A preliminary investigation into multi-agent trading simulations using a genetic algorithm

Cyril Schoreels, Jonathan M. Garibaldi
Automated Scheduling, Optimisation and Planning Research Group
School of Computer Science and Information Technology
University of Nottingham
czs@cs.nott.ac.uk

Abstract

This paper investigates the effect of using varying amounts of training data on the specificity and robustness of a multi-agent based solution for use in trading simulations using historical equity market data. Three separate amounts of training data were used in five experiments to evolve 15 groups of agents under varying conditions. These groups were then exposed to three separate test environments to determine whether differences in performance could be related back to their training environment. The results indicate that larger training data sets lead to more general solutions and overall better performance when tested in environments with varying conditions. This will lead onto future research focusing on agent behavior and decisions in financial markets.

1 Introduction

The complexity of modeling socio-economic environments raises questions of interest in how to best design systems for machine simulations. In recent years, research has explored a multitude of approaches emulating environments such as whole economies or more specific areas such as equity markets, hoping to create better understanding of how they operate and what the main factors are that determine their behavior, as these underlying factors are of key interest to governments and corporations.

Multi-agent technologies have lent themselves well to exploring this domain, as they are naturally suited to duplicate market environments, equipped with various level of autonomy and being able to simulate interactions between people and the environment well. A vast array of literature has explored the application and use of agents in any number of fields [6, 14], frequently being coupled with artificial intelligence methods, in particular genetic algorithms [4], to evolve solutions to problems.

One of the most well known projects in artificial markets is the Santa Fe Stock Market, in which simple market mechanics that allow agents to make trading decisions using classifier systems are used [3]. By generating demand functions for shares and other forecasting methods, agents are able to trade within the market, creating similar patterns observed in real markets. However, due to the complexity of the artificial market the actual causes are again hard to determine and increase the difficulty of relating this back to real markets [9]. Similarly, there have been many other projects looking at simulating agents as traders in virtual markets using historical market data, such as the Penn-Lehman Automated Trading Project [8]. Many others have focused on various aspects of artificial markets, ranging from studying agent behavior within markets [11] to analyzing market behavior and interactions using agents trading across national boundaries [2].

As markets represent either physical or virtual meeting places for buyers and sellers to interact and exchange property of various forms, it is becoming increasingly attractive to not only use computer simulations to study market characteristics, but also to develop systems that interact directly or indirectly with human counterparts in the long-term. For instance, as automated trading systems in investment banks are gaining prominence and straight-through-processing is becoming common place [13], a humans role in data analysis is becoming less prominent.

In this investigation a system was developed aimed at exploring how different training settings affect the performance of agents in the test environment. It is hypothesized that due to the nature of financial time series, solutions can very easily suffer from over-fit. Though a high level of specificity of a particular situation may be desirable where training and testing environments are likely to be a continuation of each other, the disadvantages of over-fit would become apparent if training and testing environments were discontinuous. Specificity being defined as the level of specialization of an agent group for a particular environment. In a changing environment, a more general solution able to operate effectively, if not optimally, in a multitude of different environments would be preferable. A general solution being defined as one that is able to operate successfully in a multitude of environments and not being particularly suited to any one specific situation. To assess these conjectures, the following hypotheses are formulated.

Hypothesis 1 (H1): A solution evolved using little training data will be highly specific to the environment encountered during training and display above average performance in such environments, but fare considerably worse when in an unknown environment.

Hypothesis 2 (H2): A solution evolved using large amounts of training data will be fairly general in nature as it will have encountered a multitude of different environments, making it able to cope with changing environments during testing.

Hypothesis 3 (H3): Increasing the quantity of training data will improve generality and robustness up to a point, after which no significant further gains can be made.

In other words, using several highly specialized solutions in the appropriate environment would display better performance, while evolving a more general solution should perform moderately well across the board. Furthermore, increasing the quantity of training data eventually leads to diminishing marginal returns and increasing it further will not improve the solutions performance, as illustrated in Figure 1.

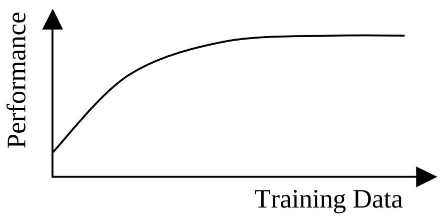

Figure 1. Diminishing marginal returns graph.

2 Agent Model and System Design

In order to assess these ideas, a system was designed using an agent population representative of traders, with training using a variety of settings but with three identical training data sets. The environment used historical equity market data from the German DAX-30 Index, with no transaction costs, no interest paid on capital holdings and the environment is assumed discrete and deterministic in a liquid market, meaning that an agent's actions cannot affect prices.

2.1 Agent Design

The system is based on the concept of simulating traders in an equity market, with each agent representing a trader with their own personal portfolio. Each agent is provided with a set starting capital and is allowed to conduct as many trades as desired at the end of each day using the securities closing prices, with the aim of maximizing profit at the end of each generation cycle. An agents characteristics and behavior are exclusively determined by its genome.

Agents can perform three types of actions, which are to buy, sell or hold any security. In order to make a decision for any particular security, a range of technical indicators are used to generate trading signals for the security, using the agent's genome as parameters. Essentially, an agent assesses every security using technical indicators, interprets their results and then makes a decision to buy, sell or hold following the pattern prescribed by its decision type. The amount to invest in a series of purchases is again determined by the genome and split equally between all securities flagged for acquisition. For all other securities no action is taken, which translates into holding the security if it is currently held in the portfolio.

In order to allow for differing approaches to selecting securities for purchase or sale, four decision types were implemented in the decision process. Decision type 1 performs a simple comparison between the number of buy and sell signals, taking the appropriate action if one is greater than the other for any particular security. For instance, out of 8 possible signals for a security, if 3 are buy and 2 are sell, an agent of this type would want to purchase this security. Decision type 2 follows the same principle as decision type 1, however also stipulates that for a buy or sell action to occur, at least half of all signals must be in favor. In this instance for example, out of 8 possible signals, if only 3 are buy signals even though no sell signals exist, no action will be taken as it failed to reach the minimum buy threshold. Decision type 3 sums the indicators by taking buy signals as positive and sell signals as negative, as well as including a weighting process on each signal, increasing or decreasing its magnitude and hence impact on the final sum. Therefore a positive sum would translate into an overall buy signals, while a negative sum into an overall sell signal. Decision type 4 follows the same principle as decision type 3, except for adjusting the final sum to create a threshold value which it needs to exceed prior to resulting in an overall buy or sell decision.

2.2 Genetic Algorithm

For selection, elitism [7] is used, whereby a portion of the most successful agents carries forward unaltered every generation, and immigration [5], where a portion of the worst performers are killed off and replaced by a new randomly generated group of agents immigrating into the system, constantly introducing new genetic material. This ensures a gradual move towards a global optimum or optima, while also avoiding premature convergence and non-exclusion of other possible solutions not present in the original base populations gene pool. The mediocre part of the population, those not killed or part of the elite, engage in reproduction among themselves and the elite using simple crossover. Two randomly selected agents of the same decision type can mate, with a random part of the first agents genome being replaced by the equivalent section from the second agents genome, forming a new genome combination. The only restriction in this process is that there exists a 25% chance of an agent mating with an agent of a different decision type, as for instance weighting genes will not have been relevant to types 1 or 2 previously. The entire process is illustrated in Figure 2 below.

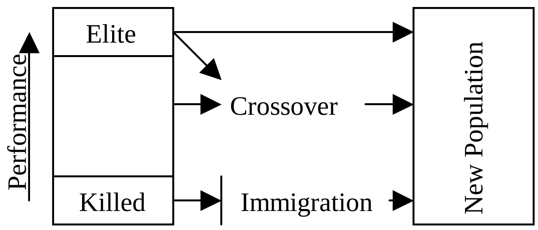

Figure 2. Genetic algorithm representation.

In these experiments, 25% of the population was considered to be elite while 25% of the bottom population was killed and replaced through immigration, leaving the middle 50% to reproduce with the elite and form part of the new population for the next generation. Fitness was determined by simply considering the total assets held by each agent at the end of every generation, or in other words, the sum of its capital and value of all holdings at the last day's closing price.

The genome for every agent consists of a string of integers taking values 1-10, of length 28 with cardinality 10. The exceptions to this were genes 1, 2, 21 and 28, defining decision types with a cardinality of 4, the risk averseness factor with a cardinality of 2 as well as the SO D variable and BB deviation variable with a cardinality of 5. All genes are defined as in the Table 1 below. The search space consisted of $2x10^{26}$ possible gene combinations.

Gene	Range	Description/Function
G_1	1-4	Decision type
G_2	1-2	Risk averseness factor
G_3	1-10	Capital investment proportion
G_4	1-10	Moving Average weight
G_5	1-10	RSI weight
G_6	1-10	Short-term ROC weight
G_7	1-10	Long-term Price ROC weight
G_8	1-10	SO interpretation 1 weight
G_9	1-10	SO interpretation 2 weight
G_{10}	1-10	MACD weight
G_{11}	1-10	BB weight
G_{12}	1-10	MA short-term value
G_{13}	1-10	MA long-term value
G_{14}	1-10	RSI time period
G_{15}	1-10	RSI buy threshold
G_{16}	1-10	RSI sell threshold
G_{17}	1-10	ROC level
G_{18}	1-10	ROC short-term value
G_{19}	1-10	ROC long-term value
G_{20}	1-10	SO K variable value
G_{21}	1-5	SO D variable value
G_{22}	1-10	SO buy threshold
G_{23}	1-10	SO sell threshold
G_{24}	1-10	MACD short-term value
G_{25}	1-10	MACD long-term value
G_{26}	1-10	MACD signal line
G_{27}	1-10	BB time period value
G_{28}	1-5	BB deviations number

Table 1. Gene descriptions.

2.3 Technical Indicators

Financial technical analysis is *the study of prices to make better investments* [1], where technical indicators represent the tools used for analysis. In other words, indicators are mathematical models used indicate buy or sell conditions. They are commonly based on closing price or volume data, though in this system the former was used exclusively. The following is a brief description of the indicators used, and how the agent genome is used to personalize analysis and interpretation of the securities. The following descriptions are primarily based on a summary presented by Achelis [1]. They include calculations and interpretations used in the system itself, including the translation of a number of gene values. When an agent is initialized, some of its gene values would not be appropriate for direct use in the system and need to be modified. For instance, though G_{17} can be used directly without translation for use in the technical indicator, this is not possible for G_{13}. The translation is based on achieving an average value approximate to that suggested in literature.

A simple Moving Average (MA) shows the average value of a securities price over time. If the moving average over the short-term is larger than that over the long-term, it indicates an upward trend and a buy signal would be generated or vice versa. Additionally, the short-term moving average can be compared to the current price of the security, which if greater, would indicate a downward trend and hence a sell signal should be generated. In this implementation, if the agent is risk averse as determined by G_2, it bases its interpretation on a logical AND between those two interpretations and be more reluctant to generate a buy signal. On the other hand, if the agent is risk taking, a logical OR is used and either interpretation suggesting a buy would suffice for the agent to consider this a buy signal. The short-term and long-term moving average values are calculated as:

$$(1)\quad MA = \sum_{i=1}^{N} price(i) \Big/ N$$

where $N=4G_{12}$ for the short-term and $N=(5G_{12})+50$ for the long term.

The Relative Strength Index (RSI) is a price-following oscillator that compares the internal strength of a single security. Quite simply, if the calculated value lies above or below the sell or buy threshold respectively, the appropriate signal will be generated. For instance, if the calculated value is 80, a sell signal is generated. The threshold values were $5G_{15}$ for buy and $(4G_{16})+50$ for sell. The RSI is calculated as:

$$A = \sum_{i=1}^{N} UpCloses \Big/ NumberOfUpCloses$$

$$B = \sum_{i=1}^{N} DownCloses \Big/ NumberOfDownCloses$$

$$(2)\quad RSI = 100 - \left(100 / \left(\left(A/B\right) + 1\right)\right)$$

where $N=2.5G_{14}$.

The Price Rate-of-Change (ROC) indicator is based on the assumption of cyclical price movements, and considers the relative change of prices over time to indicate trends. The same principles apply for long- or short-term analysis, where if the calculated value lies below the negative threshold value it indicates a buy, while a calculated value above the positive threshold value indicates a sell. It is calculated by:

$$(3)\quad ROC = price(i) - price(i-N)$$

where $N=2G_{18}$ for the short-term and $N=4G_{19}$ in the long-term.

The Stochastic Oscillator (SO) compares a security's price relative to its price range over a given time period. Multiple interpretations are possible, though the following two are used in this instance. First, it can be considered a buy signal if the K value is larger than the D value or vice versa. Second, threshold values can be used for both K and D. In that case, if K and/or D is smaller than the buy threshold a buy signal is generated, or equally, if K and/or D is larger than the sell threshold a sell signal is generated. Threshold values are $3.5G_{22}$ for buy and $(4G_{23})+50$ for sell. K is calculated in equation 4 over $1.5G_{20}$ days, while D is a moving average of K over G_{21} days.

$$A = price(i) - LowestClose$$

$$B = HighestClose - LowestClose$$

$$(4)\quad K = \left(A/B\right) * 100$$

The Moving Average Convergence Divergence (MACD) *is a trend following momentum indicator that shows the relationship between two moving averages of prices* [1]. It compares its calculated value to a moving average of itself over a time period, whereby a buy signal is generated if the moving average is smaller, and a sell signal if the moving average is larger. The MACD is calculated:

$$(5)\quad MACD = \exp MA(2G_{24}) - \exp MA(4.5G_{25})$$

where exponential moving average is:

$$(6)\quad \exp MA = \left(\left(price(i) - prevMA\right) * \left(2/(N+1)\right)\right) + prevMA$$

where N represents the time period it was measured over (G_{24} or G_{25}) and prevMA is the previous exponential moving average apart from the first instance, where a simple moving average is used. The signal line is translated as $1.5G_{26}$.

Lastly, Bollinger Bands (BB) are not a standalone indicators as they do not generate explicit buy or sell signals and are generally used to provide a form of guideline, indicating possible trend reversals. In this case, if the current price breaks through the lower bollinger band it is considered a buy signal, while if it breaks through the upper band it is considered a sell signal. The upper and lower bands are calculated as:

$$stdDev = \sum_{i=1}^{N} \left(price(i) - MA(N)\right)^2$$

$$(9)\quad upperBand = MA(3.5G_{27}) + \left(stdDev * G_{28}\right)$$

$$(10)\quad lowerBand = MA(3.5G_{27}) - \left(stdDev * G_{28}\right)$$

where $N=3.5G_{27}$.

2.4 Training and Testing Data

For consistency each experiment used identical data sets, defined as Training A-C and Testing A-C, as detailed in Table 2 and Table 3.

	Training A	Training B	Training C
Days	521	1304	2609
Time period	01.01.88 to 29.12.89	01.01.85 to 29.12.89	01.01.80 to 29.12.89

Table 2. Training data.

	Test A	Test B	Test C
Days	522	1305	2610
Time period	01.01.90 to 31.12.91	01.01.90 to 30.12.94	01.01.90 to 31.12.99

Table 3. Testing data.

Due to changes in the constituents of the Index, daily closing price information was only available for 20 securities over the desired 20 year time span. The selected securities used here are listed in Table 4.

840400	648300	519000	717200	515100
760080	823212	803200	723610	575200
802200	703712	761440	766400	695200
593700	843002	750000	543900	514000

Table 4. Securities used as denoted by their WKN.

3 Experiments Outline and Training

With three variant factors, only five experiments were run as part of this preliminary investigation, each analyzing the impact of a different factor from training on the test performance of those agents. If the same observations are made for each experiment, it can be assumed that any conclusions drawn therefrom should be valid for any type of training setup. The experiments are outlined in Table 6, with the altered factor in bold.

	Population	Generations	Time	Shares
1	100	1000	Variable	20
2	100	Variable	**24 hrs**	20
3	**40**	1000	Variable	20
4	**200**	1000	Variable	20
5	100	1000	Variable	**10**

Table 6. Description of experiments run.

For every experiment three groups of agents were trained on the three training data sets A-C, resulting in 15 trained groups of agents. The hypotheses claim that solutions trained on little data are very specific but highly successful in their environment and that solutions trained on large quantities of data are more general and only moderately successful across multiple environments. If this is true, then it should be possible to make the same observations in every experiment for every group of trained agents once exposed to test data.

Assuming that training A and test A are sufficiently similar environments, H1 will be tested by assessing whether groups of agents from training A performed better than groups from training B or C, who are assumed to be increasingly more general solutions. H2 will be tested by assessing whether groups from training C manage to perform better than groups from training A in tests B and C. Additionally, both assessments of H1 and H2 will have groups from trainings A and C compared to the Index as a rough benchmark, indicating whether they performed comparatively well to the market overall rather than just among each other.

4 Test Results

The elite from each group of agents was exposed to three sets of data as described earlier, with their daily and final assets recorded for analysis. The results

	Test A		Test B		Test C	
	Elite	Area	Elite	Area	Elite	Area
Group 1-A	-11.81%	38,765.14	16.40%	118,607.27	126.16%	340,023.74
Group 1-B	-11.01%	40,356.74	15.71%	121,031.97	178.60%	368,067.68
Group 1-C	-06.70%	42,867.32	19.65%	125,592.90	173.46%	379,558.51
Group 2-A	-14.66%	37,265.53	10.69%	113,506.18	125.46%	328,236.78
Group 2-B	-09.10%	40,846.50	26.96%	126,171.19	219.79%	404,081.68
Group 2-C	-10.28%	40,747.50	19.95%	123,713.16	187.12%	382,207.47
Group 3-A	-21.05%	35,994.10	07.42%	108,716.59	178.38%	336,954.16
Group 3-B	-08.31%	39,230.41	22.93%	124,637.77	150.84%	382,296.93
Group 3-C	-01.92%	42,019.77	29.88%	130,873.66	188.65%	399,495.86
Group 4-A	-18.99%	41,404.81	00.78%	128,312.01	145.30%	428,760.34
Group 4-B	-11.81%	45,828.66	21.96%	150,854.94	172.52%	526,632.97
Group 4-C	-09.02%	41,623.79	19.01%	123,039.05	183.96%	374,293.85
Group 5-A	-13.99%	43,474.73	17.75%	147,868.67	177.87%	487,859.25
Group 5-B	-08.56%	45,027.80	30.06%	153,570.47	166.82%	507,197.18
Group 5-C	-08.86%	43,798.64	38.39%	151,968.58	225.43%	549,159.89
DAX	-11.86%	37,555.01	16.01%	118,665.61	288.64%	392,537.68

Table 5. Summary of test results.

from this are shown in Table 5.

For each test, the average performance of the elite from training is recorded as a percentage. Additionally, the area under the daily total average assets graph is recorded as an alternative measure of performance. The importance of this becomes apparent when an agent or group of agents performance is consistently better than that of its competitors, but at one point suffers unusually large losses. If performance is measured at this point, that particular agent will then appear to be a below average trader. Hence, by using the area under their daily performance graph for comparison, this eliminates the possibility of misinterpreting overall performance by taking only one instance in time. For the purposes of this paper, absolute performance will refer to the average performance of the elite at the end of each test while overall performance will be defined as the area under the graph, based on the average assets of the elite population for every day of each test. These measurements were done for every group in every test set, with the DAX Index used as a reference.

As can be seen from the above, in the majority of cases the elite population outperformed the Index in both performance measures. The main exception relates to results from test C, where every agent group failed to attain comparative gains to the DAX and even in terms of area did not consistently manage to achieve better performance.

Additionally, nine agents, that were not trained but designed with a fixed genome, were tested in the same manner. This was done to establish an internal comparative measure to determine whether optimization of the genome and deviating from parameters recommended by literature for technical analysis lead to any improvements in performance. It was also assumed that G_3, responsible for allocation of funds to a purchase, is set to 7, representing an investment of 70% each time. Lastly, weightings were assumed to be equal for every indicator. A brief description of each agent and its results is in Table 7.

As is immediately evident, virtually all results are inferior to that of the evolved agents, with the exception of fixed agent 6, which managed to outperform the Index in two out of the three instances in addition to outperforming the majority of all evolved groups. However, on the whole it could be argued that employing one of the above fixed strategies without any learning does not appear to be very reliable and produces varied results.

5 Analysis and Discussion

Within test A, a significant number of groups outperformed the Index in both absolute performance

	Description	Test A	Test B	Test C
Fixed 1	Pursues a simple ‘no-trade’ strategy	00.00%	00.00%	00.00%
Fixed 2	Decision type 1 and risk-averse	-20.24%	-08.13%	86.16%
Fixed 3	Decision type 2 and risk-averse	-18.43%	04.84%	151.60%
Fixed 4	Decision type 3 and risk-averse	-19.71%	-09.98%	81.74%
Fixed 5	Decision type 4 and risk-averse	-21.04%	-16.62%	62.67%
Fixed 6	Decision type 1 and risk taking	-01.88%	25.24%	33.39%
Fixed 7	Decision type 2 and risk taking	-19.97%	04.59%	203.12%
Fixed 8	Decision type 3 and risk taking	-10.28%	09.00%	157.30%
Fixed 9	Decision type 4 and risk taking	-05.13%	07.28%	-03.72%

Table 7. Summary of test results for fixed agents.

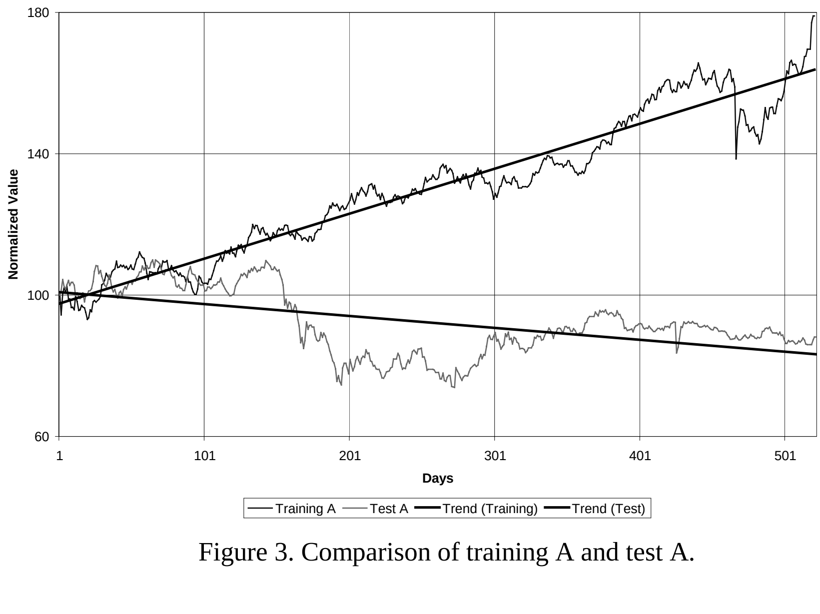

Figure 3. Comparison of training A and test A.

and overall performance. Most notable is the apparent better performance for groups from training C, which is surprising as H1 predicted groups from training A to fare better for test A. However, this claim does not appear substantiated in this instance, as not a single time does a group from training A outperform a group from trainings B or C. Furthermore, all groups from training A, apart one, are the only ones that fail to outperform the Index. Even when considered in terms of overall performance, groups from training A fare consistently worse than any other. This indicates that H1 appears to be false. However, to assess it correctly, the assumption that test A is a continuation of the environment from training A needs to be verified as well. As can be seen in Figure 3 using a normalized DAX graph and accentuated by trend lines, the environment for training and testing appear fundamentally different as agents from training A experienced a continuous growth environment while the test A environment is clearly downward trending. As hypothesized by H1, groups from training A fared worse given such a change in environment, but on the other hand no testing was done to assess whether it would have performed substantially better if no change in environment had occurred after training. Hence the validity of H1 remains inconclusive.

In test B, as for test A, groups trained from training A again fail to be very competitive. Additionally, there is no apparent superiority between groups from training B to groups from training C, in particular when considering both absolute and overall performances. For instance, group 5-C outperforms group 5-B, however in terms of overall performance 5-B is slightly larger than that for 5-C.

For test C the most significant difference to previous tests is the failure of any agent group to outperform the Index, and only when considering overall performance did groups 2-B, 3-C and 5-C post better results. However, groups from training C appear to perform better in the latter three experiments than those from training A or B. This is in support of H2, as it hypothesized that the most general solutions, in this case groups from training C, should fare best in the most varied environment, in this case test C.

Finally, several Mann-Whitney U tests [12] were used to statistically assess the hypotheses using overall performance. For every test, a comparison was made between groups trained on A to groups trained on C, groups trained on A to groups trained on B as well as groups trained on B to groups trained on C. In total 9 tests were conducted, with the null hypothesis stating that no difference exists between each group’s performance, while the alternative hypothesis stating that the latter class of group performed better. For example, when comparing groups from training A to groups from training C, if the null hypothesis is rejected, performance from groups trained on C will be considered significantly higher than that of groups from training A in a one-tailed test.

	A to C	A to B	B to C
Test A	0.00397	0.00397	0.07540
Test B	0.00397	0.00397	0.21032
Test C	0.00397	0.00397	0.21032

Table 8. Mann-Whitney U Test P values.

As seen in Table 8, groups from training B and C are always significantly different from those of training A when using a 0.01 level of significance, while there exists no significant difference between groups from B or C in any of the tests.

It is possible to infer from these results that extensive training periods appear more desirable than short-term training, indicating a high level of robustness and performance, as suggested in H2. However, as performance of groups from training B and C are not significantly different it can be inferred that H3 is true, as doubling data quantity from B to C had no significant effect.

6 Conclusions

The claim that an environment specific solution would perform best in a similar environment could not be substantiated by these results, as the test environment differed significantly from that found in training, only showing that a group trained for a specific environment did indeed not perform well in a different environment. Therefore H1 remains partly inconclusive. However, H2 appears valid as groups of agents with fairly general solutions managed to outperform groups with more specific solutions in tests where the trading environment varied significantly, regardless of any training related settings found in the five different experiments. In respect to H3, it can be concluded that it appears true, as increasing training data is subject to diminishing marginal returns on agents performance as no difference was exhibited by groups from training B or C, and a significant amount of training time could be saved if this is taken into account for future experiments.

These results in themselves are not astonishing, however they form the basis of some future work and present an interesting methodology following on from previously published material [10]. It will form the basis of further research into agent behavior and decision making in a financial market environment.

Acknowledgments

The authors would like to thank Dr. Djamila Ouelhadj for her contributions as well as the department of Computer Science and Information Technology in the University of Nottingham for the funding provided to conduct this research.

References

[1] S.B. Achelis, *Technical Analysis from A to Z*, McGraw-Hill, New York, 2000.

[2] T. Ankenbrand and M. Tomassini, “Agent based simulation of multiple financial markets”, *Neural Network World*, 1997, Vol. 4, pp. 397-405.

[3] W.B. Arthur, J. Holland, B. LeBaron, R. Palmer and P. Tayler, “Asset pricing under endogenous expectations in an artificial stock market”, *The Economy as an Evolving Complex System II*, Addison-Wesley, Reading, 1997, pp. 15-44.

[4] W. Banzhaf, P. Nordin, R.E. Keller, and F.D. Francone, *Genetic Programming - An introduction: on the automatic evolution of computer programs and its applications*, Kaufmann, San Francisco, 1998.

[5] J. Branke, "Evolutionary Approaches to Dynamic Optimization Problems - A Survey", *GECCO Workshop on Evolutionary Algorithms for Dynamic Optimization Problems*, 1999, 134-137.

[6] J. Ferber, “Multi-Agent Systems - An Introduction to Distributed Artificial Intelligence”, Addison-Wesley, 1999.

[7] D. Goldberg, *Genetics Algorithm in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, USA, 1989.

[8] M. Kearns and L. Ortiz, “The Penn-Lehman Automated Trading Project”, *IEEE Intelligent Systems*, 2003.

[9] B. LeBaron, “Agent-based computational finance: Suggested readings and early research”, *Journal of Economic Dynamics and Control*, Elsevier, 2000, Vol. 24, pp. 679-702.

[10] C. Schoreels, B. Logan, J.M. Garibaldi, “Agent based genetic algorithm employing financial technical analysis for making trading decisions using historical equity market data”, To appear in Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology 2004, Beijing, China.

[11] S. Schulenburg and P. Ross, “An Adaptive Agent Based Economic Model”, *Learning Classifier Systems: From Foundations to Applications*, Springer-Verlag, Berlin, 2000, Vol. 1813, pp. 265-284.

[12] S. Siegel, *Nonparametric statistics for the behavioral sciences*, McGraw-Hill, New York, 1956.

[13] F. Remacle, “Straight Through Processing: A Forum for Change”, Financial Publishing International, London, UK, 2000.

[14] M. Wooldridge, *An Introduction to MultiAgent Systems*, Wiley, 2002.