The Market Fraction Hypothesis under different GP algorithms

Michael Kampouridis
School of Computer Science and Electronic Engineering
University of Essex, Wivenhoe Park, CO4 3SQ, UK

Shu-Heng Chen
AI-Econ Center, Department of Economics
National Cheng Chi University, Taiwan 11623

Edward Tsang
School of Computer Science and Electronic Engineering
University of Essex, Wivenhoe Park, CO4 3SQ, UK

Abstract

In a previous work, inspired by observations made in many agent-based financial models, we formulated and presented the Market Fraction Hypothesis, which basically predicts a short duration for any dominant type of agents, but then a uniform distribution over all types in the long run. We then proposed a two-step approach, a rule-inference step and a rule-clustering step, to testing this hypothesis. We employed genetic programming as the rule inference engine, and applied self-organizing maps to cluster the inferred rules. We then ran tests for 10 international markets and provided a general examination of the plausibility of the hypothesis. However, because of the fact that the tests took place under a GP system, it could be argued that these results are dependent on the nature of the GP algorithm. This chapter thus serves as an extension to our previous work. We test the Market Fraction Hypothesis under two new different GP algorithms, in order to prove that the previous results are rigorous and are not sensitive to the choice of GP. We thus test again the hypothesis under the same 10 empirical datasets that were used in our previous experiments. Our work shows that certain parts of the hypothesis are indeed sensitive on the algorithm. Nevertheless, this sensitivity does not apply to all aspects of our tests. This therefore allows us to conclude that our previously derived results are rigorous and can thus be generalized.

Introduction

There are several types of models in agent-based financial markets literature. One way of categorizing them is to divide them into the N-type models and the Santa-Fe Institute (SFI) like ones (Chen, Chang, & Du, 2010). The former type of model focuses on the mesoscopic level of markets, by allowing agents to choose between different types of strategies. A typical example is the fundamentalist-chartist model. Agents in this model are presented by these two strategy types and at any given time they have to choose between these two. Examples of such N-type models in the literature are (Brock & Hommes, 1998; Amilon, 2008; Kirman, 1991, 1993; Lux, 1995, 1997, 1998; Winker & Gilli, 2001; Boswijk, Hommes, & Manzan, 2007). A typical area of investigation, using these models, is fraction dynamics, i.e., how the fractions of the different strategy types change over time. However, what is not presented in most of these models is novelty-discovering agents. For instance, in the fundamentalist-chartists example, agents can only choose between these two types; they cannot create new strategies that do not fall into either of these types. On the other hand, the SFI-like models overcome this problem by focusing on the microscopic level of the markets. By using tools such as Genetic Programming (Koza, 1992), these models allow the creation and evolution of novel agents, which are not constrained by pre-specified strategy types. Such examples from the literature are (LeBaron, Arthur, & Palmer, 1999; Chen & Yeh, 2001; Arifovic & Gencay, 2007; Martinez-Jaramillo & Tsang, 2009).[1] However, these kinds of models tends to focus on price dynamics, rather than fraction dynamics (Chen et al., 2010).

In a previous work (Chen, Kampouridis, & Tsang, 2011; Kampouridis, Chen, & Tsang, 2009), we combined properties from the N-type and SFI-like models into a novel financial model. We first used Genetic Programming (GP) as a rule inference engine, which created and evolved autonomous agents; we then used Self-Organizing Maps (SOM) (Kohonen, 1982) as a clustering machine, and thus re-created the mesoscopic level that the N-type models represent, where agents were categorized into different strategy types. This allowed us to test for the plausibility of the *Market Fraction Hypothesis (MFH)*(Chen et al., 2010, 2011), which states that the fractions of trading strategy types that exist in a financial market constantly change over time.

However, because of the fact that an important part of our financial model was based on GP, we are interested in examining whether our previous derived results can hold under different GP algorithms. Therefore in this chapter, we test the MFH under two different GP algorithms, in order to prove that the previous results are *rigorous* and are thus not dependent on the choice of the GP algorithm. Proving this is a very important task, because it allows us to generalize our previously derived results.

The rest of this chapter is organized as follows: we first present the MFH, along with its testing methodology, and discuss the two different GP algorithms that are going to be used for the tests in this chapter. We then present the experimental designs, and also present and discuss the results of our experiments. Finally, we conclude this chapter and discuss future work.

The Market Fraction Hypothesis

Within a market there exist different types of trading strategies. For instance, the fundamentalist-chartist model informs us that in the market there are two types of

[1] We refer the reader to (Chen et al., 2010), which provides a thorough review on both N-type and SFI-like models, along with a detailed list of them.

strategies, the fundamental and the chartist. The MFH tells us that the fraction among these types of strategies keeps changing (swinging) overtime. The following two statements are the basic constituents of the MFH, as presented in (Chen et al., 2011; Kampouridis et al., 2009).

1. *In the short run, the fraction of different clusters of strategies keeps swinging over time, which implies a short dominance duration for any cluster.*

2. *In the long run, however, different clusters are equally attractive and thus their market fractions are equal.*

The first statement means that it is not possible for a single strategy type to dominate the market by attracting an overwhelming fraction of market participants for many consecutive periods. In other words, according to the MFH there is no such thing as a 'winner type'. Thus, an ex ante characterization of winners simply does not exist. Let us give an example by again using the fundamentalist-chartist model. If at time t fundamentalists dominate the market, the first MFH property says that this should not happen for too long. Eventually there should be a "switch", and chartists would take over as the dominant strategy in the market. The term 'dominance' will become technical for testing the MFH, and we shall make it precise later in the chapter.

Let us now move to the second statement. If the above continuous happening, then in the long run both fundamentalists and chartist should have occupied about the same market share, i.e., about one half.[2]

What we shall do in this chapter is to test the above two MFH properties against our empirical data, under the two new GP algorithms, and then compare these results with the ones from the original chapter (Chen et al., 2011; Kampouridis et al., 2009). The methodology followed in this chapter is the same as in the original one: we first infer trading strategies over different time periods. Afterwards, each strategy is clustered with Self Organizing Maps (SOM) (Kohonen, 1982). Each cluster denotes a trading strategy type. Finally, we test the above two statements.

Tools

This section presents the two main tools used in order to test the MFH. These two tools are GP and SOM. There is also a third tool, which we created for the purposes of these tests, the Time-Invariant SOM. However, as this tool is derived from SOM, it will be presented later in this chapter. The next sections present the GP and SOM tools.

GP algorithms

In this chapter, we use two different versions of a GP algorithm, namely EDDIE 7 and EDDIE 8, for testing the Market Fraction Hypothesis. Both of these versions should be considered as extensions and improvements of the *simple GP* presented in Chen et al. (2011); Kampouridis et al. (2009). For this reason, we first present the *simple GP*. We then

[2]This idea is first made rigorous by Kirman (Kirman, 1993), who attempted to solve a puzzling entomological problem, i.e., ants swinging among themselves within two identical sources of food.

continue by presenting EDDIE 7 and EDDIE 8, and how they have extended the *simple GP*.

Simple GP algorithm. The *simple GP* was inspired by a financial forecasting tool, EDDIE (Tsang, Li, & Butler, 1998; Tsang et al., 2000; Li & Tsang, 1999; Li, 2001), which learns and extracts knowledge from a set of data. This set of data is composed of the daily closing price of a stock, a number of attributes and signals. The attributes are indicators commonly used in technical analysis (Edwards & Magee, 1992) and are considered by the user to be relevant to the prediction.[3] Table 1 presents the technical indicators that our *simple GP* used.[4]

Table 1: Technical Indicators used by the *simple GP*. Each indicator uses 2 different periods, 12 and 50, in order to take into account a short-term and a long-term period. Formulas of our interpretation for these indicators are provided in the appendix.

Technical Indicators (Abbreviation)	Period
Moving Average (MA)	12 & 50 days
Trade Break Out (TBR)	12 & 50 days
Filter (FLR)	12 & 50 days
Volatility (VIX)	12 & 50 days
Momentum (Mom)	12 & 50 days
Momentum Moving Average (MomMA)	12 & 50 days

The signals are calculated by looking ahead of the closing price for a time horizon of n days, trying to detect if there is an increase of the price by $r\%$ (Tsang et al., 2000). For this set of experiments, n was set to 1 and r to 0. In other words, the GP was trying to use some of the indicators above in order to forecast whether the daily closing price was going to increase in the following day.

Furthermore, Figure 1 presents the Backus Naur Form (BNF) (Backus, 1959) (grammar) of the GP. As we can see, the root of the tree is an If-Then-Else statement. Then the first branch is a boolean (testing whether a technical indicator is greater than/less than/equal to a value). The 'Then' and 'Else' branches can be a new Genetic Decision Tree (GDT), or a decision, to buy or not-to-buy (denoted by 1 and 0).

Thus, each individual in the population is a GDT and its output is a recommendation to buy or not-buy. Each GDT's performance is evaluated by a fitness function, presented below.

If the prediction of the GDT is positive (1), and also the signal in the training data for this specific entry is also positive (1), then this is classified as True Positive (TP). If the

[3]'User' here refers to the user of the GP system. Such a user could, for example, be a financial expert, who uses the system as an advisor for his predictions in the stock market.

[4]We use these indicators because they have been proved to be quite useful in developing investment opportunity decision rules for forecasting rises and drops of the price in previous works like (Allen & Karjalainen, 1999), (Austin, Bates, Dempster, Leemans, & Williams, 2004) and (Martinez-Jaramillo, 2007). Of course, there is no reason not to use other information like fundamentals or limit order book information. However, the aim of this work is not to find the ultimate indicators for financial forecasting.

<Tree> ::= If-then-else <Condition> <Tree> <Tree> | Decision
<Condition> ::= <Condition> “And” <Condition> |
<Condition> “Or” <Condition> |
“Not” <Condition> |
Variable <RelationOperation> Threshold
<Variable> ::= MA_12 | MA_50 | TBR_12 | TBR_50 | FLR_12 |
FLR_50 | VIX_12 | VIX_50 | Mom_12 | Mom_50 |
MomMA_12 | MomMA_50
<RelationOperation> ::= “>” | “<” | “=”
Decision is an integer, Positive or Negative implemented
Threshold is a real number

Figure 1. The Backus Naur Form of the *simple GP*

prediction is positive (1), but the signal is negative (0), then this is False Positive (FP). On the other hand, if the prediction is negative (0), and the signal is positive (1), then this is False Negative (FN), and if the prediction of the GDT is negative (0) and the signal is also negative (0), then this is classified as True Negative (TN). These four together give the familiar confusion matrix (Provost & Kohavi, 1998), which is also presented in Table 2.

Table 2: Confusion Matrix

	Actual Positive	Actual Negative
Positive Prediction	True Positive (TP)	False Positive (FP)
Negative Prediction	False Negative (FN)	True Negative (TN)

As a result, we can use the following 3 metrics:

Rate of Correctness

$$RC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Rate of Missing Chances

$$RMC = \frac{FN}{FN + TP} \tag{2}$$

Rate of Failure

$$RF = \frac{FP}{FP + TP} \tag{3}$$

Li (Li, 2001) combined the above metrics and defined the following fitness function:

$$ff = w_1 * RC - w_2 * RMC - w_3 * RF \tag{4}$$

where w_1, w_2 and w_3 are the weights for RC, RMC and RF respectively. Li states that these weights are given in order to reflect the preferences of investors. For instance, a conservative investor would want to avoid failure; thus a higher weight for RF should be used. However, Li also states that tuning these parameters does not seem to affect the performance of the GP. For our experiments, we chose to include strategies that mainly focus on correctness and reduced failure. Thus these weights have been set to 1, 1/6 and 1/2 respectively, and are given in this way in order to reflect the importance of each performance measure for our predictions.

This concludes our brief presentation on the *simple GP*. Let us now continue by presenting EDDIE 7 and how it extends its predecessor.

EDDIE 7. EDDIE 7 extends the *simple GP* by the introduction of a constrained fitness function, which *allows to achieve lower RF at price of a higher RMC.* The effectiveness of this constrained fitness function has been discussed in details in (Li, 2001; Tsang, Markose, & Er, 2005). The constraint is denoted by R, which consists of two elements represented by percentage, given by

$$R = [Cmin, Cmax],$$

where $C_{min} = \frac{P_{min}}{N_{tr}} \times 100\%$, $C_{max} = \frac{P_{max}}{N_{tr}} \times 100\%$, and $0 \leq C_{min} \leq C_{max} \leq 100\%$. N_{tr} is the total number of training data cases, P_{min} is the minimum number of positive position predictions required, and P_{max} is the maximum number of positive position predictions required.

Therefore, a constraint of $R = [50, 65]$ would mean that the percentage of positive signals that a GDT predicts[5] should fall into this range. When this happens, then w_1 remains as it is (i.e. 1 in our experiments). Otherwise, w_1 it takes the value of zero.

The BNF grammar of EDDIE 7 and the rest of the algorithm characteristics remain the same as in the *simple GP*.

This concludes this short presentation of EDDIE 7. However, because of the fact that both EDDIE 7 and the *simple GP* used pre-specified periods, we are also going to use another algorithm, EDDIE 8, which allows the GP to search in the search space of the technical indicators. This thus allows the GP to come up with solutions that its predecessors cannot. The next section explains how EDDIE 8 does this.

EDDIE 8. Let us consider a function $[y = f(x)]$, where the input x is the indicators and the output y is the prediction made by EDDIE. The function f is unknown to the user and is the GDTs that EDDIE generates in order to make its prediction. As we said earlier, EDDIE 7 and the *simple GP* uses a number of indicators, with different pre-specified periods (12 and 50 days). This therefore means that the input x consists of constants. EDDIE 8 uses another function $z = g(x)$, which determines which indicator and which period should be used. EDDIE 8 is not only searching in the space of GDTs, but also in the space of indicators. It can thus return Genetic Decision Trees (GDTs) that are using any period within a range that is defined by the user.

[5]As we have mentioned, each GDT makes recommendations of buy (1) or not-to-buy (0). The former denotes a positive signal and the latter a negative. Thus, within the range of the training period, which is t days, a GDT will have returned a number of positive signals

```
<Tree> ::= If-then-else <Condition> <Tree> <Tree> | Decision
<Condition> ::= <Condition> "And" <Condition> |
                <Condition> "Or" <Condition> |
                "Not" <Condition> |
                VarConstructor <RelationOperation> Threshold
<VarConstructor> ::= MA period | TBR period | FLR period | VIX period |
                     Mom period | MomMA period
<RelationOperation> ::= ">" | "<" | "="
Terminals:
          MA, TBR, FLR, VIX, Mom, MomMA are function symbols
          Period is an integer within a parameterised range, [MinP, MaxP]
          Decision is an integer, Positive or Negative implemented
          Threshold is a real number
```

Figure 2. The Backus Naur Form of EDDIE 8

As we can see from the new syntax at Figure 2, there is no such thing as a *Variable* symbol in EDDIE 8. Instead, there is the *VarConstructor* function, which takes two children. The first one is the indicator, and the second one is the Period. Period is an integer within the parameterized range [MinP, MaxP] that the user specifies.

As a result, EDDIE 8 can return decision trees with indicators like 15 days Moving Average, 17 days Volatility, and so on. The period is not an issue anymore, and it is up to EDDIE 8, and as a consequence up to the GP and the evolutionary process, to decide which lengths are more valuable for the prediction.

The immediate consequence of this is that now EDDIE 8 is not restricted only to the 12 indicators that EDDIE 7 and the simple GP use (which are still part of EDDIE 8's search space); on the contrary, it now has many more options available, thanks to this enlarged search space. If the reader is interested in a comparative analysis between EDDIE 7 and EDDIE 8, we refer him to Kampouridis and Tsang (2010). Let us now continue by presenting the other main tool used for the MFH experiments, which is Self Organizing Maps.

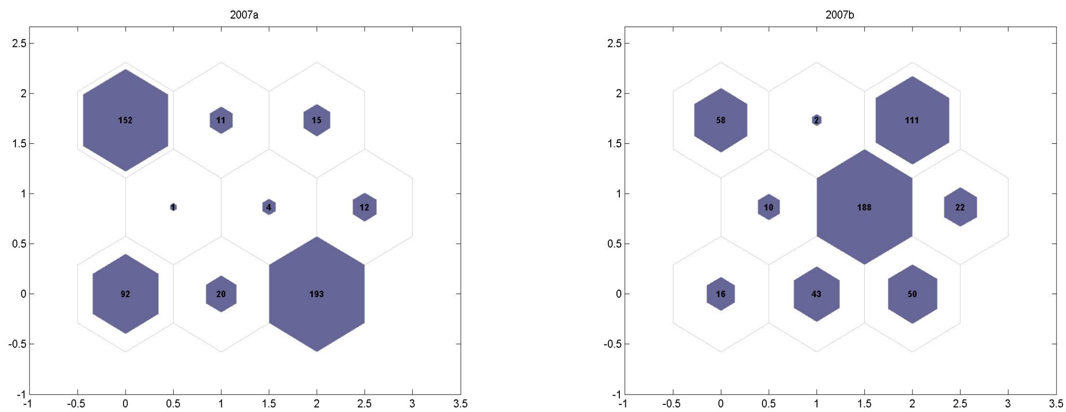
Self Organizing Maps (SOM)

In the previous section we mentioned that each GDT is a trading strategy. If, for instance the population of GDTs is 500, this means we have potentially 500 trading strategies. However, this does not mean they are all completely different from each other. We know very well from the GP bibliography (Koza, 1992; Poli, Langdon, & McPhee, 2008) that towards the end of the training period, the population might have converged and thus some trees might be the same. Moreover, it is much easier for computing purposes to have types of strategies, rather than individual ones. In addition, BH also had types of strategies/beliefs, rather than having a pool with many strategies. For these reasons, the 500 trading strategies derived from the GP were classified into 9 clusters.[6] In order to classify

[6]The number of clusters at this point was set arbitrarily. Later in this work we examine the sensitivity of the results if we tune this number.

them, we used 3×3 Self-Organizing Feature Maps (SOM).

SOM is a type of artificial neural networks that is trained using unsupervised learning, in order to return a low-dimensional representation of the input layer, which in our case is the recommendations of the GDTs. Associated with each cluster is a weight vector, which has the same dimensions as the input data. During this procedure the centroid of each cluster (hence the membership of each instance) is dynamically adjusted via a competitive learning process. Eventually, the whole population of GDT recommendations is assigned to different clusters and this is how we classify the trading strategies. Thus, the SOM output will be 9 neurons (or clusters) in a two-dimensional lattice, presenting the input data in an organised way, so that the similar strategies are clustered together.



(a) First semester of the year 2007

(b) Second semester of the year 2007

Figure 3. Two self-organized map constructed from the rules inferred using the daily data of TAIEX, the first half (the left panel) and the second half (the right panel) of 2007, respectively.

The main advantage of SOMs over other clustering techniques, such as K-means (MacQueen, 1967), is that the former can present the result in a visualizable manner so that we can not only identify these types of traders but also locate their 2-dimensional position on a map, i.e., a distribution of traders over a map. This becomes more obvious if we see Figure 3, which presents the results after running 3×3 SOM for a population of 500 individuals for the daily TAIEX[7] data for the first and second semester of 2007. As we can see in these cases, there are usually a few strategies that are occupying the majority of the population, whereas the rest of the strategies have significantly less members. We can also observe how the market fraction dynamics change from period to period.

Experimental Designs

This section summarizes the experimental designs. The experiments were conducted for a period of 17 years (1991-2007) and the data was taken from the daily closing prices of 10 international market indices. These 10 markets are: CAC 40 (France), DJIA (USA), FTSE 100 (UK), HSI (Hong Kong), NASDAQ (USA), NIKEI 225 (Japan), NYSE (USA),

[7]Taiwan Stock Exchange Capitalization Weighted Stock Index. Available from http://finance.yahoo.com

S&P 500 (USA), STI (Singapore) and TAIEX (Taiwan). For each of these markets, we run each experiment 10 times. To make it easier for the reader, we first present the testing methodology and results for a single run of the TAIEX dataset. Figure 4 presents the daily closing price of TAIEX. We then proceed with presenting summary results over the 10 runs for all datasets.

Each year was split into 2 halves (January-June, July-December), so in total, out of the 17 years, we have 34 periods.[8] The GP was therefore implemented 34 times. Table 3 presents the GP parameters for our experiments, which are the same as the ones in the previous chapter. The behavior of each GDT can be represented by its series of market timing decisions over the entire trading horizon. Therefore, the behaviour of each rule is a binary string or a binary vector of 1s and 0s (buy and not-to-buy). The length or the dimensionality of these strings or vectors is then determined by the length of trading horizon, which in this study is 6 months, i.e., 125 days long; hence, the market timing vector has 125 dimensions. Once each trading rule is concretized into its market timing vector, we can then cluster these rules by applying SOM to the associated clusters.

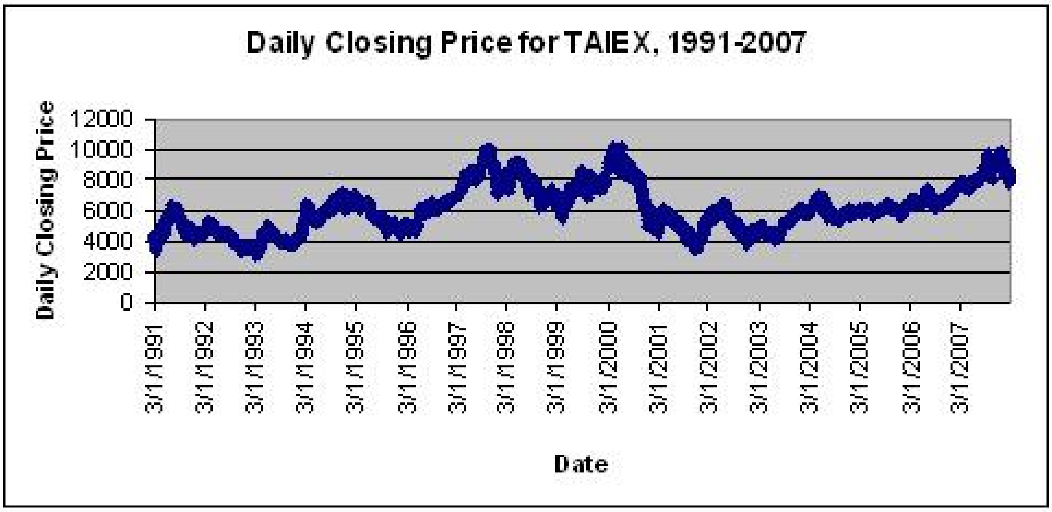


Figure 4. Daily Closing Price for TAIEX:1991-2007

Here we should emphasize that the GP was only used for creating and evolving the trading strategies. No validation or testing took place, as happens in the traditional GP approach. The reason for this is that we were not using the GP for forecasting purposes; instead, we were interested in using the GP as a *rule inference engine* which would evolve profitable trading strategies for a certain period of time. To be more specific, the GP was used for each of the 34 periods to create and evolve trading strategies. After the evolution of the strategies under a specific period, these strategies are not tested against another set. This approach is consistent with Lo's Adaptive Market Hypothesis (Lo, 2004, 2005), which

[8] At this point the length of the period was chosen arbitrarily to 6 months. We leave it to a future research to examine if and how this time horizon can affect our results.

Table 3: GP Parameters. The GP parameters for our experiments are the ones used by Koza. Only the tournament size has been changed (lowered), because we were observing premature convergence. Other than that, the results seem to be insensitive to these parameters.

GP Parameters	
Max Initial Depth	6
Max Depth	17
Generations	50
Population size	500
Tournament size	2
Reproduction probability	0.1
Crossover probability	0.9
Mutation probability	0.01
$\{w_1, w_2, w_3\}$	$\{1, \frac{1}{6}, \frac{1}{2}\}$
Period (EDDIE 8)	[2, 65]

states that the heuristics of an old environment are not necessarily suited to the new ones.[9] Furthermore, our no-testing approach is also consistent with the well-tested *overreaction hypothesis* (De Bondt & Thaler, 1985), which essentially states that top-ranked portfolios are outperformed by bottom-ranked portfolios during the next period. Thus, after evolving a number of generations (50 in this chapter), what stands (survives) at the end (the last generation) is, presumably, a population of financial agents whose market-timing strategies are financially rather successful. This population should, therefore, interest us in the spirit of Lo's adaptive market process; therefore, we use them to test how those competitive strategies perform in the future periods.

Finally, as we mentioned earlier, we used 3×3 SOM; therefore, we obtained a total 34 SOM (one per period), with 9 clusters each. In other words, in every period the 500 GDTs were placed in one of the nine clusters (i.e. the categories of the trading strategies) of that SOM. From this point on, whenever we use the term 'trading strategy type' we will be referring to one of the nine clusters and each GDT will be a member of one of these nine clusters.

Testing Methodology

After having presented the necessary tools and the experimental designs, we can now proceed to present the testing methodology. Our methodology consists of three parts: GP, SOM and the time-invariant SOM.

Let us start with GP. As we have already seen, we have used a simple GP in order to generate and evolve trading strategies. However, there is a problem with comparing trading strategies for different periods. This happens because we cannot compare the fitness

[9] Lo names this possible situation as "maladaptive". He also uses the example of the flopping of a fish for a better understanding of behaviors under different environments: on dry land the flopping might seem meaningless, but under water, it is the flopping that protects the fish from its enemies.

function of a trading strategy (GDT) from one period with the fitness function of a strategy (GDT) for another period, since they were presented with different datasets (environments).

This also applies to clusters' comparison among different SOMs. *Maps over time are not directly comparable*. In order to better understand this, consider Figure 3. The way SOM works is that it creates the clusters after it is given a specific population of, in this case, GDTs. When we have different periods with different populations, the nine clusters from different periods will generally be different, because they represent different populations of investment behaviors generated by different data environments. For example name the bottom-left cluster of each SOM (Figure 3) as 'Cluster 1'. Then we are saying that 'Cluster 1' of the SOM derived using the data of 2007a in general will not be the same as 'Cluster 1' of the SOM derived from the data using 2007b. Quite likely, they have different centroids (weighting vectors), representing different investment behaviors. This, therefore, makes the strategy types incomparable crossing different periods.

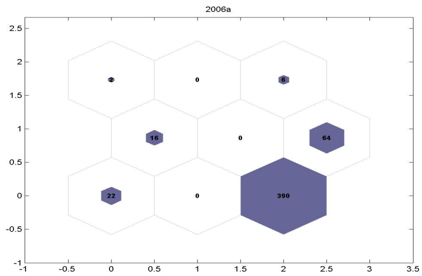
In order to tackle this problem, we introduce a time-invariant SOM based on the idea of *emigrating* and *reclustering*, which is the third and last part of our testing methodology. The remainder of this section thus presents these "translations" needed in order to make SOM from different periods comparable so as to our tests for the MFH.

Emigrating As we just mentioned, after obtaining the trading strategies from GP, we cannot directly compare them with strategies from other periods, because each period has its own dataset. Therefore it is necessary to have all periods use the same dataset as a common base. In other words, all GDTs that derived from each period emigrate to a common base period. What we mean by "emigrating" is that all GDTs from each period are applied to the dataset of the common base period. Therefore, after applying a GDT to the new dataset, new signals are created. In this way, the behavioral vectors of all GDTs derived from different periods are re-built based on the same ground and hence become comparable. In this chapter, we choose the second half of 2007 (2007b) as the base period.[10]

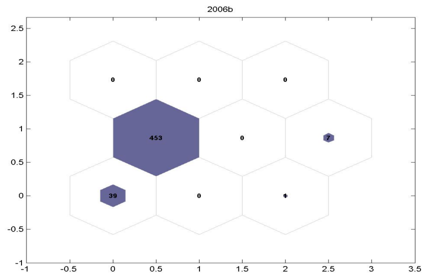
Reclustering Reclustering or time-invariant SOM is the second part of the translations, which allows SOM clusters to be compared throughout different periods. We again use 2007b as the common base period. This time, we keep the centroids of the clusters, originally derived from the common base period, fixed and assign the behavior vectors from other periods (emigrated GDTs) to one of the clusters of 2007b. This reclustering is conducted in the following way: the behavior vector of each emigrated GDT is compared to each centroid of the nine clusters and it will then be assigned to the one with minimum Euclidean distance. We do this period by period from 1991a to 2007a. 33 SOM are constructed in this way[11], and now these SOM can be directly compared to each other given that they all share the same centroids. Figure 5 presents 4 of these 34 SOMs. We will then use them to test Hypothesis 1 and 2.

[10]The base period was chosen arbitrarily. However, which base period is chosen does not affect the results.

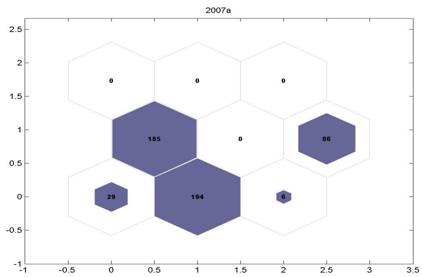
[11]2007b does not need reclustering, since we use it as the base period.



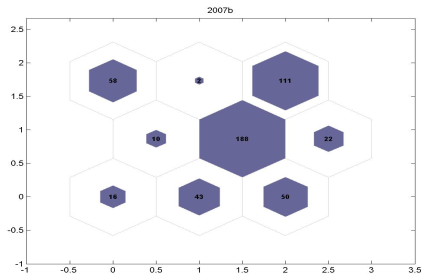
(a) First semester of the year 2006



(b) Second semester of the year 2006



(c) First semester of the year 2007



(d) Second semester of the year 2007

Figure 5. SOM of Trading Strategies After "Translating": Samples from TAIEX Cells (1,1), (1,2), (2,1) and (2,2) correspond to period 2006a, 2006b, 2007a, and 2007b, respectively. Except for the last one, 2007b, the other three are reconstructed by using 2007b as the base period.

Results

This section presents the results of our tests. In addition, because we do not know how many types of strategies exist in the markets, we also experiment with the number of clusters. Thus, we repeat the two Tests for 8 times, for clusters 2-9. Furthermore, since we are interested in comparing our results with the ones under *simple GP*, let us briefly mention those original results from (Chen et al., 2011; Kampouridis et al., 2009): overall, evidence for the MFH was weak for the majority of the datasets tested. More specifically, Test 1 showed that a small number of clusters drives the short-dynamics further away from the expectations of the MFH. Thus, evidence for the MFH gets stronger as the number of clusters increases. On the other hand, Test 2 showed the opposite picture: MFH evidence gets weaker, when the number of clusters increases. More details about each Test's individual results follow in their respective sections. Let us now see what happens if instead of the *simple GP*, we test the MFH under EDDIE 7 and EDDIE 8.

Test 1-The Short-Run Test

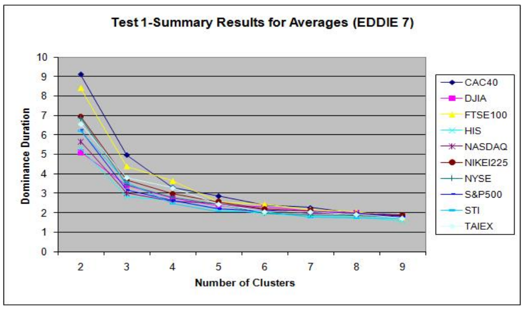
The first hypothesis regards the short-run dynamics of market fraction. Each type of trader can be a dominant group (majority) for some of the time, but the duration of its dominance can only be temporal. It is desirable to have an objective measure of how persistent a dominant cluster can be. To do so, we need an operational meaning of dominance. Even though there is no unique way of doing this, we find the following threshold to be quite general and useful.

$$DS = \frac{1+p}{N+p}, \tag{5}$$

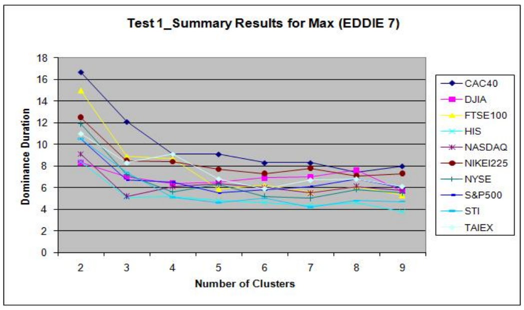
where DS denotes a threshold, N the number of clusters and p is a free parameter to manipulate the degree of dominance. By varying the parameter p, one can therefore have an operational meaning that is consistent with our intuition regarding dominance. For example, if $N = 2$ (a two-type model) and $p = 2$, a cluster can be dominant only if its market fraction is greater than a DS of 75% , a standard much higher than just breaking the tie (one half). Clearly, the higher the p, the higher the threshold. We set $p = 2$, as in the original chapter.

Furthermore, we need to be precise on what we mean by *short duration* for a dominant type. Here, any specific number may be arbitrary; after all, short is only a matter of degree. We, therefore, first present the statistics of duration observed for each type. Figure 6 presents the averages, over 10 runs, for the average and maximum dominance duration for number of clusters 2-9. *Average duration* refers to the average number of consecutive periods that the clusters of a certain period remained dominant. If for instance we have $N = 3$ clusters and Cluster 1 remained dominant for 5 periods, Cluster 2 for 8, and Cluster 3 for 8, then the *average duration* would be equal to $(5 + 8 + 8)/3 = 21/3 = 7$ periods. Regarding the *maximum duration*, the answer would be 8 in the example above. After this process is repeated for all periods and for all 10 runs, we then calculate the averages of *average duration* and *maximum duration.* What we observe in Figure 6 is that the dominance duration decreases as the number of clusters increases, for both EDDIE 7 (top) and EDDIE 8 (bottom). This suggests that while we move from lower to higher numbers of clusters, the empirical results show stronger evidence of the MFH. Of course, we should not neglect that both maximum duration graphs suggest that there is always a cluster that remains dominant for a number of periods greater than the average. Thus the short-run dynamics indicates the appearance of long-lasting dominant clusters. Furthermore, what seems to happen for all datasets is that there are always a few clusters that have strong dominance over the 34 periods, whereas the rest have very low. The low average dominance duration we see for a high number of clusters can be therefore explained by the extremely low dominance duration of the majority of clusters.

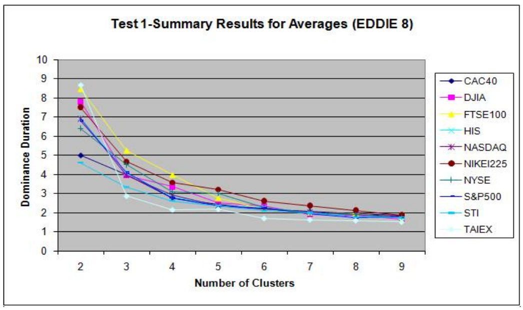
It should also be noted that these observations are inline with the original ones from Chen et al. (2011); Kampouridis et al. (2009). However, it should be stated that the GP algorithms in this chapter seem to have slightly worsened the results. Especially in the case of average maximum duration, both EDDIE 7 and 8 results show a significant increase in the duration period. In the original paper, results started in the range of 4-11 periods for the 2×1 SOM, for all datasets, and finished in the range of 3-8 periods, under the 3×3 SOM. However, the results in this chapter start in the range of 8-17 (EDDIE 7) and 7-15 (EDDIE 8) periods, and finish in the range of 4-9 (EDDIE 7) and 3-8 (EDDIE 8). What we can thus conclude from this test is that both versions of EDDIE have increased the dominance duration of the few clusters with strong dominance.



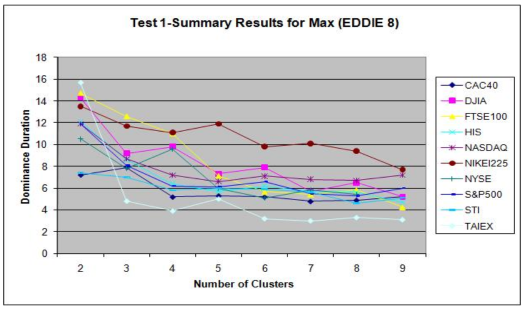
(a) Test 1-Average-EDDIE 7



(b) Test 1-Max-EDDIE 7



(c) Test 1-Average-EDDIE 8



(d) Test 1-Max-EDDIE 8

Figure 6. Test 1: Summary Results for Averages (left) and Maximums (right). The x-axis presents the summary results under different numbers of clusters. The graphs at the top present the EDDIE 7 results, whereas the ones at the bottom present the EDDIE 8 results.

Test 2-The Long-Run Test

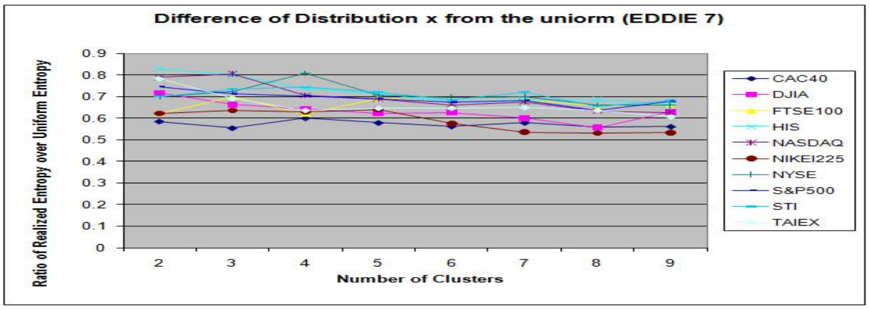
The second hypothesis that we can form relates to the long-term distribution of market fraction behavior. Here we expect that the long-term market fraction is even. In other words, if we have N types of traders, their long-term frequency of appearance should be close to $\frac{1}{N}$. Let $Card_{it}$ be the number (cardinality) of traders in Cluster i in time period t.

$$\sum_{i=1}^{N} Card_{it} = M, \forall t \tag{6}$$

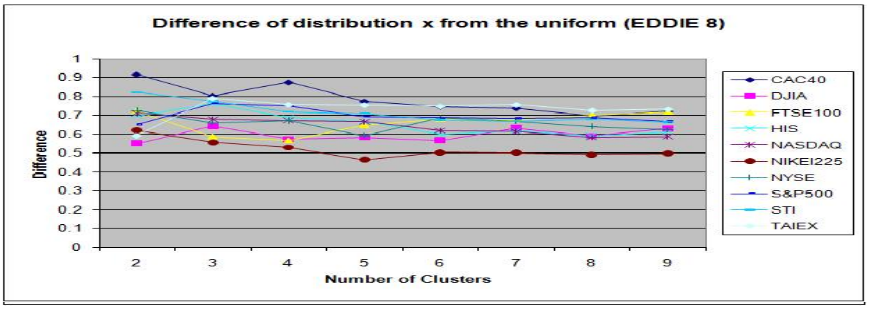
In our current setting, M, the total number of traders, is 500. The long-term histogram can be derived by simply summing the number of traders over all periods and dividing it by a total of $M \times T$ (# of periods),

$$w_i = \frac{\sum_{t=1}^{T} Card_{it}}{M \times T} \tag{7}$$

What we want to investigate here is whether the clusters have equal fractions in the long run. In order to test this, we need to observe the empirical distribution of w_i. Let us denote this distribution as f_X. The optimal case is of course when all fractions' cardinality is exactly equal. Let us call this distribution f_Y, which is the uniform one; thus, $f_Y = \frac{1}{N}$,



(a) Test 2-EDDIE 7



(b) Test 2-EDDIE 7

Figure 7. Test 2: Difference of distribution f_X from the uniform distribution for EDDIE 7 (top) and EDDIE 8 (bottom). The x-axis presents the results under different numbers of clusters.

where N is the number of clusters. In order to give a measure on how far f_X is away from the uniform, we simply use *entropy* as a metric. Hence what we do is to calculate the entropy of both distributions. For the discrete random variable, the entropy is defined as:

$$Entropy = -\sum_{i=1}^{N} P_i \ln P_i \tag{8}$$

where P_i is the fraction of each cluster. It is well known that for the uniform distribution $Entropy(Y) = \ln N$. Thus, the closer $Entropy(X)$ is to $Entropy(Y)$, the closer X is to the uniform distribution. After calculating X's entropy, we then find its difference from the uniform distribution. After obtaining the entropies over 10 runs for each dataset, we first calculate the average of these runs. We then divide each one of these averages with the benchmark entropy and thus obtain 10 different ratios (one per dataset). Of course, this ratio is maximized when the two entropies are equal, and therefore their ratio is equal to 1. Thus, the higher the ratio, the closer to the uniform the empirical distribution will be. Figure 7 presents these ratios for all datasets.

What we observe from this figure is that the ratios tend to decrease as the number of clusters increases (thus, the difference between the two distributions increases). This observation holds for both EDDIE 7 and 8. Furthermore, when only a small number of strategy types exist, we get closer to the uniform distribution, which implies that in the long run the clusters are equally competitive. Such a high diversity of the two distributions

Table 4: Minimum number of clusters required to pass the fraction thresholds of 90, 95, and 99%

		CA40	DJIA	FTSE100	HSI	NASDAQ	NIKEI	NYSE	S&P	STI	TAIEX
EDDIE 7	90%	4	4	4	4	4	4	5	5	5	4
	95%	5	5	5	5	5	5	6	6	6	5
	99%	7	7	8	7	7	7	8	8	8	8
EDDIE 8	90%	5	5	5	4	4	3	4	5	4	5
	95%	6	6	6	5	5	5	5	6	5	6
	99%	8	8	8	7	7	7	7	7	7	8

(especially with a high number of clusters) indicates again that the strong dominance for a few clusters continues to exist, even in the long run. Therefore, after combining Tests 1 and 2, we can have a quite clear picture. Clusters tend to dominate for long periods and this dominance is usually interchanged between a few clusters.

To make this argument even clearer, we also present Table 4, which shows the minimum number of clusters required to pass certain fraction thresholds. The three targeted values given in the table are 90%, 95% and 99%. Since the purpose is to see whether only a small number of clusters is required, we started with a larger number clusters, namely, nine, and see how much reduction we can make. If the target is to cover 90% of the market participants, then most markets need four to five types, and if target is even higher to 95%, then most markets need five to six types. This basically agrees with our previous findings that only a few clusters dominate the market. The fractions of the remaining clusters are marginal.

The above observations are also very similar to those made in the original paper (Chen et al., 2011). However, once again the results in this chapter seem to have got slightly worse. In our previous work, the above ratios for low SOM dimensions (e.g. 2×1) were in the range of 0.82-0.99, for all datasets. Hence, some of the empirical distributions were very close to the uniform, which of course indicates a support for Test 2. On the other hand, this does not happen here, where the difference of f_X from f_Y is quite big for both GPs, with their maximum ratios not exceeding 0.83 (EDDIE 7), and 0.91 (EDDIE 8), respectively. Thus there is no support for Test 2 in this work.

Conclusion

This work used two different GP algorithms, EDDIE 7 and EDDIE 8, to test previously derived results of the Market Fraction Hypothesis (Chen et al., 2011; Kampouridis et al., 2009). Our goal was to show that our previously derived results were rigorous and can thus be generalized, regardless the choice of the GP algorithm. Overall, results have shown the same patterns across all three GP algorithms. We found that, even in the long run, the market tends to favor few types of agents, hence the property of the long-term uniform distribution does not hold. Secondly, while most types of agents cannot be dominant consecutively for more than 2 years, a few exceptions can sustain up to 8 years. Finally, while the results above are qualitatively insensitive to the number of types, a parameter set in the test, we only need four to five types (five to six types), to account for the behavior of 90% (95%) of market participants, a result which is first established in Aoki (2002). These observations are quite significant, because they allow us to draw some important conclusions about the microstructure of financial markets.

In general, we can say that the number of clusters does not really affect the markets' microstructure dynamics. As we move to a higher number of strategy types, evidence for Test 1 gets stronger, while evidence for Test 2 becomes weaker. This applies to all three algorithms. Nevertheless, the introduction of both EDDIE versions has shown a minor deterioration in the individual test results. Thus, evidence for the MFH has got even weaker. While the results for a high number of clusters do not seem to be affected, since all 3 algorithms suggest a rejection of the hypothesis, this is not the case with a lower number of clusters. For instance, under the *simple GP* tests, there were a few datasets that would show support for Test 2 under a very low SOM dimension, e.g. 2×1. However, this does not happen any more, with any of the EDDIE versions. There is even less support for the MFH under EDDIE 7 and 8. Our explanation for this is that the constrained fitness function that both versions share is responsible for this, although we cannot yet explain how. This will be the next step of our research. More specifically, if we completely loosen the constraints in the fitness function of EDDIE 7, we will end up with the *simple GP* algorithm. We can thus start from there and then fine-tune the constraint parameters, so that we can experiment with looser and tighter constraints. This should allow us to observe how the MFH results are affected by the constrained fitness function. We believe that this will give us a better idea of why the MFH results in this chapter are different to the original ones (under low SOM dimensions). Finally, as there are no significant differences between EDDIE 7 and EDDIE 8, we have concluded that EDDIE 8's new grammar is not responsible for the changes in the MFH results.

Appendix

Technical Indicators

The following section presents the technical indicators that the GP is using, along with their formulas. We performed a sort of standardization in order to avoid having a very large range of numbers generated by GP, because this would increase the size of the search space even more. Given a price time series $[P(t),\ t \geq 0]$, and a period of length L, Equations 1-6 present these formulas.

Moving Average (MA)

$$\text{MA(L,t)} = \frac{P(t) - \frac{1}{L}\sum_{i=1}^{L} P(t-i)}{\frac{1}{L}\sum_{i=1}^{L} P(t-i)} \tag{9}$$

Trade Break Out (TBR)

$$\text{TBR(L,t)} = \frac{P(t) - max\{P(t-1), \ldots, P(t-L)\}}{max\{P(t-1), \ldots, P(t-L)\}} \tag{10}$$

Filter (FLR)

$$\text{FLR(L,t)} = \frac{P(t) - min\{P(t-1), \dots, P(t-L)\}}{min\{P(t-1), \dots, P(t-L)\}} \tag{11}$$

Volatility (VIX)

$$\text{VIX(L,t)} = \frac{\sigma(P(t), \dots, P(t-L+1))}{\frac{1}{L}\sum_{i=1}^{L} P(t-i)} \tag{12}$$

Momentum (Mom)

$$\text{Mom(L,t)} = P(t) - P(t-L) \tag{13}$$

Momentum Moving Average (MomMA)

$$\text{MomMA(L,t)} = \frac{1}{L}\sum_{i=1}^{L} Mom(L, t-i) \tag{14}$$

References

Allen, F., & Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. *Journal of Financial Economics*, *51*, 245-271.

Amilon, H. (2008). Estimation of an adaptive stock market model with heterogeneous agents. *Journal of Empirical Finance*, *15 (2)*, 342-362.

Aoki, M. (2002). Open models of share markets with two dominant types of participants. *Journal of Economic Behavior and Organization*, *49(2)*, 199-216.

Arifovic, J., & Gencay, R. (2007). Statistical properties of genetic learning in a model of exchange rate. *Journal of Economic Dynamics and Control*, *24*, 981-1005.

Austin, M., Bates, G., Dempster, M., Leemans, V., & Williams, S. (2004). Adaptive systems for foreign exchange trading. *Quantitative Finance*, *4(4)*, 37-45.

Backus, J. (1959). The syntax and semantics of the proposed international algebraic language of Zurich. In *International conference on information processing* (p. 125-132). UNESCO.

Boswijk, P., Hommes, C., & Manzan, S. (2007). Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics and Control*, *31(6)*, 1938-1970.

Brock, W., & Hommes, C. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, *22*, 1235-1274.

Chen, S.-H., Chang, C.-L., & Du, Y.-R. (2010). Agent-based economic models and econometrics. *Journal of Knowledge Engineering Review (forthcoming)*.

Chen, S.-H., Kampouridis, M., & Tsang, E. (2011). Microstructure dynamics and agent-based financial markets. In T. Bosse, A. Geller, & C. M. Jonker (Eds.), *Multi-agent-based simulation xi, 11th international workshop, revised papers, lnai* (Vol. 6532, p. 121-135). Berlin Heidelberg: Springer-Verlag. (forthcoming)

Chen, S.-H., & Yeh, C.-H. (2001). Evolving traders and the business school with genetic programming: A new architecture of the agent-based artificial stock market. *Journal of Economic Dynamics & Control*, *25*, 363-393.

De Bondt, W., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, *40*, 793-805.

Edwards, R., & Magee, J. (1992). Technical analysis of stock trends. *New York Institute of Finance*.

Kampouridis, M., Chen, S.-H., & Tsang, E. (2009). Market fraction hypothesis: A proposed test. In *The 9th asia-pacific complex systems conference.* Tokyo. (Also available from http://kampouridis.net/publications/)

Kampouridis, M., & Tsang, E. (2010). EDDIE for investment opportunities forecasting: Extending the search space of the GP. In *Proceedings of the IEEE conference on evolutionary computation* (p. 2019-2026). Barcelona, Spain.

Kirman, A. (1991). Epidemics of opinion and speculative bubbles in financial markets. In (p. 354-368). M. Taylor (Hrsg.), "Money and Financial Markets", London, Macmillan.

Kirman, A. (1993). Ants, rationality and recruitment. *Quarterly Journal of Economics*, *108(1)*, 137-156.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Journal of Biological Cybernetics*, *43*, 59-69.

Koza, J. (1992). *Genetic programming: On the programming of computers by means of natural selection.* Cambridge, MA: MIT Press.

LeBaron, B., Arthur, W., & Palmer, R. (1999). Time series properties of an artificial stock market. *Journal of Economic Dynamics and Control*, *23*, 1487-1516.

Li, J. (2001). *FGP: A genetic programming-ased financial forecasting tool.* Unpublished doctoral dissertation, Department of Computer Science, University of Essex.

Li, J., & Tsang, E. (1999). Improving technical analysis predictions: An application of genetic programming. In *The 12th international flairs conference* (p. 108-112). USA.

Lo, A. (2004). The adaptive market hypothesis: market efficiency from an evolutionary perspective. *Journal of Portfolio Management*, *30*, 15-29.

Lo, A. (2005). Reconciling efficient markets with behavioral finance: The adaptive markets hypothesis. *Journal of Investment Consulting*, *2*, 21-44.

Lux, T. (1995). Herd behavior, bubbles and crashes. *Economic Journal*, *105*, 880-896.

Lux, T. (1997). Time variation of second moments from a noise trader/infection model. *Journal of Economic Dynamics and Control*, *22*, 1-38.

Lux, T. (1998). The socio-economic dynamics of speculative markets: Interacting agents, chaos and the fat tails of return distributions. *Journal of Economic Behavior and Organization*, *33*, 143-165.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th berkeley symposium on mathematical statistics and probability* (p. 281-297). University of California Press.

Martinez-Jaramillo, S. (2007). *Artificial financial markets: An agent-based approach to reproduce stylized facts and to study the red queen effect.* Unpublished doctoral dissertation, CFFEA, University of Essex.

Martinez-Jaramillo, S., & Tsang, E. (2009). A heterogeneous, endogenous and co-evolutionary GP-based financial market. *IEEE Transactions on Evolutionary Computation*, *13(1)*, 33-55.

Poli, R., Langdon, W., & McPhee, N. (2008). *A field guide to genetic programming.* Lulu.com.

Provost, F., & Kohavi, R. (1998). Glossary of terms. *Journal of Machine Learning*, *30(2-3)*, 271-274.

Tsang, E., Li, J., & Butler, J. (1998). EDDIE beats the bookies. *International Journal of Software, Practice & Experience, Wiley*, *28(10)*, 1033-1043.

Tsang, E., Li, J., Markose, S., Er, H., Salhi, A., & Iori, G. (2000). EDDIE in financial decision making. *Journal of Management and Economics*, *4(4)*.

Tsang, E., Markose, S., & Er, H. (2005). Chance discovery in stock index option and future arbitrage. *New Mathematics and Natural Computation, World Scientific*, *1(3)*, 435-447.

Winker, P., & Gilli, M. (2001). Indirect estimation of the parameters of agent-based models of financial markets. *FAME Research paper, University of Geneva*, *38*.