Advances in Complex Systems

Microstructure Dynamics and Agent-Based Financial Markets: Can Dinosaurs Return?

MICHAEL KAMPOURIDIS

School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, CO4 3SQ, UK
mkampo@essex.ac.uk

SHU-HENG CHEN

AI -Econ Center, Department of Economics, National Cheng Chi University, Taiwan 11623, R.O.C.
chen.shuheng@gmail.com

EDWARD TSANG

Centre for Computational Finance and Economic Agents, University of Essex, Wivenhoe Park, CO4 3SQ, UK
edward@essex.ac.uk

Received (received date)
Revised (revised date)

This paper formalizes observations made under agent-based artificial stock market models into a concrete hypothesis, which is called the Dinosaur Hypothesis. This hypothesis states that the behavior of financial markets constantly changes and that the trading strategies in a market need to continuously co-evolve with it in order to remain effective. After formalizing the hypothesis, we suggest a testing methodology and run tests under 10 international financial markets. Our tests are based on a framework that we recently developed, which used Genetic Programming as a rule inference engine, and Self-Organizing Maps as a clustering machine for the above rules. However, an important assumption of that study was that maps among different periods were directly comparable with each other. In order to allow this to happen, we had to keep the same clusters throughout the different time periods of our experiments. Nevertheless, this assumption could be considered as strict or even unrealistic. In this paper, we relax this assumption. This makes our model more realistic. In addition, this allows us to investigate in depth the dynamics of market behavior and test for the plausibility of the Dinosaur Hypothesis. The results show that indeed markets' behavior constantly changes. As a consequence, strategies need to continuously co-evolve with the market; if they do not, they become obsolete or dinosaurs.

Keywords: Genetic Programming; Self-Organizing Feature Map, Market Microstructure; Market Behavior Dynamics; Dinosaur Hypothesis

1. Introduction

From a mesoscopic viewpoint, the market dynamics can be regarded as the collective interacting behavior of a number of types (clusters) of agents. The question is whether these types or clusters remain fixed, and what will change over time is simply the fraction of these types, or, alternatively, both the types and fraction are also subject to change. This issue is motivated by the recent literature on agent-based financial markets, specifically, the co-existence of the two classes of models; one is the H-type models[a] and the other is the SFI[b] or SFI-like models [2].

For the former class, a typical model is the fundamentalist-chartist model. In this model, there are only two types of agents, fundamentalists and chartists, in the market, and only the fractions of these two types of agents change over time. On the other hand, their behavioral specification remains unchanged. For the latter case, the models are not built upon this constancy: types are not given. Instead, the evolutionary force works down to each individual and all the way up to their types and clusters. Similar in vein to Alfred Marshall's famous quotation, these models deal with the agents whose nature and constituents as well as the outer form are constantly changing. [16]

Brian Arthur, a pioneer of the SFI artificial markets, used dinosaurs as a metaphor to describe this constantly-changing property [1]. More specifically, Arthur stated:

"We find no evidence that the market behavior ever settles down; the population of predictors continually co-evolves. One way to test this is to take agents out of the system and inject them in again later on. If market behavior is stationary they should be able to do as well in the future as they are doing today. But we find that when we "freeze" a successful agent's predictors early on and inject the agent into the system much later, the formerly successful agent is now a dinosaur. His predictions are unadapted and perform poorly. The system has changed" (Ibid, p.24).

Similar observations were also made by Chen and Yeh in their artificial stock market framework [4]. In addition, Chen and Yeh found that *a dinosaur's performance decreases monotonically.* Lastly, Raberto et al. [19] also showed that within their artificial financial market framework, the dominance of a particular trading strategy is not fixed, but depends on the particular condition of the market.

The above motivates us to *investigate whether these observations can also hold in the 'real' world, under empirical datasets.*[c] In order to do this, we use the framework from a previous work [3], where we developed an approach to empirically examine the dynamics of market fractions of different types of agents. That study used Genetic Programming (GP) [13, 17] as a rule-inference engine to find out the behavioral rules of agents, and then used Self-Organizing Maps (SOM) [12] to clus-

[a]H-type models refer to the models that have H types of fixed, and pre-specified types of agents. For instance, in a 2-type model, there would be 2 types of agents, like fundamentals and chartists.
[b]Santa Fe Institute
[c]Both Arthur's and Chen and Yeh's works were under an artificial stock market framework.

ter these agents into types of trading strategies. Once after the clusters and the associated map were obtained from period to period, the dynamics of market fractions were studied based on a time series of these maps. For example, we studied two main properties motivated by the Market Fraction Hypothesis [3]: the short-term horizon and the long-term frequency.

However, that study rested upon an important assumption, i.e., the maps derived from each period were comparable with each other. This comparability assumption itself required that the types (clusters), as well as their operational specification, would not change over time. If this were not the case, then the subsequent study would be questioned. This was mainly due to one technical step in our analysis called translation. The purpose of translation was to place the behavior of agents observed in one period in a different period and to recluster it for the further cross-period comparison. We could not meaningfully have done this without something like topological equivalence, which could not be sustained without the constancy of the types. However, this assumption can be considered as strict and unrealistic. Strategy types do not necessarily remain the same over time. For instance, if a chartist strategy type exists in time t, it is not certain it will also exist in $t + 1$. If market conditions change dramatically, the agents might consider other strategy types as more effective and choose them. The chartist strategy would then stop existing.

In this paper, we are interested in extending our previous work by investigating the market microstructure when both the types and the fractions of strategies are subject to change. Thus, we again employ GP as a rule inference engine and SOM as a clustering machine, with the only difference being that we relax the above assumption of static clusters over time. This makes our model more realistic and allows us to test the existence of the constantly-changing property in the behavior of financial markets that Arthur first observed [1], or in other words what we call the *Dinosaur Hypothesis (DH)*.

In addition, another contribution of this paper is the attempt to formalize the DH, by presenting its main constituents and by suggesting a testing methodology. We run tests for 10 international markets and hence provide a general examination of the plausibility of the DH. One goal of our empirical study is to use the DH as a benchmark and examine how well it describes the empirical results which we observe from the various markets. More importantly, these tests will allow us to observe the dynamics of market behavior and investigate how they change in the long run.

The rest of this paper is organized as follows: Section 2 elaborates on the DH. Section 3 presents the basic tools used for our tests, namely Genetic Programming (GP) and Self-Organizing Feature Maps (SOM). Section 4 presents the experimental designs. Section 5 addresses the methodology employed to test the DH and explains the technical approaches needed to be taken to facilitate the tests of the DH. Section 6 presents the test results. First it presents the results over a single run for a single dataset. Then it summarizes the results over 10 runs for this single dataset and it finally presents summary results for all 10 datasets. Section 7 concludes this paper and also discusses possible directions for further research.

2. The Dinosaur Hypothesis

The following statements form the basic constituents of the DH and are based on Arthur's work [1]:

(1) The market behavior never settles down
(2) The population of predictors that exists in the market continuously co-evolves with it

These two statements indicate the non-stationary nature of financial markets and imply that strategies cannot remain effective unless they continuously co-evolve with the market; if they do not, they become obsolete or *dinosaurs*.

The second statement also informs us that a population of predictors exists in a market. According to Arthur [1], an agent's basic problem is to profit in the next period, and the only way to do that is by predicting the direction of the market. Thus, predictors are models that map the patterns which are formed by the current price set into a forecast for the next period's price. In other words, agents in the market use different predictors to forecast future price movements. In our framework, we refer to predictors as 'trading strategies'.

However, since, as we have already said, the above observations were made under an artificial stock market framework, we are interested in testing them against empirical data. We will first use GP to create trading rules and then SOM to cluster them. The next section presents these two tools and also explains our motivation for using them.

Finally, in order to make the reading of this paper more comprehensive, we present two definitions, inspired by Arthur's work; *Dinosaur*, which refers to a training strategy that has performed well in some periods, but then ceased performing well in the periods that followed. This means that this strategy may or may not become effective again. If it does, then it is called a *returning dinosaur*.

3. Tools

This section presents the two main tools used in order to test the DH. These two tools are GP and SOM. Sections 3.1 and 3.2 first present the motivation behind the choice of these two tools, and then present the two tools, GP and SOM, respectively.

3.1. *Genetic Programming (GP)*

3.1.1. *Motivation for using GP*

In this work, we assume that traders' behavior, including price expectations and trading strategies, is either not observable or not available. Instead, their behavioral rules have to be *estimated* by the observable market price. Using macro data to estimate micro behavior is not new, as many H-type empirical agent-based models have already performed such estimations [2]. However, such estimations are based on very strict assumptions upon which a formal econometric model can be built. Since we no longer keep these assumptions, an alternative must be developed, and in this paper we recommend Genetic Programming (GP).

The use of GP is motivated by considering the market as an evolutionary and selective process.[d] In this process, traders with different behavioral rules participate in the markets. Those behavioral rules which help traders gain lucrative profits will attract more traders to imitate, and rules which result in losses will attract fewer traders. Genetic programming is a formalization which does not rest upon any pre-specified class of behavioral rules. Instead, in GP, a population of behavioral rules is randomly initiated, and the survival-of-the-fittest principle drives the entire population to become fitter and fitter in relation to the environment. In other words, given the non-trivial financial incentive from trading, traders are aggressively searching for the most profitable trading rules. Therefore, the rules that are outperformed will be replaced, and only those very competitive rules will be sustained in this highly competitive search process.[e]

Hence, even though we are not informed of the behavioral rules followed by traders at any specific time horizon, GP can help us infer what these rules are approximately by simulating the evolution of the microstructure of the market. Without imposing tight restrictions on the inferred behavioral rules, GP enables us to create more realistic, and possibly more complex behavioral rules.[f]

For the experiments in this paper, we are using a GP algorithm called 'simple GP'. The next subsection presents this algorithm.

3.1.2. *Simple GP presentation*

Our simple GP is inspired by a financial forecasting tool, EDDIE [21–23, 11], which applies genetic programming to evolve a population of financial advisors, or, alternatively, a population of market-timing strategies, which guide investors on when to buy or hold. These market-timing strategies are formulated as decision trees, which, when combined with the use of GP, are referred to as *Genetic Decision Trees* (GDTs). Our GP uses indicators commonly used in technical analysis: Moving Average (MA), Trader Break Out (TBR), Filter (FLR), Volatility (Vol), Mo-

[d]See [14, 15] for his eloquent presentation of the *Adaptive Market Hypothesis*.
[e]It does not necessarily mean that the types of traders surviving must be smart and sophisticated. They can be dumb, naive, randomly behaved or zero-intelligent. Obviously, the notion of rationality or bounded rationality applying here is ecological [20, 9]
[f][8] provides the first illustration of using genetic programming to infer the behavioral rules of human agents in the context of ultimatum game experiments. Similarly, [10] uses genetic algorithms to infer behavioral rules of agents from market data

mentum (Mom), and Momentum Moving Average (MomMA).[g] Each indicator has two different periods, a short- and a long-term one, 12 and 50 days respectively.

Each of these market-timing strategies (GDTs) is syntactically (grammatically) produced by the Backus Naur Form (BNF). Figure 1 presents the BNF 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 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 recommendation is to buy (1) or not-buy (0). Figure 2 presents a sample GDT generated by the GP.

```
<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 | Vol_12 | Vol_50 | Mom_12 | Mom_50 |
               MomMA_12 | MomMA_50
<RelationOperation> ::= ">" | "<" | "="
Decision is an integer, 1 or 0 (to-buy or not-to-buy)
Threshold is a real number
```

Fig. 1. The Backus Normal Form of the GP

Fig. 2. Sample GDT generated by the GP.

[g]We use these indicators because they have been proved to be quite useful in previous works like [11]

Given a recommendation made by a GDT, we can construct a confusion matrix [18] and report whereas this recommendation was a True Positive (TP), a True Negative (TN), a False Positive (FP), or a False Negative (FN). Based on these classifications, we can then record the rate of correct recommendations made by the GP, the rate of the to-buy opportunities that existed in the market but the GP falsely identified as not-to-buy, and the rate of the to-buy recommendations that were falsely identified as to-buy. These three rates are defined as Rate of Correctness (RC), Rate of Missing Chances (RMC), and Rate of Failure (RF), respectively. Their formulas are presented below, in Equations (1)-(3):

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

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

The above metrics combined give the GP's 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, and 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. For our experiments, we chose to include GDTs that mainly focus on correctness and reduced failure. Thus these weights have been set to 1, $\frac{1}{6}$ and $\frac{1}{2}$ respectively, and are given in this way in order to reflect the importance of each performance measure for our predictions.

Given a set of historical data and the fitness function, GP is then applied to evolve these market-timing strategies in a standard way. After evolving a number of generations (50 in this paper), what stands (survives) at the end (the last generation) is, presumably, a population of financial agents whose market-timing strategies are financially rather successful. We therefore use these strategies to infer what those competitive strategies may be in the period coinciding with the data period.

Finally, we should again remind the reader that "GDTs", in our framework, is a term equivalent to "predictors" in Arthur's work. However, because a GDT makes recommendations to an agent of whether to buy or not-to-buy, they could also be seen as *trading strategies*. GDTs and trading strategies are thus two terms that are often used interchangeably in this work.

3.2. *Self-Organizing Feature Maps (SOM)*

3.2.1. *Motivation for using SOM*

Once a population of rules is inferred from GP, it is desirable to cluster them based on a chosen similarity criterion so as to provide a concise representation of the microstructure. The similarity criterion which we choose is based on the *observed trading behavior*.[h] Based on this criterion, two rules are similar if they are *observationally equivalent* or *similar*, or, alternatively put, they are similar if they generate the same or similar market-timing behavior.[i]

Given the criterion above, the behavior of each trading rule can be represented by its series of market timing decisions over the entire trading horizon, for example, 6 months. Therefore, if, as we said earlier, we denote the decision "buy" by "1" and "not-buy" by "0", then the behavior of each rule is a binary vector. The dimensionality of these vectors is then determined by the length of the trading horizon. For example, if the trading horizon is 125 days long, then the dimension of the market timing vector is 125. Once each trading rule is concretized into its market timing vector, we can then easily cluster these rules by applying Kohonen's Self-Organizing Maps to the associated clusters.

The main advantage of SOMs over other clustering techniques such as K-means is that the former can present the results in a visualizable manner. In this way, 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.

Let us now present SOM in detail.

3.2.2. *SOM presentation*

SOM is a type of artificial neural network 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 a number of neurons (or clusters) in a two-dimensional lattice, presenting the input data in an organized way, so that the similar strategies are clustered together.

[h] Other clustering criteria could take place, too, such as risk averseness or statistical significance of rules' performance in out-of-sample data. However, in this paper we are interested in focusing on the behavioral aspects of the rules, since we are investigating the behavior of financial markets. We leave it as a future work to examine if different clustering criteria can affect our results.

[i] One might question the above similarity criterion, since very different rules might be able to produce the same signals. However, this does not pose a problem in this work, since we are interested in the behavior of the market (and thus the rules' behavior). We are not interested in the semantics aspect of the rules.

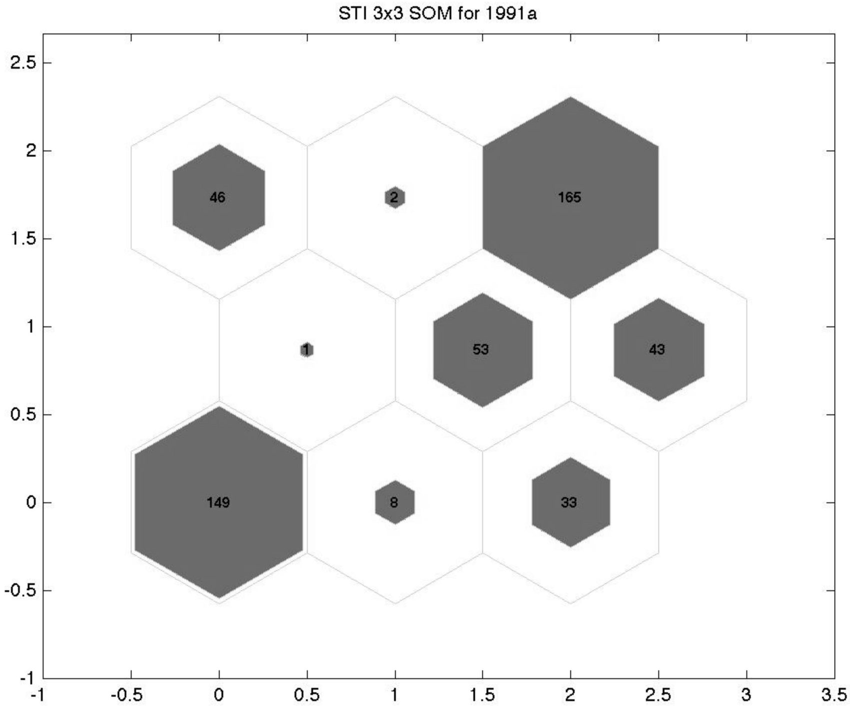


Fig. 3. 3 by 3 Self-Organizing Feature Map of the first semester of the year 1991 for STI.

For our experiments, we use MathWorks' Neural Network toolbox[j], which is built-in the MATLAB environment. Figure 3 presents the results after running 3×3 SOM[k] for a population of 500 individuals for the daily STI (Singapore) data for the first semester of 1991.[l] As we can observe in this case, there are two strategy types that occupy a big fraction of the population, whereas the rest of the strategy types have significantly fewer members.

This concludes the presentation of GP and SOM, which are the two techniques we use in our framework. Let us continue in the next section by presenting the experimental designs.

[j] *http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/self_or4.html*
[k] 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.
[l] Available from http://finance.yahoo.com

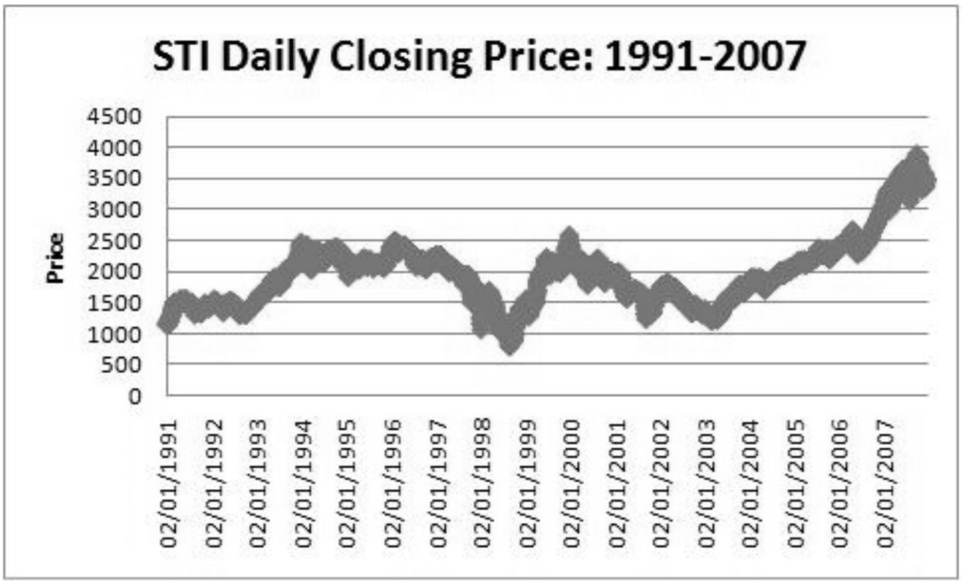


Fig. 4. Daily Closing Price for STI:1991-2007

4. 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), 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 STI dataset. Figure 4 presents the daily closing price of STI. We then proceed by presenting summary results over the 10 runs for all datasets.

Each year was split into 2 semesters (January-June, July-December), so in total, out of the 17 years, we have 34 periods.[m] The GP was therefore implemented 34 times. Table 1 presents the GP parameters for our experiments. The GP parameters for our experiments are the ones used by Koza [13]. Only the tournament size has been lowered, because we were observing premature convergence. Other than that, the results seem to be insensitive to these parameters.

It is also important to say that the GP was only used for creating and evolving the trading strategies. No validation or testing took place, as is the case in the traditional GP approach. The reason for this is because we were not using the GP

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

Table 1. GP 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
$\{w_1, w_2, w_3\}$	$\{1, \frac{1}{6}, \frac{1}{2}\}$

for forecasting purposes; instead, as we have already said we were interested in usng the GP as a *rule inference engine* so that it can help us to see what were the strongest species during a certain period. To be more specific, the GP was used for each of the 34 periods and each time created and evolved trading strategies. After the evolution of the strategies under a specific period, these strategies (GDTs) were not tested against another set. This approach is consistent with Andrew Lo's Adaptive Market Hypothesis [14, 15], as it states that the heuristics of an old environment are not necessarily suited to the new one. Lo refers to 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. Furthermore, our no-testing approach is also consistent with the well-tested Overreaction Hypothesis [5, 6], which essentially states that winner portfolios are outperformed by loser portfolios during the next period.

Finally, as we mentioned earlier we used 3×3 SOM. Therefore, we obtained a total of 34 SOMs (one per period), with 9 clusters each. In other words, in every period the GDTs were placed in 1 of the 9 clusters (i.e., the types of the trading strategies) of that SOM. From this point on, whenever we use the term 'trading strategy type' we will be referring to 1 of the 9 clusters and the GDTs will be members of any of these 9 clusters.

5. Testing Methodology

After having presented the necessary tools and the experimental designs, we can now continue with presenting the testing methodology. But before we do this, let us first present some frequently used terms:

- *Base period*, is the period during which we create and evolve GDTs that are going to be used for testing the DH
- *Future period(s)*, is a period or periods which follow the base period (in chronological order)

Furthermore, as we said earlier, our data are divided into semesters. Thus the first semester of a year will be denoted with an 'a' at the end (e.g., 1991a), and the second semester of a year will be denoted with a 'b' at the end (e.g., 1991b). Let us now present the tests and our suggested testing methodology. Each of the following subsections presents each one of the suggested tests, followed by its testing methodology.

5.1. *Dissatisfaction test*

In order to investigate whether the behavior of markets is non-stationary, we do the following. After generating and evolving GDTs for each one of the 34 periods, and clustering them into strategy types via SOM, we recluster the GDTs of each base period to the maps of their future periods.[n] If the Dinosaur Hypothesis holds, we should observe that GDTs (strategies) from past periods have more and more difficulties fitting into the environments of the future periods. This is because these strategies have remained unadapted to the new environments and have thus turned into dinosaurs.

Let us give an example, and assume that there is a strategy type (cluster) in time t, which represents 'fundamentalists'; then, all GDTs which follow a fundamental strategy are placed in that cluster. When we then take the GDTs from that period (base period) and recluster them to strategy types of future periods, it is not guaranteed that there will again be a cluster that represents fundamentalists. If market behavior constantly changes, there is a possibility that this type of strategy does not exist any more. Thus, the GDTs find themselves *unadapted* to the new environment (clusters) and have to choose another cluster, which represents them as closely as possible. This cluster will be the one that has the centroid with the smallest Euclidean distance from the market-timing vectors of these GDTs.[o] Of course, since now the SOM of the future period is formed by different clusters, the GDTs might not fit as well as they did in the base period. In order to measure this 'unfitting', we use a 'dissatisfaction rate', i.e., how dissatisfied these GDTs will be when placed into a future period's cluster that does not represent their strategy. *If the market is non-stationary, the GDTs' dissatisfaction rate will be high*, as a result of the changes that took place in the market. The dissatisfaction rate is defined as the Euclidean distance of a GDT's market-timing vector to the centroid of the cluster in which it is placed, after the reclustering procedure.

In addition, motivated by Chen and Yeh's stricter test results [4], who found that a dinosaur's performance decreases monotonically, we want to investigate if this also applies to our empirical datasets. Hence, under the Dinosaur Hypothesis

[n]The process of reclustering is explained in the next section.
[o]One may wonder if the choice of the Euclidean distance as a distance metric, when the vectors of the GDTs are binary, is an appropriate one. However, this does not pose a problem, because the vectors of the clusters' centroids are real valued and thus calculations by using the euclidean distance are meaningful.

the following two statements should hold:

The average dissatisfaction rate of the population of trading strategies (GDTs) from future periods should

(1) *Not be less than or equal to the dissatisfaction of the base period* (**Test 1**)
(2) *Increase continuously, as the testing period moves further away from the base period* (**Test 2**)

As we can see, we have a population of GDTs and we then monitor their future performance in terms of their dissatisfaction rate, in accordance with Arthur's and Chen and Yeh's experiments. Test 1 is basically the procedure we have just described above and is inspired by Arthur's work [1]. Test 2 is inspired by the observation that Chen and Yeh made [4], regarding the monotonic decrease of a GDT's performance. However, in our framework we do not require that changes in the dissatisfaction rate are monotonic. The reason for this is because when Chen and Yeh tested for the Dinosaur Hypothesis (they did not explicitly use this term), they only tested it over a period-window of 20 days, which is relatively short, and hence makes it easier to observe an example of monotonically decreasing behavior. Thus, requiring that the GDTs' dissatisfaction rate increases monotonically in the long run would be very strict, and indeed hard to achieve. For that reason, Test 2 requires that the dissatisfaction changes are continuous, but not monotonic. It should also be mentioned that we are interested in qualitative results, meaning that we want to see how close the real market behaves in comparison with what is described by the DH.

Let us now explain the process of reclustering.

5.2. *Reclustering process*

Let us give an example when 1991a is the base period. Each evolved GDT would first be moved to the next period, 1991b, and reclustered into one of the clusters of that period. In order to 'decide' which cluster to choose, the GDT compares the Euclidean distance of its market-timing vector with each cluster; it is then placed into the cluster with the smallest Euclidean distance. The same procedure follows for all GDTs of the population. At the end, the population of evolved GDTs from the base period of 1991a has been reclustered into the clusters of period 1991b. We also follow the same procedure with all future periods. This means that the GDTs from 1991a are also reclustered in 1992a, 1992b, ..., 2007b.

The same process is followed for all other base periods (i.e., 1991b, 1992a, ..., 2007a). This means that when 1991b is the base period, the GDTs that were created and evolved during 1991b will be reclustered into the SOMs of all future periods. After 1991b, 1992a takes over as the base period and the same procedure takes place again. We do this until 2007a. We obviously cannot do this for 2007b, since there are no periods available after this year. The reader should also bear in mind that we only apply the evolved GDTs to the SOMs of future periods; for instance, when the base period is 2000a, we do not apply the evolved GDTs backwards in time, only forwards. We are not interested in looking at what would happen in the past; we are only interested to observe how the dissatisfaction of the GDTs is affected in the future.

Once the process of reclustering is complete, we calculate the dissatisfaction rate of each GDT in the population. Next, we calculate the population's average dissatisfaction rate. We do the same for all 34 periods. Given a base period, the population average dissatisfaction of all periods is normalized by dividing those population average dissatisfaction rates by the population average dissatisfaction rate in the base period. Hence, each base period has its normalized average dissatisfaction rate equal to 1. In order to prove that the market is non-stationary, we need to show that the normalized average dissatisfaction rate of the GDTs increases in the future periods, and never returns to its initial value of 1, which was during the base period. If, on the other hand, this rate reaches 1 or below, *this is an indication of a cyclic market behavior*, since the GDTs have found the same conditions with the base period, and as a result feel as 'satisfied' as before. In this case, we refer to these GDTs as *returning dinosaurs*.

Therefore, if the DH holds, we should make the following two observations: first, that the normalized average dissatisfaction rate in future periods does not drop to 1 or lower (Test 1). Second, we should observe that this rate increases continuously over time (Test 2).

6. Results

This section presents the experimental results. In order to make it easier for the reader, we first start by presenting the results of a single dataset (STI) in Section 6.1. We then continue by presenting the summary results for all 10 datasets in Section 6.2.

6.1. *Results of a single run of a single dataset*

Because of the fact that this test is connected with SOM, we are also experimenting with different numbers of clusters (i.e., different SOM dimensions). This happens because we want to see if our results are sensitive to the number of clusters that can exist in a market. Thus, we first begin by presenting the results for the 3x3 SOM (9 clusters), and then present the results under different SOM dimensions.

6.1.1. *Results under 3×3 SOM (9 clusters)*

As Test 1 states, we want to make sure that the population dissatisfaction from future periods will not return to the range of dissatisfaction of the base period. In other words, we want to make sure that there are no *returning dinosaurs*. Then, Test 2 requires that we examine whether this population dissatisfaction from future periods continues to increase, as we move further away from the base period.

Test 1 We test this statement for one period at a time. The subject period forms our base period. As already mentioned, a returning dinosaur is observed if any future normalized population dissatisfaction rate is less than or equal to 1. Let us now have a look at Figure 5. The x-axis presents the 34 periods and the y-axis presents the dissatisfaction rate for each period. Period 1991a acts here as the base period.

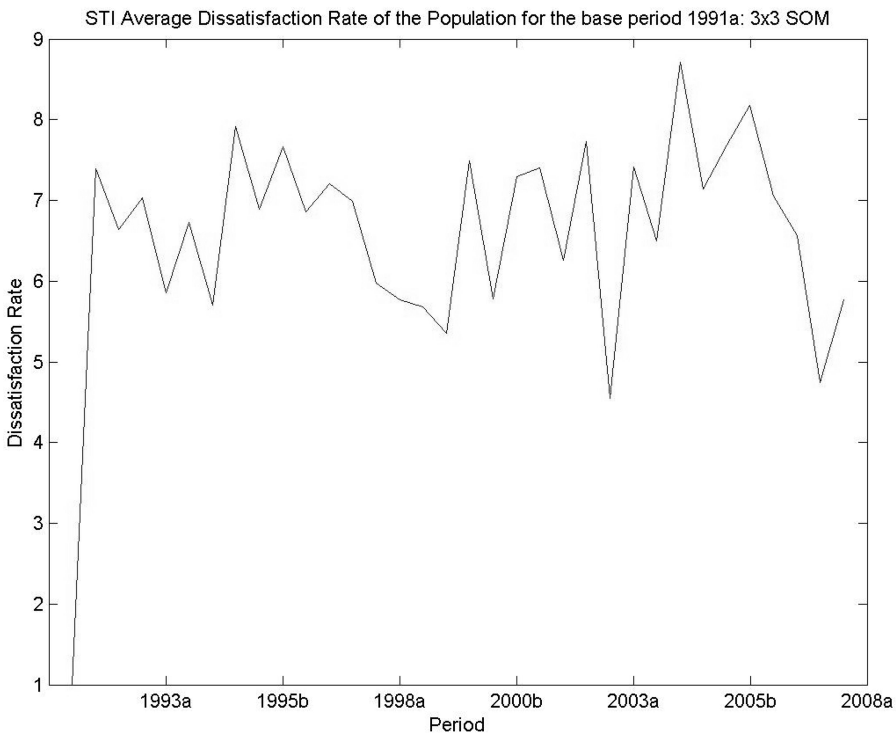


Fig. 5. Average dissatisfaction rate of the population of GDTs that were evolved in 1991a, over the future 34 periods. Period 1991a acts here as the base period, because this is the period in which the evolved trees were initially clustered. After calculating the dissatisfaction rate of each GDT in 1991a, we calculated the average dissatisfaction rate of the population. We then reclustered these GDTs to all future periods. In this way, we are able to observe how the average dissatisfaction rate of the population is affected while it is moving away from the base period. The results are for the 3×3 SOM dimension.

As we can see, the dissatisfaction rate immediately increases after we move away from the base period 1991a. It starts from around 7.5, and never drops again to a range close to 1. On average, the dissatisfaction rate is 6.73, which is actually far away from the threshold of 1. Thus no returning dinosaurs are observed.

Nevertheless, this result is when 1991a is the base period. Let us now present what happens after we have followed the same procedure for all periods, i.e., after all periods have acted as the base period. In order to do that, we iterate through each base period and calculate the minimum dissatisfaction rate among its future periods. For instance, in Figure 5, the minimum dissatisfaction rate is around 4.5 in period 2002b. This means that, for the base period of 1991a, the lowest dissatisfaction rate that any future period managed to get was 4.5. So we extract this value. We do the same for all 33 base periods and thus end up with a 1×33 vector, $MinAvgDis$, which shows the potential returning dinosaur per base period. The plot of the $MinAvgDis$ vector is shown in Figure 6.

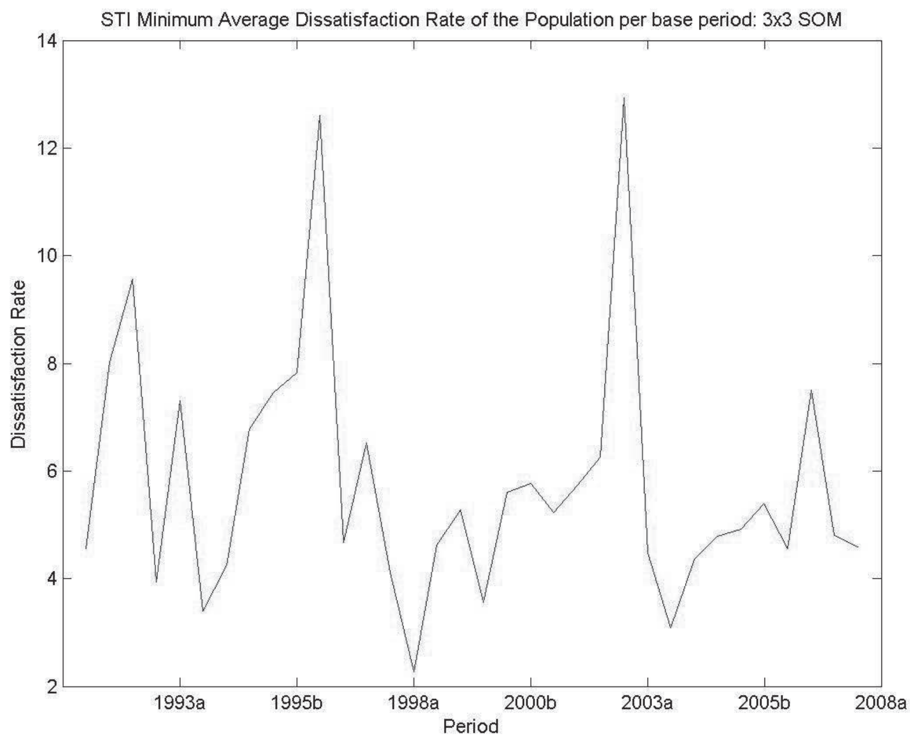


Fig. 6. Minimum average dissatisfaction rate of the population of GDTs per base period, for 3×3 SOM. After obtaining the series of the average dissatisfaction rate for each base period, over its future periods, we calculate the minimum average dissatisfaction rate per base period. In this way, we have an indication of whether there are any returning dinosaurs in each base period.

As we can observe, there is no base period with a minimum dissatisfaction rate of

1 or below. There are a few periods with relatively low dissatisfaction rates (1992b, 1993a, 1999b, 2003b). However, the lowest rate any period gets is 2, in period 1998a. In addition, the average minimum dissatisfaction rate for STI for the 3×3 SOM is 5.79. We can therefore conclude that strategies from past periods are on average not 'satisfied' in future years.

Test 2 Let us now move to Test 2. To show a continuous increase in the population dissatisfaction rate, we calculate the sum of the dissatisfaction rates of all those future periods that have distance from the base period equal to 1, then the sum of those future periods with distance equal to 2, and so on, up to distance equal to 33. In order to do this, we first need to create a table of distances, like the one in Table 2. Each row of this table presents the distances of the future periods from their base period. For instance, if 91a is the base (first row), then future period 91b has a distance equal to 1, future period 92a has a distance equal to 2, and so on. Table 3 presents an example of a series of normalized population dissatisfaction rates for the future periods of each base period. For example, when the base period is 91a (first row), the normalized dissatisfaction rate starts from 1 in 91a, then rises to 7.66 (91b), then goes to 6.82 (92a), and so on, until it reaches dissatisfaction equal to 5.94 in future period 07b. Let us now denote the sum of the dissatisfaction rates we mentioned at the beginning of this section by $\sum_{|i-j|=m} Dis(i,j)$, where i, j are the base and future period, respectively, $|i-j|$ is their absolute distance, as presented in Table 2, and m is the distance from the base period that takes values from 1 to 33. We divide this sum by the number of occurrences where $|i-j| = m$. This process hence returns the average of the normalized population dissatisfaction, and allows us to observe how it changes, as the distance m from the base period increases. We call this metric D_m and it is presented in Equation 5.

Table 2. Distance of future periods from their base period, over the 17 years 1991-2007. The further away we move from a period, a single unit of distance is added.

				j			
		91a	91b	92a	92b	...	07b
	91a	0	1	2	3	...	33
	91b	1	0	1	2	...	32
i	92a	2	1	0	1	...	31
	92b	3	2	1	0	...	30
	...	...	...	...	...	...	...
	07b	33	32	31	30	...	0

$$D_m = \frac{\sum_{|i-j|=m} Dis(i,j)}{\{\#(i,j), |i-j| = m\}} \qquad (5)$$

Table 3. Example of series of future population dissatisfaction rates per base period. Each base period's series is presented as a horizontal line of this table. Dissatisfaction rates have been normalized, so that the rate in the base period is always equal to 1.

		j					
		91a	91b	92a	92b	...	07b
	91a	1	7.66	6.82	7.08	...	5.94
	91b		1	3.76	5.72	...	4.70
i	92a			1	9.74	...	7.77
	92b				1	...	8.80
	...	...	...	...	...	...	...
	07b					...	1

Let us give an example: if we want to calculate D_{32}, we need to sum up the population dissatisfaction rates that have distance $m = 32$. This happens with $Dis(91a, 07a)$ (dissatisfaction rate of GDTs from base period 91a, when applied to future period 07a) and $Dis(91b, 07b)$ (dissatisfaction rate of GDTs from base period 91b, when applied to future period 07b). Therefore D_{32} would be equal to the sum of these two rates over 2, as there are only 2 periods we are interested in.[p] By calculating D_m for all m values, we can have a clear idea of how the average of the population dissatisfaction changes when we move from periods that are close to the base (low m), to periods that are further away (high m), and thus observe whether there is a continuous increase.

Figure 7 presents the results. It is interesting that for $m = 1{-}22$, D_m experiences a continuous increase. However this changes for higher values of m. Overall, Test 2 is not justified for STI, under 3×3 SOM.

To summarize, the experimental results of Test 1 for the STI dataset under the 3x3 SOM showed that there are no returning dinosaurs. Regarding Test 2, a continuous increase in the average dissatisfaction rate of the population of GDTs could only be observed for some periods.

6.1.2. *Experimenting with the number of clusters*

Let us now examine if the above results hold under different numbers of clusters. We experiment with the following SOM dimensions: 2×1, 3×1, 2×2, 5×1, 3×2, 7×1, 4×2, 3×3, for a range of trading strategy types 2-9. All graphs in this section include information for all of the different dimensions.

First let us start with the average dissatisfaction rate for the base period 1991a, which is presented in Figure 8. As we can observe, the dissatisfaction rate follows

[p] The distance $m = 32$ can also be found in 07a91a and 07b91b. However, we do not take them into account because, as we said earlier in Section 5, we are not interested in applying the evolved GDTs of a base period (here 07a and 07b) backwards in time (91a and 91b, respectively).

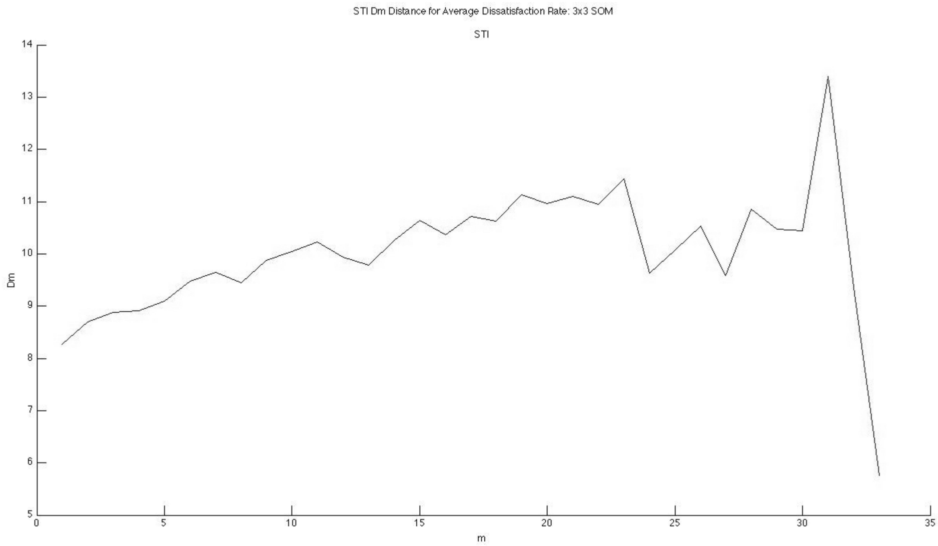


Fig. 7. STI D_m distance of the dissatisfaction rate for 3×3 SOM.

the exact same pattern we observed earlier: it immediately increases after we move away from the base period. It is within a range of 3 to around 9, depending on the SOM dimension, and never drops again to a range close to 1. On average, the average dissatisfaction rate for 1991a varies in the range of 4.29 under 2 types of trading strategies (clusters), to 6.73 (9 trading strategy types). Thus there are again no returning dinosaurs observed, for any SOM dimension. In addition, we again do not observe a continuous increase in the dissatisfaction rate. Hence, Test 1 holds for all SOM dimensions, whereas Test 2 does not. One final observation we can make is that the dissatisfaction rate increases as the number of clusters increases. This is something that does not come to us as a surprise, since the increased number of clusters has increased the sum of the dissatisfaction rates among the 500 GDTs, and thus has increased the average rate, too. What is important to state, however, is that the pattern of non-stationary behavior we observed earlier remains, under all trading strategy types. The number of clusters does not affect the test's results.

Let us now present the same results for all base periods. Figure 9 presents these results. The horizontal line indicates when the dissatisfaction rate is equal to 1, and is given as a reference.

As we can observe, nothing really changes when compared to the results presented earlier, for the 3×3 SOM. There is again no base period, for any SOM dimension, reaching an average dissatisfaction rate of 1. The lowest rate we observe is again for period 1998a, which is around 2. On average, the minimum dissatisfaction rate varies in the range of 3.56 (2 clusters) to 5.79 (9 clusters). Therefore, we

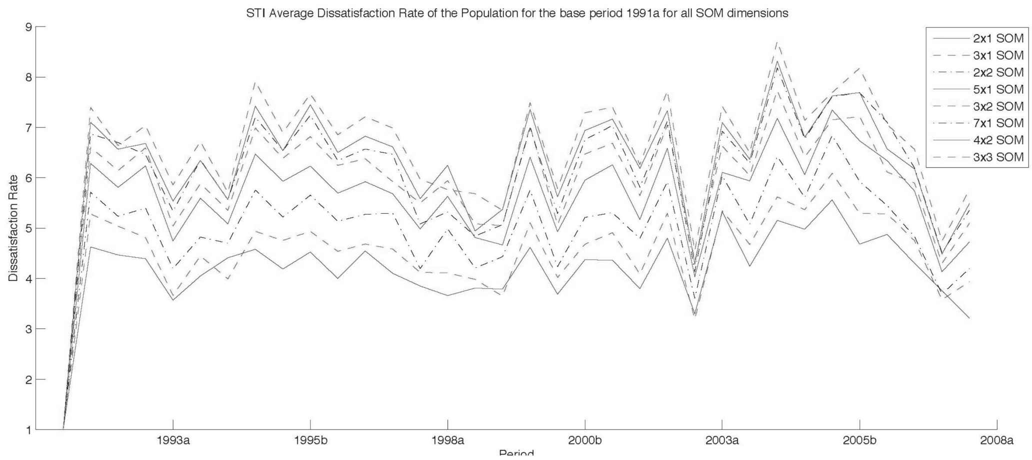


Fig. 8. Average dissatisfaction rate of the population of GDTs that were evolved in 1991a, over the future 34 periods. Period 1991a acts here as the base period, because this is the period in which the evolved trees were initially clustered. After calculating the dissatisfaction rate of each GDT in 1991a, we calculated the average dissatisfaction rate of the population. We then reclustered these GDTs to all future periods. In this way, we are able to observe how the average dissatisfaction rate of the population is affected while it is moving away from the base period. Each line in the figure represents a different SOM dimension.

can again see that dinosaurs do not return.

Regarding Test 2, we again use the D_m metric. Figure 10 presents the results. Once again, we observe the same pattern for all SOM dimensions, as with the ones we saw earlier for the 3×3 SOM: a continuous increase up to a point, and then upward and downward movements. In addition, we again see that as the number of clusters increases, there is also an increase in the dissatisfaction rate.

To summarize, the experimental results of Test 1 for the STI dataset show that there are no returning dinosaurs, under all different SOM dimensions tested. The market seems to be constantly changing and hence strategies are not 'satisfied' in their new environments. Furthermore, changing the SOM dimension (and thus the number of strategy types in the market) does not seem to significantly affect the results. The only difference we can observe is that the dissatisfaction rate increases when more types of strategies exist in the market (more SOM dimensions). Finally, there is again no continuous increase in the average dissatisfaction rate of the GDTs, after we move away from the base period. Therefore Test 2 is not justified.

Let us now examine if the above results generalize under different international markets.

6.2. *Summary results for all datasets*

So far we have presented the results for a single dataset, STI. In this section we will present the results for all 10 datasets. As a reminder, these datasets are: CAC40,

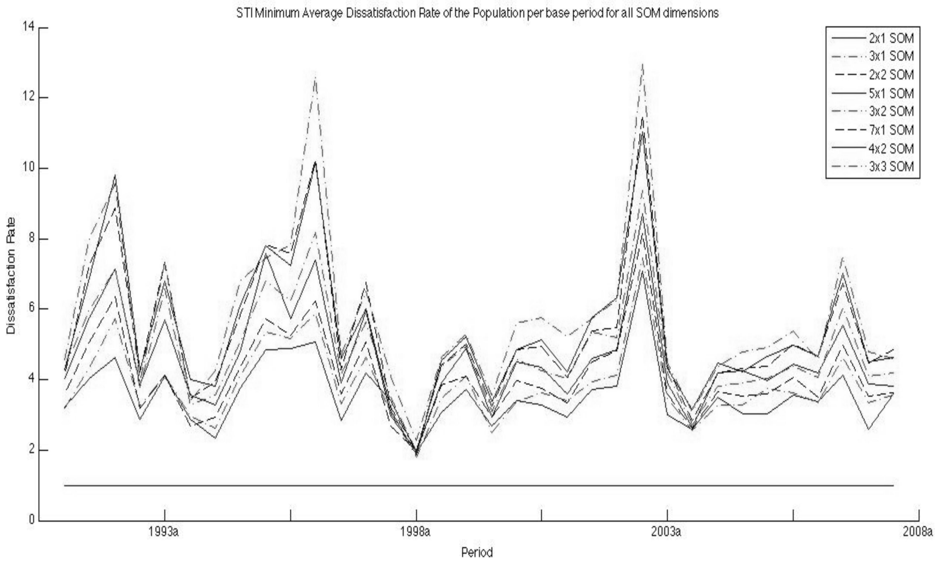


Fig. 9. STI Minimum average dissatisfaction rate of the population of GDTs per base period, for all SOM dimensions. After obtaining the series of the average dissatisfaction rate of each base period, over its future periods, we calculate the minimum average dissatisfaction rate per base period. In this way, we have an indication of whether there are any returning dinosaurs in each base period. The horizontal line indicates a dissatisfaction rate equal to 1, and is given as a reference.

DJIA, FTSE100, HSI, NASDAQ, NIKEI, NYSE, S&P500, STI, and TAIEX.

6.2.1. *Test 1*

Figures 11 and 12 present the results for Test 1. There are ten subfigures, each one presenting the minimum average dissatisfaction rate for all SOM dimensions, for each of the datasets. We recall that the horizontal line at each subfigure indicates when the average dissatisfaction rate is equal to 1; as we already know, whenever a period's rate is less than or equal to 1, then this period has a returning dinosaur. From the graphs it is clear that no dataset reaches a minimum average dissatisfaction rate of 1. Tables 4 and 5 also present the average of the average and minimum dissatisfaction rates, per cluster, per dataset. As we can observe, on average all datasets have an average dissatisfaction rate of 4.78 for the 2×1 SOM, which climbs to 7.95 for the 3×3 SOM. There seems to be a lower boundary, which does not allow the average dissatisfaction rate to go below it. This lower boundary is around 4.5-5 for the 2×1 SOM and rises with the number of clusters, reaching around 7-8 for the 3×3 SOM. However, even if we want to take into account the outliers, Table 5 informs us that on average the minimum dissatisfaction rate does not go below 3.40 for 2×1 SOM and 5.50 for 3×3 SOM. It is thus obvious that

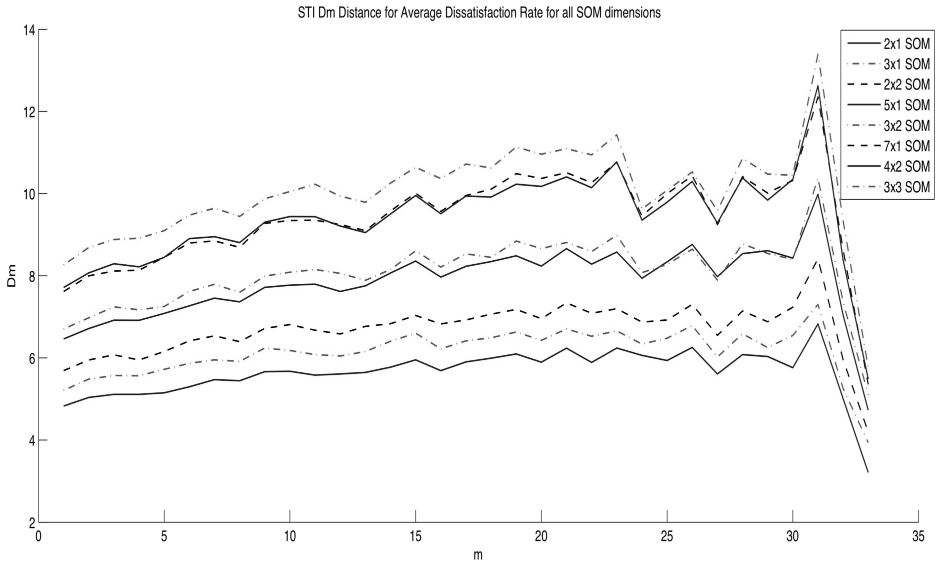


Fig. 10. STI D_m distance of the dissatisfaction rate for all SOM dimensions.

GDTs do not find a familiar environment in future periods. Hence, *dinosaurs do not return or return only as lizards*. The latter refers to the exceptional cases when dissatisfaction becomes relatively low, for instance around 2. Market conditions have allowed strategies from past time periods to feel 'satisfied' again, although this satisfaction is not as high as in the base period. Thus market conditions have some similarities to the ones of the base period, but are by no means the same nor do they offer the same 'satisfaction' to the strategies.

Table 4. Average of Average Dissatisfaction Rate per Cluster per Dataset

	2×1	3×1	2×2	5×1	3×2	7×1	4×2	3×3
CAC40	4.69	5.04	5.32	6.03	6.27	7.09	7.13	7.46
DJIA	4.93	5.46	5.88	6.69	7.05	7.76	7.98	8.51
FTSE100	4.77	5.12	5.55	6.37	6.62	7.43	7.52	8.05
HIS	5.06	5.49	5.79	6.68	6.90	7.75	7.87	8.32
NASDAQ	4.42	4.84	5.21	5.94	6.16	6.91	6.93	7.50
NIKEI	4.88	5.26	5.50	6.54	6.69	7.68	7.64	8.28
NYSE	4.44	4.90	5.22	5.92	6.17	6.88	6.96	7.27
S&P500	4.36	4.61	4.93	5.53	5.79	6.45	6.56	6.86
STI	5.17	5.65	6.11	6.98	7.19	8.30	8.33	8.88
TAIEX	5.04	5.48	5.74	6.54	6.96	7.76	7.78	8.40
Mean	4.78	5.19	5.53	6.32	6.58	7.40	7.47	7.95

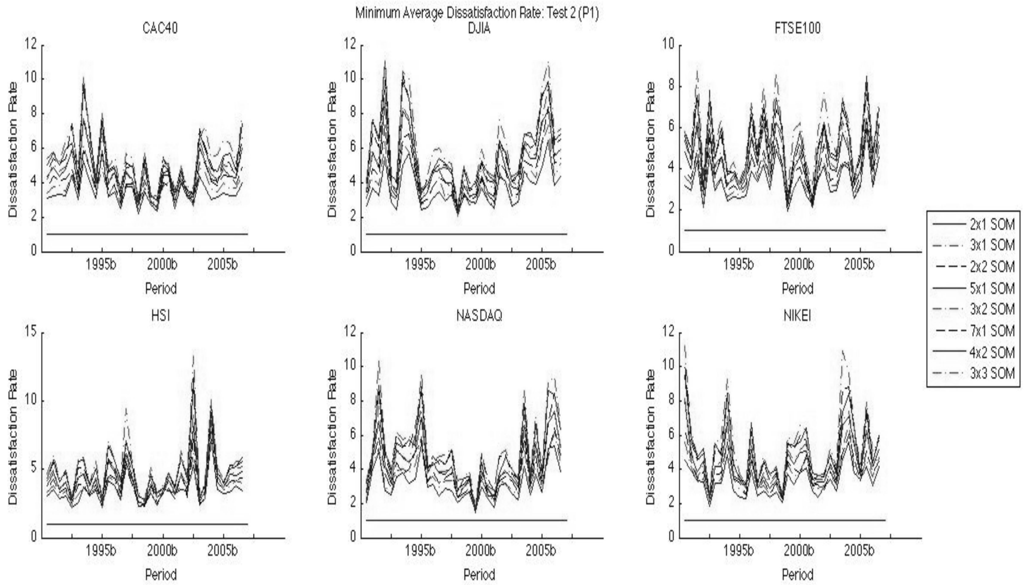


Fig. 11. Test 1: Minimum average dissatisfaction rate of the population of GDTs per base period, for all SOM dimensions, for all datasets. Each subfigure represents a single dataset. From left to right, top to bottom: CAC40, DJIA, FTSE100, HSI, NASDAQ, NIKEI.

Table 5. Average of Minimum Dissatisfaction Rate per Cluster per Dataset

	2×1	3×1	2×2	5×1	3×2	7×1	4×2	3×3
CAC40	3.51	3.77	4.01	4.53	4.68	5.10	5.13	5.51
DJIA	3.60	4.00	4.37	4.85	5.14	5.58	5.77	6.14
FTSE100	3.35	3.56	3.88	4.34	4.50	5.01	5.11	5.54
HSI	3.38	3.58	3.77	4.15	4.30	4.80	4.83	5.15
NASDAQ	3.16	3.51	3.74	4.20	4.35	4.89	4.84	5.25
NIKEI225	3.36	3.74	3.88	4.40	4.53	5.13	5.18	5.57
NYSE	3.52	3.94	4.14	4.69	4.98	5.33	5.53	5.56
SP&500	3.28	3.48	3.77	4.18	4.31	4.78	4.80	5.07
STI	3.56	3.83	4.09	4.61	4.79	5.34	5.41	5.79
TAIEX	3.29	3.64	3.83	4.25	4.44	4.88	5.00	5.43
Mean	3.40	3.71	3.95	4.42	4.60	5.09	5.16	5.50

6.2.2. *Test 2*

Let us now move to Test 2, which is presented in Figure 13. Each subfigure presents the results under a different market. DJIA, FTSE 100 and NIKEI seem to experience a continuous increase in the D_m metric. On the other hand, evidence for a

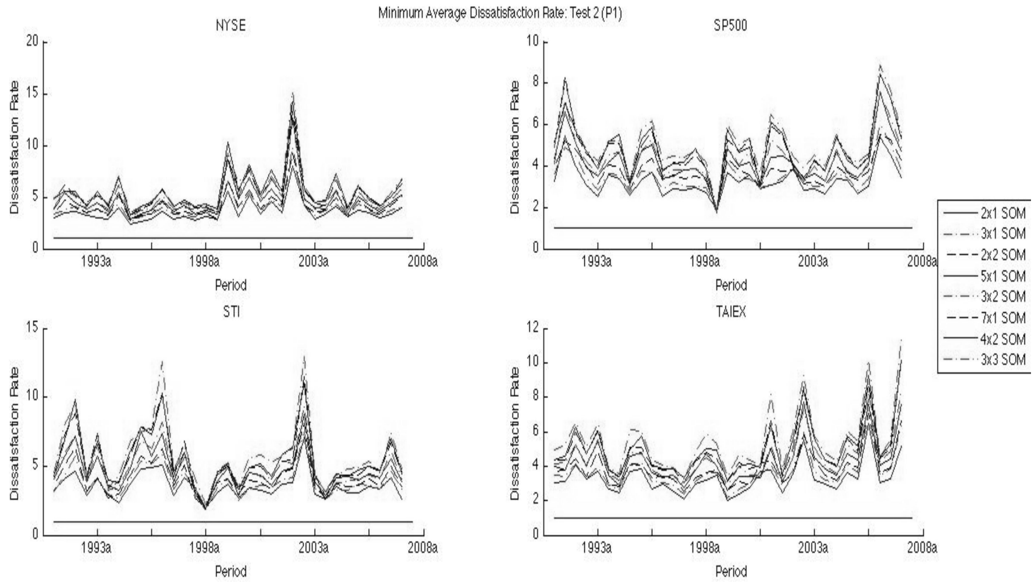


Fig. 12. Test 1: Minimum average dissatisfaction rate of the population of GDTs per base period, for all SOM dimensions, for all datasets. Each subfigure represents a single dataset. From left to right, top to bottom: NYSE, S&P500, STI, TAIEX

continuous increase in the other markets seems weak.

In order to confirm this eye-browsing observation, we run a linear trend regression for each dataset. Our goal is to examine if the slope of the trend is positive; if this happens, then it denotes a continuous increase in the metric. One more thing we should say is that we only run the linear regression for a single SOM dimension, the 3×3 one. The reason for this is because for each dataset, the D_m metric experiences the same patterns, under all SOM dimensions. Thus it would not make a big difference if we took another dimension and applied linear regression. It should also be mentioned that because the variance of the D_m observations seems relatively high, we apply variance stabilizing transformation, and use the logarithmic values instead of the raw values of the metric. Figure 14 presents the results of the linear regression. From the graphs we can clearly see that HSI and NYSE experience a continuous *decrease* in the metric. The majority of the remaining datasets' metric seems to remain stable over time, with the exception of DJIA, FTSE 100 and NIKEI.

Table 6 presents the coefficients, which denote the slope of the trend. In order to see whether the coefficients are statistically significant, we run a t-test and present the respective p-values. As we can observe, the p-value is below 5% (significance level) for the following datasets: DJIA, FTSE 100, HSI, NIKEI and NYSE. This means that DJIA, FTSE 100 and NIKEI 's positive slope is statistically significant.

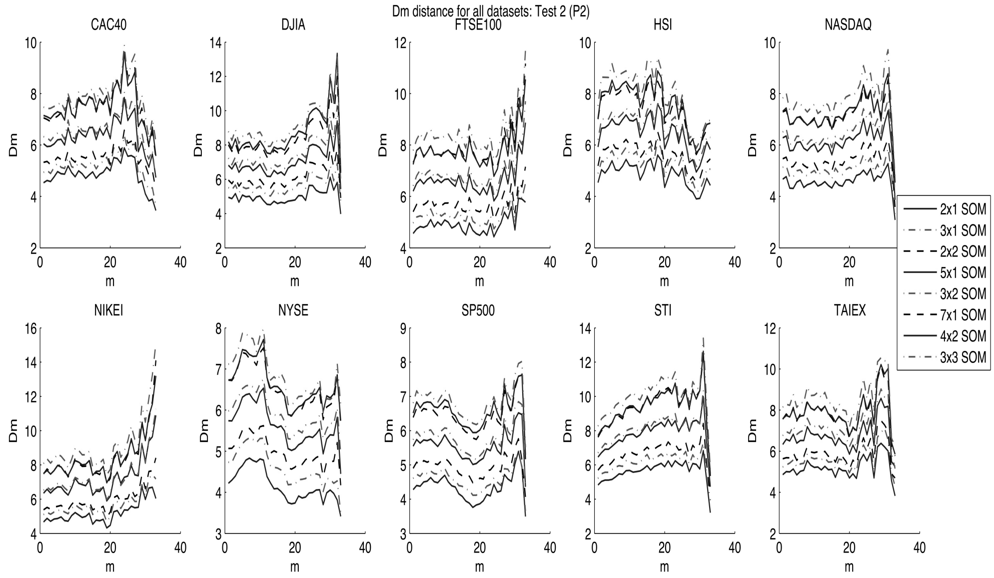


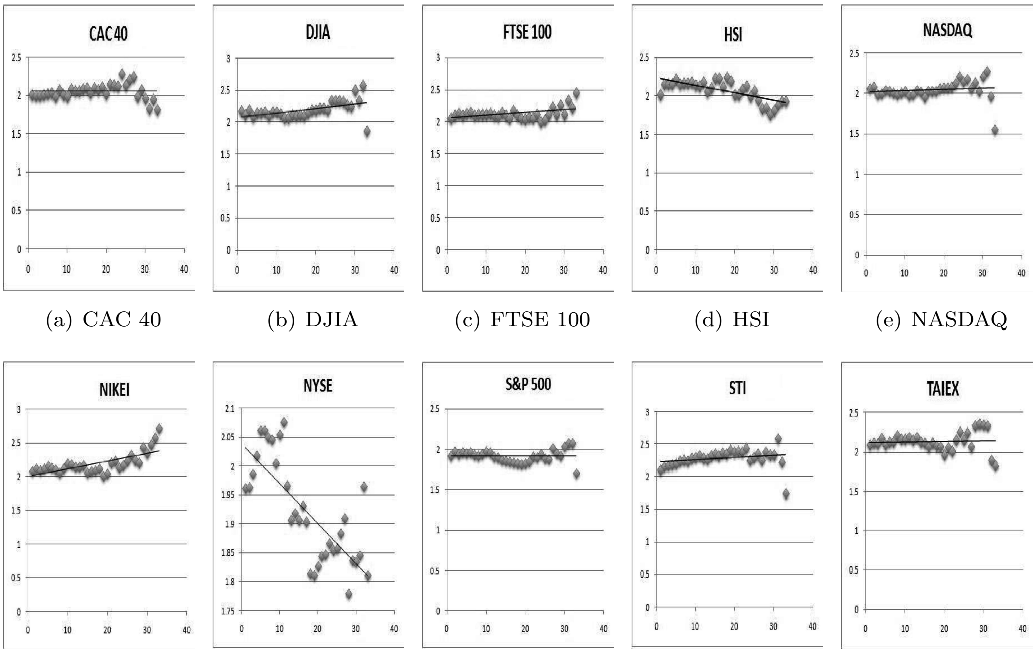
Fig. 13. Test 2: D_m distance of the dissatisfaction rate for all SOM dimensions for all datasets. Each subfigure represents a single dataset.

Thus *these markets experience a continuous increase* in their D_m metric. The remaining markets have either a statistically significant negative coefficient, thus a continuous decrease (HSI, NYSE), or no statistically significant positive/negative coefficient, thus no evidence for continuous increase (CAC 40, NASDAQ, S&P 500, STI, TAIEX).

Table 6. Slope of the trend and p-values per dataset

	Coefficient	p-value
CAC 40	0.000256	0.8905560
DJIA	0.006977	0.0027034
FTSE 100	0.003843	0.0154329
HSI	-0.009849	0.0000024
NASDAQ	0.001508	0.4836810
NIKEI	0.011413	0.0000053
NYSE	-0.006919	0.0000004
S&P 500	-0.000095	0.9468793
STI	0.003035	0.2231986
TAIEX	0.000558	0.7977585

The above result thus does not support Test 2, for the majority of the datasets tested in this paper (7 out of 10). However, as we saw there are 3 markets where we



(a) CAC 40 (b) DJIA (c) FTSE 100 (d) HSI (e) NASDAQ

(f) NIKEI (g) NYSE (h) S&P 500 (i) STI (j) TAIEX

Fig. 14. Linear Regression for the 3×3 SOM observations for the D_m metric.

actually observe statistically significant continuous increase in the D_m metric. This is very interesting, because it implies that over the period of 17 years that were examined in this work, the market behavior was changing continuously in such way that dinosaurs were having increased difficulties fitting in their respective market.

Lastly, it is also interesting that this behavior could only be observed in 3 markets only. What are the differences between these 3 markets and the other 7 remains a question, which we leave for future work. More research should take place in the statistical properties of these markets; this could lead us to a better understanding of the differences between DJIA, FTSE 100 and NIKEI with the rest of the markets.

7. Conclusion

To summarize, this paper presented the Dinosaur Hypothesis (DH) and suggested a testing methodology for it. This hypothesis says that the behavior of a market never settles down and that the strategies in this market continuously co-evolve with it. This observation was first made by [1] and later by [4]. However, these models made this observation based on simulations under artificial stock markets. In our current work, we were interested in examining whether these observations could hold as well in the real world and thus tested the hypothesis with empirical data. In order to do this, we extended a framework from a previous work [3], where we had empirically examined the dynamics of market fractions of different types of agents. In that study, we used Genetic Programming (GP) as a rule-inference engine to find out the behavioral rules of agents, and Self-Organizing Maps (SOM) to cluster these agents. However, because of an important assumption in that work, we had to require that SOM clusters, as well as their operational specification, would remain the same over time. In this paper, we relaxed that assumption. This offered more realism to our model and allowed us to investigate market behavior dynamics and test for the Dinosaur Hypothesis.

We used two tests, which were inspired by the works of Arthur [1], and Chen and Yeh [4]. The first test investigated whether strategies from the past can successfully be re-applied to the future. The results showed that this is not possible and thus verified this test. Strategies that have not co-evolved with the market become dinosaurs and cannot fit new environments. The second test investigated whether the dissatisfaction rate of such strategies, which are being applied to new environments, can follow a continuous increase. This was verified only by 3 out of the 10 datasets tested. Strategies are dissatisfied, or in other words do not fit their environments if they do not adapt to the changes that have taken place; however, this dissatisfaction does not necessarily increase continuously with time. Nevertheless, it is quite interesting that results regarding the continuous increase of the dissatisfaction rate are not homogeneous. This certainly deserves further investigation and we thus leave it as a future work.

What we can safely conclude from our experimental results is that trading strategies always need to follow changes that take place in the markets, and co-evolve with them. The implications of our results are very important. If strategies do not co-evolve with the market, they eventually become ineffective, even if they were very successful in the past. Markets constantly change and we need to follow and adapt to these changes if we want to remain successful.

Future research will include some changes in our model. At the moment, we have used GP and SOM as our two main tools. A question that arises is whether our results can be affected by different choices of GP and SOM algorithms, or even by different settings of these algorithms. In addition, it would be interesting to investigate whether our results would stand under a different framework, where different rule-inference machines and clustering techniques would be used. For example, Genetic Algorithms could be used instead of GP. Moreover, standard hierarchical clustering [24] or growing hierarchical self-organizing maps [7] could be used instead of SOM and provide us more details of the hierarchical structure of the market participants. Finally, another possible direction is to extend our testing methodology to other types of markets, like the Foreign Exchange Market (FOREX); this would allow us further generalization of our results.

Acknowledgments

The EPSRC grant No. EP/P563361/01 and the grant provided by AI-Econ Center, National Cheng Chi University, Taiwan are gratefully acknowledged. The first author is also grateful to Christos Michalopoulos from National Taiwan University for his valuable comments on earlier drafts of this paper.

References

[1] Arthur, B., On learning and adaptation in the economy (1992), working paper 92-07-038, Santa Fe Institute.

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

[3] Chen, S.-H., Kampouridis, M., and Tsang, E., Microstructure dynamics and agent-based financial markets, in *Multi-Agent-Based Simulation XI, 11th International Workshop, Revised Papers, LNAI*, eds. Bosse, T., Geller, A., and Jonker, C. M., Vol. 6532 (Springer-Verlag, Berlin Heidelberg, 2011), pp. 121–135, forthcoming.

[4] Chen, S.-H. and Yeh, C.-H., 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** (2001) 363–393.

[5] De Bondt, W. and Thaler, R., Does the stock market overreact?, *Journal of Finance* **40** (1985) 793–805.

[6] De Bondt, W. and Thaler, R., Further evidence on investor overreaction and stock market seasonality, *Journal of Finance* **42(3)** (1987) 557–581.

[7] Dittenbach, M., Rauber, A., and Merkl, D., Recent advances with the growing hierarchical self-organizing map, in *Proceedings of the 3rd Workshop on Self-Organizing Maps*, eds. Allinson, N., Yin, H., Allinson, L., and Slack, J., Advances in Self-Organizing Maps (Springer, Lincoln, England, 2001), pp. 140–145.

[8] Duffy, J. and Engle-Warnick, J., *Using symbolic regression to infer strategies from experimental data* (Springer, 2002), pp. 61–82.

[9] Gigerenzer, G. and Todd, P., *Fast and Frugal Heuristics: The Adaptive Toolbox*, Gigerenzer, G. Todd, P. and the ABC Research Group (Oxford University Press, 1999), pp. 3–34.

[10] Izumi, K. and Ueda, K., Analysis of dealers' processing financial news based on an artificial market approach, *Journal of Computational Intelligence in Finance* **7** (1999) 23–33.

[11] Kampouridis, M. and Tsang, E., EDDIE for investment opportunities forecasting: Extending the search space of the GP, in *Proceedings of the IEEE Conference on Evolutionary Computation* (Barcelona, Spain, 2010), pp. 2019–2026.

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

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

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

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

[16] Marshall, A., *Principles of Economics* (MacMillan, New York, 1924).

[17] Poli, R., Langdon, W., and McPhee, N., *A Field Guide to Genetic Programming* (Lulu.com, 2008).

[18] Provost, F. and Kohavi, R., Glossary of terms, *Journal of Machine Learning* **30(2-3)** (1998) 271–274.

[19] Raberto, M., Cincotti, S., Focardi, S.M. and Marchesi, M., Traders' Long-Run Wealth in an Artificial Financial Market, *Computational Economics*, **22(2)**, Springer, (2003) 255-272

[20] Simon, H., Rational choice and the structure of environments, *Psychological Review* **63** (1956) 129–138.

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

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

[23] Tsang, E., Yung, P., and Li, J., EDDIE-Automation, a decision support tool for financial forecasting, *Journal of Decision Support Systems, Special Issue on Data Mining for Financial Decision Making* **37(4)** (2004) 559–565.

[24] Xu, R. and Wunsch, D., *Clustering* (Wiley-IEEE Press, 2008).