Machines and Markets: Assessing the Impact of Algorithmic Trading on Financial Market Efficiency

EC331: Research in Applied Economics

Department of Economics, University of Warwick

ID: 1828708

April 2021

Word Count: 4991

Abstract

The rise of machine learning has revolutionised finance. Institutions across the world have increasingly turned to data science and machine learning to create trading models without the need for human intervention. This has had various implications for the financial markets that they operate in, including market efficiency. This paper simulates a financial market with agent-based modelling and Monte-Carlo style simulations, to motivate a qualitative discussion about the implications of increased algorithmic trading on financial market efficiency. It finds that algorithmic traders (ATs) can seemingly increase market efficiency through better liquidity management and more complete extraction of information from prices. However, this also comes with increased instability and potential convergence to an unstable equilibrium. The Adaptive Market Hypothesis (Lo, 2004) is suggested as an alternative framework for analysing AT behaviour.

Acknowledgements

I would like to thank my academic supervisor, Professor Jonathan Cave, for his insightful comments and the invaluable conversations we have had. I would also like to thank Tomás Mella Pickersgill and Max Woolterton for their assistance with Python debugging and Wian Stipp for our productive discussions regarding recurrent neural networks. The project has been done in Python, with TensorFlow and Keras used for the neural networks; Numpy and Pandas used for data manipulation and Matplotlib used for graphs. All errors are my own.

Contents

1 Introduction

In 1965, Eugene Fama and Paul Samuelson independently arrived at a conclusion that would go on to become one of the most foundational, yet widely-debated theories in financial economics: The Efficient Markets Hypothesis (EMH). Its premise is simple, past prices and information cannot be used to generate a trading strategy that will consistently outperform the market (Fama, 1965; Samuelson, 1965). Although they both agreed on this, they presented different mechanisms for this phenomenon, with Fama suggesting that prices converge towards the "true value" of the underlying asset, and Samuelson arguing that the randomness of asset prices is itself a phenomenon, and is independent of the fundamental value of the asset. Although Fama's definition of efficient markets is generally more accepted (particularly due to his seminal 1970 empirical analysis), both definitions are distinct and shall be referred to as the "Fama-EMH" and "Samuelson-EMH" respectively for the purpose of this paper.

The idea of an efficient market is foundational to many theories in financial economics. The Capital Asset Pricing Model (CAPM) is just one example of an extremely widely used financial theory that relies on markets being efficient, with over 74% of CFOs still using it regularly (Harvey & Graham, 2001). As a result, it is imperative to understand to what extent markets are truly efficient, and whether or not our understanding needs to be updated. A perfectly efficient market would imply that the entire industry of asset management is essentially worthless – a rather alarming notion to clients who have placed over $85 trillion in the hands of asset managers (PricewaterhouseCoopers, 2020). Furthermore, in a more general economic sense, highly inefficient markets mean a sub-optimal allocation of resources, an outcome that could limit productivity and economic growth.

Over the last few decades, data science and machine learning techniques have become increasingly common among financial institutions, with some choosing to exclusively use quantitative and statistical methods to perform their trades. In the foreign exchange market from 2006 to 2019, the proportion of trades executed by algorithmic traders (ATs) increased from 25% to 92% (Kissell, 2020), demonstrating their recent proliferation.

Assessing the impact of these ATs on market efficiency has been difficult. As they are mostly proprietary, finding accurate, primary-level datasets is difficult without insider privileges. While it is still possible to assess various factors which contribute to market efficiency (which has been done fairly robustly in previous literature) such as volatility and liquidity, casual analysis of ATs and market efficiency is sparse in literature – a gap which this paper aims to fill. This paper employs agent-based modelling to simulate a financial market, where ATs are sequentially added to the market over successive "generations". This produces a dataset with visible, incremental impacts of increasing the proportion of ATs over time. It should be noted that these simulations are not quantitatively robust – a market containing multiple algorithms interacting can lead to various types of emergent behaviour (of which this paper presents one), and so the results are used qualitatively to motivate a discussion about the potential impact of ATs on market efficiency.

This paper finds that it is possible for ATs to present the illusion of increased Fama-EMH efficiency, as they learn to expect prices around the "true price" of the asset, yet have no underlying incentive to maintain this price. This expectation comes from the fact that initially during the early generations, arbitrageurs (who ensured price did not deviate from the true price) dominated. Resultantly, the role of regulators changes. They should aim to ensure that ATs converge to a desired strategy (as we see in this market) and then remain vigilant to mitigate the potential effects arising from increased instability. The ATs do, however, ensure that the mechanisms for maintaining efficiency (namely, liquidity management) remain smooth, which is in line with results from previous literature. The Adaptive Market Hypothesis (Lo, 2004) is also cited as an alternative to the EMH for better explaining the emergent behaviour of ATs.

The rest of this paper is structured as follows: section 2 summarises previous literature in this field; section 3 describes the model, simulation set-up and metrics used for assessing efficiency; section 4 analyses the results of the simulation; section 5 discusses these results and their implications and finally, section 6 concludes.

2 Literature Review

There is an expansive body of data science literature detailing various model specifications for ATs, all claiming to have generated supernormal returns using just past prices (indicating markets are weak-form inefficient). While the results of these papers, such as Dash & Dash (2016); Enke & Thawornwong (2005); Patel, et al. (2015) and Schumaker & Chen (2009) are impressive, they often fail to generalise beyond the markets their models are trained on (Harper, 2016). Similarly, econometric literature has seen mixed results regarding market efficiency tests (Showalter & Gropp, 2019). It is important to note that these results do not speak to the effect of ATs on market efficiency but can still be informative, as they give an account of the development of these algorithms over time.

Due to the difficulty in directly measuring market efficiency, the literature focuses on assessing mechanisms through which efficiency may improve or deteriorate. Liquidity is often used as a key metric – a liquid market ensures that traders are able to meet their desired position sizes without affecting the trading price too significantly. This would allow the market to better adjust to new information, thereby ensuring it remains efficient. Indeed, the majority of literature claims that an increase in AT activity causes an increase in liquidity (Hendershott et. al., 2011; Menkveld, 2011; Payne & Friederich, 2011). The bid-ask spread is also often used to assess efficiency. A low spread indicates high levels of competition and reduced costs for traders, which ensures they are able to act on their private information and therefore contribute to the efficiency mechanism. Castura, et al. (2010) find that spread decreases correlate with increased AT activity, a result that is also confirmed by other papers such as Hendershott & Riordan (2009). Castura et. al. were also first to attempt a direct assessment of the change in market efficiency with increased ATs through the variance ratio test (Lo & MacKinlay, 1988) and find that there is a causal improvement in efficiency.

Overall, the literature is in favour of ATs increasing efficiency, through increased liquidity, reduced spread, better price efficiency and discovery (Brogaard et al., 2013). However, the impact of ATs on volatility remains an area of uncertainty (Zhang, 2010). Nilsson & van der Hoorn (2012) find a positive relationship between AT activity and volatility. They suggest the common belief that high-frequency traders decrease activity in volatile markets should be reconsidered. This view is opposed, however, by a larger body of literature suggesting that volatility decreases with increased AT activity (Cliff, et al., 2011; Jovanovic & Menkveld, 2010).

Linton & Mahmoodzadeh (2018) summarise the key impacts of ATs on market efficiency (discussed above) with one additional conclusion: the increased efficiency also comes with heightened instability. This results in an increasingly important role for regulators to maintain stability where possible and remain vigilant in the presence of normalisation of deviance. Linton & Mahmoodzadeh also call for further work to mitigate the issues surrounding the lack of reliable data in this field and better describe how a macroprudential regulator might approach this new phenomenon. This paper attempts to address their call to action and in fact takes their conclusions about instability one step further. It will suggest that the increased efficiency may be merely an illusion of an evolutionary algorithmic niche, masquerading as a Fama-EMH efficient market.

3 Methodology

The marketplace for the simulations contains one asset with a “true price”, which evolves linearly over time, with exogenous random shocks.[1] To begin with, the market contains four “flavours” of traders: arbitrageurs, chartists, value traders and noise traders. These four flavours of traders roughly represent four types of identifiable investors in markets, where arbitrageurs are fundamental analysts, chartists are

[1]For a complete breakdown of the details of the simulation, please see:
https://github.com/karansgarg/rae_repo/blob/main/market_simulation.ipynb

technical analysts, value traders are value investors and noise traders are unsophisticated retail traders. Arbitrageurs and chartists are modelled in a more sophisticated manner. They act to maximise their prospect-utility from each trade, whereas the other two flavours have more simple dynamics that inform their trades. This can be seen in Figure 3.1.[2] The four flavours of traders are allowed to mutate over time, as often happens in real markets, where dominated strategies are dropped for more favourable ones as time goes on. A key part of the EMH is the mechanism through which efficient prices are reached. This occurs through a collection of investors acting on their own private information, which is then reflected in the updated trading price. This continues until all available information is "priced in" by the market. To replicate this mechanism of private information, many parameters that drive trading decisions are unique to each individual, not just flavours.

A very simple market mechanism is used. The model uses discrete timesteps; in each timestep, the bids and asks are collated and ordered by the best price (bids from highest to lowest and asks from lowest to highest). They are then divided into orders of unit volume and iteratively matched, descending down the order book until a pair of orders is reached where a match is not possible (the ask price exceeds the bid price). The most recent match is used as the trading price for that period, and all matches are executed at that price (a simple average is used when the prices do not match perfectly). This means that placing more aggressive orders can ensure that you get matched (especially important during times of low volume), however you do not gain/lose any additional surplus beyond the market price.

In this setup, a trader's bid (ask) represents the maximum (minimum) they would be willing to pay (accept) for a transaction. This results in a negative spread as the matching bid price is always at least as large as the matching ask price. While this may be somewhat unrealistic, the spread can still be used as a useful metric – a lower (more negative) spread can still imply greater liquidity as more matches are able to take place between the greatest bid and lowest ask.

A simple long-short term memory (LSTM) network is used to model the algorithmic traders. While it is not as sophisticated as many contemporary alternatives, it has been used extensively in timeseries predictions as it does not suffer from the "vanishing gradient" problem[3] that plagues many other recurrent neural networks (Hochreiter & Schmidhuber, 1997). Figure 3.2 shows the architecture of the complete model. The data goes through a logarithmic transformation and is then normalised for training the models, as is standard practice with price timeseries. A 20% dropout layer[4] before the final output is also added to prevent model overfitting. If the predicted price is above

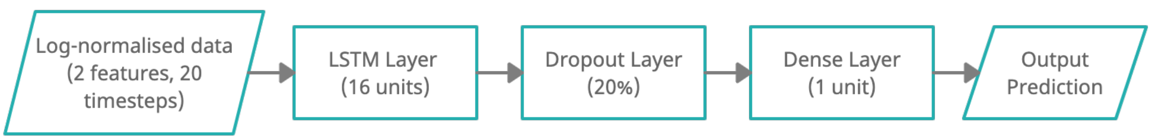
[2]Variable names and functions can be found in Appendix A.
[3]The vanishing gradient problem is one where traditional recurrent neural networks are unable to "remember" information from many periods ago.
[4]A dropout layer is one where random connections in the model are severed between training examples to avoid the model relying on any one "path" of nodes to make the majority of its predictions.

(below) the current price, a bid (ask) is placed, with the order price equal to the predicted price and the quantity dependent on how confident the AT is.

Figure 3.1: Table Summarising the Trading Mechanism for each Flavour[5]

Flavour	**Arbitrageur**	**Chartist**	**Value Trader**	**Noise Trader**
Buy Condition	$P_{t-1} < (E_{t-1} + \varepsilon^i)$	$M_t^s > M_t^l$	$A^i > A_t$ $P^{max} > P_{t-1}$	$A^i > A_t$ $\sim N(0,1) > 0$
Utility-based Trade Generation and Execution (Bid)	$x_{t,g}^i = (E_{t-1} + \varepsilon^i) - P_{t-1}$ $x_{t,b}^i = P_{t-1} - P^{min}$ $p_{t,g}^i = \frac{(E_{t-1} + \varepsilon^i) - P_{t-1}}{(E_{t-1} + \varepsilon^i) - P^{min}}$ $p_{t,b}^i = 1 - p_{t,g}^i$ $U_t^i = V(x_{t,g}^i w(p_{t,g}^i) + x_{t,b}^i w(p_{t,b}^i)$ If $U_t^i > L^i$, place bid	$x_{t,g}^i = P^{max} - P_{t-1}$ $x_{t,b}^i = M_t^l - P_{t-1}$ $p_{t,g}^i = \sigma(100({}_{t-b_i}R_t))$ $p_{t,b}^i = 1 - p_{t,g}^i$ $U_t^i = V(x_{t,g}^i w(p_{t,g}^i) + x_{t,b}^i w(p_{t,b}^i)$ If $U_t^i > L^i$, place bid		
Bid Price	$P_{t-1} + f^i$	$P_{t-1} + f^i$	$P_{t-1} + f^i$	$P_{t-1} + f^i$
Bid Quantity	$U_t^i(S^i)$	$U_t^i(S^i)$	$\lvert\sim N(3,1)\rvert$	$\lvert\sim N(3,1)\rvert$
Sell Condition	$P_{t-1} > (E_{t-1} + \varepsilon^i)$	$M_t^s < M_t^l$	$A^i > A_t$ $P^{max} \leq P_{t-1}$	$A^i > A_t$ $\sim N(0,1) < 0$
Utility-based Trade Generation and Execution (Ask)	$x_{t,g}^i = P_{t-1} - (E_{t-1} + \varepsilon^i)$ $x_{t,b}^i = P_{t-1} - P^{max}$ $p_{t,g}^i = \frac{P_{t-1} - (E_{t-1} + \varepsilon^i)}{(E_{t-1} + \varepsilon^i) - P^{max}}$ $p_{t,b}^i = 1 - p_{t,g}^i$ $U_t^i = V(x_{t,g}^i w(p_{t,g}^i) + x_{t,b}^i w(p_{t,b}^i)$ If $U_t^i > L^i$, place ask	$x_{t,g}^i = P_{t-1} - P^{min}$ $x_{t,b}^i = P_{t-1} - M_t^l$ $p_{t,g}^i = \sigma(-100({}_{t-b_i}R_t))$ $p_{t,b}^i = 1 - p_{t,g}^i$ $U_t^i = V(x_{t,g}^i w(p_{t,g}^i) + x_{t,b}^i w(p_{t,b}^i)$ If $U_t^i > L^i$, place ask		
Ask Price	$P_{t-1} - f^i$	$P_{t-1} - f^i$	$P_{t-1} - f^i$	$P_{t-1} - f^i$
Ask Quantity	$U_t^i(S^i)$	$U_t^i(S^i)$	$\frac{I^i}{2}$	$\lvert\sim N(3,1)\rvert$

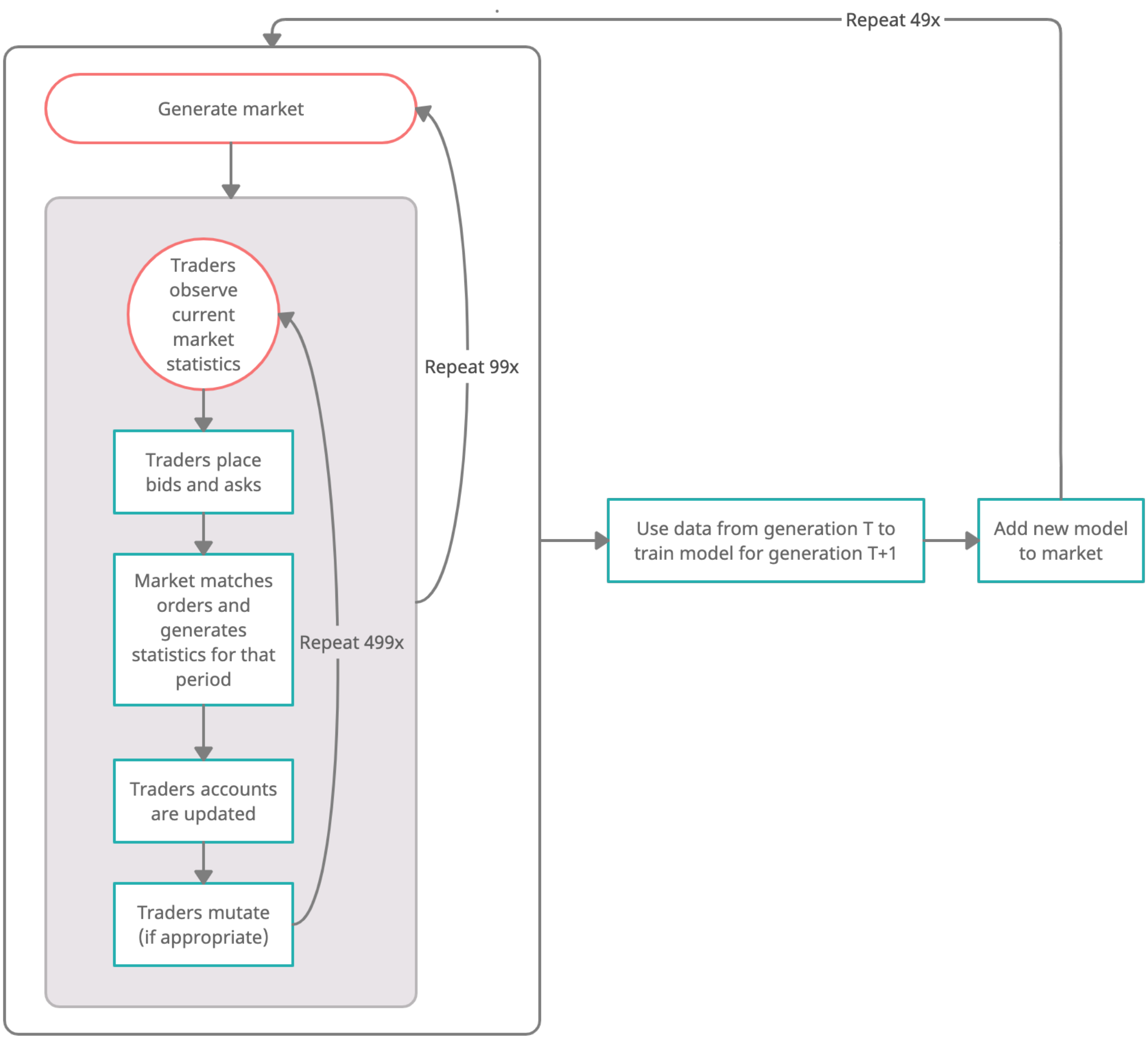
Figure 3.2: Flowchart Showing Model Structure for Algorithmic Traders



[5]The asset is not sub-divisible – any non-integer bid/ask quantities are rounded appropriately before being submitted to the market and all prices and rounded to the nearest 2 d.p.

The model begins with 100 “human” traders and no ATs in the market. This market then runs for 500 periods completing one “run”. The initial distribution of flavours among the hundred traders is determined randomly. This process is then repeated 100 times, to ensure a complete distribution of all the random variables within the model, including exogenous shocks. These 100 runs then form a “generation”. This generation forms the training dataset for the next generation’s AT, which after being trained is added to the market, beginning the next generation. Through this iterative process, 50 ATs are added, completing 51 generations of data.

Figure 3.3: Flowchart of Simulation



Once the simulations are complete, various aspects of market efficiency are analysed. Liquidity, volatility, direct measures of efficiency, as well as analysis of the ATs’ model structure are considered. Traditional econometric tests are also conducted, however the price timeseries suffer from high levels of endogeneity so these results are not reported (although can be found in Appendix B).

4 Results[6]

4.1 Returns and Earnings

The main claim of the EMH, that investors cannot consistently make above-market returns, is something that is directly testable through this simulation. Figure 4.1 shows the average return per generation of the top performing trader, and Figure 4.2 shows the proportion of traders each generation that earned above-market returns. From the combination of these two we can clearly see as generations progress, it becomes increasingly difficult to beat the market, suggesting an increase in market efficiency.

Figure 4.1: Graph Showing the Average Excess Returns of the Top Performing Trader in each Run

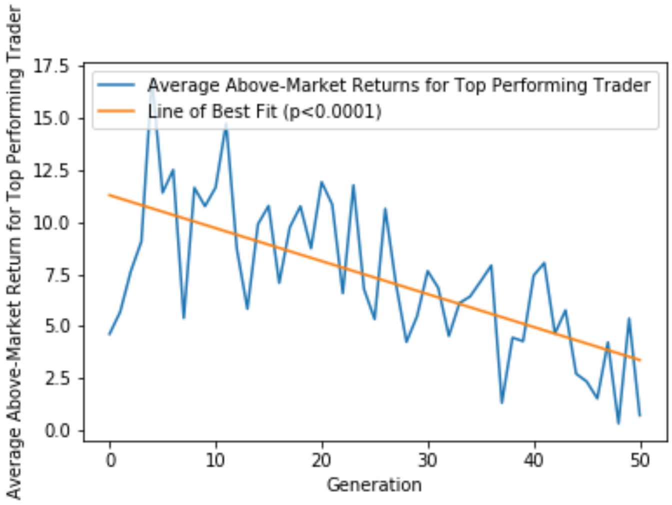
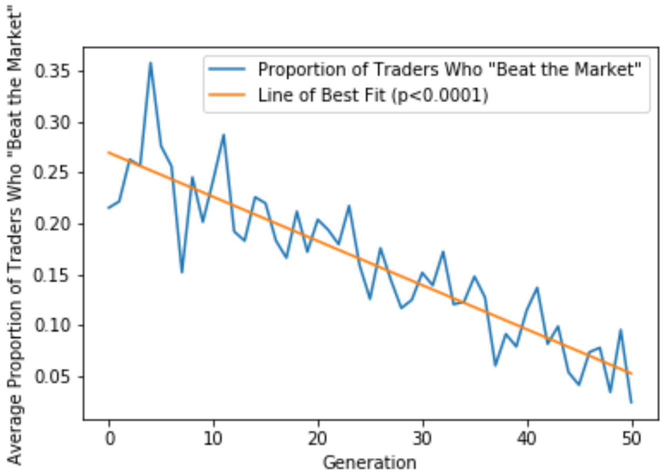


Figure 4.2: Proportion of Traders Who Beat the Market



6 In each of the following sections, graph(s) pertinent to the results are shown. In each graph, a line of best fit is drawn, with reported p-values indicating with what significance the null hypothesis of a t-test was rejected that the slope of the line of best fit is equal to 0. In cases where the line of best fit is polynomial (of order two or above), the p-values correspond to the order of coefficient (e.g. p2 refers to the quadratic term). As the data should be taken qualitatively, only the direction of the trend is discussed. An explanation of how the line of best fit was derived can be found in Appendix B.

Figure 4.3 shows the number of runs in each generation dominated by a particular flavour (the mutative nature of the traders means that more often than not the population converges to a particular strategy). We can see that the proportion of time the value traders dominate remains fairly stable, whereas the arbitrageurs seem to capture some of the time previously held by chartists. This indicates that over the generations, it becomes more profitable to be an arbitrageur (suggesting that price reverts back to the "true price" more often) resulting in an increase in Fama-EMH efficiency. This is reinforced by Figure 4.4, where it can be seen that chartists no longer feature as the top earner after 10 generations. Concurrently, ATs begin to dominate around this time too, suggesting this is the time it took to converge to the strategy they found most profitable.

Figure 4.3: Graph Showing the Number of Runs Dominated by Each Flavour

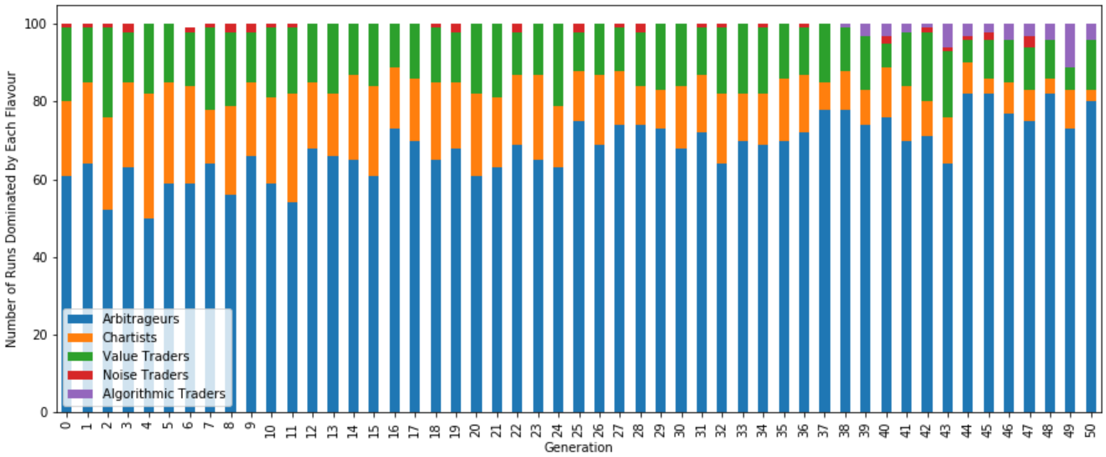
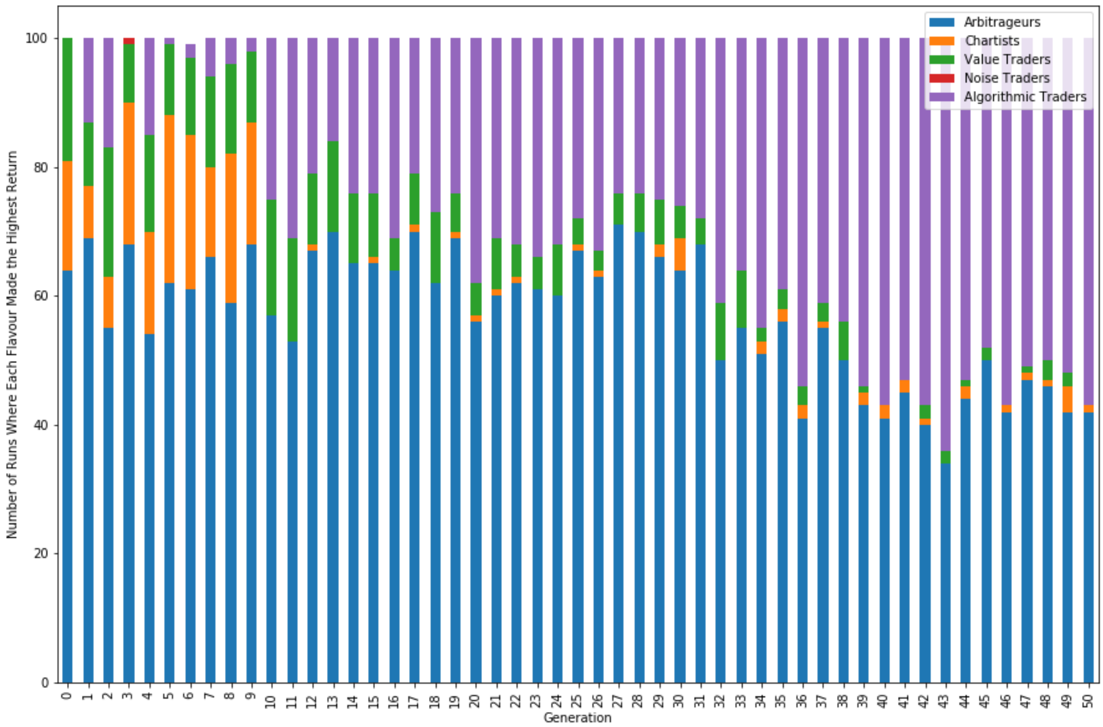


Figure 4.4: Graph Showing Average Highest Earner by Flavour



4.2 Liquidity

Figure 4.5 (average spread) implies an increase in liquidity over time. As more ATs are added, the spread decreases, indicating an increased potential book depth. Figure 4.6 (average volume) confirms this. In this market, all potential trades are executed as there are no non-market orders. An increase in traded volume indicates that if a trader wanted to alter their portfolio, they could do so without impacting price too severely as there is a higher chance of being matched with a trader on the other side of the market. These results are consistent with existing literature (Brogaard et al., 2013). It should be noted, however, that it is possible that the increase in liquidity may just be due to an increase in market participants.

Figure 4.5: Average Spread

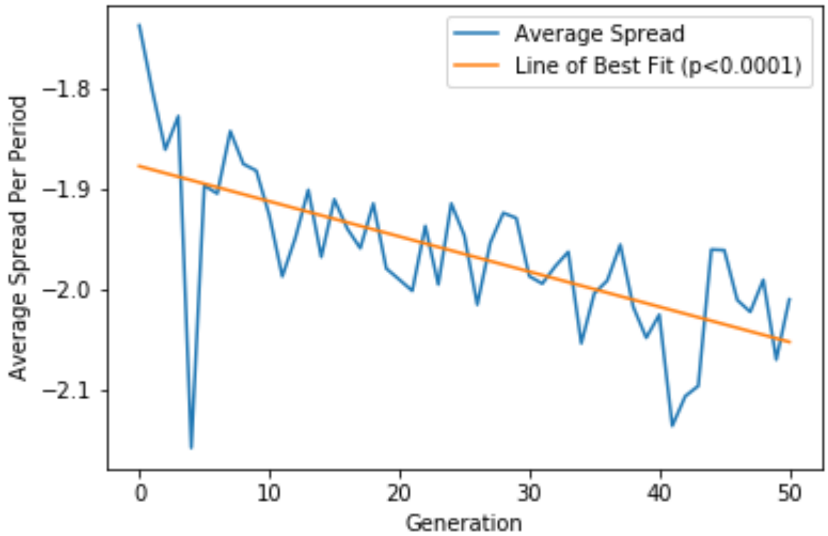


Figure 4.6: Average Volume per Period

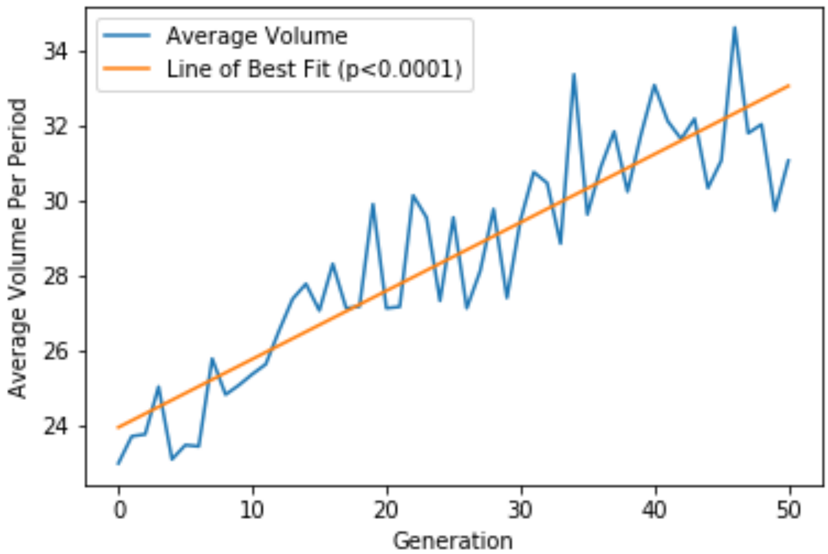
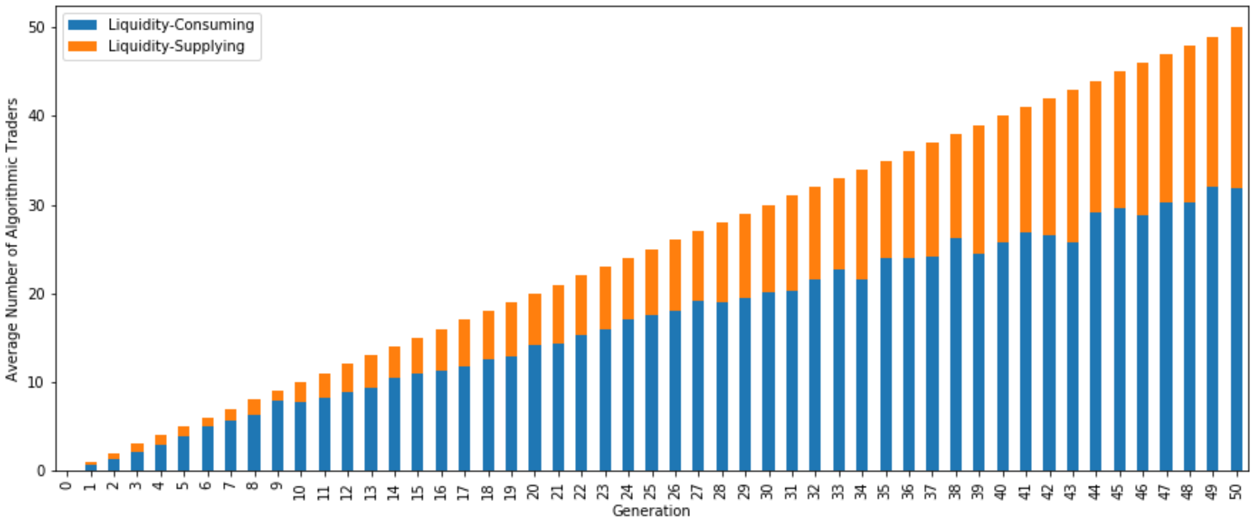


Figure 4.7 shows the average behaviour of ATs with regard to whether or not they are primarily liquidity-consuming or liquidity-generating. The prevalent view in literature is that ATs act as de facto market makers, consuming liquidity when it is cheap and supplying it when it is expensive (Hendershott & Riordan, 2009). Indeed, these results are corroboratory; we can see an increase in the proportion of liquidity-supplying behaviour as time passes, indicating that the added ATs are contributing to the increased market liquidity, and it is not entirely down to an increase in market participants.

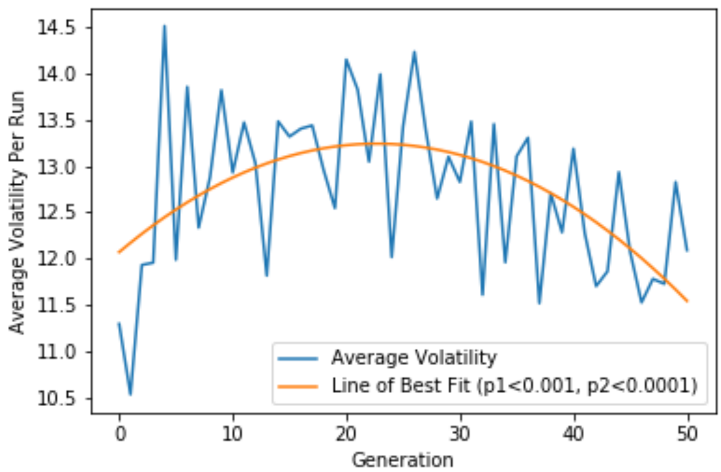
Figure 4.7: Graph Showing Average Liquidity-Consuming and Liquidity-Supplying Behaviour



4.3 Volatility

Initially when ATs are added, the volatility increases. However, we then see a steady decline, supporting the results from the relevant literature (Jovanovic & Menkveld, 2010). This may be due to the fact that it took a number of generations for the ATs to collectively converge to an effective strategy, before which orders were placed more erratically. From Figure 4.8 (combined with Figure 4.4), we can see it took about 10-15 generations for this convergence to take place.

Figure 4.8: Average Volatility Per Run



4.4 Time Inefficiency

Another way to measure overall Fama-EMH efficiency, is to assess how long, and by how much, price deviates from the “true price”. Figures 4.9 and 4.10 show this respectively. We can see a downward trend in both metrics, indicating that the price generally trended closer to the “true price” as more ATs were added to the market, indicating an increase in efficiency.

Figure 4.9: Proportion of Time Spent Away From the “True Price”

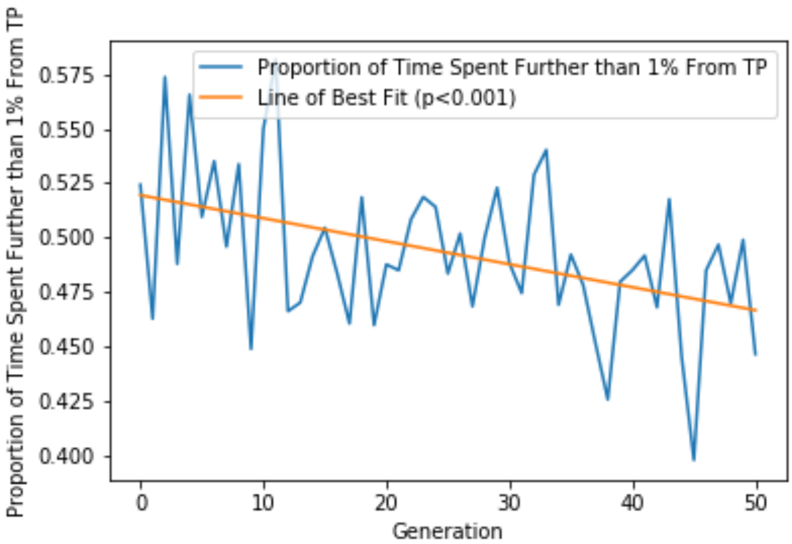
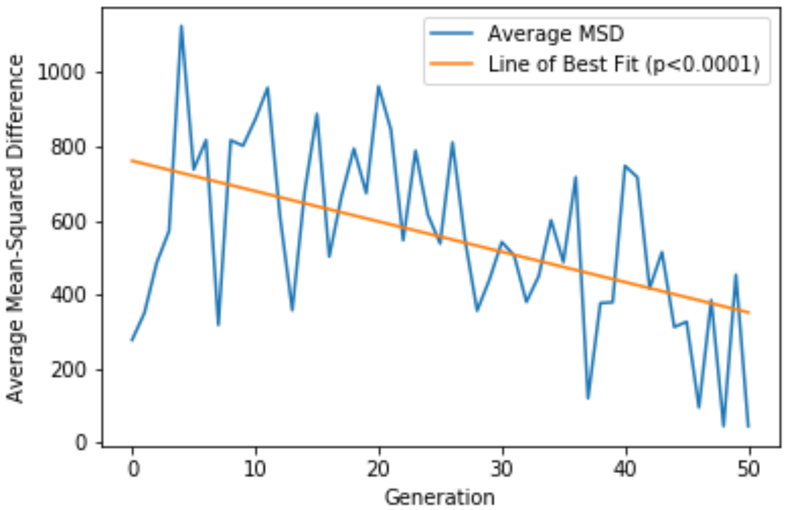


Figure 4.10: Average Mean Squared Distance Between Trading Price and “True Price” (per Run)



4.5 Deviation Analysis

We can also analyse the deviations between the trading price and the “true price”. Figures 4.11 through 4.14 show the average length and number (per run) of such deviations, which have been categorised into “returned” and “unreturned” deviations.

Figure 4.11: Average Length (in periods) of Returned Deviations

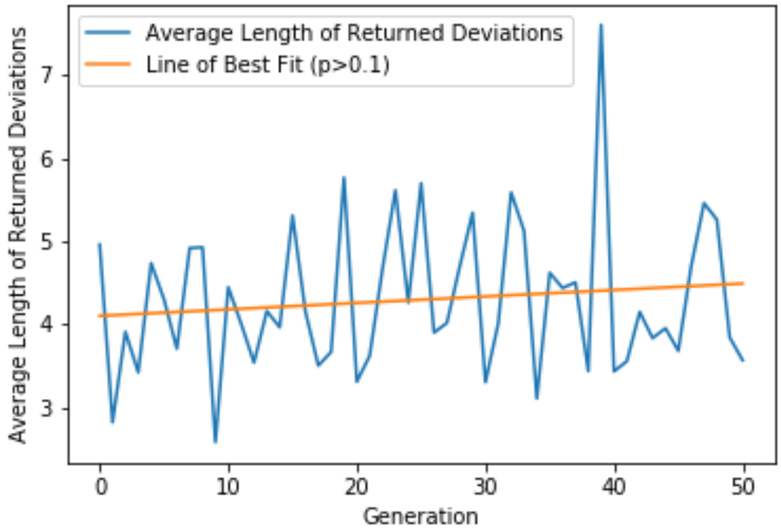


Figure 4.12: Average Length (in periods) of Unreturned Deviations

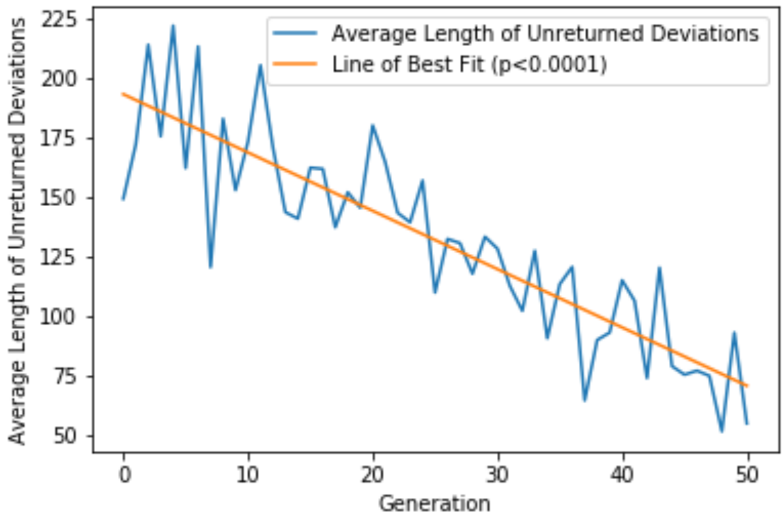


Figure 4.13: Average Number of Returned Deviations per Run

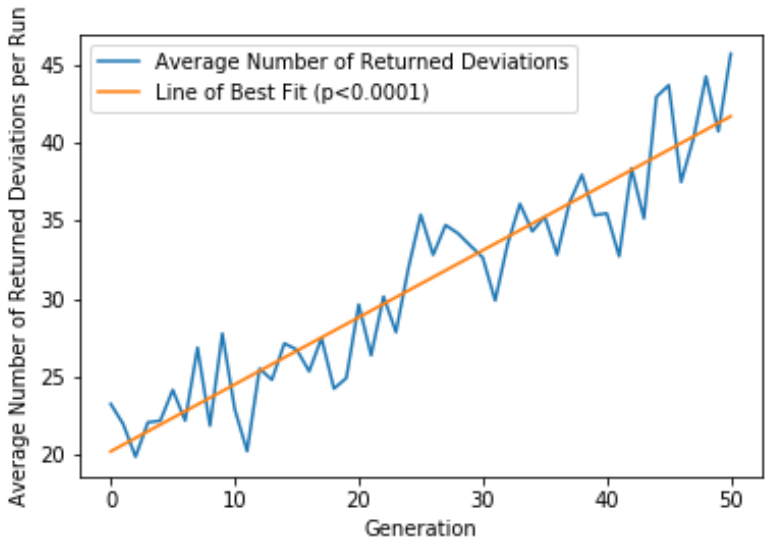
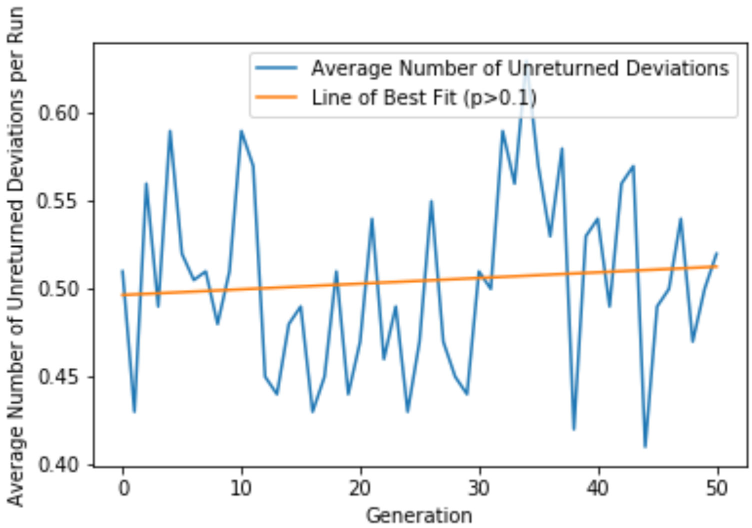


Figure 4.14: Average Number of Unreturned Deviations per Run



While the average length of returned deviations did not change, the frequency of them increased. The Fama-EMH does not strictly restrict the number of deviations around the "true price", as price can deviate substantially as long as the average of these deviations is 0. However, on closer inspection it seems that this may be indicative of increased inefficiency. It is possible that ATs are bringing price to the "true price" and then moving it away again, suggesting that they are making use of information not already priced into the market – a direct violation of the Fama-EMH. It is also possible that ATs may be completely ignoring the "true price", and the prices they predict (representative of their own private information) coincidentally align with the "true price" of the asset. This, again, violates the Fama-EMH as the "true price" of the asset is irrelevant to the trading decisions of these market participants. The feature analysis of the models comprising the ATs indicate the latter to be more likely.

The converse is true for unreturned deviations. Their frequency does not change, but the length of them decreases. This may indicate an increased stability in the market as prices take far longer to drift off (and not return) than in earlier generations. However, the feature analysis reveals this is, in fact, a self-fulfilling prophecy rather than stability. The price remains closer to the "true price" for longer because the ATs (who are completely driving the market in later generations) predict this price. This is not because they have any regard for the "true price", but rather that this price coincides with their predictions.

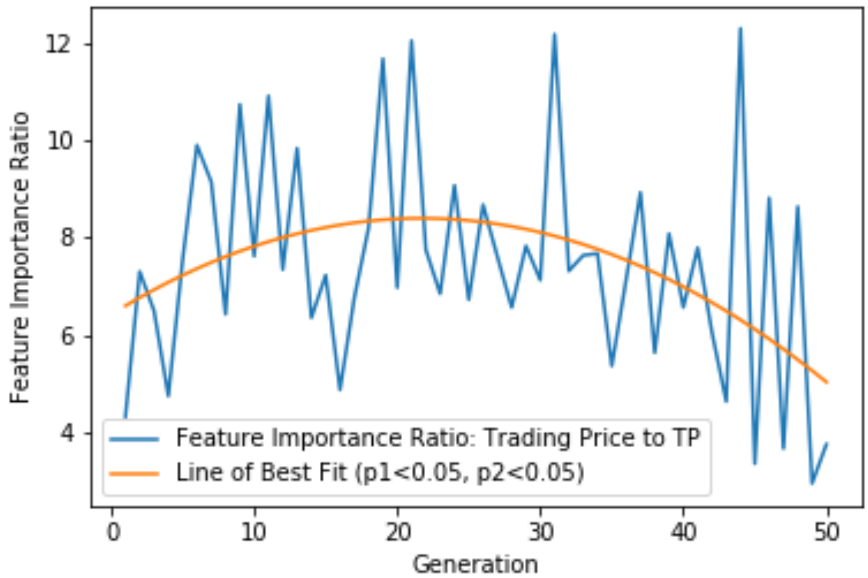
4.6 Model Analysis

One of the benefits of using an LSTM network is that it can take multiple features as inputs. In the case of this model, both the trading price and "true price" were used as inputs. A sensitivity analysis[7] of these features can reveal how much weight the models place on the trading price, as compared to the "true price", and in turn this can provide

[7] A full explanation of how the sensitivity analysis is conducted can be found in Appendix C.

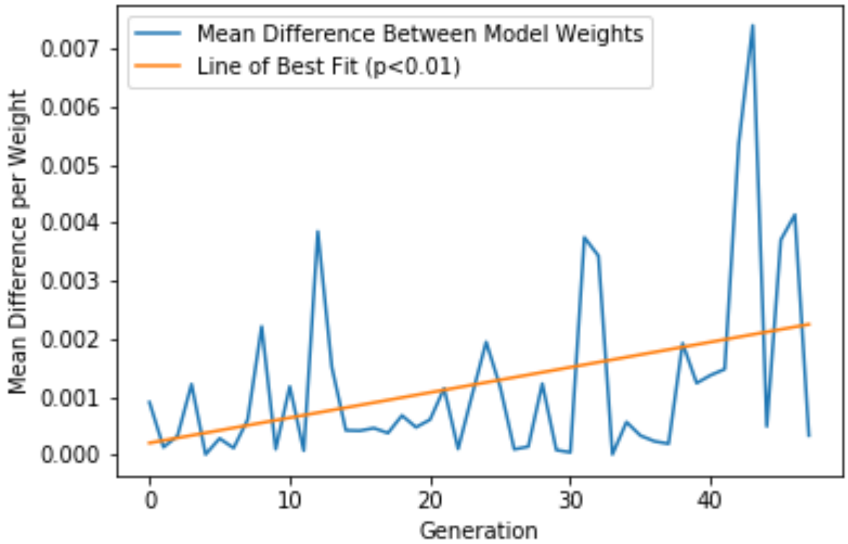
an assessment of market efficiency. Should the models converge to a strategy where they equally use both features, one could say this is indicative of high market efficiency, as it represents the investors' private information (comprising of expectations and beliefs) being in line with the predictions of the Fama-EMH. This produces a stable equilibrium at the "true price". Figure 4.15 shows that this certainly is not the case. At best, the feature importance ratio is 4; the trading price is 4 times more important in determining the model prediction than the "true price". This means that the market may present an illusion of efficiency, but this, in fact, is not a stable equilibrium, as not all market participants are operating off correct beliefs. The human traders may incorrectly view the market as efficient, falsely believing that this is maintained by the ATs. It should be noted, however, that the figure indicates this ratio is decreasing with the number of ATs. This suggests that over time, it may be possible to reach a stable version of this market where ATs do act to bring the price back to the "true price". Further research would be needed to confirm this.

Figure 4.15: Feature Importance Ratio of Trading Price to True Price



One of the biggest criticisms faced by ATs is the instability they bring. Since AT behaviour is not restricted by time and bias in the same way that human behaviour is, they have the potential to make large, correlated moves which can be devastating for a market. The constant threat of an unexpected, sharp downturn can deeply erode market confidence. Market confidence is essential for many large institutions who load and unload large positions over a length of time. Figure 4.16 shows the average difference between the weights and biases of consecutive generations of ATs, as an inverse measure of correlated behaviour. The average difference seems to trend upwards, indicating a negative correlation between consecutive models. This would imply that the mechanism for instability lessens as more ATs are added (opposing the current view in literature (Cliff, et al., 2011)). However, this analysis only considers consecutive generations, and correlative behaviour is likely to be far more complex.

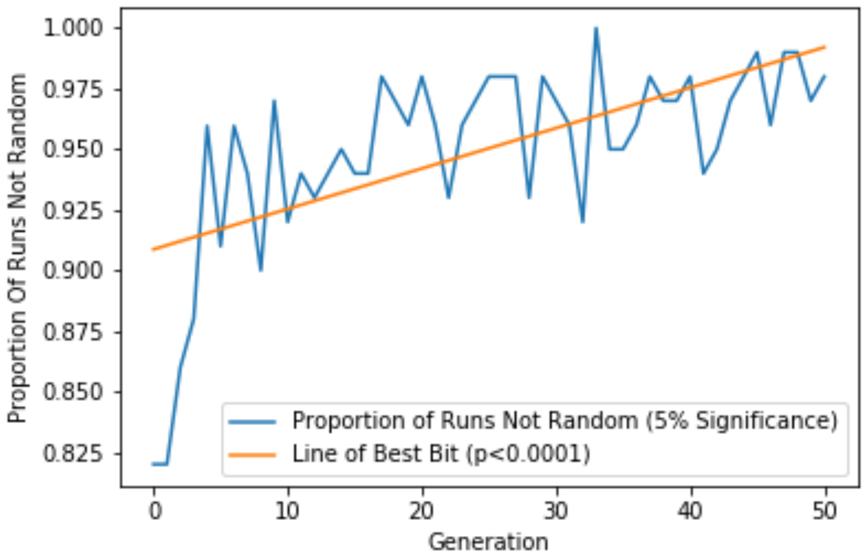
Figure 4.16: Average Difference per Weight/Bias of Consecutive Models



4.17 Runs Test

The Samuelson-EMH predicts that prices follow a martingale, a form of random walk (Samuelson, 1965). This can be tested through a Wald-Wolfowitz runs test.[8] Figure 4.17 shows the proportion of runs that reject the null hypothesis that the timeseries comes from a random distribution. A clear upward trend can be seen, suggesting that as more ATs are added, prices generally become more predictable, and less Samuelson-EMH efficient. This presents an apparent contradiction in the results, as Figures 4.1 and 4.2 suggest an increase in Samuelson-EMH efficiency. A reconciliation of these facts is presented in the next section.

Figure 4.17: Graph Showing Proportion of Runs That Reject the Null Hypothesis of the Wald-Wolfowitz Test at 5% Significance



[8] An explanation of the test can be found in Appendix D.

5 Discussion

While many indicators such as liquidity and mean investor returns indicate that ATs increase market efficiency, evidence is also presented of decreasing Fama-EMH and Samuelson-EMH efficiency. A reconciliation of these facts is offered by Martin & Nagel (2019) who, in their model, consider a cross-sectional, multi-asset economy with a large number of homogeneous algorithmic investors. They find that in-sample tests of efficiency lose their meaning as many different (some arguably irrelevant) predictors can appear to forecast asset returns, with as much explanatory power as traditional risk-based models. In an algorithmic setting, many of the assumptions underlying traditional models of efficiency do not apply. For example, algorithmic investors do not operate off the same information set, are not risk averse and react nonlinearly to changes. Resultantly, Martin & Nagel conclude that traditional risk-based (and even behavioural-based) explanations of asset returns (such as the EMH) should not be considered as appropriate. The same conclusion could be drawn from the results presented in this paper.

Traditionally, the Fama-EMH predicts that rational investors trade based on their own private information. That information is exchanged with the market in return for profit through trading. As the market makes many such trades with investors, it collects information, leading the price to reflect the net sum of this information over time. Additionally, this net sum reflects the "true price" of the underlying asset. In this market, the evolution of price towards the "true price" over time seems not to be a result of this mechanism. Instead, it seems a somewhat arbitrary outcome, driven by the fact that when the ATs initially began to converge to a strategy, arbitrageurs were the dominant flavour of traders (who deliberately acted to move price towards the "true price"). This particular flavour dominating the market is nothing more than an idiosyncracy of this model. The sensitivity analysis of the models showed that the ATs' expectations of price reverting to the "true price" was weak, and they instead based their predictions on the previous trading prices of the asset. It is not inplausible to expect prices to be far more unstable, had chartists been the dominant flavour in the earlier generations.

As a result, an alternative model to assess AT behaviour could be considered. The Adaptive Market Hypothesis (Lo, 2004) is an attempt to reconcile the EMH with behavioural economics, although a lot of the evolutionary logic applied to human investors can similarly be applied to the evolution of the algorithms in this model. Lo presents behavioural biases as by-products of evolution, arguing they are traits that were once essential for survival and have since persisted in humans past their necessity (e.g. probability-matching). Similarly, the proclivity of ATs to maintain price around the "true price" is the by-product of them developing strategies to perform well in a market full of arbitrageurs. Furthermore, Lo cites natural selection as a method for choosing top-performing trading strategies – which then "reproduce" as more investors adopt the better-performing strategy, ultimately resulting in “alpha turning to beta”. An algorithmic equivalent is plausible – once an extremely profitable strategy is discovered, ATs converge to it over time, resulting in an evolutionary niche model which perfectly captures the characteristics of that particular market. This could explain why, over time, the average above-market returns earned by traders decreased, as more ATs employed the strategy, removing the “edge” from the early adopters. It seems that, in the limit, the market may return to being Samuelson-EMH efficient as a convergent population of ATs is better able to extract information from the market prices. However, algorithmic convergence is far from guaranteed, especially in a model with more complex interactions. Whether such a convergent state even exists for particular markets is unclear, let alone whether or not it is discoverable by a system of algorithms. What may result is a system in which the evolutionary niche itself is constantly updating – as more investors adopt a particular alpha-generating strategy, their actions may give rise to new strategies which can better outperform the market.

Additionally, the stability of this market is difficult to quantify. While the correlation of AT behaviour seems to be decreasing with time, a system of complex algorithms can lead to a variety of emergent behaviours. It is possible for algorithmic interactions to produce unexpected behaviour, despite the algorithms themselves acting predictably on an individual level (Mackenzie, 2019). Macroprudential regulators, therefore, must remain vigilant to not fall into the trap of normalising deviance, as market stability can be rapidly eroded. One result of this paper is that regulators might benefit from adding their own ATs to the market, in an attempt to drive the evolutionary niche early on in model convergence (assuming the market will converge at all). If a regulator can get ATs to indirectly “care” about various indicators of market efficiency then the market may evolve in such a way, similar to the mechanism in this paper. Coordination between stock exchanges and regulators may be necessary to achieve this. Large institutions spare, quite literally, no expense in ensuring they are able to receive tick data as fast as possible – competition in these markets operates at the scale of microseconds. Tick regulation, therefore, (something which has been previously considered for other purposes), may be something that is particularly beneficial. It allows regulators to gain an edge on the market, as they can slow it down when needed to perform their operations, and speed it back up at will (Angel, 2012). Regulators also need to be aware of the increased instability in markets. Emergent behaviour is not only unstable, but in certain cases can be manipulated through other algorithms. These algorithms can use adversarial perturbations[9] to affect a subset of algorithms, leading to a cascading effect (Gleave, et al., 2019).

[9] Deliberately attempting to undermine the behaviour of other algorithms (Gleave, et al., 2019).

6 Conclusion

This paper has used a novel application of agent-based modelling to simulate a financial market with varying amounts of algorithmic traders. It finds that increased AT activity leads to improved liquidity through spread and book depth. AT activity is also not as necessarily correlated as previous literature may suggest, although further research is needed. Increased AT activity could potentially lead to increased efficiency, although it is uncertain whether this is a stable form of efficiency, where one's beliefs and actions are consistent across the population. Regulators have an increasingly important role to play, in terms of influencing the convergent path of algorithms, while simultaneously remaining alert for signs of abnormality.

A key feature leading to increased instability in markets is the positive feedback loop effect caused by algorithms who train and retrain, learning from each other's behaviour. This effect has not been explored in this paper, and is important to understand before any concrete regulatory action is taken. Additionally, the volatility of the metrics presented in this paper has been high, suggesting further work on a larger scale is needed. A larger market may present less endogeneity, allowing for more sophisticated analysis, including econometric testing and Monte-Carlo Markov chain analysis. This market contained only homogeneous (and fairly simple) ATs, restricting the range of potential emergent behaviours. Markets of heterogeneous ATs who are able to dynamically learn may provide further insight into such behaviours. Another useful extension would be to add a market-maker agent who can dynamically respond to different situations, allowing a greater level of realism for the model.

A significant implication of these results is a change in the meaning of prices. They no longer act as signals from the market to investors, but rather represent the transitory beliefs of algorithms, who view them as a single point in a sequence of inputs to produce an output with no regard to their economic significance. Further theoretical work should endeavour to explore the repercussions of this phenomenon.

The meaning of market efficiency is rapidly changing in the information age. As one of the first papers to consider the widespread implications of this, it makes a number of novel contributions. It also forms a basis for further avenues of research to increase our understanding of computer-based trading and market efficiency.

7 References

Angel, J. J., 2012. Tick size regulation: costs, benefits, and risks, London: HM Government Office for Science.

Brogaard, J., Hendershott, T. & Riordan, R., 2013. High Frequency Trading and Price Discovery, Frankfurt: European Central Bank.

Castura, J., Litzenberger, R., Gorelick, R. & Dwivedi, Y., 2010. Market Efficiency and Microstructure Evolution in U.S. Equity Markets: A High-Frequency Perspective. [Online]
Available at:
http://finance.wharton.upenn.edu/department/Seminar/micro/Litzenberger_transient_vol5_2010.pdf
[Accessed 21 April 2021].

Cliff, D., Hendershott, T. & Zigrand, J.-P., 2011. The Future of Computer Trading in Financial Markets, London: HM Goverment Office for Science.

Cliff, D. T. et al., 2011. The Future of Computer Trading in the Financial Markets, London: UK Government Office for Science.

Dash, R. & Dash, P. K., 2016. A hybrid stock trading framework integrating technical analysis with machine learning techniques. The Journal of Finance and Data Science, 2(1), pp. 42-57.

Enke, D. & Thawornwong, S., 2005. The use of data mining and neural networks for forecasting stock market returns. Expert Systems with Applications, Volume 29, pp. 927-940.

Fama, E., 1965. The Behavior of Stock-Market Prices. The Journal of Business, 38(1), pp. 34-105.

Gleave, A. et al., 2019. ADVERSARIAL POLICIES: ATTACKING DEEP REINFORCEMENT LEARNING. New York, Department of Computer Science, Cornell University.

Harper, K. M., 2016. Challenging the Efficient Market Hypothesis with Dynamically Trained Artificial Neural Networks, Jacksonville: University of North Florida.

Harvey, C. & Graham, J. R., 2001. The theory and practice of corporate finance: evidence from the field. Journal of FInancial Economics, 60(2), pp. 187-243.

Hendershott, T., Jones, C. M. & Menkveld, A. J., 2011. Does Algorithmic Trading Improve Liquidity?. The Journal of Finance, 66(1), pp. 1-33.

Hendershott, T. & Riordan, R., 2009. Algorithmic Trading and Information. [Online] Available at: http://www.netinst.org/Hendershott_Riordan_09-08.pdf
[Accessed 21 April 2021].

Hochreiter, S. & Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation, 9(8), pp. 1735-1780.

Jovanovic, B. & Menkveld, A. J., 2010. Middlemen in Limit Order Markets. [Online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1624329
[Accessed 22 April 2021].

Kahneman, D. & Tversky, A., 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty, 5(1), pp. 297-323.

Kissell, R., 2020. Algorithmic Trading Methods: Applications Using Advanced Statistics, Optimization, and Machine Learning Techniques. 2nd Edition ed. London: Elsevier.

Linton, O. & Mahmoodzadeh, S., 2018. Implications of high-frequency trading for security markets, London: Institute for Fiscal Studies.

Lo, A., 2004. The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective. The Journal of Portfolio Management, 30(5), pp. 15-29.

Lo, A. W. & MacKinlay, A. C., 1988. The Size and Power of the Variance Ratio Test in Finite Samples: A Monte Carlo Investigation. Journal of Econometrics, 40(1), pp. 203-238.

Mackenzie, D., 2019. How Algorithms Interact: Goffman's 'Interaction Order' in Automated Trading. Theory, Culture and Society, 36(2), pp. 39-59.

Martin, I. W. R. & Nagel, S., 2019. Market Efficiency in the Age of Big Data. s.l., NBER Working Paper No. w26586.

Menkveld, A. J., 2011. High Frequency Trading and The New-Market Makers, Amsterdam: Tinbergen Institute.

Nilsson, M. & van der Hoorn, D., 2012. The Relationship between High Frequency Trading and Stock Market Volatility. [Online]
Available at: https://lup.lub.lu.se/student-papers/search/publication/2837241
[Accessed 22 April 2021].

Patel, J., Shah, S., Thakkar, P. & Kotecha, K., 2015. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications, 42(1), pp. 259-268.

Payne, R. & Friederich, S., 2011. Computer based trading, liquidity and trading costs, London: HM Government Office for Science.

PricewaterhouseCoopers, 2020. Asset & Wealth Management Revolution: Embracing Exponential Change, London: PricewaterhouseCoopers.

Rieger, M. O., Wang, M. & Hens, T., 2017. Estimating cumulative prospect theory parameters from an international survey. Theory and Decision, 82(1), pp. 567-596.

Samuelson, P. A., 1965. Proof That Properly Anticipated Prices Fluctuate Randomly. Industrial Management Review, 6(2), pp. 41-49.

Schumaker, R. P. & Chen, H., 2009. Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Transactions on Information Systems, 27(2), p. 19.

Showalter, S. & Gropp, J., 2019. Greencastle: Department of Economics, DePauw University.

Zhang, F., 2010. High-Frequency Trading, Stock Volatility, and Price Discovery. [Online]
Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1691679
[Accessed 21 April 2021].

8 Appendix

8.1 Appendix A

Figure 8.1: Table explaining variables

Variable	Distribution (if applicable)	Description
P_t		Trading price in period t
E_{t-1}		True price in period t
ε^i	N(0, 1.5^2)	Individual error term in assessing true price
M_t^s		Short term (10 period) simple moving average of the true price, at time t
M_t^l		Long term (50 period) simple moving average of the true price, at time t
A^i	U(0, 1)	Individual activation frequency
A_t	U(0, 1)	Activation frequency threshold for period t
P^{max}		Maximum recorded trading price for that run
P^{min}		Minimum recorded trading price for that run
$\sim N(0,1)$		A random realisation from a N(0, 1) distribution
$x_{t,g}^i$		Expected "good state" outcome for individual i in period t
$x_{t,b}^i$		Expected "bad state" outcome for individual i in period t
$p_{t,g}^i$		Expected probability of the "good state" outcome for individual i in period t
$p_{t,b}^i$		Expected probability of the "bad state" outcome for individual i in period t
U_t^i		Prospect-utility for individual i in period t
L^i,	U(0, 0.5]	Individual threshold which utility must surpass for individual to place a trade
${}_{t-b_i}R_t$	$b_i \sim \lceil(0, 10)\rceil$	The return on the asset over some fixed period t-b and t, where b is an integer representing how backward looking the trader is
f^i	U(0, 1)	Individual faith rate, a lower faith rate trader trades less aggressively and is more likely to mutate between flavours
S^i	U(1, 2)	Scale factor to transform utility to order quantity
$\|\sim N(3,1)\|$		Absolute value of a random realisation from a N(3, 1) distribution
I^i		Inventory (stock balance) of individual i

Figure 8.2: Table explaining functions

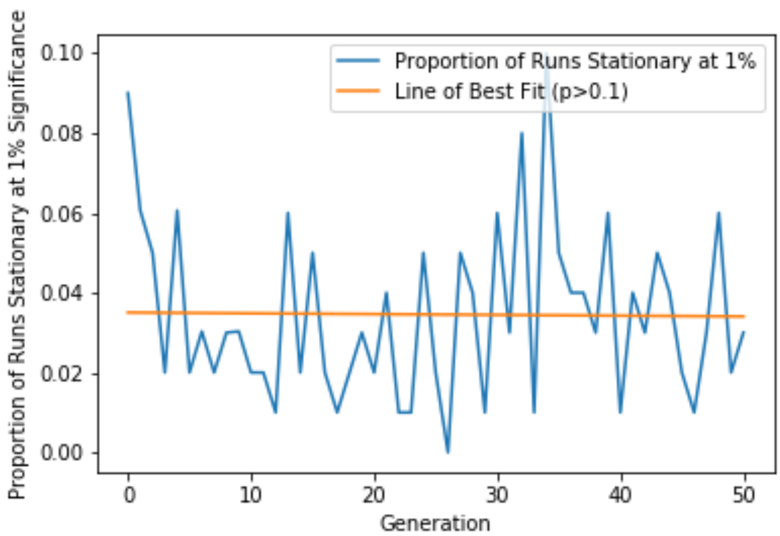
Function	Definition
$\sigma(x)$	Standard sigmoid transformation: $\sigma(x) = \frac{1}{1+e^{-x}}$
$w(p)$	Probability weighting function[10] $w(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{\frac{1}{\gamma}}}, \gamma = 0.5$[11]
$V(x)$	Value function $V(x) = \begin{cases} x^{\alpha}, \text{if } x \geq 0 \\ -\lambda(-x)^{\beta}, \text{ if } x < 0 \end{cases}, \alpha = 0.44, \beta = 0.49, \lambda = 1.06$

8.2 Appendix B

The line of best fit for each graph was derived by sequentially testing polynomials of increasing magnitude. First a linear line of best is tested. Then a quadratic one is tested. If the quadratic term fails a t-test of being significantly different from 0, then the linear one is used. If not, the polynomial increases in order until the leading term is not significant from 0.

The graphs below show the results for stationarity tests, as well as tests for cointegration between the trading price and the "true price" at various significance levels. A standard Augmented-Dickey Fuller Test is used for conducting stationary with MacKinnon critical values.

Figure 8.3: Graph Showing Proportion of Runs Stationary at 1% Significance



[10] The weighting function and value functions from (Kahneman & Tversky, 1992)
[11] Functional forms, gamma value and value function parameters from (Rieger, et al., 2017)

Figure 8.4: Graph Showing Proportion of Runs Stationary at 5% Significance

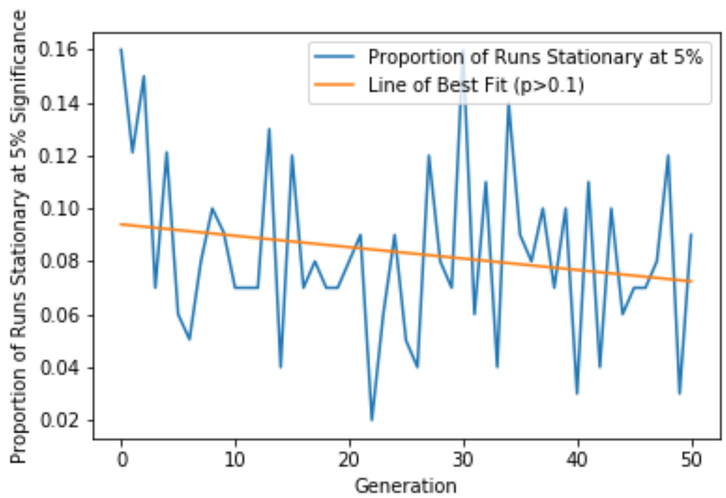


Figure 8.5: Graph Showing Proportion of Runs Stationary at 10% Significance

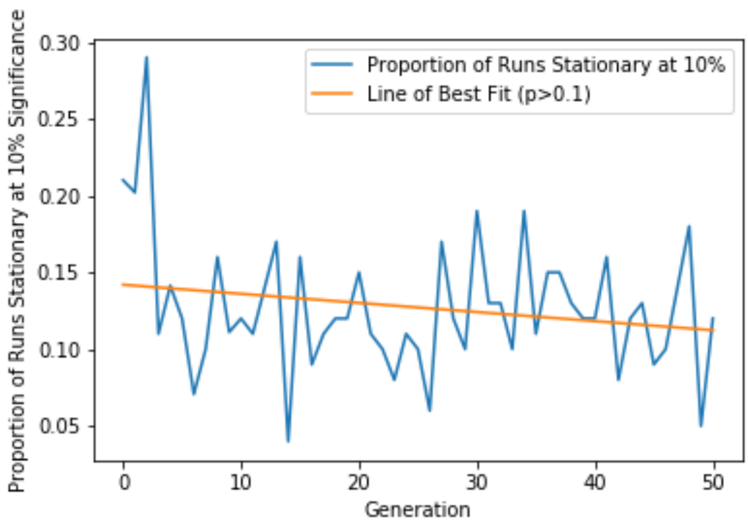


Figure 8.6: Graph Showing Proportion of Runs with Evidence for Cointegration at 1% Significance

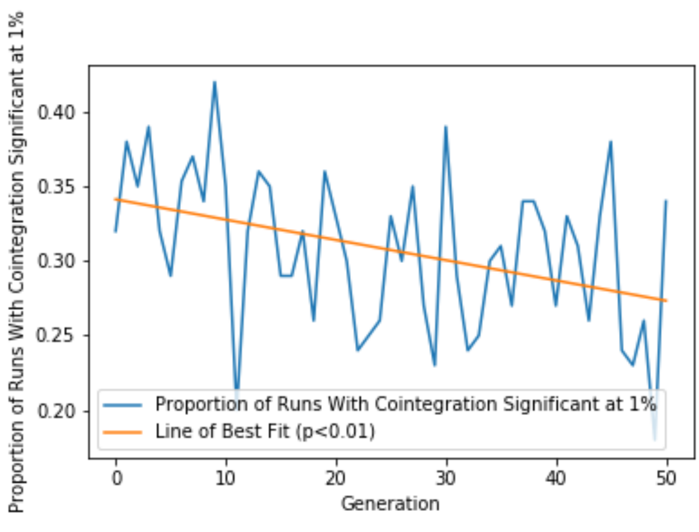


Figure 8.7: Graph Showing Proportion of Runs with Evidence for Cointegration at 5% Significance

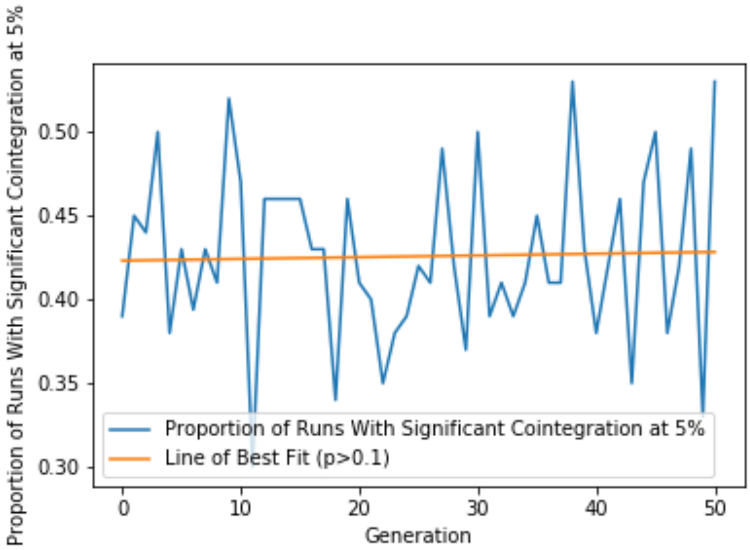
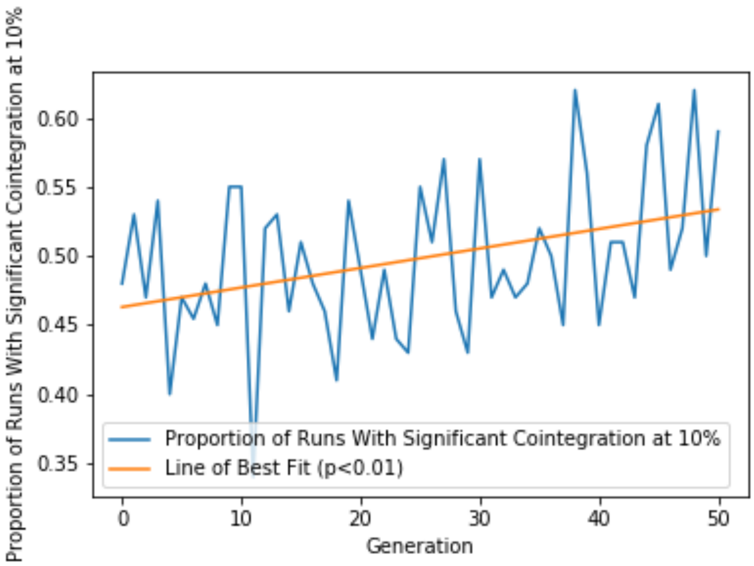


Figure 9: Graph Showing Proportion of Runs with Evidence for Cointegration at 10% Significance



8.3 Appendix C

The input data for the model contains 2 features, the true price and the trading price. To assess the importance of each feature, a sensitivity analysis is conducted. The test works by taking the model predictions after training, $\hat{y}$, and then altering each input feature by some small amount, $x_i + \varepsilon$. The perturbations used in this example are random realisations from the distribution N(0, 0.2^2). How much the prediction changes based on the small change in each input feature can be used as a measure of the relative importance of each input feature (measured by the root mean square difference between the original prediction and the new prediction $\tilde{y}_i$. As the perturbation has been chosen somewhat arbitrarily, the correct way to interpret the results from this test is as a ratio (or some measure comparing the two features, not independently).

8.4 Appendix D

The Wald-Wolfowitz test (also known as "runs test") is a two-sided hypothesis test that looks at runs of a series to assess whether or not each item in the series is an independent realisation of a random variable. To conduct the test, each market run was converted to a binary series, where each period is allocated a 1 if the market price rose that period, or a 0 if the market price fell that period. From this, a number of "runs" are determined, where a run is a sequence of any length (including 1) of consecutive draws of the same value (either 1 or 0). For example, the sequence "100010" has 4 runs. The null hypothesis of the test is that the number of runs in a sequence of N is a random variable with conditional distribution approximately normal (given the number of positive and negative observations, i.e. 1s and 0s), with parameters:

$$Mean: \mu = \frac{2n_1n_2}{n_1+n_2}$$

$$Sample\ variance: S^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

Where n_1 is the number of 1s and n_2 is the number of 0s in the sequence. The test statistic is then:

$$Z = \frac{R - \mu}{S}$$

Where R is the number of runs in the sequence. The null hypothesis is rejected when this test statistic falls within the critical regions of a standard normal distribution, depending on the chosen significance level.