

Investigating the Effect of Different GP Algorithms on the Non-Stationary Behavior of Financial Markets

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

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

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

Abstract—This paper extends a previous market microstructure model, where we used Genetic Programming (GP) as an inference engine for trading rules, and Self Organizing Maps as a clustering machine for those rules. Experiments in that work took place under a single financial market and investigated whether its behavior is non-stationary or cyclic. Results showed that the market's behavior was constantly changing and strategies that would not adapt to these changes, would become obsolete, and their performance would thus decrease over time. However, because experiments in that work were based on a specific GP algorithm, we are interested in this paper to prove that those results are independent of the choice of such algorithms. We thus repeat our previous tests under two more GP frameworks. In addition, while our previous work surveyed only a single market, in this paper we run tests under 10 markets, for generalization purposes. Finally, we deepen our analysis and investigate whether the performance of strategies, which have not co-evolved with the market, follows a continuous decrease, as it has been previously suggested in the agent-based artificial stock market literature. Results show that our previous results are not sensitive to the choice of GP. Strategies that do not co-evolve with the market, become ineffective. However, we do not find evidence for a continuous performance decrease of these strategies.

I. INTRODUCTION

There are several types of models in the 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 (see [1] for details). The former type of models focuses on the mesoscopic level of markets, by allowing agents to choose among different types of strategies. A typical example is the fundamentalist-chartists model. Agents in this model are presented with these two strategy types and at any given time they have to choose between these two. A typical area of investigation of 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 these two 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 Algorithms [2], these models allow the creation and evolution of novel agents, which are not constrained by pre-specified strategy types.¹ However, this kind of models tend to focus on price dynamics, rather than fraction dynamics [1].

In previous works [3], [4], we presented a model which combined properties from the N -type and SFI-like models: by using Genetic Programming (GP) [5] as a rule inference engine, we created and evolved autonomous agents; we then used Self-Organizing Maps (SOM) [6] 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. We then investigated aspects of markets' microstructure, such as fraction dynamics [3], and behavior dynamics [4]. In the current paper, we will focus on [4], which examined whether the behavior of financial markets is non-stationary, as first indicated by Arthur in [7].² Tests in [4] took place under a single market and results showed that such non-stationary behavior exists; in fact, unless trading strategies continuously co-evolved with the market, they became obsolete or what Arthur called *dinosaurs*.

However, a question that can arise is whether these results are sensitive to the choice of GP and SOM algorithms. Could different rule inference engines or different clustering techniques affect our results? In this paper, we focus on the GP algorithms and investigate the behavior of the financial markets under two different GP systems. Our goal is to show that the previous results are *rigorous* and thus not dependent on the choice of GP. Proving this is a very important task, because it allows us to generalize the results in [4]. Also for generalization purposes, we increase the number of markets tested from 1 to 10. Finally, we deepen our investigation by examining an observation made by Chen and Yeh [8] under their artificial stock market, where they found that a

¹We refer the reader to [1], which provides a thorough review on both N -type and SFI-like models, with a detailed list of these models.

²Arthur's work took place under SFI's artificial stock market. In [4] we examined whether his observations could hold under an empirical dataset.

dinosaur's performance follows a monotonic decrease. We test this observation under empirical datasets. This is a *new* test, which did not take place in [4].

The rest of this paper is organized as follows: Section II presents the model used for our experiments, Section III presents the two GP algorithms we use in this paper, and also compares them with the one used in the original paper [4]. Section IV discusses the experimental designs, and Section V presents the results of our experiments. Finally, Section VI concludes this paper and discusses future work.

II. MODEL

Our model combines the property of novel agents that SFI-like models offer, with the property of clusters/types of the N -type models. In order to do this, we use GP as a rule inference engine, and SOM as a clustering technique for these rules. This section describes in details how we have used these two tools, and also explains our motivation for using them.

A. Genetic Programming As A Rule-Inference Engine

The use of GP is motivated by considering the market as an evolutionary and selective process.³ 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. An advantage of GP is that it does not rest upon any pre-specified class of behavioral rules, like many other models in the agent-based finance literature [1]. 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.⁴

Hence, even though we are not informed of the behavioral rules followed by traders, GP can help us infer what these rules are approximately, by simulating the evolution of the microstructure of the market. Traders can then be clustered based on realistic, and possibly complex behavioral rules.⁵ Details of the specific GP algorithms used in this paper follow in Section III. The next subsection explains how the clustering of the rules takes place.

B. Self Organizing Maps For Clustering

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 of the

markets. The similarity criterion which we choose is based on the *observed trading behavior*. 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 timing⁶ behavior.

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 we denote the decision “enter the market” by “1” and “leave the market” by “0”, then the behavior of each rule is a binary vector, which we call market-timing 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 (SOM) to the associated clusters.

Figure 1 presents an example of a 3×3 SOM, where, 500 artificial traders are grouped into 9 clusters. In a sense, this could be perceived as a snapshot of a nine-type agent-based financial market dynamics. Traders of the same type indicate that their market timing behavior is very similar. We can observe that the largest cluster has a market share of 71.2% (356/500), whereas the smallest one has a market share of 0.2% (1/500).

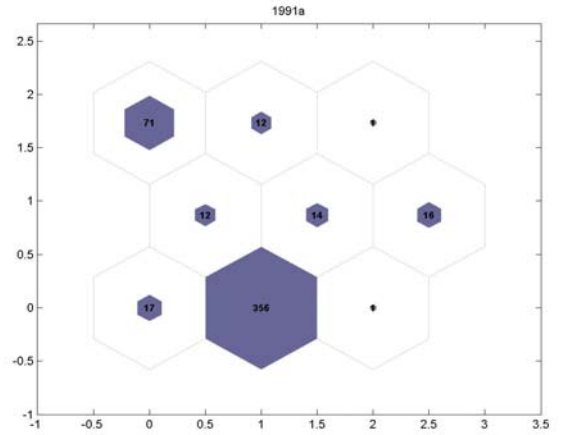


Fig. 1. Example of a 3×3 Self-Organizing Map

III. THE DIFFERENT GP ALGORITHMS

In this section, we present the two GP algorithms that we will be using for our experiments, namely EDDIE 7 and EDDIE 8. These two algorithms should be considered as extensions and improvements of the *simple GP* presented in [4]. For this reason, we start with a short presentation of the *simple GP*. We then continue by presenting EDDIE 7 and EDDIE 8, and how they have extended the *simple GP*.

⁶Market timing refers to the strategy of making buy or sell decisions of stocks, by attempting to predict future price movements.

³See [9] for his eloquent presentation of the *Adaptive Market Hypothesis*.

⁴It does not mean that all types of traders surviving must be smart and sophisticated. They can be dumb, naive, randomly behaved or zero-intelligent. Obviously, the notion of rationality applied here is *ecological* [10], [11].

⁵[12] provides the first illustration of using genetic programming to infer the behavioral rules of human agents in the context of ultimatum game experiments.

1) *Simple GP Algorithm*: The *simple GP* was inspired by a financial forecasting tool, EDDIE [13], [14], 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 used indicators commonly used in technical analysis: Moving Average (MA), Trader Break Out (TBR), Filter (FLR), Volatility (Vol), Momentum (Mom), and Momentum Moving Average (MomMA).⁷ 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) was syntactically (grammatically) produced by the Backus Naur Form (BNF). Figure 2 presents the BNF grammar of the GP. As we can see, the root of the tree was an If-Then-Else statement. Then the first branch was a Boolean (testing whether a technical indicator was greater than/less than/equal to a value). The ‘Then’ and ‘Else’ branches could be a new GDT, or a decision, to buy or not-to-buy (denoted by 1 and 0). Thus, each individual in the population was a GDT and its recommendation was to buy (1) or not-buy (0). Depending on

```

<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. 2. The Backus Naur Form of the simple GP uses to construct trees

the classification of the predictions we could have are four cases: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). As a result, we used the following 3 metrics, presented in Equations (1)-(3):

Rate of Correctness

$$RC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Rate of Missing Chances

$$RMC = \frac{FN}{FN + TP} \quad (2)$$

Rate of Failure

$$RF = \frac{FP}{FP + TP} \quad (3)$$

The above metrics combined gave the following fitness function, shown in Equation (4):

$$ff = w_1 * RC - w_2 * RMC - w_3 * RF \quad (4)$$

⁷We used these indicators because they have been proved to be quite useful in previous works like [15]

where w_1 , w_2 and w_3 were the weights for RC, RMC and RF respectively, and were 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 were set to 1, $\frac{1}{6}$ and $\frac{1}{2}$ respectively, and were given in this way in order to reflect the importance of each performance measure for our predictions.

2) *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 [14]. The constraint is denoted by R , which consists of two elements represented by percentage, given by

$$R = [C_{min}, C_{max}],$$

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.

Hence, a constraint of $R = [50, 65]$ means that the percentage of positive signals that a GDT predicts⁸ will 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 0.

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. Let us now see how EDDIE 8 works.

3) *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. Function f is unknown to the user and is the GDTs that EDDIE generates. As mentioned earlier, EDDIE 7 and the *simple GP* use 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 $y = f(g(z))$, where $x = g(z)$; in other words, x is now a function, 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 GDTs that are using any period within a range that is defined by the user.

As we can see from the new syntax in Figure 3, the *Variable* symbol has been replaced by 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 MA, 17 days TBR, and so on. The period is not an issue anymore, and it is up to EDDIE 8 to decide which lengths are more valuable for the prediction.

⁸As 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 Vol period
Mom period MomMA period
<RelationOperation> ::= ">" "<" "="
Terminals:
MA, TBR, FLR, Vol, Mom, MomMA are function symbols
Period is an integer within a parameterised range, [MinP, MaxP]
Decision is an integer, 1 or 0 (to buy or not-to-buy)
Threshold is a real number

Fig. 3. The Backus Naur Form of EDDIE 8

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.⁹

Given a set of historical data and the fitness function, the above GP algorithms are applied to evolve the market-timing strategies in a standard way. After evolving a number of generations, 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.

IV. EXPERIMENTAL DESIGNS

The experiments can be divided into two parts: in the *first part*, we investigate whether financial markets experience a non-stationary behavior. If this is the case, then strategies that do not adapt to these changes, should become dinosaurs, as Arthur states [7]. In the *second part*, we investigate whether there is a continuous decrease in the performance of dinosaurs over time, as [8] states.

The experiments are conducted for a period of 17 years (1991-2007) and the data is 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), NIKKEI 225 (Japan), NYSE (USA), S&P 500 (USA), STI (Singapore) and TAIEX (Taiwan). For each of these markets, we run each experiment for 10 times.

Each year is split into 2 halves (January-June, July-December), so in total, out of the 17 years, we have 34 periods.¹⁰ 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). The GP systems is therefore executed 34 times, i.e., one time per period. Table I presents the GP parameters for our experiments.

⁹For a comparative analysis between EDDIE 7 and EDDIE 8, see [15].

¹⁰At this point the length of the period is 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 I
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

Finally, we define as 'base period', the period during which GP creates and evolves GDTs. We also define 'future period(s)', as the period(s) which follow(s) the base period (in chronological order).

A. Non-Stationary Behavior

After generating and evolving strategies for each one of the 34 periods, we then use SOM to cluster these strategies into types. We do this for every one of the 34 periods. Thus, we end up with 34 SOMs, which represent the market in different time periods over the 17 year horizon. Then, in order to investigate whether the behavior of markets is non-stationary, we re-cluster the GDTs of each base period, to all future periods' clusters.¹¹ By applying the same GDTs (strategies) to clusters of future periods, we can observe how well these strategies fit in new environments (clusters). The logic behind this is the following: when we first evolved and clustered the GDTs (base period), these GDTs were placed in clusters that represented their respective strategies. For instance, if there was a strategy type (cluster) that represented 'fundamentalists', then all GDTs which follow a fundamental strategy would have been placed in there. When we then take the GDTs from a base period and re-cluster them to strategy types of future periods, it is not guaranteed that there will again be a cluster that represents fundamentalists. If market 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 these GDTs. Of course, since now the SOM of the future period is formed by different clusters, the GDTs might not fit in 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 to a future period's cluster that does not represent their strategy. If the market is non-stationary, then because the market conditions have significantly changed, the GDTs' dissatisfaction rate will be high. The dissatisfaction rate of is defined as the Euclidean distance of a GDT's market-timing vector to the centroid of the cluster it is placed after the re-clustering procedure.

Let us now explain the process of re-clustering. We start with 1991a as the base period. Each evolved GDT is moved to

¹¹The process of re-clustering is explained later in this section.

the next period, 1991b, and re-clustered to 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 to the centroid of each cluster; it is then placed to 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 will have been reclustered to the clusters of period 1991b. The same procedure is followed in all future periods. This means that the GDTs from 1991a are also re-clustered to 1992a, 1992b, ..., 2007b. Finally, the same process is done for all other base periods (i.e. 1991b, 1992a, ..., 2007a).

Once the process of re-clustering is completed, 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 in the base period. If, on the other hand, this rate reaches 1 or below, it is an indication of a cyclic market behavior, since the GDTs have found the same conditions with the base period, and as a result felt as ‘satisfied’ as before.

Finally, we define as *dinosaurs* the population of GDTs that has been re-clustered from a base period to future periods, because these GDTs have not adapted to the new market environment (clusters of the SOMs from future periods). When these GDTs’ normalized average dissatisfaction rate drops to less than or equal to 1, we call them *returning dinosaurs*.

B. Continuous Performance Drop

We are also interested in investigating Chen and Yeh’s [8] observation regarding the monotonic decrease in the performance of dinosaurs. However, since in this paper the performance is measured by the dissatisfaction rate, we will be investigating the nature of the *increase* of this rate, instead of decrease. But we do not require this increase to be monotonic. The reason for this is because Chen and Yeh ran tests over a period-window of 20 days, which is relatively short, hence easy to achieve monotonic changes. Thus, requiring that a predictor’s performance changes monotonically in the long run would be very strict, and indeed hard to achieve. For that reason, we will be examining whether the dissatisfaction rate’s increase is continuous, rather than not monotonic.

V. RESULTS

A. Returning Dinosaurs

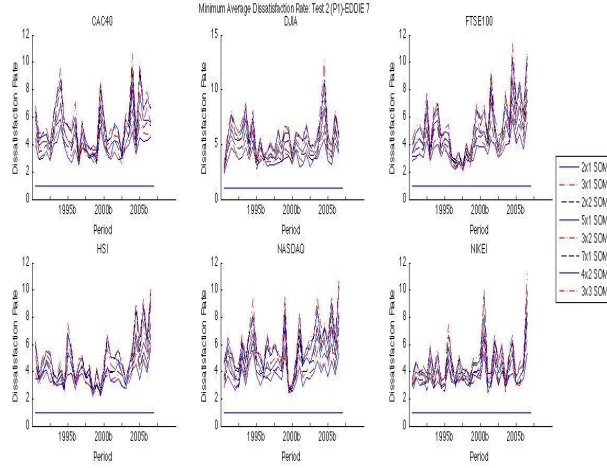
As already explained, returning dinosaurs denote a cyclic market behavior. In order to examine if dinosaurs return, we iterate through each base period and calculate the minimum normalized average dissatisfaction rate for each future period.

This will give us an indication of how many returning dinosaurs, if any, exist. If, for instance, 1991a is the base period, then there is a series of 33 population dissatisfaction values for its future periods. We obtain the minimum value among these 33 values, in order to check how close to 1 this future period is. This process is then repeated for 1991b and its 32 future periods, 1992a, and so on, until base period 2007a. We thus end up with a 1×33 vector, called *DissatisfVect* which shows whether any returning dinosaurs exist. In addition, we are interested in investigating whether the number of strategy types that exist in a market, can affect the test results. We thus run tests for 2-9 types, under the following SOM dimensions: 2×1 , 3×1 , 2×2 , 5×1 , 3×2 , 7×1 , 4×2 , and 3×3 . The graphs of the *DissatisfVect* vectors are presented in Figures 4(a) and 4(b) for EDDIE 7, and Figures 4(c) and 4(d) for EDDIE 8.

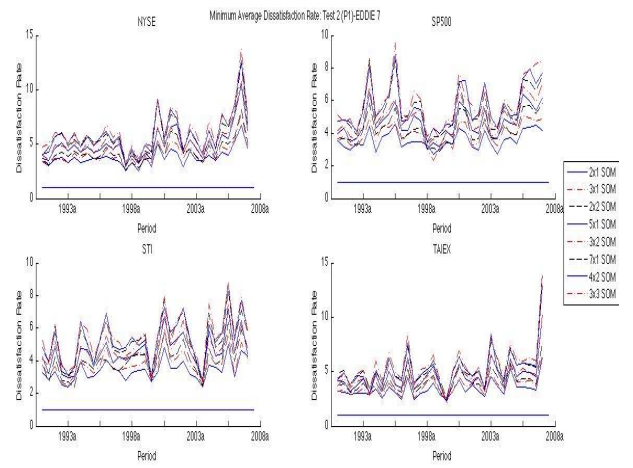
However, before discussing the results in these figures, let us first summarize the results returned by the *simple GP* in [4]. In that work, we presented results under a single market, STI (Singapore). Results showed that *there were no returning dinosaurs*. In fact, the dissatisfaction rate never went below 2 (with 1 being the threshold of a returning dinosaur, as discussed in the previous section).

Let us now go back to Figure 4. Each subfigure presents the dissatisfaction rates of a single market. Each line in the subfigures represents the results for a different SOM dimension. The horizontal line denotes a dissatisfaction rate of 1, and is given as a reference. Results are on average of 10 runs. What we can see from these graphs is that there are no base periods with a minimum normalized dissatisfaction rate below 1. In fact, the closest to 1 this rate gets is again around 2: FTSE100 (1998a), HSI (2000b), S&P500 (1998a), STI (1993a) for EDDIE 7 and NYSE (1998b), S&P500 (1998b), STI (2000a, 2003a), TAIEX(1997b) for EDDIE 8. This applies to all SOM dimensions. Furthermore, Tables II(a) and II(b) present the average of the minimum dissatisfaction rate per cluster, for EDDIE 7 and EDDIE 8, and verify this observation. As we can see, the minimum dissatisfaction rate is on average 3.58 and 3.37 for the 2×1 SOM, for EDDIE 7 and EDDIE 8, respectively; it gradually increases as the number of clusters increases, reaching 5.71 and 5.30 for the 3×3 SOM, under EDDIE 7 and EDDIE 8, respectively. Hence, the minimum dissatisfaction rate is on average quite far away from 1.¹² These results are inline with the ones under the *simple GP*. We can thus argue that *dinosaurs do not return or return only as lizards*. The latter refers to the exceptional cases when dissatisfaction goes relatively low, for instance around 2. Market conditions have allowed unadapted strategies to feel ‘satisfied’ again, although this satisfaction is not as high as in the base period. Thus market conditions from future periods show similarities to the ones of the base period; however, these conditions do not by any means become the same, at any stage.

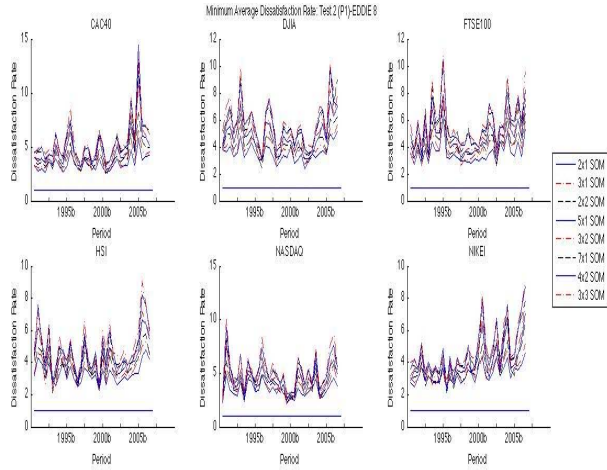
¹²One might support that using an average measure would not show the whole picture, because some the fluctuations in the rates are a result of the whole population, rather than individual strategies. However, this does not pose a problem to our research, because we are interested in investigating the market as a whole, and not just individual strategies.



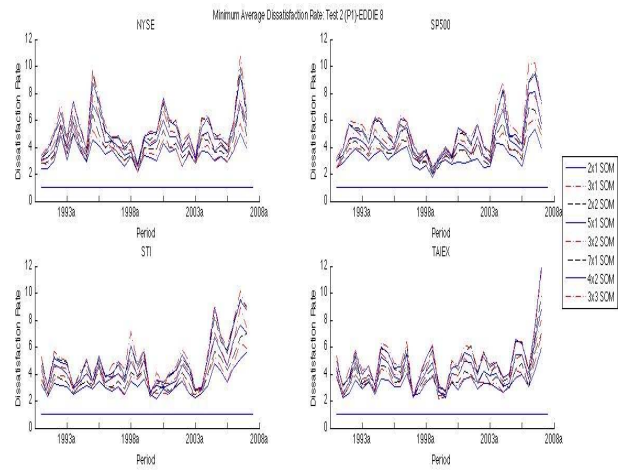
(a) CAC 40, DJIA, FTSE 100, HSI, NASDAQ, NIKEL. (EDDIE 7)



(b) NYSE, S&P 500, STI, TAIEX. (EDDIE 7)



(c) CAC 40, DJIA, FTSE 100, HSI, NASDAQ, NIKEL. (EDDIE 8)



(d) NYSE, S&P 500, STI, TAIEX. (EDDIE 8)

Fig. 4. Minimum normalized population dissatisfaction rate among all future periods, per base period. Each line represents a different SOM dimension. Each subfigure presents a different international market. Results are for EDDIE 7 (top) and EDDIE 8 (bottom).

The above observation leads us to conclude that market behavior constantly changes. Nevertheless, sometimes it can resemble old environments; when this happens, unadapted strategies might perform relatively well again (i.e. dinosaurs return as lizards). However, these strategies can never become successful again, unless they follow the changes that have happened in the market.

One final observation we can make is that the number of clusters does not affect the test's results. The dissatisfaction rate of each market follows always the same pattern, regardless the number of clusters. No returning dinosaurs are observed, under any number of the trading strategy types tested.

B. Continuous Increase In The Dissatisfaction Rate

As mentioned in Section I, the test in the following section is a contribution of this paper, and was not presented in [4]. 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 III(a). Each row of this table presents the distance of the future periods from their base period. For instance, if 91a is the base (first row), then future period 91b has distance equal to 1, future period 92a has distance equal to 2, and so on. Table III(b) presents an example of 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 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,

TABLE II
AVERAGE OF MINIMUM DISSATISFACTION RATE PER CLUSTER, PER DATASET, FOR EDDIE 7 AND EDDIE 8.

(a) EDDIE 7									(b) EDDIE 8								
	2x1	3x1	2x2	5x1	3x2	7x1	4x2	3x3		2x1	3x1	2x2	5x1	3x2	7x1	4x2	3x3
CAC40	3.66	3.89	4.31	4.77	4.95	5.54	5.65	5.91	CAC40	3.44	3.68	3.99	4.54	4.66	5.23	5.32	5.50
DJIA	3.74	4.11	4.34	4.88	5.07	5.65	5.79	6.09	DJIA	3.56	3.94	4.18	4.66	4.92	5.49	5.49	5.84
FTSE100	3.55	3.78	4.20	4.62	4.82	5.38	5.47	5.80	FTSE100	3.54	3.89	4.24	4.65	4.83	5.51	5.56	5.84
HSI	3.50	3.58	3.93	4.12	4.49	4.92	4.95	5.13	HIS	3.29	3.56	3.78	4.20	4.19	4.51	4.66	4.88
NASDAQ	3.75	4.14	4.48	4.92	5.19	5.72	5.87	6.26	NASDAQ	3.43	3.72	3.92	4.26	4.44	4.91	4.95	5.24
NIKEI	3.47	3.65	3.76	4.11	4.22	4.85	4.67	5.03	NIKEI	3.28	3.47	3.58	3.99	4.14	4.54	4.64	4.77
NYSE	3.80	4.10	4.33	4.99	5.11	5.74	5.90	6.22	NYSE	3.44	3.72	4.02	4.44	4.56	5.13	5.26	5.47
S&P500	3.55	3.91	4.21	4.61	4.89	5.41	5.61	5.83	S&P500	3.25	3.72	3.86	4.30	4.49	4.93	5.04	5.34
STI	3.37	3.72	3.95	4.45	4.54	5.11	5.22	5.34	STI	3.18	3.52	3.70	4.19	4.28	4.81	4.86	5.10
TAIEX	3.43	3.71	4.07	4.53	4.58	5.31	5.23	5.54	TAIEX	3.26	3.48	3.68	4.09	4.27	4.68	4.85	5.01
Mean	3.58	3.86	4.16	4.60	4.79	5.36	5.43	5.71	Mean	3.37	3.67	3.90	4.33	4.48	4.97	5.06	5.30

as presented in Table III(a), and m is the distance from the base period and 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).

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

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 GDT from base period 91a, when applied to future period 07a) and $Dis(91b, 07b)$ (dissatisfaction rate of GDT 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.¹³ 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.

Figures 5(a) and 5(b) present the results for all datasets, under EDDIE 7 and EDDIE 8, respectively. Each line represents again a different SOM dimension. As we can see, results under both algorithms can lead us to the same conclusion: none of the 10 international markets experiences a continuous increase in the dissatisfaction rate. This thus allows us to conclude that although the dissatisfaction rate increases once we move away from the base period, it does not follow a continuous increase. The performance of the trading strategies does not continuously deteriorate over time.

¹³The distance $m = 32$ can also be found in 07a91a and 07b91b. However, we do not take them into account because we are not interested in applying the evolved GDT of a base period (here 07a and 07b) backwards in time (91a and 91b, respectively).

VI. CONCLUSION

To conclude, this paper extended a previous work [4], where we presented a framework for investigating the behavior dynamics of financial markets. However, because part of the experiments were based on a GP algorithm, we were interested in presenting rigorous tests which would show that our previously derived results were not sensitive to the choice of GP. We thus repeated the tests under two more GP algorithms. In addition, we increased the number of markets tested in [4], from 1 to 10, to generalize our results. Finally, we also extended the original paper by testing an observation made by Chen and Yeh [8] under their artificial stock market framework; this observation said that the performance of strategies that have not been co-evolving with the market, decreases monotonically.

A first observation that can be made is that homogeneity exists across the results. All three GP algorithms have shown similar patterns and have not shown any returning dinosaurs for any of the 10 markets tested. Therefore, the first goal of this paper has been successful. Our results seem to be independent to the choice of GP. The fact that there are no returning dinosaurs of course indicates that unless the strategies continuously adapt to the changes happening in the market, they cannot survive. Markets' behavior constantly changes. This is an important observation, because it implies that trading strategies have a finite lifetime. They cannot be used for a long period of time, unless they co-evolve with the market. Finally, results regarding the continuous increase in the dissatisfaction rate showed that this does not happen under any of the markets or algorithms tested in this paper. This means that future market environments can resemble environments in the past, although they never become the same again. Thus, dinosaurs (unadapted strategies) can occasionally return as lizards, meaning that they can show some relatively good performance, but they fail to become as successful as they initially were.

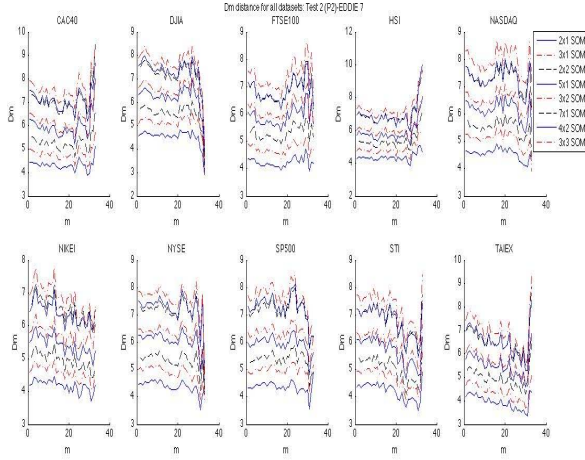
Future work will focus on investigating the behavior of the markets under different SOM algorithms. In addition, we are planning to inject specific strategy types in the markets

TABLE III

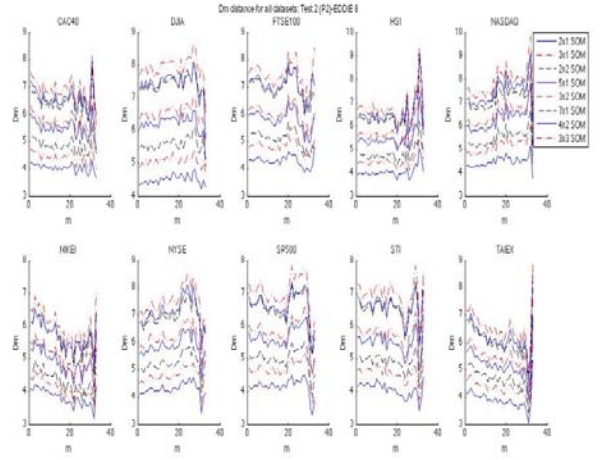
(A) DISTANCE OF FUTURE PERIODS FROM THEIR BASE PERIOD, OVER THE 17 YEARS 1991-2007. THE FURTHER AWAY WE MOVE FROM A PERIOD, A SINGLE UNIT OF DISTANCE IS ADDED. (B) EXAMPLE OF SERIES OF FUTURE POPULATION DISSATISFACTION RATES PER BASE PERIOD. EACH BASE PERIOD'S SERIES IS PRESENTED AS A HORIZONTAL LINE OF THIS TABLE. DISSATISFACTION RATES HAVE BEEN NORMALIZED, SO THAT THE RATE IN THE BASE PERIOD IS ALWAYS EQUAL TO 1.

(a)							(b)								
		91a	91b	j 92a	92b	...	07b			91a	91b	j 92a	92b	...	07b
i	91a	0	1	2	3	...	33	i	91a	1	7.66	6.82	7.08	...	5.94
	91b	1	0	1	2	...	32		91b		1	3.76	5.72	...	4.70
	92a	2	1	0	1	...	31		92a			1	9.74	...	7.77
	92b	3	2	1	0	...	30		92b				1	...	8.80

	07b	33	32	31	30	...	0		07b					...	1



(a) EDDIE 7



(b) EDDIE 8

Fig. 5. D_m values for all m from 1 to 33. Each subfigure represents a single dataset. Each line shows a different SOM dimension.

(e.g. chartists) and observe how the behavior of the market is affected by such strategies, and how such dinosaurs might act in the long run.

ACKNOWLEDGMENT

This work was supported by an EPSRC grant (No. EP/P563361/01) and by AI-Econ Center of National Cheng Chi University, Taiwan.

REFERENCES

- [1] S.-H. Chen, C.-L. Chang, and Y.-R. Du, "Agent-based economic models and econometrics," *Journal of Knowledge Engineering Review* (forthcoming), 2010.
- [2] J. Holland, *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, 1975.
- [3] S.-H. Chen, M. Kampouridis, and E. Tsang, "Microstructure dynamics and agent-based financial markets," in *Multi-Agent-Based Simulation XI, 11th International Workshop, Revised Papers, LNAI*, T. Bosse, A. Geller, and C. M. Jonker, Eds., vol. 6532. Berlin Heidelberg: Springer-Verlag, 2011, pp. 121–135, forthcoming.
- [4] M. Kampouridis, S.-H. Chen, and E. Tsang, "Market microstructure: Can dinosaurs return? A self-organizing map approach under an evolutionary framework," in *EvoApplications 2011*, ser. Lecture Notes in Computer Science. Springer, 2011, p. forthcoming.
- [5] J. Koza, *Genetic Programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT Press, 1992.
- [6] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Journal of Biological Cybernetics*, vol. 43, pp. 59–69, 1982.
- [7] B. Arthur, "On learning and adaptation in the economy," 1992, working paper 92-07-038, Santa Fe Institute.
- [8] S.-H. Chen and C.-H. Yeh, "Evolving traders and the business school with genetic programming: A new architecture of the agent-based artificial stock market," *Journal of Economic Dynamics & Control*, vol. 25, pp. 363–393, 2001.
- [9] A. Lo, "The adaptive market hypothesis: market efficiency from an evolutionary perspective," *Journal of Portfolio Management*, vol. 30, pp. 15–29, 2004.
- [10] H. Simon, "Rational choice and the structure of environments," *Psychological Review*, vol. 63, pp. 129–138, 1956.
- [11] G. Gigerenzer and P. Todd, *Fast and Frugal Heuristics: The Adaptive Toolbox*, ser. Gigerenzer, G. Todd, P. and the ABC Research Group. Oxford University Press, 1999, pp. 3–34.
- [12] J. Duffy and J. Engle-Warnick, *Using symbolic regression to infer strategies from experimental data*, ser. Evolutionary Computation in Economics and Finance. Springer, 2002, pp. 61–82.
- [13] E. Tsang, J. Li, S. Markose, H. Er, A. Salhi, and G. Iori, "EDDIE in financial decision making," *Journal of Management and Economics*, vol. 4(4), 2000.
- [14] E. Tsang, S. Markose, and H. Er, "Chance discovery in stock index option and future arbitrage," *New Mathematics and Natural Computation, World Scientific*, vol. 1(3), pp. 435–447, 2005.
- [15] M. Kampouridis and E. Tsang, "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.