45	1.05
$\sigma^2\left(HFT^{ask}\right)$	$100,000	0.96	1.01	0.92	0.78	0.90	0.32	0.95
$\left(k^{bid} * \sigma\left(HFT^{bid}\right)\right)^2$	bps.^{2}	28.55	39.49	29.93	30.04	28.54	20.39	31.12
t-stat		29.91	37.47	30.39	30.03	28.92	20.21	(30.19)
$\left(k^{ask} * \sigma\left(HFT^{ask}\right)\right)^2$	bps.^{2}	26.73	37.97	29.17	28.11	26.45	17.66	28.38
t-stat		(39.90)	(50.93)	(41.41)	(39.17)	(40.12)	(24.48)	(42.47)
$\sigma^2\left(w_{i,t}\right)$	bps.^{2}	314.28	433.44	327.01	311.73	300.81	267.34	368.81
Transitory price component								
ϕ		0.18	0.25	0.22	0.19	0.21	0.10	0.22
ψ^{bid}	bps/$100,000	-3.01	-4.11	-3.29	-4.31	-3.30	-2.22	-3.36
t-stat		(-20.20)	(-29.97)	(-21.95)	(-19.59)	(-20.19)	(-11.93)	(-22.10)
ψ^{ask}	bps/$100,000	5.29	6.03	4.92	5.00	3.84	2.99	5.47
t-stat		(21.99)	(37.19)	(29.90)	(26.60)	(24.20)	(19.18)	(26.73)
$\sigma^2\left(HFT^{bid}\right)$	$100,000	0.99	1.18	1.05	1.00	1.06	0.50	1.07
$\sigma^2\left(HFT^{ask}\right)$	$100,000	0.97	1.10	0.96	0.83	0.95	0.38	0.97
$\left(\psi^{bid} * \sigma\left(HFT^{bid}\right)\right)^2$	bps.^{2}	9.20	16.81	10.17	9.87	10.01	5.94	10.79
t-stat		(28.88)	(41.19)	(30.57)	(29.92)	(31.84)	(20.21)	(31.28)
$\left(\psi^{ask} * \sigma\left(HFT^{ask}\right)\right)^2$	bps.^{2}	9.05	16.77	10.73	9.80	10.00	5.90	10.66
t-stat		(41.09)	(51.57)	(38.22)	(34.38)	(37.91)	(24.26)	(40.18)
$\sigma^2\left(s_{i,t}\right)$	bps.^{2}	200.73	301.20	198.03	203.19	202.39	144.82	203.11

The model is estimated for each currency pair, every minute, using HFT variables to decompose the observable historical price $p_{i,t}$ for currency pair i at time t (in one-minute increments) into two components: the permanent price component $m_{i,t}$ and the transitory component $s_{i,t}$: $p_{i,t} = m_{i,t} + s_{i,t}$; $m_{i,t} = m_{i,t-1} + w_{i,t}$ $w_{i,t} = k^{bid}{}_i \widetilde{HFT}^{bid}{}_{i,t} + k^{ask}{}_i \widetilde{HFT}^{ask}{}_{i,t} + \mu_{i,t}$;
$s_{i,t} = \phi s_{i,t-1} + \psi^{bid}{}_i HFT^{bid}{}_{i,t} + \psi^{ask}{}_i HFT^{ask}{}_{i,t} + \upsilon_{i,t}$ Where $HFT^{bid}{}_{i,t}$ and $HFT^{ask}{}_{i,t}$ represent bid and ask order flows; $\widetilde{HFT}^{bid}{}_{i,t}$ and $\widetilde{HFT}^{ask}{}_{i,t}$ are the surprise components of those order flows. *T*-statistics are calculated using standard errors double-clustered on currency pair and one-minute data.

Table 78.0 Disaggregated State Space Model of HFT^{bid} and HFT^{ask} for each of the six currency pairs based on the whole sample of 206,413 one-minute high frequency data.

I observed a negative relation between ψ^{bid} and the transitory component across all currency pairs, indicating overall that HFT^{bid} trades are orientated opposite to the transitory component of prices. This important finding suggests that when currency prices deviate from their fundamental values, HFT initiate trades to restore prices to their efficient levels. As a result, the distance between quoted prices of the six currency pairs and the efficient price has been reduced. On the other hand, the coefficients of ψ^{ask} are all positive, suggesting adverse selection of coefficients based on uninformed transitory price component. HFT^{bid} trades are associated with substantially more information being incorporated into the prices of the six currency pairs, and smaller pricing errors in comparison with HFT^{ask} trades.

Overall, I found evidence that HFT has a beneficial role in the price discovery process in terms of smaller pricing errors and information being incorporated into prices of the six currency pairs. More informative FX currency prices means a better allocation of resources. Also, reduced pricing errors in my experiment helped to improve the efficiency of prices and lowered the transaction costs.

However, the information content of HFT in my experiment lasts for one minute only, and it is unlikely to become public in such a short period of time. Hence, HFT's informational advantage has a short-term nature. The same principle is valid for the reduced pricing errors. It is still unclear whether one-minute reductions in pricing errors lead to significantly more efficient resource allocation and more effective investment decisions. The picture of whether high frequency trading is harmful to the economy is also incomplete. The overall HFT is negatively related to pricing errors which is a sign that no manipulation strategies are in place. On the other hand, HFT^{ask} is positively connected to pricing errors, indicating the presence of either manipulating trading strategies or inappropriate risk management practices. This finding is consistent with the empirical results of Brogaard *et al.* (2013).

From another point of view HFT imposes difficulties for investors and pension funds to purchase large blocks of financial instruments, because HFT software detects and front-runs the order. Big institutional investors, on the other hand, have a substantial advantage based on their financial strength in buying HFT algorithms, their ability to locate their servers in close proximity to the exchanges, and eventually to manipulate the FX market.

Chordia *et al.* (2013) argue that buy-side investors could struggle to trade large positions, and the speed disadvantage reduces their ability to supply liquidity leading to increased costs. Low-latency connections to FX markets enable high frequency traders to operate at much faster speeds, disinclining traditional traders to invest or forcing them to spend excess amounts of money to keep up with all the technological innovations. Substantial investments in computer and communication power to reduce latency in trading posed the question of whether HFT adds value (Chordia *et al.,* 2013). Hansbrouck and Saar (2013) found that some HFT algorithms need only 2-3 milliseconds to identify the arrival of an order, analyse it, and generate an order. This very high operational speed prevent human traders from appropriately observing the limit order book, indicating that market dynamics might be dominated entirely by the interplay between trading algorithms. Individual investors have witnessed the transformation of stock markets to more complex entities, and have restricted access to the same type of trading equipment as large, institutional investors. HFT could potentially create negative externalities on other market participants due to continuous generation of submissions and cancellations of limit orders that increases the need for costly equipment updates and worsens market regulation (Gai *et al.,* 2012). Moreover, poor programming of HFT algorithms has the potential to disrupt markets (for instance, the Knight Capital case).

The occurrence of HFT has forced traditional investors to conduct big transactions into 'dark pools'. Nowadays, investors are extra cautious about the possibility of having their traders detected and headed off by high frequency traders. This is the reason why traditional investors allocate their orders in 'dark pools'. 'Dark pools' are off-exchange trading platforms administrated by brokers where financial instruments are executed anonymously and the prices are not announced in advance. According to the recent UK Foresight report (2012), two-thirds of investors went into 'dark pool' transactions, tripling their market share in the last few years to 3.3% of the total trading volume. Trading in 'dark pools' is having several negative consequences, such as: increased expenses which affect the transparency of the market by imposing price obstacles for the other investors. The ASIC report (2013) suggests that HFT performed in off-market 'dark pools' are adversely affecting the quality of asset price information and widening the bid-ask spread for several assets.

Moreover, orders executed on crossing systems (automated services that match or executes orders away from exchange markets) are leading to uninformed investment decisions, and listed companies are unaware of where their assets are trading.

I illustrate 'dark pool' trading by the following practical example. The currency pair GBP/USD is trading at a bid of $1.5231 and ask of $1.5235 and an investor would like to sell the pair. Under normal trading conditions an investor could generate a sell order at $1.5234 and hope that other investors will buy, although there is no guarantee that anyone will purchase at this particular price. If the same investor was to allocate a hidden offer at $1.5234 in a 'dark pool', which is not prohibited by law and nobody can see it. Then he/she could generate a large number of public and visible bids at prices around $1.5229 and $1.5228. The investor does not intend to purchase at those prices and in fact does not want to purchase at all. But for the other market participants with access to the public order book it looks like there is significant interest in GBP/USD. This will create an illusion that an investor is making a big investment in GBP/USD greatly enhancing the buying pressure. The other investors put their bids at $1.5232 and $1.5233 in order to overcome the new buyer in the market. When all traders bid at $1.5234 they discover the hidden sell order at this particular price. The hidden order actually receives a very good price on its trade and counts as a liquidity provider. Then the unrealistic bids of $1.5229 and $1.5228 are removed from the market. Unfortunately the victims of the whole process are traditional investors mislead by unrealistic public bid orders and someone trading against these orders.

To limit the severe consequences of 'dark pools', I think that policymakers need to introduce stricter circuit breakers that halt trading when prices experience large and sudden movements. Current regulatory rules include circuit breakers, but they are not tight enough- allowing around ten per cent drops in prices before they come into force. In the US, circuit breakers halt trading as soon as stocks change by ten per cent or more within five minutes (SEC, 2010g). To avoid the danger of systemic risks, I think that limits on the number of orders traded and minimum holding periods to reduce the pace of HFT are needed. I propose a minimum holding period of 700 milliseconds (0.7 seconds) for small orders of $500 or less comparing to the current holding peiod of 900 milliseconds.

9.7. Conclusions

Over the past ten years, rapid technological advances transformed the trading of all financial instruments. In the last few years in particular, HFT has increased significantly and liquidity has experienced substantial improvement. However, correlation is not necessarily causation.

This chapter represents an unusual empirical study of the efficiency of FX markets by investigating the profitability of HFT strategies based on one-minute historical data and discusses its implications on market quality. Whilst the regulatory arguments for and against HFT continue, I believe that my study contributes to the recent debate by providing appropriate empirical evidence. I investigated the efficiency of FX markets by analysing the profitability of technical trading rules at the frequency at which this trading actually takes place in the real world. I developed a real-life intraday technical trading scenario which involved six of the most traded currency pairs and found evidence of HFT predictability and profitability after taking into account appropriate transaction costs. The STGP technique outperformed *ex-ante* traditional econometric forecasting models.

I also examined the impact of HFT on market quality and integrity. The ability to observe all trading activities in my experiments enabled the investigation of the impact of HFT bid and ask orders on market quality. Consistent with Brogaard *et al.* (2013), I have found evidence that HFT enhances the efficiency of prices and play a positive role in the price discovery process by trading in the direction of permanent price changes and in the opposite direction of transitory pricing errors. However, the fact that HFT^{ask} orders are positively associated with pricing errors could be a sign of manipulatory trading strategies, inappropriate risk management practices, or adverse order selection and anticipation. I think that further investigation is needed to examine the exact reason for this positive association due to the short-term nature of one-minute high frequency data. The information content of HFT in my experiment lasts for one minute only, and therefore is unlikely to become public in such a short period of time. Hence, it is still unclear whether the short-term nature of one-minute high frequency data is a good indicator of informational advantage leading to more efficient resource allocation or whether reduced pricing errors constitute better investment decisions.

Due to its recent emergence the HFT discussion is not supported by solid academic research (Chordia *et al.,* 2013). This combined with the uncertainty surrounding HFT, the debate as to whether this particular type of trading is beneficial or harmful to market efficiency is likely to continue long into the future. HFT platforms are not machines with minds of their own. They have been developed by highly skilled humans who profit from advances in technology. Chaboud (2009) provides evidence that computer trades are more highly correlated with each other than human trades, indicating that strategies generated by machines are not as diverse as those developed by humans.

Although there is no definite conclusion as to whether HFT plays a beneficial or harmful role in the market, my empirical findings are one step towards a better understanding of the underlying principles of HFT and its implications on market structure and performance. HFT offers appropriate platforms for making forecasts in just milliseconds to seize arbitrage opportunities and provide new frameworks and challenges for high frequency forecasting. Hence, I think that future research involving non-public data for long-term trading, and even data measured in milliseconds and nanoseconds, is needed. My next research goal is to investigate whether HFTs have different impact on different trader populations by constructing markets composed of different number of artificial traders. Meanwhile, the individual investor will be at an increasing disadvantage, being unable to keep up with the necessary investments in trading technology, and big institutional investors will to continue to benefit from the rise of the machines.

10.0 Conclusions of the thesis

The aim of this thesis is to examine the behaviour of financial markets by using agent-based computational technique named Strongly Typed Genetic Programming (STGP). I applied this adaptive computational-based learning algorithm to real-life historical data of different stocks, indices, and currency pairs to analyse various stylised facts of financial returns, market efficiency and stock market forecasts.

Chapter 4 of this thesis demonstrate that stock market dynamics and nonlinearities are better represented by the AMH rather than the EMH. The presence of enhanced heterogeneity and greater genetic diversity leads to higher market efficiency measured by the AMH. Moreover, individual trader learning, adaptation and evolution reinforced the notion of market efficiency which seems to exist simultaneously with the need for adaptive flexibility. In Chapter 5, I have found more herding behaviour in a group of stocks than in individual stocks. Moreover, the behaviour of the market populated by greater number of trader's exhibit less herding. In the next chapter of my thesis I have found a mixed picture of positive and negative impacts from individual intelligence and market performance. Chapter 7 illustrate the predictive ability of the STGP in comparison with traditional forecasting models. The STGP technique was able to generate statistically and economically significant excess returns after taking into account appropriate transaction costs. Chapter 8 demonstrate that small-cap stocks are more predictable than large-cap stocks and their dominance is not period-specific. The same chapter highlight the very weak causal relationship between trading volume and stock returns. In the last empirical chapter I offer in depth analysis of high frequency trading profitability and market regulation. The STGP technique outperforms the traditional parametric and non-parametric forecasting models. I have also found evidence of positive impact of high frequency trading on price discovery and market dynamics.

In Chapters 4 and 5, I pointed out some of the limitations in my experiments – the difference between human intelligence and artificial trader intelligence and the need of precise knowledge of the demand elasticity of the three financial instruments to measure herding behaviour.

These limitations provided the following future areas of research. Future research could possibly include an investigation of the sensitivity of models to parameters in determining the role of intelligence and market structure. This can be combined with the degree of traders' prudence – the time horizon which traders look back at while they make investment decisions. Social learning can also be added on in order to analyse the adaptive switch between social and individual learning.

I believe further research should include a direct comparison of the predictive abilities of Genetic Programming and Strongly Typed Genetic Programming as well as a comparison of Strongly Typed Genetic Programming versus a combination of forecasts.

Future experiments involving data measured in milliseconds and nanoseconds is needed. My next research goal is to obtain such data and investigate whether high frequency trading has different impact on different trader populations.

References

Ahlstedt, J., Villysson, J. (2012). High Frequency Trading.Working Paper.

Aitken, M.J., Frino, A., McCorry, M.S. Swan, P. 1998. Short sales are almost instantaneously bad news: Evidence from the Australian stock exchange. *The Journal of Finance,*53(6): 2205-2223.

Alam, S.J., Geller, A., Tsvetovat, M. (2012). *Networks in agent-based social simulation*. In A.J. Heppenstall, A.T.Crooksa, L.M.See, M.Batty (eds.), Agent-based models of geographical systems, 199-216. Dordrecht: Springer.

Allen, F., Karjalainen, K. (1999). Using genetic algorithms to find technical trading rules. Working paper. The Wharton School, University of Pennsylvania.

Alon, I., Qi M., Sadowski, R.J. (2001). Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services,*8: 147-156.

Alpert, M., Howard, R. (1982). *A progress report on the training of probability assessors*, In Kahneman, D., Slovic, P., Tversky, A., eds, Judgement Under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press.

Amilon, H. (2005). Estimation of an adaptive stock market model with heterogeneous agents. Sveriges Riskbank Working Paper Series No. 177.

Andreassen, P., Kraus, S. (1988). *Judgemental prediction by extrapolation*. Harvard University Mimeo.

Angel, J., Harris, L., Spatt, C. (2011). Equity trading in the 21st century. *Quarterly Journal of Finance* ,1: 1-53.

Antoniewicz, R. (1992). A casual relation between stock returns and volume (208). Board of Governors of the Federal Reserve System Finance and Economics Discussion Series. August,50.

Aoki, M. (2002). Open models of share markets with two dominant types of participants. *Journal of Economic Behaviour and Organization*, 49: 199-216.

Arifovic, J. (1996). The behaviour of the exchange rate in the genetic algorithm and experimental economies. *Journal of Political Economy,* 104: 510-541.

Arifovic, J., Masson, P. (1999). Heterogeneity and evolution of expectations in a model of currency crisis, Technical report, Simon Fraser University, Vancouver, BC, Canada.

Armstrong, J.S. (1985). *Long-range forecasting*. 2nd edn, New York: Wiley.

Arthur, W.B. (1994). Inductive reasoning and bounded rationality, *American Economic Review*, 84: 406-411.

Arthur, W.B. (2005). *Out-of-equilibrium economics and agent-based modelling*. Handbook of Computational Economics, vol2, Agent-Based Computational Economics, K.Judd and L. Tesfatsion, eds, Elsevier, North-Holland.

Arthur, W.B., Holland, J., LeBaron, B., Palmer, R., Tayler, P. (1997). Asset Pricing Under Endogenous Expectations in an Artificial Stock Market. In: Arthur, W.B., Durlauf, S., Lane, D., (Eds.), *The Economy as an Evolving Complex Systems II.*Addison-Wesley, Reading, MA, 15-44.

Arrow, K.J. (1982). Risk perception in psychology and economics. *Economic Inquiry*, 20: 1-9.

Aslund, A.(1995).How Russia Became a Market Economy, Washington, DC: Brookings Institution.

Athale, C.A., Deisboeck, T.S. (2006). The effects of EGF-receptor density on multi-scale tumour growth patterns. *Journal of Theoretical Biology*, 238(4): 771-779.

Audet, N., Gravelle, T., Yang, J. (2001). Optimal market structure: Does one shoe fit all? Technical report, Bank of Canada, Ottawa, CA.

Australian Securities & Investments Commission. (2013). Dark liquidity and high frequency trading. Report 331.

Avery, C., Zemsky, P. (1998). Multidimensional Uncertainty and Herd Behaviour in Financial Markets.*The American Economic Review*,88:724-748.

Avramov, D. (2002). Stock returns predictability and model uncertainty. *Journal of Financial Economics*, 64: 423-458.

Axelrod, R. (1997). *The complexity of co-operation: Agent-based models of competition and collaboration*. Princeton: Princeton University Press.

Axelrod, R. (2006). *Agent-based modelling as a bridge between disciplines*. In L.Tesfatsion, K.L. Judd (eds.), Handbook of computational economics: Agent-based computational economics, 2: 1583-1598. Amsterdam, North Holland.

Axtell, R. (1999). The emergence of firms in a population of agents: Local increasing returns, unstable nash equilibria, and power law size distributions, Working paper No.3. Washington DC: Center on Social and Economic Dynamics (The Brookings Institute).

Bachelier, L. (1900). *Theorie de la speculation.* Gauthiers-Villars, Paris. Reprinted in English (translated by Boness, A.J.). *In the random character of stock market prices*, Coother, P.H (ed.), 17-87. Cambridge, MA, MIT Press, 1964.

Baek, E., Brock, W.A. (1992). A general test for nonlinear Granger causality: bivariate model, unpublished manuscript, University of Wisconsin, Madison, WI.

Baillie, R.T., Bollerslev, T. (1991). Intra-day and inter-market volatility in foreign exchange rates. *Review of Economic Studies*, 58: 565-585.

Bak, P., Paczuski, M., Shubik, M. (1999). Price variations in a stock market with many agents, Discussion paper 1132. New Haven, Cowles Foundation, (Yale University).

Ball, R., Kothari, S.P. (1989). Nonstationary expected returns: implications for tests of market efficiency and serial correlations in returns. *Journal of Financial Economics*,25:51-74.

Ball, R., Kothari, SP., Shanken, J.(1995). Problems in measuring portfolio performance: an application to contrarian investment strategies. *Journal of Finance*,38:79-107.

Banerjee, A. (1992). A Model of Herd Behaviour. *Quarterly Journal of Economics*, 107(3):724-748.

Bank for International Settlements. (1996). Central Bank Survey of Foreign Exchange and Derivatives Market Activity,Basle, Switzerland.

Bank for International Settlements. (2005). Triennial Central Bank Survey: Foreign Exchange and Derivatives Martket Activity in April 2004. Basel: BIS.

Banz, R.W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*,9: 3-18.

Barnett, W.A, A.R, Gallant, M.J.Hinich., J.A.Jungeilges (1998). "A single-blind controlled competition among tests for nonlinearity and chaos". *Journal of Econometrics*,82:157-192.

Barnes, J. (1982). *Programming in Ada*. Reading, MA: Addison-Wesley.

Becker, G.(1962). Irrational Behaviour and Economic Theory. *Journal of Political Economy,*70(1):1-13.

Becker, G.(1976). The Economic Approach to Human Behaviour, Chicago: University of Chicago Press.

Becker, L.A., Seshadri, M. (2003). GP-evolved technical trading rules can outperform buy and hold. 3rd International Workshop on Computation.

Beckmann, K., Werding, M.(1994). Markets and the use of Knowledge. Thesting the "Hayek Hypothesis"in Experimental Sock Markets. Universitat Passau,94032,Passau,Germany.

Bekiros, S.D. (2010). Heterogeneous trading strategies with adaptive fuzzy Actor-Critic reinforcement learning: A behavioural approach. *Journal of Economic Dynamics & Control*. 34: 1153-1170.

Beltratti, A., Margarita, S. (1992). *Evolution of trading strategies among heterogeneous artificial economic agents*, in J.A. Meyer, H.L. Roitblat, W. Wilson, eds, 'From Animals to Animals 2' MIT Press, Cambridge, MA.

Beltratti, A., Margarita, S., Terna, P. (1996). *Neural networks for economic and financial modelling,* International Thomson Computer Press, London, UK.

Benink, H.A., J.L.Gordillo., J.P.Pardo, C.R.Stephens (2010). "Market efficiency and learning in an artificial stock market: a perspective from Neo-Austrian economics". *Journal of Empirical Finance*,17:668-688.

Benitez,J.M., Castro, J.L., Requena, I. (1997). Are artificial neural networks black boxes*, IEEE Transactions on Neural Networks*, 8: 1156-1164.

Bernard, R.N. (1999). Using adaptive agent-based simulation models to assist planners in policy development: The case of rent control, Working paper 99-07-052. Santa Fe, NM: Santa Fe Institute.

Bessembinder, H., Seguin, P.J. Price volatility, trading volume, and market depth: Evidence from futures markets. *Journal of Financial and Quantitative Analysis,*28(1): 21-38.

Bhattacharyya, S., Pictet, O.V., Zumbach, G. (1998).Representational semantics for Genetic Programming based learning in high frequency financial data. Proceedings of the 3rd Annual Genetic Programming Conference.

Bhattacharyya, S., Pictet, O.V., Zumbach, G. (2002). Knowledge-intensive genetic discovery in foreign exchange markets*. IEEE Transactions on Evolutionary Computation* ,6(2): 169-178.

Biais, B., Foucault, T., Moinas, S. (2013). Equilibrium fast trading. HEC Paris Research Paper No.968/2013, available from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2024360. Accessed on 18/05/2013.

Biais, B., Foucault, T., Moinas, S. (2011). Equilibrium high-frequency trading. Working paper, available at http://ssrn.com/abstract=2024360. Accessed on 02/05/2013.

Bick, B., Kraft, H., Munk, C. (2013). Solving constrained consumption-investment problems by simulation of artificial market strategies. *Management Science*, 59(2): 485-503.

Bikhchandani, S., Hirshleifer. D., Welch, I. (1992) A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, 100(5): 992-1026.

Bikhchandani, S., Sharma, S. (2001). Herd Behaviour in Financial Markets. *IMF Staff Papers*, 47 (3).

Binswanger,M.(1999).*Stock markets, speculative bubbles and economic growth: New dimensions in the co-evolution of real and financial markets*. Edward Elgar.

Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford.

Blanchard, O., Watson,M. (1983) Bubbles, Rational Expectations and Financial Markets.

Blume, L., Easley, D. (1992). Evolution and market behaviour. Journal of Economic Theory, 58: 9-40.

Blume, L., Easly, D.& O'Hara, M. (1994). Market statistics and technical analysis: the role of volume. *Journal of Finance*, 49: 153-181.

Boehmer, E., Fong, K., Wu, J. (2012). International evidence on algorithmic trading. Working paper,EDHEC.

Bogle, J.C. (2000). *Common sense of mutual funds: New imperatives for the intelligent investor*. Wiley. New York.

Bollen, J., Mao, H., Zeng, X-J. (2010). Twitter mood predicts the stock market. *Journal of Computational Science,*44: 26-45.

Bollerslev, T., Chou, RC., Kroner, K.(1992).ARCH Modelling in Finance. *Journal of Econometrics,*52:5-59.

Bossaerts, P., Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: what do we learn? *The Review of Financial Studies*. 12(2): 405-428.

Bottazzi, G., Dosi, G., Rebesco, I. (2003). Institutional architectures and behavioural ecologies in the dynamics of financial markets, Technical report, S. Anna School of Advanced Studies, Pisa, Italy.

Bouchaud, J-P. (2008). Economics needs a scientific revolution, *Nature*, 455: 1181-1190.

Box, G.E.P., Jenkins, G.M. (1970). *Time series analysis, forecasting and control*. Holden-Day, San Francisco.

Brabazon, A., O'Neil, M. (2004). Evolving technical trading rules for spot foreign-exchange markets using grammatical evolution. *Computational management Science*, 1: 311-327.

Brandouy, O., Mathieu, P., Veryzhenko, I. (2013). CCIS 271: 350-364.

Brandouy, O., Mathieu, P., Veryzhenko, I. (2013). Risk aversion impact on investment strategy performance: A multi agent-based analysis. Working paper.

Breen, W., Glosten, L.R., Jagannathan, R. (1990). Predictive variations in stock index returns. *Journal of Finance.* 44: 1177-1189.

Brewer, P., Huang, M., Nelson, B., Plott, C.(2002). On the Behavioural Foundations of the Law of Supply and Demand: Human Convergence and Robot Randomness. *Experimental Economics,*5(3):179-208.

Brock, WD., Dechert,D., Scheinkman,J., LeBaron, B. (1996).A Test for Independence Based on the Correlation Dimension, *Econometric Reviews,*15:197-235.

Brock, W., Hommes, C. (1997). A rational route to randomness, *Econometrica,* 65: 1059-1095.

Brock, W., Hommes, C. (1998). Heterogeneous beliefs and bifurcation routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, 22: 1235-1274.

Brock, W., Lakonishok, J., LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*. 47: 1731-1764.

Brock, W., LeBaron,B. (1996). A dynamical structural model for stock return volatility and trading volume. *Review of Economics and Statistics*,78: 197-235.

Brock, W.A., Hommes,C. (1997a). A rational route to randomness. *Econometrica*,65:1059-1095.

Brock, W.A.,Hommes, C. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing models. *Journal of Economic Dynamics and Control*,22:1235-1274.

Brogaard, J. (2011). High Frequency Trading and Volatility. Working Paper.

Brogaard, J., Hendershott, T., Riordan, R. (2013). High frequency trading and price discovery. Working Paper.

Bullard, J., Dufyy, J. (2001). Learning and excess volatility, *Macroeconomic Dynamics*, 5: 272-302.

Caballero, R.J. (1992). A fallacy of composition. *American Economic Review*, 82(5): 1279-1292.

Campbell, J.Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*. 18: 373-399.

Cao, Q., Leggio, K.B., Schniederjans.(2005). A comparison between Fama and French’s model and artificial neural networks in predicting the Chinese stock market. Computers and Operations Research. 32: 2499-2512.

Caram, L.F., Caiafa, C.F., Proto, A.N., Ausloos, M. (2010). Dynamic peer-to-peer competition. *Physica A: Statistical Mechanics and its Applications*, 389: 2628-2636.

Carrion, A. (2013). Very fast money: high-frequency trading on NASDAQ. *Journal of Financial Markets*, forthcoming.

Cecen, A.A., Erkal, C. (1996). Distinguishing between stochastic and deterministic behaviour in high frequency foreign exchange rate returns: can non-linear dynamic help forecasting? *International Journal of Forecasting*, 12: 465-473.

Cipriani, M., Guarino, A. (2002). *Herding and Price Convergence in a Laboratory Financial Market* [online]. Available from http://scholar.googleusercontent.com/scholar?q=cache:Of46BBIzP3MJ:scholar.google.com/+cipriani+and+guarino+(2002)&hl=en&as_sdt=0,5 Accessed on 03/12/2011.

Chaboud, A., Chiquoine, B., Hjalmarsson, E., Vega, C. (2010). Rise of the machines: algorithmic trading in the foreign exchange market, International Finance Discussion Papers,Board of Governors of the Federal Reserve System.

Chakrabarti, R. (1999). Just another day in the inter-bank foreign exchange market. *Journal of Financial Economics*, 56(1): 29-64.

Chakrabarti, R., Roll, R. (1999). Learning from others, reacting and market quality. *Journal of Financial Markets*, 2: 153-178.

Chakraborti, A., Toke, I.M., Patriarca, M., Abergel, F. Econophysics review: II. Agent-based models. *Quantitative Finance*, 11(7): 1013-1041.

Chakradhara, P., Narasimhan, V. (2006). Predicting stock returns: an experiment of the artificial neural network in Indian stock market. *South Asia Economic Journal*, 7:2.

Challet, D., Zhang, Y.C. (1997). Emergence of cooperation and organization in an evolutionary game, *Physica A,* 246: 153-178.

Chan, KC. (1988). On the contrarian investment strategy. *Journal of Business*,61:147-163.

Chan, K.C., Chen N-F. (1991). Structural and return characteristics of small and large firms. *Journal of Finance*, 46: 1467-1484.

Chan, K.C., Jegadeesh, N., Lakonishok, J. (1996). Momentum strategies. *Journal of Finance*, 51(5): 1681-1714.

Chan, L., Hamao, Y., Lakonishok, J. (1991). Fundamentals of stock returns in Japan. Journal of Finance.46.1739-1764.

Chan, N.T., Shelton, C. (2001). An electronic market-maker, Technical report, MIT, Cambridge, MA.

Chan,N.T.,LeBaron,B.,Lo,A.W.Poggio,T.(1999).Agent-based models of financial markets: A comparison with experimental markets. Working paper.

Chan, W-S., Ng, M. (2004). Robustness of alternative non-linearity tests for SETAR models. *Journal of Forecasting*. 23(3): 215-231.

Chang, P.H.K., Osler, C.L. (1999). Methodical madness: technical analysis and the irrationality of exchange-rate forecasts*. Economic Journal, 109*: 636-661.

Chatfield, C. (1995a). Model uncertainty, data mining and statistical inference (with discussion). *Journal of the Royal Statistical Society*, A, 158: 419-466.

Chatfield, C. (1996). Model uncertainty and forecast accuracy. *Journal of Forecasting*. 15: 495-508.

Charest, G. (1978a). Split information, stock returns and market efficiency-I. *Journal of Financial Economics*, 6: 265-296.

Chen,S-H.(2002). *Genetic algorithms and genetic programming in computational finance*. Kluwer Academic Publishers.

Chen, G., Firth, M., Rui O.M. (2001). The dynamic relation between stock returns, trading volume, and volatility. *The Financial Review*, 38: 153-174.

Chen, A-S., Leung, M.T., Daouk, H.(2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index*. Computers and Operations Research*, 30: 901-923.

Chen,S.H., Liao C.C. (2005). Agent-based computational modeling of the stock price-volume relation. *Information sciences* ,170: 75-100.

Chen, S.H., Lux,T., Marchesi,M. (2000). Testing for non-linear structure in an artificial financial market. *Journal of Economic Behaviour and Economic Organization*, 46(3): 327-342.

Chen, S., Tai, C.(2003). Trading Restrictions, Price Dynamics and Allocative Efficiency in Double Action Markets: Analysis based on Agent-Based Modeling and Simulations. *Advances in Complex Systems*,6(3):283-302.

Chen, S.H.,Yeh, C.H. (1999).On the Consequences of 'Following the Herd': Evidence from the Artificial Stock Market. *International Conference on Artificial Intelligence-IC-AI*, 388-394.

Chen, S.H., Yeh, C.H. (1996). Toward a computable approach to the efficient market hypothesis: an application of genetic programming. *Journal of Economic Dynamics and Control*,21:1043-1063.

Chen, S.H., Yeh, C.H. (2001). Evolving traders and the business school with genetic programming: A new architecture of the agent-based stock market. *Journal of Economic Dynamics and Control*, 25: 363-393.

Chen, S.H., Yeh, C.H. (2002). On the emergent properties of artificial stock markets: the efficient market hypothesis and the rational expectations hypothesis. *Journal of Economic Behaviour and Organization,*49:217-239.

Chen, S.H.,Yeh, C.H., Lee, W.C. (1998). Option pricing with Genetic Programming. *Proceedings of the 3rd Annual Genetic Programming Conference.*

Chen, S.H., Yeh, C.H., Liao C-C. (2002). Testing for Granger causality in the stock price-volume relation: A perspective from the agent-based model of stock markets [online]. Available from http://www.aiecon.org/staff/shc/pdf/icai992.pdf. Accessed on 03/11/2012.

Chen, S., Liao, C. (2005). Agent-Based Computational Modelling of the Stock-Price Volume Relation*. Information Sciences* , 170: 75-100.

Cheng, B., Titterington, D.M. (1994). Neural networks: A review from a statistical perspective. *Statistical Science*, 9(1): 2-54.

Cheung, Y.W., Chinn, M.D. (2001). Currency traders and exchange rate dynamics: a survey of the US market. *Journal of International Money and Finance*, 20: 439-471.

Chiarella, C., He, X. (2002a). An adaptive model on asset pricing and wealth dynamics with heterogeneous trading strategies. Research paper 84, Quantitative Finance Research Group, University of Technology, Sydney.

Chiarella, C.,He, X. (2002b). Heterogeneous beliefs, risk and learning in a simple asset pricing model, *Computational Economics*, 19: 95-132.

Chiarella, C.,He, X. (2003). Dynamic beliefs and learning under a $_L$-process-the heterogeneous case. *Journal of Economic Dynamics and Control*, 27: 503-531.

Chidambaran, N.K., Lee.C.H.J., Trigueros, J.R. (1998). Adaptive evolutionary approach to option pricing via Genetic Programming. *Proceedings of the 3rd Annual Genetic Programming Conference.*

Chidambaran, N.K.Lee,C.H.J.,Trigueros,J.R.(1998).An adaptive evolutionary approach to option pricing via Genetic Programming. Proceedings of the 3rd Annual Genetic Programming Conference.

Chordia, T., Goyal, A., Lehmann, B.N., Saar, G. (2013). High-frequency trading. *Journal of Financial Markets*, forthcoming.

Chordia, T., Roll, R., Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics,* 87: 249-268.

Chordia, T., Roll, R., Subrahmanyam, A. (2011). Recent trends in trading activity and market quality. *Journal of Financial Economics*, 101: 243-263.

Christie, W., Huang, R. (1995).Following the Pied Piper: do Individual Returns Herd Around the Market? *Financial Analysts Journal*, July-August: 31-37.

Clement, R.T. (1989). Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting,* 5: 559-583.

Clements, M.P., Smith, J. (1997). A Monte Carlo study of the forecasting performance of empirical SETAR models. Working paper. University of Warwick, UK.

Cliff, D.,Bruten, J.(1997). Zero is not enough: On the Lower Limit of Agent Intelligence for Continuous Double Action Markets. *HP Technical Report,*HPL-97-141.

Cincotti, S., Focardi, S.M., Marchesi, M., Raberto, M. (2003). Who wins? Study of long-run trader survival in an artificial stock market. *Physica A*, 324: 227-233.

Cipriani, M.,Guarino, A.(2002). Herding and Price Convergence in a Laboratory Financial Market [online]. Available from http://scholar.googleusercontent.com/scholar?q=cache:Of46BBIzP3MJ:scholar.google.com/+cipriani+and+guarino+(2002)&hl=en&as_sdt=0,5 Accessed on 03/01/2012.

Coleman, R.D. (1997). A history of the size effect. Working paper.

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

Conrad, J.S. (1995). The new finance: The case against efficient markets. *Journal of Finance*, 50: 1348-1352.

Conrad, J.S. Kaul, G. (1993). Long-term market overreaction or biases in computed returns? *Journal of Finance*,48:39-63.

Cont, R. (2001). Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance,*1: 223-236.

Cont, R. (2007). Volatility clustering in financial markets: empirical facts and agent-based models. *Long Memory in Economics*:289-309.

Cont, R., Bouchaud. (2000).Herd behaviour and aggregate fluctuations in financial markets. *Macroeconomic Dynamics,*4: 170-196.

Cooke, D., Kugh, D. (2012). Investigating general equilibria in artificial stock markets. MEng Final Year Project.

Corcos, A., Eckmann, J-P., Malaspinas, A., Malevergne, Y. (2002). Imitation and contrarian behaviour: Hyperbolic bubbles, crashes and chaos. *Quantitative Finance*, 2: 264-281.

Corrado, C., Truong, C.(2007). Forecasting stock index volatility: Comparing implied volatility and the intraday high-low price range. *The Journal of Financial Research*, 30(2): 201-215.

Coyle, EJ.(1988). Stack filters and the mean absolute error criterion. *IEEE Transactions on Acoustics, Speech, and Signal Processing*,36(8):1244-1254.

Crooks, A.T., Heppenstall, A. (2012). *Introduction to Agent-Based Modelling*. In Agent-Based Models of Geographical Systems, eds, A.J. Heppenstall et al. Springer Science+ Business Media.

Crouch, R.L. (1970). A nonlinear test of the random walk hypothesis. *American Economic Review,*199-202.

Cucio, R., Goodhart, C., Guillaume, D., Payne, R. (1997). Do technical trading rules generate profits? Conclusiosn from the intra-day foreign exchange market. *International Journal of Finance & Economics,* 2(4): 267-280.

Daigler, R., Wiley, M. (1999). The impact of trader type on the futures volatility-volume relation. *The Journal of Finance*,6:2297-2316.

Davenow, A., Welch, I. (1996).Rational Herding in Financial Economics. *European Economic Review*,40 (3-5): 603-615.

Dawid, H. On the convergence of genetic learning in a double auction market. *Journal of Economic Dynamics and Control*, 23: 1545-1567.

Day, R.,Huang,W. (1990). Bulls, bears, and market sheep. *Journal of Economic Behaviour and Organization*,14: 299-329.

Decamps, JP., Lovo, S. (2002) Risk Aversion and Herd Behaviour in Financial Markets. *Finance and Economics Department,* The University of Toulouse.

De Bondt, W.F.M., Thaler, R.H. (1985). Does the stock market overreact? *Journal of Finance*, 40(3): 793-805.

De Bondt, W.F.M.,Thaler, R.H. (1987). Further evidence on investor overreaction and stock market seasonality. *Journal of Finance*, 42(3): 557-581.

De Fontnouvelle, P. (2000). Information dynamics in financial markets. *Macroeconomic Dynamics*, 4: 139-169.

DeFusco, R.A.,Johnson,R.R.,Zorn,T.S.(1990).The effect of executive stock option plans on stockholders and bondholders. *The Journal of Finance*, 45(2): 617-627.

De Grauwe, P., Dewachter, H., Embrechts, M. (1993). *Exchange rate theory: Chaotic models of foreign exchange markets*, Blackwell, Oxford.

De Lima, P.J.F (1998). Nonlinearities and nonstationarities in stock returns. *Journal of Business and Economic statistics*,16(2): 227-236.

DeLong, JB., Shleifer, A., Summers, L., Waldman, R. (1990) Noise Trader Risk in Financial Markets, *Journal of Political Economy* 98(4): 703-738.

Dempster, M.A.H., Jones, C.M. (2001). A real-time adaptive trading system using genetic programming. *Quantitative Finance*, 1: 397-413.

Dempster, M.A.H., Jones, C.M. (2002). Can channel pattern trading be profitably automated? *The European Journal of Finance,* 8: 275-301.

Diaz, M.A., Alvarez, A. (2007). Forecasting exchange rates using an evolutionary neural network. *Applied Financial Economics Letters,*3: 5-9.

Dickey, D. A.,W. A. Fuller. (1981).Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica* ,49: 1057-1-72.

Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*,13: 253-263.

Dillon, T.S. (1991). Artificial neural network applications to power systems and their relationships to symbolic methods. *Electrical Power & Energy Systems*, 13: 66-72.

Dissanaike, G. (1997). Do stock market investors overreact? *Journal of Business Finance and Accounting*, 42(3): 27-50.

Domowitz, I. (2010). Take heed the lessons from the 1962 flash crash. Available at http://www.advancedtrading.com/exchanges/225700888. Accessed on 03/03/2013.

Doukas, J.A., Kim, C., Pantzalis, C. (2004). Divergent opinions and the performance of value stocks. *Financial Analysts Journal*, 60(6): 55-64.

Dragulescu, A., Yakovenko, V.M. (2000). Statistical mechanics of money. *The European Physical Journal B*, 17: 723-729.

Dragulescu, A., Yakovenko, V.M. (2001). Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A*, 299: 213-221.

Drake,A.E.,Marks,R.E.(2002).*Genetic algorithms in economics and finance: forecasting stock market prices and foreign exchange – a review,* In genetic algorithms and genetic programming in computational finance (Chen, S-H, eds). Kluwer Academic Publishers.

Ehrenstein, G. (2002). Cont-Bouchaud percolation model including tobin tax. *International Journal of Modern Physics C*, 13 (10).

Ehrenstein, G., Westerhoff, F., Stauffer, D. (2005). Tobin tax and market depth. *Quantitative Finance*, 5: 213-218.

Ehrentreich, N. (2006). Technical trading in the Santa Fe Institute artificial stock market revisited. *Journal of Economic Behaviour and Organization*, 61: 599-616.

Ehrentreich, N. (2008). *Lecture notes in Economics and Mathematical Systems*. Springer-Verlag Berlin Heidelberg.

Embrechts, P.,Kluppelberg, C.,T.Mikosch,T. (1997). *Modelling Extremal Events for Insurance and Finance*,Springer-Verlag, Berlin Heidelberg.

Engle, R.F. (2000). The econometrics of ultra-high-frequency data. *Econometrica,* 68(1): 1-22.

Engle,R.W.,Tuholski,S.W.,Laughlin,J.E.,Conway,A.R.(1999).Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*,3: 309-331.

Epps, T.W. (1975). Security price changes and transactions volumes: theory and evidences. *American Economic Review* ,65: 586-597.

Epps,T.W., Epps, M.L. (1976). The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distribution hypothesis. *Econometrica*, 44: 305-321.

EViews.com (2011). *EViews7.0 Feature List* [online]. Available from: http://www.eviews.com/EViews7/ev7features.html. Accessed on 28 July 2011.

European Bank for Reconstruction and Development (EBRD) ,1996. *Transition Report 1996,*London: EBRD.

Fama, E (1963). Mandelbrot and the stable Paretian hypothesis. *Journal of Business,*36:420-429.

Fama, E (1965a). The behaviour of stock market prices. *Journal of Busniess*,38: 34-105.

Fama, E (1965b). Random walks in stock market prices. *Financial Analysts Journal*,21: 55-59.

Fama, E (1970). Efficint capital markets: a review of theory and empirical work. *Journal of Finance*,25: 238-417.

Fama, E.F. (1991). Efficient capital markets: II. *Journal of Finance*. 46: 1575-1618.

Fama,E.F., French, K.R. (1992). The cross-section of expected stock returns. *The Journal of Finance* XLVII(2): 427-465.

Fang, Y. (2003). Forecasting combination and encompassing tests. *International Journal of Forecasting,* 19: 87-94.

Farmer, J.D., Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460: 685-686.

Farmer, J.D. Joshi, S. (2002). The price dynamics of common trading strategies. *Journal of Economic Behaviour and Organization*, 49: 149-171.

Farmer, D.J., Lo, A.W. (1999). Frontiers of finance: Evolution and efficient markets. *Proceedings of the National Academy of Sciences of the United States of America*, 96: 9991-9992.

Federal Reserve. (2012). Selected Interest Rates-H15 [online]. Available from http://www.federalreserve.gov/releases/h15/current/. Accessed on 14/09/2012.

Ferber, J. (1999). *Multi-Agent Systems: An Introduction to Distributed Artificial Intell*igence. Addison-Wesley Longman, Harlow, UK, 509 pp.

Fernandez-Rodriguez, F., Sosvilla-Rivero, S., Andrada-Felix, J. (2003). Technical analysis in foreign exchange markets: evidence from the EMS. *Applied Financial Economics*,13: 113-122.

Foresight, UK. (2012). The UK Foresight Report [online]. Available from: www.biz.gov.uk/foresight. Accessed on 12/01/2013.

Forsythe, R., Nelson, F., Neumann, G., Wright, J. (1992) Anatomy of an Experimental Political Stock Market. *The American Economic Review*,82(5): 1142-1161.

Foster, D.F., Smith, T., Whaley, R.E. (1997). Assessing goodness-of-fit of asset pricing models: the distribution of the maximal R^2. *The Journal of Finance*, 2: 591-607.

Foucault, T., Hombert, J., Rosu, I. (2013). News trading and speed. HEC Paris Research Paper No.975/2013, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2188822. Accessed 01/05/2013.

Frank, M.Z., Gencay,R.,Stengos,T. (1988). "International chaos?" *European Economic Review*,32: 1569-1584.

Frankel, J.A., Froot, K.A. (1986). The dollar as an irrational speculative bubble: A table of fundamentalists and chartists. *Marcus Wallenberg Papers in International Finance*, 1: 27-55.

Frankel, J.A., Froot, K.A. (1986). Understanding the U.S dollar in the eighties: The expectations of chartists and fundamentalists. *Economic Record*, Special Issue, December: 24-38.

Frankel, J.A., Froot, K.A. (1988). *Explaining the demand for dollars: International rates of return and expectations of chartists and fundamentalists*, in R. Chambers and P. Paarlberg, eds, 'Agriculture, Macroeconomics, and the Exchange Rate', Westview Press, Boulder, CO.

Froot, K., Obstfeld, M. (1989).Intrinsic Bubbles: the Case of Stock Prices. *NBER Working Paper*, No. 3091.

Froot, K.,Scharfstein,D.,Stein,J.(1992). Herd on the street: Informational inefficiencies in a market with short-term speculation. *Journal of Finance*, 47:1461-1494.

Fujihara, R.A., Mougoue M. (1997). An examination of linear and nonlinear casual relationships between price variability and volume in the petroleum futures markets. *Journal of Futures Markets*, 17(4): 385-416.

Fung, H-G., Patterson, G.A. 1999. The dynamic relationship of volatility, volume, and market depth in currency futures markets. *Journal of International Financial Markets, Institutions and Money*, 9: 33-59.

Fyfe, C., Marney, J.P., Tarbert, H. (2005). Risk adjusted returns from technical trading: a genetic programming approach. *Applied Financial Economics*, 15: 1073-1077.

Gai, J., Yao, C., Mao, Y. (2012). The externalities of high-frequency trading. Working paper,University of Illinois.

Gallant, A.R., Rossi, P.E. (1992). Stock prices and volume. *The Review of Financial Studies*,5:199-242.

Galai, D. (1978). Empirical tests of boundary conditions for CBOE options. *Journal of Financial Economics*, 6: 187-211.

Garman, M.B. (1976). Market microstructure. *Journal of Financial Economics*, 3: 257-272.

Gaunersdorfer, A., Hommes, C. (2000). A nonlinear structural model for volatility clustering, Working paper 00-04, CeNDEF, University of Amsterdam.

Geanakoplos, J., Axtell, R., Farmer, D.J., Howitt, P., Conlee, B., Goldstein, J., Hendrey, M., Palmer, N.M., Yang, C-Y. (2012). Getting at systemic risk via an agent-based model of the housing market. Cowles Foundation Discussion Paper No. 1852.

Gehrig, T., Menkhoff, L. (2003). Technical analysis in foreign exchange-the workhorse gains further ground. Discussion Paper,University of Hannover.

Gencay, R., Ballocchi, G., Dacorogna, M., Olsen, R. Pictet, O. (2002). Real-time trading models and the statistical properties of foreign exchange rates. *International Economic Review*, 43(2): 463-487.

Gencay, R., Dacorogna, M., Olsen, R. (2003). Foreign exchange trading models and market behaviour. *Journal of Economic Dynamics & Control,*27: 909-935.

Georges, C. Learning with misspecification in an artificial currency market. *Journal of Economic Behaviour and Organization*, 60: 70-84.

Giannitsarou, C. (2003). Heterogeneous learning. *Review of Economic Dynamics,*6**:**885-906.

Gigerenzer, G.,Gaissmaier,P.(2011). Heuristic decision making. *The Annual Review of Psychology.* 62:451-482.

Glosten, L. , Milgrom, P. (1985). Bid, Ask, amd Transaction Prices in a Specialist Market with Heterogeneously Informed Traders, *Journal of Financial Economics* 14: 71-100.

Gode, D.,Sunder, S.(1993). Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality. *Journal of Political Economy,*101:119-137.

Gompers, P., Matrick, A. (2001). Institutional Investors and Equity Prices. 116 *Quarterly Journal of Economics* 229.

Good, N., Schafer, JB., Konstan, JA., Borchers, AI., Sarwar, B., Herlocker, J., Riedl, J. (1999). Combining collaborative filtering with personal agents for better recommendations. *Amaricam Association of Artificial Intelligence Association*,AAAI-99 Proceedings.

Goodhart, C., O'Hara, M. (1997). High frequency data in financial markets: issues and applications. *Journal of Empirical Finance,* 4(2-3): 73-114.

Gomber, P., Arndt, B., Lutat, M., Uhle, T. (2011). High Frequency Trading. Working Paper*.*

Granger, C.W.J. (1969). Investigating casual relations by econometric models and cross-spectral methods. *Econometrica*, 37: 424-438.

Granger, C.W.J. (1980). *Forecasting in business and economics*. Academic Press, Inc., New York.

Granger, C.W.J., Newbold, P. (1972). Economic forecasting-the atheist's viewpoint. Paper presented to Conference on the Modelling of the UK Economy, London Graduate School of Business Studies.

Granger, C.W.T., Newbold, P. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society*, A. 137 (2): 131-165.

Grossman, S., Stiglitz, J. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, 70: 393-415.

Gulyas, L., Balazs, A., Kiss, A. (2004). An early agent-based stock market: replication and participation. Working paper.

Guo, X. (2012). High Frequency Trading. Working Paper.

Hackel, K.S., Livnat. J., Rai A. (1994). The free cash flow/small-cap anomaly. *Financial Analysis Journal,* September-October,33-52.

Hagströmer, B., Nordèn, L., 2013. The diversity of high frequency traders. *Journal of Financial Markets*, forthcoming.

Hansbrouck, J., Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, forthcoming.

Hansen, P.R., Lunde, A., Nason, J.M. (2005). Testing the significance of calendar effects. Federal Reserve of Atlanta Working Paper No. 2005-02.

Haugen, R.A. (1999). *The new finance: the case against efficient markets*, 2^{nd} Ed. Upper Saddle River, NJ: Prentice Hall.

Hendershott, T.C., Jones, C.M., Menkveld, A.J. (2011). Does algorithmic trading improve liquidity? *Journal of Finance*, 66(1): 1-33.

Hendershott, T.C., Menkveld, A.J. (2011). Price pressures. Working Paper.

Hendershott, T.C., Riordan, R. (2011). High frequency trading and price discovery. Working Paper.

Henriksson, R.D., Merton, R.C. (1981). On the market timing and investment performance II: statistical procedures for evaluating forecasting skills. *Journal of Business*, 54: 513-533.

Hansen, P.R. (2005). A test for superior predictive ability. Journal of Business and Economic Statistics ,23: 365-380.

Harris, L. (1983). The joint distribution of speculative prices and daily trading volume. Working paper, University of Southern CA.

Harris, M., Raviv, A. (1993). Differences of opinion make a horse race. *Review of Financial Studies*,6:473-506.

Harris, L., Gurel, E. (1986). Price and volume effects associated with changes in the S&P 500 list: new evidence for the existence of price pressures. *The Journal of Finance*, XLI(4): 815-829.

Hart, M.L., Lamper, D., Johnson, N.F. (2003). An investigation of crash avoidance in a complex system, Technical report, Physics Department, Oxford University, Oxford, UK.

Hartley, J.E. (1997). *The Representative Agent in Macroeconomics*, Routledge, London, New York.

Hayek, F.(1945). The Use of Knowledge in Society (first published in: *American Economic Review,* 35: 519-530, reprinted in: Nishiyama and Leube [1984].211-224.

Hayek, F.(1968). Competition as a Discovery Procedure, reprinted in: Nishiyama and Leube [1984].254-265.

Henry, O.T., McKenzie, M. 2006. The impact of short selling on the price volume relationship: Evidence from Hong Kong. *Journal of Business*,79(2): 671-691.

Hibon, M., Evgeniou, T. (2005). To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting*. 21: 15-24.

Hiemstra,Y.(1996).Applying neural networks and genetic algorithms to tactical asset allocation.*Neuro Vest Journal*, May/June.

Hiemstra, C., Jones, J.D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance*, 49(5): 1639-1664.

Hirshleifer,D., Subrahmanyam,A.,Titman,S.(1994). Security analysis and trading patterms when some investors receive information before others. *Journal of Finance*,49:1665-1698.

Hjorth, U. (1994). *Computer intensive statistical methods- validation model selection and bootstrap*, London: Graham and Hall.

Hommes, C.H (2000). Financial Markets as nonlinear adaptive evolutionary systems. *Quantitative Finance*,1: 149-167.

Hommes, C. (2001). Financial markets as nonlinear adaptive evolutionary systems. *Quantitative Finance*, 1: 149-167.

Holmes, C.C., Adams, N.M. (2002). A probabilistic nearest neighbour method for statistical pattern recognition. *Journal of the Royal Statistical Society*, 64(2): 295-306.

Hommes, C. (2011). The heterogeneous expectations hypothesis: some evidence from the lab. *Journal of Economic Dynamics and Control*, 35: 1-24.

Hong, H., Kubik, J.D., Fishman, T. 2012. Do arbitrageurs amplify economic shocks? *Journal of Financial Economics,*103: 454-470.

Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feedforward networks are universal approximations. *Neural Networks*, 2(5): 359-366.

Horowitz, J.L., Loughran, T., Savin, N.E. (2000). *The disappearing size effect*. Research in Economics, 54: 83-100.

How, J., Ling, M., Verhoeven P. (2010). Does size matter? A genetic programming approach to technical trading. *Quantitative Finance*, 10(2): 131-140.

Hsiao, C.(1981). Autoregressive modelling and money-income causality detection. *Journal of Monetary Economics*,7: 85-106.

Hsieh, D.A (1989). Testing for nonlinear dependencein daily foreign exchange rates. *Journal of Business*,62: 339-368.

Hsu, P-H., Kuan, C-M. (2005). Reexamining the profitability of technical analysis with data snooping checks. *Journal of Financial Econometrics*, 3(4): 606-628.

Hsu, P-H., Hsu, Y-C.,Kuan C-M. (2010). Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. *Journal of Empirical Finance ,*17: 471-484.

Huang, W., Nakamori, Y., Wang, S-Y. (2005). Forecasting stock market movement direction with support vector machine. Computers and Operations Research. 32: 2513-2522.

Huang, Y-P. Chen, S-H. Hung, M-C., Yu, T. (2012). Liquidity cost of market orders in the Taiwan Stock Market: A study based on an order-driven agent-based artificial stock market. *International Review of Financial Analysis*, 23: 72-80.

Huang, Z.F., Solomon, S. (2000). Power, levy, exponential and Gaussian-like regimes in autocatalytic financial systems. *European Physical Journal B*, 20: 601-607.

Huang, Z.F., Solomon, S. (2001). Finite market size as a source of extreme wealth inequality and market instability. *Physica A*, 294: 503-513.

Huang, Z.F., Solomon, S. (2002). Stochastic multiplicative processes for financial markets. *Physica A*, 306: 412-422.

Hudson, R., Atanasova, C. (2009). Equity Returns at the end of the month: Further Confirmation and Insights. *Financial Analysts Journal,*65(4): 14-16.

Hurst, H.E (1951). Long-term storage of reservoirs: an experimental study. *Transactions of the American Society of Civil Engineers,*116: 770-799.

Hyndman, R. J., Koehler, A.B. (2006). (2006). Another look at measures of forecasting accuracy. *International Journal of Forecasting*, 22(4): 679-688.

Iba, H., Nikolaev, H. (2000). Genetic programming polynomial models of financial data series. *In Proceedings of CEC,* IEEE Press: 1459-1466.

Iba, H., Sasaki, T. (1999). Using Genetic Programming to predict financial data. Evolutionary Computation. CEC99. *Proceedings of 1999 IEEE.*

Ilachinski, A. (1997). Irreducible semi-autonomous adaptive combat (ISAAC): An artificial-life approach to land combat. Alexandria: Center for Naval Analyses.

Iori, G. (2002). A microsimulation of traders activities in the stock market: The role of heterogeneity agents' interactions and trade frictions. *Journal of Economic Behaviour and Organization*, 49: 269-285.

Jacobs, B.I. (1999). *Capital ideas and market realities: option replication, investor behaviour, and stock market crashes*. Malden, MA: Blackwell.

Jacobs, B.I. (2004). Risk avoidance and market fragility. *Financial Analysts Journal*, 60(1): 26-30.

Jacobs, B.I. (2009). Tumbling tower of babel: Subprime securitization and the credit crisis. *Financial Analysts Journal*, 65(2): 17-30.

Jacobs, B.I., Levy, K.N., Markowitz, H.M. (2010). Simulating security markets in dynamic and equilibrium modes. *Financial Analysts Journal*, 66(5): 42-67.

Jaffe, J., Westerfield, R. (1985). Patterns in Japanese common stock returns: Day of the week and turn of the year effects. *Journal of Financial and Quantitative Analysis*. 20: 261-272.

Jegadeesh, N.,Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48: 65-91.

Jegadeesh, N., Titman, S. (2001). Profitability of momentum strategies: An evolution of alternative explanations. *Journal of Finance*, 56: 699-720.

Jensen, M.C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, 6: 95-101.

Johnson, R., Chittenden, W.T., Jensen, G.R. (1999). Residential politics, stocks, bonds, bills, and inflation. The Journal of Portfolio Management, 26(1): 27-31.

Johnson, N.F. Lamper, D., Jefferies, P., Hart, M.L., Howison, S. (2001). Application of multi-agent games to the prediction of financial time-series. *Physica A*, 299: 222-227.

Jones, C.M. (2002). *A century of stock market liquidity and trading costs*. Graduate School of Business Columbia University. SSRN313681.

Jones, C.M. (2013). What do we know about high-frequency trading? Columbia Business School Research Paper No.13-11. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2236201. Accessed on 01/05/2013.

Joshi, S., Parker, J., Bedau, M.A. (2002). Financial markets can be at sub-optimal equilibria. *Computational Economics*, 19: 5-23.

Jovanovic, B., Menkveld, A.J. (2012). Middleman in limit-order markets. Working Paper,available at http://ssrn/com/abstract=1624329. Accessed on 10/05/2013.

Kaboudan, M.A. (1999). A measure of time series' predictability using genetic programming applied to stock returns. *Journal of Forecasting*,18: 345-357.

Kaboudan, M.A. (2000). Genetic programming prediction of stock prices. *Computational Economics*, 16: 207-236.

Kaboudan, M.A. (2001). Genetically evolved models and normality of their fitted residuals. *Journal of Economic Dynamics and Control*,25(11): 1719-1749.

Kahneman, D., Tversky, A. (1982). Intuitive prediction: biases and corrective procedures, Reprinted in Kahneman et al., *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kaizoji, T.,Bornholdt, S.,Fujiwara,Y. (2002). Dynamics of price and trading volume in a spin model of stock markets with heterogeneous agents. *Physica A: Statistical Mechanics and its Applications*,316(1-4): 441-452.

Kaplan, D.T (1994). Exceptional events as evidence for determinism. *Physica D*,73: 38-48.

Kareken, J., Wallace, N. (1981). On the indeterminacy of equilibrium exchange rates, *Quarterly Journal of Economics*, 96: 207-222.

Kato, K., Schallheim, J. (1985). Seasonal and size anomalies in the Japanese stock market. *Journal of Financial and Quantitative Analysis,* 20: 243-260.

Kato, K., Ziemba, W., Schwartz, S. (1990*). Day of the week effects in the Japanese stocks*. In: Elton, E., & Grubber, M. (Eds.), Japanese capital markets. Harper and Row, New York.

Kearns, M., Kulesza, A., Nevmyvaka, Y. (2010). Empirical limitations on high frequency trading profitability. Working Paper.

Keim, D.B., Stambaugh, R.F. (1986). Predicting returns in stock and bond markets. *Journal of Financial Economics,* 17: 357-39.

Keleher, M.A. (2010). *Small versus large-will small cap stocks continue to outperform?* BNY Mellon Beta Management.

Khan, S.U., Rizwan F.(2008). Trading volume and stock returns: evidence from Pakistan's stock market. *International Review of Business Research Papers*,4: 151-162.

Khosravi, A., Nahavandi, S., Creighton, D., Atiya, A.F. (2010). A comprehensive review of neural network-based prediction intervals and new advances. *IEEE Transactions on Neural Networks*, 1-15.

Kim, Y.S. (2009). Exchange rates and fundamentals under adaptive learning. *Journal of Economics Dynamics and Control*, 33: 843-863.

Kim, G., Markowitz, H.M. (1989). Investment rules, margin and market volatility. *Journal of Portfolio Management*, 16: 45-52.

Kim, T-H., White, H. (2004). Pne more robust estimation of skewness and kurtosis. *Finance Research Letters,*1:56-73.

Kimoto, T., Asakawa, K. (1990). Stock market prediction system with modular neural networks. In: *Proceedings of the international joint conference on neural networks*, 1-6.

Kimura, H., Lima, F.G., Perera, L.C.J., Kerr, R.B. (2012). Stock prices in an artificial stock market with optimistic and pessimistic agents. Working paper.

Kirilenko, A., Kyle, A.S., Samadi, M., Tuzun, T. (2011). The flash crash: the impact of high frequency trading on an electronic market. Working Paper*,* University of Marylend.

Kirman, A. (1991). Epidemics of opinion and speculative bubbles in financial markets. Chapter 17 in: Taylor, M.P. ed., *Money and Financial Markets*, Blackwell: Cambridge.

Kirman, A.P. (1992). Whom or what does the representative individual represent? *Journal of Economic Perspectives*, 6(2): 117-136.

Kirman, A. (1993). Ants, rationality and recruitment. *Quarterly Journal of Economics*, 108: 137-156.

Kirman, A.P., Teyssiere, G. (2001). Microeconomic models for long-memory in the volatility of financial time series. *Studies in Nonlinear Dynamics and Econometrics*, 5: 281-302.

Klein,R.W.,Brown,S.J.(1984).Model selection when there is minimal prior information.Econometrica,52:1291-1312.

Kollman, K., Miller, J.H., Page, S.E. (1992). Adaptive parties in spatial elections. *American Political Science Review*, 86(4): 929-937.

Korczak, J.J., Lipinski, P. (2004). Evolutionary buiding of stock trading experts in a real-time system. *Proceedings, 2004 Congress on Evolutionary Computation: 940-947.*

Koza, J. (1992).*Genetic Programming: On the programming of computers by natural selection*. MIT Press.Cambridge,MA,USA.

Koza. J.R. (1995). *Genetic programming for economic modelling*. In S.Goonatilaje and P. Treleaven (eds). Intelligent Systems for Finance and Business, London: John Wiley & Sons, 251-269.

Ladley, D.(2004). Zero Intelligence in Economics and Finance. *The Knowledge Engineering Review,*27:273-286.

Lakonishok, J.& Shapiro, A.C. (1986). Systematic risk, total risk and size as determinants of stock market returns. *Journal of Banking and Finance*,10: 115-132.

Lakonishok, J., Shleifer, A., Vishny, R. (1992).The Impact of Institutional Trading on Stock Prices. *Journal of Financial Economics*, 32: 23-43.

Lakonishok, J., Shleifer, A.,Vishny R.W. (1994). Contrarian investments, extrapolation, and risk. *Journal of Finance*, 49: 1541-1578.

Lam, E. (2010). Low-risk TSX stocks have outearned riskiest peers over 30-year period: Analyst [online]. Available from http://business.financialpost.com/2010/06/22/low-risk-tsx-stocks-have-outearned-riskiest-peers-over-30-year-period-analyst/. Accessed on 03/12/2012.

Langevoort, D.C. (1992). Theories, assumptions, and securities regulation: market efficiency revisited. *University of Pennsylvania Law Review*, 67: 127-138.

LaPorta, R., Lakonishok, J., Shleifer, A., Vishny, R. (1997). Good news for value stocks: further evidence of market efficiency. *Journal of Finance*, 50: 1715-1742.

Lawrenz, C.,Westerhoff F. (2003). Modeling exchange rate behaviour with a genetic algorithm. *Computational Economics*,21: 209-229.

Lawson, T. (1997). *Economics and Reality*. Rouledge, London, New York.

LeBaron, B., Arthur, W.B., Palmer, R. (1999). Time Series Properties of an Artificial Stock Market. *Journal of Economic Dynamics and Control* 23: 1487-1516.

LeBaron, B. (2000). Agent-based computational finance: suggested readings and early research. *Journal of Economic Dynamics and Control*, 24 (5-7): 679-702.

LeBaron, B. (2000). Empirical regularities from interacting long and short memory investors in an agent-based stock market. *IEEE Transactions on Evolutionary Computation*, 5: 442-455.

LeBaron, B. (2001). Evolution and time horizons in an agent-based stock market. *Macroeconomic Dynamics*, 5: 225-254.

LeBaron,B.(2001).A biulder's guide to agent-based financial markets. *Quantitative Finance*, 1: 254-261.

LeBaron, B. (2001a). Empirical regularities from interacting long and short memory investors in an agent-based stock market. *IEEE Transactions on Evolution Computation*, 5: 442-455.

LeBaron, B. (2002). Calibrating an agent-based financial market, Technical report, International Business School, Brandeis University, Waltham, MA>

LeBaron, B.(2004). Agent-Based Financial Markets: Matching Stylezed Facts with Style. *Post Walrasian Macroeconomics: Beyond the DSGE Model,* Edited by D.Colander, Cambridge University Press, forthcoming 2006.

LeBaron, B. (2004a). Agent-based computational finance. Working paper, Brandeis University.

LeBaron,B.(2005).*Building the 'Santa Fe artificial stock market'*. In: Luna,F.,Perrone,A(Eds.),Agent-based theory, languages, and experiments. Routledge Publishing.

LeBaron, B.(2006). Agent-Based Computational Finance. In L.Tesfatsion, & K.L.Judd (Eds.), *Handbook of Computational Economics*,(2):1187-1233.

LeBaron, B. (2012). A real Minsky moment in an artificial stock market. Working paper, International Business School, Brandeis University.

LeBaron, B., Yamamoto, R. (2007). Long-memory in an order-driven market. *Physica A: Statistical mechanics and its Applications*,383:85-89.

Lee, B-S., Rui, O. (2002). The dynamic relationship between stock returns and trading volume: Domestic and cross-country evidence. *Journal of Banking and Finance*,26:51-78.

Lee, C-C., Lee, J-D., Lee, C-C. (2010). Stock prices and the efficient market hypothesis: evidence from a panel stationary test with structural breaks. *Japan and the World Economy*,22:49-58.

Lee, C.I., Mathur, I. (1996b). A comprehensive look at the efficiency of technical trading rules applied to cross-rates. *European Journal of Finance*, 2: 389-411.

Lee, C.I., Pan,M.S., Liu, Y.A. (2001). On market efficiency of Asian foreign exchange rates: evidence from a joint variance ratio test and technical trading rules. *Journal of International Financial Markets,* 11: 199-214.

Lettau, M. (1997). Explaining the facts with adaptive agents: the case of mutual fund flows. *Journal of Economic Dynamics and Control*, 21: 1117-1147.

Levich, R., Thomas, L. (1993). The significance of technical trading rule profits in the foreign exchange market: a bootstrap approach. *Journal of International Money and Finance*,12: 451-474.

Levy, M., Levy, H., Solomon, S. (1994). A microscopic model of the stock market: Cycles, booms, and crashes. *Economics Letters*, 45: 103-111.

Leung, M.T.,Daouk, H., Chen, A-S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting*, 16: 173-190.

Li, J. (2001). FGP: A genetic programming based tool for financial forecasting. PhD thesis, University of Essex, UK.

Lindley, D.V.(1957).A statistical paradox.*Biometrica,*44:187-192.

Lilien, G.L., Kotler, P. (1983). *Marketing decision making: a model building approach*. Harper and Row Publishers, New York.

Lillo, F.,Farmer,J.D. (2004). The long memory of the efficient market. *Studies in Nonlinear Dynamics and Econometrics*,8(3): 1-20.

Lim, K-P. (2007). Ranking market efficiency for stock markets: A non linear perspective. *Physica A,* 376: 445-454.

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 market hypothesis. Journal of Investment Consulting, 7(2): 21-44.

Lo, A.W., MacKinlay,A.C. (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies*,3: 175-205.

Lucas, J., Robert, E. (1976). *Econometric Policy Evaluation: A Critique*. In K.Brunner and A.H. Meltzer, eds, The Phillips curve and labour markets, volume 1 of Carnegie-Rochester Conference Series on Public Policy, 19-46, North Holland.

Lustick, I. (2002). PS-L: A user-friendly agent-based modelling platform for testing theories of political identity and political stability. *Journal of Artificial Societies and Social Simulation*, 5(3).

Lux, T., Marchesi, M. (2000). Scaling and sriticality in a stochastic multi-agent model of a financial market. *Nature*, 397: 498-500.

Lux, T., Marchesi, M. (2000). Volatility clustering in financial markets: A microsimulation of interacting agents. *International Journal of Theoretical and Applied Finance*, 3(4): 675-702.

Lux, T., Schornstein. (2005). Genetic learning as an explanation of stylized facts of foreign exchange markets. *Journal of Mathematical Economics*, 41: 169-196.

Lux, T., Westerhoff, F. (2009). Economics crisis. *Nature Physics*, 5: 2-3.

Macintosh, N.B., Shearer, T., Thornton, D.B., Welker, M. (2000). Accounting as simulacrum and hyper-reality: perspectives on income and capital. *Accounting Organizations and Society*, 25(1): 13-50.

Mahfoud, S., Mani, G. (1997). Financial forecasting using genetic algorithms. *Journal of Applied Artificial Intelligence*, 10(6): 543-565.

Makridakis, S., Winkler, R.L. (1989). Sampling distribution of post-sample forecasting errors. *Applied Statistics*,38: 331-342.

Malik, N. (2005). Artificial neural networks and their applications. *National Conference on Unearthing Technological Developments & Their Transfer for Serving Masses*, Mathura, India.

Malkiel, B.G (1987). *Efficient market hypothesis,* in J.Eatwell, M.Milgate and P.Newman (eds). The New Palgrave: Finance,Norton,127-134.

Malkiel, B.G. (2003). The efficient market hypothesis and its critics. *Journal of Economic perspectives*, 17: 59-82.

Managers Insight. (2007). Large cap vs. Small cap- a historical comparison of two distinct classes[online]. Available from http://www.investmentnews.com/assets/docs/CI48311613.PDF Accessed on 01/12/2012.

Mandelbrot, B. (1963).The Variation of Certain Speculative Prices. *Journal of Business,*36*:* 392-417.

Mandelbrot, B.(1997). *Fractals and Scaling in Finance.* Berlin: Springer.

Mannaro, K., Marchesi, M., Setzu, A. (2006). Using an artificial financial market for assessing the impact of Tobin-like transaction taxes. *Journal of Economic Behaviour and Organization*, forthcoming.

Maillet, B., Michel, T. (2000). Further insights of the puzzle of technical analysis profitability. *European Journal of Finance*, 6: 196-224.

Marcucci, J. (2005). Forecasting stock market volatility with regime-switching GARCH models. *Studies in Non-linear Dynamics & Econometrics*, 9(4): 1-42.

Marney,J.P.Tarbert,H.,Fyfe,C.(2000).Technical trading versus market efficiency – a genetic programming approach. Computing in Economics and Finance, Society for Computational Economics (paper #169).

Marney,J.P.,Fyfe,C.,Tarbert,H.,Miller,D.(2001).Risk adjusted returns to technical trading rules: a genetic programming approach. Computing in Economics and Finance, Society for Computational Economics (paper #147).

Markose, S., Tsang, E.P.K., Martinez, S. (2003). The red queen principle and the emergence of efficient markets: An agent-based approach. *Proceedings, 8th Workshop on Economic and Heterogeneous Interacting Agents (WEHIA)*. Kiel, Germany.

Marshall, A. (1890). *Principles of Economics*. Macmillan & Co, Ltd, London.

Martin, A.D. (2001). Technical trading rules in the spot foreign exchange markets of developing countries. *Journal of Multinational Financial Management*, 11: 59-68.

Martinez, C.H., Rosu, I. (2013). High frequency traders, news and volatility. Working Paper*,* available at http://ssrn.com/abstract=1859265. Accessed on 11/05/2013.

Mathieu, P., Secq, Y. (2012). Simulating artificial stock markets with efficiency. Working paper.

Mathieu, P., Brandouy, O. (2012). Introducing ATOM. Working paper.

Mathur, I., Lee, C.I., Gleason, K.C. (2001). Trading rule profits in Latin American currency spot rates. *International Review of Financial Analysis*, 10: 135-156.

Matthews, R.B., Gilbert, N.G. Roach, A., Polhill, J.G., Gotts, N.M. (2007). Agent-based land-use models: A review of applications. *Landscape Ecology*, 22: 1447-1459.

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

Meese, R.A., Rose, A. (1991). An empirical assessment of non-linearities in models of exchange rate determination. *Review of Economic Studies.* 58: 603-619.

Menkveld, A.J. (2013). High frequency trading and the new-market makers. *Journal of Financial Markets*, forthcoming.

Menkveld, A.J. (2011). High frequency trading and the new-market makers. Working Paper.

Menkveld, A.J., Koopman, S.J., Lucas, A. (2007). Modeling around-the-clock price discovery for cross-listed stocks using state space methods. *The Journal of Business and Economics Statistics*,25(2): 213-225.

Mian, G.M., Teo, T.G.L. (2003). Do errors in expectations explain the cross-section of stock returns? *Pacific-Basin Finance Journal*,12: 197-217.

Minsky, H.P. (1986). *Stabilizing an Unstable Economy*. Yale University Press.

Mitchell, T.M. (1997). *Machine Learning*. WCB/McGraw-Hill, New York.

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

Montana, D.J. (2002). Strongly Typed Genetic Programming. Working paper.

Moody, J., Lizhong, W. (1997). What is the true price-state space models for high frequency FX data. Engineering (CIFEr).

Muth, J.F. (1961). Rational expectations and the theory of price movements. *Econometrica*, 29(3): 315-335.

Nagel, K., Rasmussen, S. (1994). *Traffic at the edge of Chaos*. In R.Brooks (ed.), Artificial life IV: 222-236. Cambridge, MA: MIT Press.

Nason, J.A., Diebold, F.X. (1990). Nonparametric exchange rate prediction? *Journal of International Economics,* 28: 315-332.

Narendranath, U.A., Thurakam, D., Parthasarathy, K. (1997). An ANN based approach for voltage stability assessment. *International Conference of Computer Applications in Electrical Engineering, Recent Advances*, 666-670.

Neely, C.J., Weller, P.A. (1999). Technical trading rules in the European monetary system. *Journal of International Money and Finance,* 18 (3): 429-458.

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., Ulrich, J.M. (2009). The adaptive market hypothesis: Evidence from the foreign exchange market. *Journal of Financial and Quantitative Analysis*, 44 (2): 467-488.

Nichols, N.A. (1993). Efficient? Chaotic? What's the new finance? *Harvard Business Review*, 71: 50-56.

Noah, S. (2011) Individual Trader Behaviour in Experimental Asset Markets, *The Journal of Electronic Publishing,* The University of Michigan: 1-69.

Noe, T., Rebello, M., Wang, J. (2003). Corporate financing: An artificial agent-based analysis. *Journal of Finance*, 63: 943-973.

North, M.J., Macal., C.M., Vos, J.R. (2004). Terrorist organization modelling. *North American Association for Computational Social and Organizational Science Conference 2004*, Pittsburgh.

Oberlechner, T.(2001). The Importance of Technical and Fundamental Analysis in the European Foreign Exchange Market. *International Journal of Finance & Economics*,6(1):81-93.

O'Hara, M., Ye, M. (2011). Is market fragmentation harming market quality? *Journal of Financial Economics,* 100: 459-474.

Oharazawa, N. (2010). Global Markets Exploration. Comparing small cap stocks to large cap stocks in developed markets[online]. Available from http://www.russell.com/indexes/documents/research/global-markets-exploration-tailored-view-asian-markets-russell-custom-indexes.pdf . Accessed on 02/12/2012.

Opdyke, J.D. (2010). Small caps loom large [online]. Available from http://online.wsj.com/article/SB10001424052748703572504575214722949936724.html. Accessed on 26/11/2012.

Osler, C. (2000). *Support for resistance: technical analysis and intraday exchange rates*. FRBNY Economic Policy Review.

Othman, A. (2008). Zero-intelligence agents in prediction markets [online]. Available from http://www.cs.cmu.edu/~aothman/zipred.pdf . Accessed on 12/11/2012.

Oussaidene, M., Chapard, B., Pictet, O.V., Tomassini, M. (1997). Parallel genetic programming and its application to trading model induction. *Parallel Computing*, 23: 1183-1198.

Pagan, A. (1996).The Econometrics of Financial Markets. *Journal of Empirical Finance,*3*:* 15-102.

Pai, P-F., Lin, C-S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *The International Journal of management Science*, 33: 497-505.

Panayi, E., Harman, M., Wetherilt. (2013). Agent-based modelling of stock markets using existing order book data. Multi-agent-based Simulation XIII, Working paper.

Park, C.H., Irwin, S.H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys,* 21(4): 786-826.

Parunak, H., Savit, R., Riolo, R.L. (1998). *Agent-based modelling vs. equation-based modelling: A case study and users' guide*. In MABS, Sichman, C. and Gilbert (eds.) Multi-Agent Systems and Agent-Based Simulation. Berlin: Springer-Verlag.

Persaud, A. (2000) Sending the Herd Off the Cliff Edge: The Disturbing Interaction Between Herding and Market-Sensitive Risk Management Practices. *The Journal of Risk Finance*, 59-65.

Pesaran, M.H., Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics,*10: 461-465.

Peters, E.E (1991). A chaotic attractors for the S&P 500. *Financial Analysis Journal*,March-April:55-62.

Petersen, M. (2009). Estimating standard errors in finance panel data sets: comparing approaches. *Review of Financial Studies*, 22: 435-480.

Picard, R.R., Cook, R.D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, 79: 575-583.

Pisedtasalasai ,A.,Gunasekarage ,A. (2007). Casual and dynamic relationships among stock returns, return volatility and trading volume: Evidence from emerging markets in South-East Asia. Asia-Pacific Financial Markets**,**14: 277-297.

Poggio,T.,Lo,A.W.,LeBaron,B.,Chan,N.T.(2001).Agent-based models of financial markets: A comparison with experimental markets. Sloan Working Paper 4195-01.

Ponta, L., Pastore, S., Cincotti, S. (2011). Information-based multi-assets artificial stcok market with heterogeneous agents. *Nonlinear Analysis: Real World Applications*, 12: 1235:1242.

Ponta, L., Raberto, M., Cincotti, S. (2011). A multi-asset artificial stock market with zero-intelligence traders. *A Letters Journal Exploring the Frontiers of Physics*, 93: 202-208.

Potters, M., Cont, R., Bouchaud, J-P. (2008). Financial markets as adaptive ecosystems [online]. Available from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.8786 Accessed on 18/08/2013.

Potvin, J.Y., Soriano, P., Vallee M. (2004). Generating trading rules on the stock markets with genetic programming. *Computers and Operations Research***,**31: 1033-1047.

Qi, M., Wu Y. (2006). Technical trading-rule profitability, data snooping, and reality check: evidence from foreign exchange market. *Journal of Money, Credit and Banking*,38: 2135-2158.

Qian, B., Rasheed, K. (2007). Stock market prediction with multiple classifiers. *Applied Intelligence*, 26: 25-33.

Raberto, M., Cincotti, S. (2005). Modeling and simulation of a double auction artificial financial market. *Physica A*, 355: 34-45.

Raberto, M., Cincotti, S., Focardi, S., Marchesi, M. (2001). Agent-based simulation of a financial market. *Physica A: Statistical Mechanics and its Applications,*299:319-327.

Ready, M.J.(2002). Profits from technical trading rules. *Financial Management*, 31: 43-61.

Reid, D.J. (1969). A comparative study of prediction techniques on economic data. PhD thesis. University of Nottingham.

Reid, D.J. (1971). A comparison of the forecasting techniques on economic time series. Paper given at conference organised by the Society for Long-Range Planning and Forecasting Study Group of the Operational Research Society.

Reinganum, M. (1981). Misspecification of Capital Asset Pricing: Empirical anomalies based on earnings' yields and market values'. *Journal of Financial Economics*,9: 19-46.

Reinganum, M. (1983). Portfolio strategies based on market capitalization. *Journal of Portfolio Management*, 9: 29-36.

Reitz, S., Taylor, M.P. (2006). The coordination channel of foreign exchange intervention: a nonlinear microstructural analysis. Discussion Paper, Series 1:Economic Studies*,* No. 08/2006.

Richards, A.J. (1997). Winner-loser reversals in national stock market indices: can they be explained? *Journal of Finance*,52:2129-2144.

Rieck, C. (1994). *Evolutionary simulation of asset trading strategies,* in E. Hillebrand and J. Stender, eds, 'Many-Agent Simulation and Artificial Life, IOS Press.

Riordan, R., Storkenmaier, A. (2012). Latency, liquidity and price discovery. Journal of Financial Markets 15, 416-437.

Rogalski ,R.J. (1978). The dependence of prices and volume. *Review of Economics and Statistics* ,60(2): 268-274.

Roll,R.(1977). A critique of the asset pricing theory's test Part I: On past and potential testability of the theory. Journal of Financial Economics, 4(2): 129-176.

Routledge, B.R. (2001).Adaptive learning in financial markets. *Review of Financial Studies*, 12: 1165-1202.

Routledge, B.R. (2001). Genetic algorithm learning to choose and use information. *Macroeconomic Dynamics*, 5: 303-325.

Rubinstein, M. Rational markets: Yes or No? The affirmative case. *Financial Analysts Journal*. 57(3): 15-29.

Rutledge, D.J.S. (1984). *Trading volume and price variability: New evidence on the price effects of speculation*. In Peck AE (ed.), Selected Writings on Futures Markets: Research Directions in Commodity Markets*.* (237-251). Chicago: Chicago Board of Trade.

Rutkauskas, A.V., Ramanauskas, T. (2013). Building an artificial stock market populated by reinforcement-learning agents. *Journal of Business Economics and Management*, 34: 11-21.

Saacke, P. (2002). Technical analysis and the effectiveness of central bank intervention. *Journal of International Money and Finance,* 21: 459-479.

Saatcioglu. K., Starks, L.T. (1998). The stock price-volume relationship in emerging stock markets: the case of Latin America*. International Journal of Forecasting ,*14: 215-225.

Samanidou, E., Zschischang, E., Stauffer, D., Lux, T. (2007). Agent-based models of financial markets. Working paper.

Samuelson, P (1965). *Foundations of Economic Analysis*. Cambridge, MA: Harvard University Press.

Sargent, T.J. (1979). *Macroeconomic Theory*. Academic Press, New York.

Sargent, T. (1993). *Bounded rationality in macroeconomics*. Clarendon Press, Oxford.

Savit, R. (1988). When random is not random: an introduction to chaos in market prices. *Journal of Futures Markets,*8: 271-290.

Savit, R. (1989). Nonlinearities and chaotic effects in option prices. *Journal of Futures Markets*,9: 507-518.

Savit, R., Manuca, R., Riolo, R. (1999). Adaptive competition, market efficiency, and phase transitions, *Physical Review Letters*, 82: 2203-2206.

SEC, 2010g. Securities and Exchange Commission, Circuit Breakers and Other Market Volatility Procedures. http://www.sec.gov/answers/circuit.htm .Accessed 15/02/2013.

Scharfstein, D.,Stein, J. (1990)Herd Behaviour and Investment. *American Economic Review,*

80: 465-79.

Scheinkman, J.A., LeBaron, B. (1989). Nonlinear dynamics and stock returns. *Journal of Business,*3: 311-337.

Schleifer, A. (2000). *Inefficient markets: An introduction to behavioural finance*. Oxford, England: Oxford University Press.

Schwartz, L.A., Norris, K.D. 2012. The impact of temporary short selling restrictions on the volatility of financial stock prices: Does firm size matter? Working paper. Wingate University.

Seamon,J.G.,Kenrick,D.T.(1994).*Psychology*.Englewood Cliffs,NJ:Prentice Hall.

Shanken, J., Smith, C.W. (1996). Implications of capital markets research for corporate finance. *Financial Management*, 25: 98-104.

Shefrin, H. (2000). *Beyond greed and fear: Understanding behavioural finance and the psychology of investing*. Harvard Business School Press, Boston.

Shiller, R.J. (2000). From efficient markets theory to behavioural finance. *Journal of Economic Perspectives*, 17: 83-104.

Shiller,R.J.,Pound.J.(1989).Survey evidence on diffusion of interest and information among investors. *Journal of Economic Behaviour and Organization*,10: 47-66.

Shynkevich, A. (2012). Performance of technical analysis in growth and small cap segments of the US equity market*. Journal of Banking and Finance*, 36: 193-208.

Siegel, J.J. (2002). *Stocks for the long run*.Third edition. McGraw Hill. New York.

Silvapulle, P., Choi J-S. (1999). Testing for linear and nonlinear Granger causality in the stock price-volume relation: Korean evidence*. The Quarterly Review of Economics and Finance*,39(1): 59-76.

Simon, H.(1973). The Structure of Ill Structured Problems. *Artificial Intelligence,*4:181-201.

Shalen, C.T. (1993). Volume, volatility and the dispersion of beliefs. *Review of Financial Studies*,6:405-434.

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

Smirlock, M., Starks L. (1988). An empirical analysis of the stock price-volume relationship. *Journal of Banking and Finance*,12: 31-41.

Smith, N. (2011) Individual Trader Behaviour in Experimental Asset Markets, *The Journal of Electronic Publishing,* The University of Michigan: 1-69.

Smith,V.L.(1982).Microeconomic systems as an experimental science. American Economic Review, 72: 923-955.

Sosvilla-Rivero, S., Andrada-Felix, J., Fernandez-Rodriguez, F. (2002). Further evidence on technical trade profitability and foreign exchange intervention. *Applied Economics Letters,* 9: 827-832.

S&P Capital IQ Financial Communications. (2012). Good things may come in small packages: small-cap stocks[online]. Available from [http://www.axa-equitable.com/investments/small-cap-stocks.html. Accessed on 28/11/2012](http://www.axa-equitable.com/investments/small-cap-stocks.html).

Steele, G. (1984). *Common Lisp*. Burlington, MA: Digital Press.

Stigler, G.J. (1964). Public regulation of the securities market. *Journal of Business*, 37: 117-142.

Subrahmanyam, A. (2007). Behavioural finance: A review and synthesis. *European Financial Management*, 34: 78-90.

Sullivan, R., Timmermann, A., White. H. (1997). Data-snooping, technical trading rule performance, and the bootstrap, working paper, University of California, San Diego.

Sunder, S.(2006). Economic Theory:Structural Abstraction or Behaviour Reduction? *History of Political Economy,* 38.Duke University Press.

Sunder, S.(2006). Determinants of Economic Interaction: Behaviour or Structure. *Journal of Economic Interaction and Coordination*. Springer-Verlag.

Sunder, S.(2007). Exploring Market Structures with Zero-Intelligence Agents. Eight Trento Summer School. Intensive course in agent-based finance.

Switzer, L.N. (2010). The behaviour of small cap vs. Large cap stocks in recessions and recoveries: Empirical evidence from the United States and Canada. *North American Journal of Economics and Finance*,21: 332-346.

Switzer, L.N., Fan, H. (2007). Spanning tests for replicable small-cap indexes as separate asset classes: Evidence from G-7 countries. *Journal of Portfolio Management*, 33(4): 102-110.

Switzer ,L.N., Tang, M.(2009). The impact of corporate governance on the performance of US small-cap firms. *International Journal of Business*, 14(4): 343-357.

Szpiro, G.G. (1997). The emergence of risk aversion. *Complexity*, 2: 31-39.

Takahashi, T., Tamada, R. & Nagasaka, K. (1998). Multiple line-segments regression for stock prices and long-range forecasting system by neural networks. In: *Proceedings of the 37th SICE annual conference,* 1127-1132.

Tauchen, G.E., Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica*,51(2):485-505.

Tay, N.S.P. (2013). *Agent-based models of financial markets*. In Encyclopedia of Finance, eds, C-F. Lee and A.C. Lee, Springer Science+ Business Media, New York.

Taylor, M., Allen, H. (1992). The use of technical analysis in the foreign exchange markets. *The Journal of International Money and Finance,* 11: 304-314.

Tesfatsion, L. (2006). Agent-based computational economics: A constructive approach to economic theory. In L.Tesfatsion, K.L. Judd (eds.), Handbook of computational economics: Agent-based computational economics, 2: 831-880. Amsterdam, North Holland.

Tesfatsion, L., Judd, K.L. (2006). *Handbooks in Economic Series*. North-Holland.

Tesfatsion, L. (2003).Agent-based computational economics: modelling economies as complex adaptive systems. *Information Sciences*, 149: 263-269.

Thaler, R.H. (1993). *Advances in behavioural finance*. New York, NY: Russell Sage Foundation.

Thielemann, U. (2000). A brief theory of the market- ethically focused. *International Journal of Social Economics*, 27(1): 6-31.

Thomas, J., Sycara, K. (1999). The importance of simplicity and validation in genetic programming for data mining in financial data. AAAI Technical Report WS-99-06.

Thompson, R. (1978). The information contents of discounts and premiums on closed-end fund shares. Journal of Financial Economics, 6: 151-186.

Thompson, S. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics*, 99: 1-10.

Timmermann, A., Granger, C.W.J. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting,*20: 15-27.

Tsang, E.P.K., Li,J., Butler, J.M. (1998). EDDIE beats the bookies. *International Journal of Software, Practice & Experience*, 28(10): 1033-1043.

Tsang, E.P.K., Martinez-Jaramillo, S. (2004). Computational Finance. *IEEE Computational Intelligence Society, 8-13.*

Tversky, A., Kahneman, D. (1982). *Evidential impact of base rates*, in Kahneman, D., Slovic, P., Tversky, A., ads, Judgement Under Uncertainty: Heuristics and Business. Cabmridge: Cambridge University Press.

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.

Varian, H.R. (1987). *Differences of opinion in financial markets*, In Financial Risk: Theory, Evidence, and Implications. Peoceedings of the 11th Annual Economic Policy Conference of the Federal Reserve Bank of St. Loius, 3-37.

Wallace,T.D. (1972).Weaker criteria and tests for linear restrictions in regression. Econometrica, 40:689-698.

Wallis, K. (1980). Econometric implications of the Rational Expectations Hypothesis. *Econometrics*, 48(1): 49-73.

Wang, J. (1994). A model of competitive stock trading volume. *Journal of Political Economy*,102: 127-168.

Wang, H. 2004. Dynamic volume-volatility relation. Working paper. School of Economics and Finance, the University of Hong Kong.

Weigand, R.A, Belden, S., Zwirlein, T.J. (2004). Stock selection based on mutual fund holdings: Evidence from large-cap funds. *Financial Services Review*, 13: 139-150.

Welch, I. (1992).Sequential Sales, Learning, and Cascades. *Journal of Finance*, 47: 695-732.

Wermers, R. (1999). Mutual Fund Herding and the Impact on Stock Prices. *Journal of Finance* 58(2): 581-622.

Wescrfietd, R. (1977). The distribution of common stock price changes: An application of transactions time and subordinated stochastic models. *Journal of Financial and Quantitative Analysis* ,12: 743-765.

Westerhoff, F. Heterogeneous traders and the Tobin tax. *Journal of Evolutionary Economics*, 13: 53-70.

Westerhoff, F. (2003c). Speculative markets and the effectiveness of price limits. *Journal of Economic Dynamics and Control*, 28: 493-508.

White, H. (2000). A reality check for data snooping. *Econometrica,* 68: 1097-1126.

Wiener, N. (1956). *The theory of prediction*, in: EF Beckenbach (ed.) Modern Mathematics for the engineers, Series 1, McGraw-Hill, New York (Chapter 8).

Willey, T (1992). Testing for nonlinear dependence in daily stock indices. *Journal of Economics and Business*,44(1): 63-74.

Wilson, J.H., Keating, B. (1990). *Business forecasting*. Richard Irwin, Inc, Homewood, IL.

Winker, P., Gilli, M. (2001). Indirect estimation of the parameters of agent-based models of financial markets. Technical report 38, FAME.

Witkam, J. (2011) Altreva Adaptive Modeller*, User's Guide* [online]. Available from http://altreva.com/Adaptive_Modeler_Users_Guide.htm. Accessed on 25 July, 2011.

Witkam, J., 2013. Altreva Adaptive Modeler, User's Guide [online]. Available from http://altreva.com/Adaptive_Modeler_Users_Guide.htm. Accessed on 25 January, 2013.

Yang, H-J., Sun, G-P. (2013). Study on the stability of an artificial stock option market based on bidirectional conduction. *Entropy*, 15: 700-720.

Yeh, C. H. (2003). Tick size and market performance, Technical report, Yuan Ze University, Chungli, Taiwan.

Yeh, C. (2007). The Role of Intelligence in Time Series Properties. *Computational Econimics*,30: 95-123.

Yeh, C.H. (2008). The Effects of Intelligence on Price Discovery and Market Efficiency. *Journal of Economic Behaviour & Organization,*68: 613-625.

Yeh, C.H.,Yang, C. (2010). Examining the Effectiveness of Price Limits in an Artificial Stock Market. *Journal of Economic Dynamics and Control*, 34: 2089-2108.

Ying, C.C. (1966). Stock market prices and volumes of sales. *Econometrica,*34: 676-686.

Yooon, Y., Kozaki, M. (1992). Predicting stock price performance: a neural network approach. In: *Proceedings of the international joint conference on system sciences*, 156-162.

Youki, K., Mardyla, G., Takenaka, S., Tsutsui, Y., Tsusui, Y. (2013). Disposition effect and loss aversion: An analysis based on a simulated experimental stock market. Working paper.

Youssefmir, M., Huberman, B.A. (1997). Clustered volatility in multiagent dynamics. *Journal of Economics Behaviour and Organization,* 32: 101-118.

Zarowin, P. 1990. Size, seasonality, and stock market overreaction. *Journal of Financial and Quantitative Analysis*,25: 113-125.

Zhang, F. (2010). High-frequency trading, stock volatility, and price discovery. Working Paper.

Zhang, M., Ciesielski, V.(1999). Genetic Programming for multiple class object detection. *Advanced Topics in Artificial Intelligence*, Lecture notes in computes science, 1747:180-192.

Zhou, A. (2004). *Enhanced emerging market stock selection*. In Genetic Programming Theory and Practice I.

Zschischang,E.,Lux,T.(2001).Some new results on the Levy, Levy, and Solomon microscopic stock market model. *Physica A*, 291: 563-573.

Zumbach, G., Pictet, O.V., Masutti, O. (2002). Genetic programming with syntactic restrictions applied to financial volatility forecasting. Olsen & Associates Working Paper No.GOZ.2000-07-28.

Appendix

Content published in advance of the thesis submission on 8th of October 2013.

Herd behaviour experimental testing in laboratory artificial stock market settings. Behavioural foundations of stylized facts of financial returns. *Physica A: Statistical Mechanics and its Applications,* 392: 4352-4372 **(Chapter 5)**

The implications of trader cognitive abilities on stock market properties. *Intelligent Systems in Accounting, Finance and Management*. Accepted-forthcoming (in press). **Chapter 6.**