Market Fraction Hypothesis: A Proposed Test

Michael Kampouridis
School of Computer Science and Electronic Engineering
University of Essex
UK
mkampo@essex.ac.uk

Shu-Heng Chen
Department of Economics
National Chengchi University
Taipei, Taiwan 116
chchen@nccu.edu.tw

Edward Tsang
School of Computer Science and Electronic Engineering
University of Essex
UK
edward@essex.ac.uk

August 27, 2010

Abstract

Within a market, we can have different types of trading strategies. The proportion among these strategies is called the market fraction. According to the Market Fraction Hypothesis (MFH), this fraction swings overtime. This swinging has been observed in several theoretical and empirical models. However, a common assumption of these models is that the trading strategies are static and prespecified. Our paper therefore focuses on a more general version of this issue. After presenting the MFH and formalizing it with its main constituents, we propose a testing methodology, under the settings that the strategies are both non-static and not prespecified. Our technical analysis is composed of three parts, namely, genetic programming, self-organizing maps (SOM) and time-invariant self-organizing maps. The first two parts are straightforward given a large pile of well-established literature on each of them. The last part is novel and plays a critical role in this analysis, because it allows us to compare SOMs among different time periods. We run 3 different tests for 10 international markets and the experimental results show that some parts of the hypothesis is not well supported by the data.

Keyword: Market Fraction Hypothesis, Genetic Programming, Self-Organizing Feature Map, Time-Invariant Self-Organizing Feature Map
1 Introduction

A widely and well-known idea in economics and finance is the Efficient Market Hypothesis (EMH) (Fama, 1965). EMH states that market prices fully reflect all available information. A basic assumption that the EMH makes is that the market participants are rational. However, objections have been arisen by a number of economists (Gervais and Odean, 2001; Shefrin, 2000; Huberman and Regev, 2001; De Bondt and Thaler, 1985; Laibson, 1997), because they observed deviations from market rationality and therefore questioned this assumption.\(^1\) Recently, Lo (Lo, 2004, 2005) suggested an extension of the EMH, which he called the Adaptive Market Hypothesis (AMH). The AMH applies the principles of evolution (i.e. competition, adaptation, and natural selection) to financial interactions. According to Lo, the Adaptive Markets Hypothesis can be viewed as a new version of the EMH, derived from evolutionary principles. Individuals can be considered as organisms that learn and try to survive. The AMH, therefore, implies that the behaviour of market participants “is not necessarily intrinsic and exogenous but evolves by natural selection and depends on the particular environment through which selection occurs (Ibid, p. 30)”\(^2\).

Lo also states that “Prices reflect as much information as dictated by the combination of environmental conditions and the number and nature of ‘species’ in the economy. (Ibid, p. 31)” By ‘species’, Lo means groups of market participants that each one behaves in a common manner. Examples of these groups are pension funds, retail investors, market makers and hedge-fund managers. If many species are competing for rather scarce resources within a single market, then, according to Lo, that market is likely to be highly efficient (e.g. the market for 10-Year US Treasury Notes). However, if it is only a small number of species competing for rather abundant resources in a market, then that market will be less efficient (e.g. the market for oil paintings from the Italian Renaissance). Finally, Lo suggests that the degree of market efficiency is also related to factors such as the magnitude of profit opportunities available and the adaptability of the market participants. The last factor is what concerns us in this paper.

Another related and prominent idea regarding the adaptation and group dynamics is Brock and Hommes (Brock and Hommes, 1998) (for the rest of this paper it is going to be referred to as BH), where they developed a simple asset pricing model with heterogeneous beliefs. In this model, agents have beliefs about the future price of the risky asset and update these beliefs based on a fitness measure, which is the past realised profits. Agents are thus clustered according to their beliefs. BH, by using dynamical systems analysis (Guckehneimer and Holmes, 1983; Arrowsmith and Place, 1994; Eckmann and

---

\(^1\)A good summary of these deviations from market rationality can be found at Lo (2005).

\(^2\)The components of the AMH can be summarised in the following six points:

- Individuals act in their own self-interest
- Individuals make mistakes
- Individuals learn and adapt
- Competition drives adaptation and innovation
- Natural selection shapes market ecology
- Evolution determines market dynamics
Ruelle, 1985; Brock, 1986), provide analytical evidence and numerical demonstration for circumstances under which chaos exists and therefore the fractions of the clusters keep changing and even swinging. In their paper, the BH test for the existence of chaos and prove that it can exist under the two and four belief types.3

Furthermore, this swinging has also been observed and proved in many other agent-based financial models (Kirman, 1991, 1993; Lux, 1995, 1997, 1998). In addition, it has also been observed in empirical studies of agent-based models (Gilli and Winker, 2003; Winker and Gilli, 2001; Boswijk, Hommes, and Manzan, 2007; Amilon, 2008). For a more detailed description of agent-based financial models that have this swinging property, the reader is referred to Chen, Chang, and Du (2010).

Based on these observations, Chen (Chen, 2008; Chen, Chang, and Du, 2010) suggested a new hypothesis, called the Market Fraction Hypothesis (MFH). The MFH states that there is a constant swinging among the fractions of the types of trading strategies that exist in a market.4 In this approach, instead of using an asset pricing model as BH did in order to test for chaos, the MFH uses Lo’s idea of adaptive markets. More details about the MFH and its testing methodology will follow later on this paper, in Sections 2 and 5, respectively. The MFH has theoretically been demonstrated by BH, since they presented evidence under which chaos exists (and thus swinging among different trading strategy types).5

Furthermore, like what we mentioned earlier, the MFH has also been empirically observed. However, all of the above models assume that the trading strategy types are static and prespecified. By this we mean that these models endow their agents with a specific number of trading strategies (types) which they have to choose from. To the best of our knowledge, the MFH has not been empirically examined under a more dynamic environment that strategies are not static and are not exogenously given. Therefore, inspired by the idea of Lo (Adaptive Markets) and BH (Adaptive Belief Systems), we present a new MFH model which incorporates this more general setting and test it.

The second contribution of this paper is the attempt to formalize the MFH by presenting its main constituents and to suggest a testing methodology. We run tests for 10 international markets and hence provide a general examination of the plausibility of the MFH. One goal of our empirical study is to use the MFH as an benchmark and examine how well it describes the empirical results which we observe from various markets. In particular, we are interested in knowing how this benchmark performs when we tune the key parameter, i.e., the number of types in the market. Here we should also mention that we do not restrict our tests only to few pre-specified strategy types, as it happens in most current studies of the swinging fractions using the Adaptive Belief Systems or the like (see Chen, Chang, and Du, 2010). More details about tuning the number of trading strategies can be found at Section 6.

The rest of this paper is organized as follows: Section 2 elaborates on the MFH. Section 3 presents the basic tools used for our tests, namely Genetic Programming (GP) (Koza, 1992; Poli, Langdon, and McPhee, 2008) and Self-Organizing Feature Maps (SOM) (Koho-

---

3The belief types in BH are like strategy types: fundamentalists and different types of trend chasers (upward or downward trend chaser, contrarian).

4In the MFH, the “trading strategy type” is equivalent to BH’s “agent belief type”

5However, the term Market Fraction Hypothesis is not used in BH or any other relevant studies surveyed in Chen, Chang, and Du (2010).
nen, 1982). Section 4 presents the experimental designs. Section 5 addresses the methodology employed to test the MFH and explains the technical approaches needed to be taken for facilitating the test of the MFH. These proposed approaches play an important role for our experiments, since they allow comparison of trading strategies displayed in SOMs throughout different periods. Section 6 presents the test results. First it starts by presenting the results over a single run for a single dataset. Then it continues by presenting the summary results over 10 runs for this dataset and it finally presents summary results for all datasets. Section 7 concludes this paper and briefly discusses possible directions for further research.

2 The Market Fraction Hypothesis

As we have already said, within a market there exist different types of trading strategies. The MFH tells us that the fraction among these types of strategies keeps changing (swinging) overtime. The following three statements are the basic constituents of the MFH, and are based on the summary of the empirical development of the agent-based financial models, presented in Chen et al (Chen, Chang, and Du, 2010).

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

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

3. The size (fraction) of each type of trading strategies is positively correlated to its earning performance.

The first statement means that it is not possible for a single strategy to dominate the market by attracting an overwhelming fraction of market participants for many consecutive periods. In other words, according to the MFH there is no such thing as a ‘winner type’. Thus, an ex ante characterisation of winners simply does not exist. The term ‘dominance’ will become technical for testing the MFH, and we shall make it precise in Section 6.

The second statement means that if for instance the market has two trading strategies (like the traditional fundamentalists-chartists model), their fraction should keep on swinging. For instance, at time \( t \) Strategy A could occupy 20% of the market participants and Strategy B 80%, whereas later on at another time period, this proportion switches to 90% and 10%, respectively. So eventually in the long run both types of traders will have occupied about the same market share, i.e., about one half.\(^6\)

The third and final statement mentions that strategies that occupy many traders should also be the most profitable ones. This positive correlation is a result which is in spirit consistent with evolutionary dynamics.\(^7\) Traders, in order to survive, adapt to changes and

\(^6\)This idea is first made rigorous by Kirman (Kirman, 1993), who attempted to solve a puzzling entomological problem, i.e., ants swinging among themselves within two identical sources of food.

\(^7\)While mathematical representation of this evolutionary process is not unique, they share a common essence: survival of the fittest.
follow the strategies that give them higher profits. We therefore expect to see that whenever many traders are using the same strategy, this strategy will be returning a very high profit. 

Hence, what we shall do in this paper is to test the above three MFH properties against our empirical data. As we mentioned earlier, we are interested in qualitative results, meaning that we want to see how close the real market behaves as what are described by the MFH under different number of clusters. The tests take place in an evolutionary environment, with the use of GP. First of all, we create and evolve trading strategies for different periods. Afterwards, each strategy is clustered with SOM. Each cluster denotes a trading strategy type. Finally, we test the above 3 statements. The next section presents the basic tools we used for testing the MFH.

3 Tools

This section presents the two main tools used in order to test the MFH. These two tools are GP and SOM. There is also a third tool, which we created for the purposes of these tests, the Time-Invariant SOM. However, as this tool is derived from SOM, it will be presented later, in Section 5. Sections 3.1 and 3.2 present the GP and SOM, respectively.

3.1 Simple GP

As we said earlier at Section 1, the MFH follows the AMH’s idea of adaptive markets, where evolutionary principles prevail. Trading strategies are therefore non-static and not pre-specified (compared to BH’s approach where beliefs/strategies are specified at the beginning of their work and remain the same until the end), but evolve throughout time. In order to create and evolve these trading strategies we used a simple GP.

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

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

Furthermore, Figure 1 presents the Backus Normal Form (BNF) (Backus, 1959) (grammar) of the GP. As we can see, the root of the tree is an If-Then-Else statement. Then

---

8This positive relation has to been taken with some cautions. What may be oversimplified here is the complex dynamics. While profitable strategies can get popular, their dominance may soon reverse.

9We use these indicators because they have been proved to be quite useful in developing investment opportunity decision rules for forecasting rises and drops of the price in previous works like Allen and Karjalainen (1999), Austin et al. (2004) and Martinez-Jaramillo (2007). Of course, there is no reason why not use other information like fundamentals or limit order book information. However, the aim of this work is not to find the ultimate indicators for financial forecasting.
Table 1: Technical Indicators used by the GP. Each indicator uses 2 different periods, 12 and 50, in order to take into account a short-term and a long-term period. Formulas of our interpretation for these indicators are provided in the appendix.

<table>
<thead>
<tr>
<th>Technical Indicators (Abbreviation)</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moving Average (MA)</td>
<td>12 &amp; 50 days</td>
</tr>
<tr>
<td>Trade Break Out (TBR)</td>
<td>12 &amp; 50 days</td>
</tr>
<tr>
<td>Filter (FLR)</td>
<td>12 &amp; 50 days</td>
</tr>
<tr>
<td>Volatility (Vol)</td>
<td>12 &amp; 50 days</td>
</tr>
<tr>
<td>Momentum (Mom)</td>
<td>12 &amp; 50 days</td>
</tr>
<tr>
<td>Momentum Moving Average (MomMA)</td>
<td>12 &amp; 50 days</td>
</tr>
</tbody>
</table>

the first branch is a boolean (testing whether a technical indicator is greater than/less than/equal to a value). The ‘Then’ and ‘Else’ branches can be a new Genetic Decision Tree (GDT), or a decision, to buy or not-to-buy (denoted by 1 and 0).

\[
\text{<Tree>} ::= \text{If-then-else} \text{<Condition>} \text{<Tree>} \text{<Tree>} | \text{Decision}
\]

\[
\text{<Condition>} ::= \text{<Condition>} \text{"And"} \text{<Condition>} | \text{"Not"} \text{<Condition>} | \text{VarConstructor} \text{<RelationOperation>} \text{Threshold}
\]

\[
\text{<Variable>} ::= \text{MA}_{12} | \text{MA}_{50} | \text{TBR}_{12} | \text{TBR}_{50} | \text{FLR}_{12} | \text{FLR}_{50} | \text{Vol}_{12} | \text{Vol}_{50} | \text{Mom}_{12} | \text{Mom}_{50} | \text{MomMA}_{12} | \text{MomMA}_{50}
\]

\[
\text{<RelationOperation>} ::= \text{">"} | \text{"<"} | \text{"="}
\]

Decision is an integer, Positive or Negative implemented
Threshold is a real number
Figure 1: The Backus Normal Form of EDDIE

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

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

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

Rate of Correctness

\[
\text{RC} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}
\]
Li (Li, 2001) combined the above metrics and defined the following fitness function:

$$ff = w_1 \times RC - w_2 \times RMC - w_3 \times RF$$

(4)

where $w_1$, $w_2$ and $w_3$ are the weights for RC, RMC and RF respectively. Li states that these weights are given in order to reflect the preferences of investors. For instance, a conservative investor would want to avoid failure; thus a higher weight for RF should be used. However, Li also states that tuning these parameters does not seem to affect the performance of the GP. For our experiments we chose to include strategies that mainly focus on correctness and reduced failure. Thus these weights have been set to 0.6, 0.1 and 0.3 respectively.

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. This population should, therefore, interest us in spirit of Lo’s adaptive market process; therefore, we use them to infer what those competitive strategies may be in the period coinciding with the data period.

### 3.2 Self-Organizing Feature Maps (SOM)

In the previous section we mentioned that each GDT is a trading strategy. So, if for instance the population of GDTs is 500, this means we have potentially 500 trading strategies. However, this does not mean they are all completely different from each other. We know very well from the GP bibliography (Koza, 1992; Poli, Langdon, and McPhee, 2008) that towards the end of the training period, the population might have converged and thus some trees might be the same. Moreover, it is much easier for computing purposes to have types of strategies, rather than individual ones. In addition, BH also had types of strategies/beliefs, rather than having a pool with many strategies. For these reasons, the 500 trading strategies derived from the GP were classified in 9 clusters.\(^10\) In order to classify them, we used $3 \times 3$ Self-Organizing Feature Maps (SOM).

\(^{10}\)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.
SOM is a type of artificial neural networks that is trained using unsupervised learning, in order to return a low-dimensional representation of the input layer, which in our case is the recommendations of the GDTs. Associated with each cluster is a weight vector, which has the same dimensions as the input data. During this procedure the centroid of each cluster (hence the membership of each instance) is dynamically adjusted via a competitive learning process. Eventually, the whole population of GDT recommendations is assigned to different clusters and this is how we classify the trading strategies. Thus, the SOM output will be 9 neurons (or clusters) in a two-dimensional lattice, presenting the input data in an organised way, so that the similar strategies are clustered together.

Figures 2 present the results after running $3 \times 3$ SOM for a population of 500 individuals for the daily TAIEX data for first and second half of 1991, respectively. As we can see in these cases, there are usually 1-3 strategies that are occupying the majority of the population, whereas the rest of the strategies have significantly less members. We can also observe how the market fraction dynamics change from period to period.

4 Experimental Designs

This section summarises 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, DJIA, FTSE 100, HSI, NASDAQ, NIKEI 225, NYSE, S&P 500, STI and TAIEX. For each of these markets, we run each experiment for 10 times. To make it easier to the reader, we first present the testing methodology and results for a single run of the TAIEX dataset. Figure 3 presents the daily closing price of TAIEX. We then proceed with presenting summary results over the 10 runs for all datasets.

Each year was split into 2 halves (January-June, July-December), so in total, out of the

---


12The countries where these markets belong are presented in the appendix.
17 years, we have 34 periods. The GP was therefore implemented 34 times. Table 3 presents the GP parameters for our experiments. The behavior of each GDT can be represented by its series of market timing decisions over the entire trading horizon. Therefore, the behaviour of each rule is a binary string or a binary vector of 1s and 0s (buy and not-to-buy). The length or the dimensionality of these strings or vectors is then determined by the length of trading horizon, which in this study is 6 months, i.e., 250 days long; hence, the market timing vector has 250 dimensions. Once each trading rule is concretised into its market timing vector, we can then cluster these rules by applying SOM to the associated clusters.

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 it happens in the traditional GP approach. The reason for this is because we were not using the GP for forecasting purposes; instead, what we were interested in is to use 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) are not tested against another set. This approach is consistent with the AMH, as it states that the heuristics of an old environment are not necessarily suited to the new ones. Furthermore, our no-testing approach is also consistent with the well-tested overreaction hypothesis (De Bondt and Thaler, 1985, 1987), which essentially states that top-ranked portfolios were outperformed by bottom-ranked portfolios during the next period.

Finally, as we mentioned earlier, we used 3×3 SOM; therefore, we obtained a total
Table 3: GP Parameters. The GP parameters for our experiments are the ones used by Koza (Koza, 1992). Only the tournament size has been changed (lowered), and the reason for that was because we were observing premature convergence. Other than that, the results seem to be insensitive to these parameters.

<table>
<thead>
<tr>
<th>GP Parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Initial Depth</td>
<td>6</td>
</tr>
<tr>
<td>Max Depth</td>
<td>17</td>
</tr>
<tr>
<td>Generations</td>
<td>50</td>
</tr>
<tr>
<td>Population size</td>
<td>500</td>
</tr>
<tr>
<td>Tournament size</td>
<td>2</td>
</tr>
<tr>
<td>Reproduction probability</td>
<td>0.1</td>
</tr>
<tr>
<td>Crossover probability</td>
<td>0.9</td>
</tr>
<tr>
<td>Mutation probability</td>
<td>0.01</td>
</tr>
<tr>
<td>{w_1, w_2, w_3}</td>
<td>{0.6, 0.1, 0.3}</td>
</tr>
</tbody>
</table>

34 SOMs (one per period), with 9 clusters each. In other words, in every period the 500 GDTs were placed in one of the nine clusters (i.e. the categories of the trading strategies) of that SOM. From this point on, whenever we use the term ‘trading strategy type’ we will be referring to one of the nine clusters and each GDT will be a member of one of these nine clusters.

5 Testing Methodology

After having presented the necessary tools and the experimental designs, we can now proceed in presenting the testing methodology. Our methodology consists of three parts: GP, SOM and time-invariant SOM. However, there is a minor difference in the methodology among our proposed 3 Tests. Tests 1 and 2 need to compare clusters throughout over different periods, whereas Test 3 is only applied with each single period. There is thus a slight different approach in the designing of the testing methodology and it is presented in the following sections.

5.1 Testing Methodology for Tests 1 and 2

As we have already seen, we used a simple GP in order to generate and evolve trading strategies. However, there is a problem with comparing trading strategies among different periods. This happens because we cannot compare the fitness function of a trading strategy (GDT) from one period with the fitness function of a strategy (GDT) of another period, since they were presented with different datasets (environments).

In order to better understand this, consider Figures 2. The way SOM works is that it creates the clusters after it is given a specific population of, in our cases, GDTs. When we have different periods with different populations, the nine clusters from different periods will generally be different, because they represent different populations of investment behaviors generated by different data environments. For example name the bottom-left...
cluster of each SOM (Figures 2) as ‘Cluster 1’. Then we are saying that ‘Cluster 1’ of the SOM derived using the data of 1991a in general will not be the same as ‘Cluster 1’ of the SOM derived from the data using 1991b. Quite likely, they have different centroids (weighting vectors), representing different investment behaviors. This, therefore, makes the strategy types incomparable crossing different periods.

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

5.1.1 Translations

Emigrating As we just mentioned, after obtaining the trading strategies from GP, we cannot directly compare them among different periods, because the dataset of each period is different. What, therefore, needs to be done is to apply the same dataset as a base to all periods. In other words, all GDTs that derived from each period need to be applied to the dataset of the base period. For convenience, call these emigrant GDTs. Therefore, after applying these emigrant GDTs to the base period, new signals are created. In this way, the behavioral vectors of all GDTs originally derived from different periods are rebuilt based on the same ground and hence become comparable. In this paper, we choose the second half of 2007 (2007b) as the base period.15

Reclustering Reclustering or time-invariant SOM is the second part of translations, which allows the SOM clusters to be compared throughout different periods. We again use 2007b as the common base period. This time, we keep the centroids of the clusters originally derived from the common base period fixed and assign the behavior vectors from other periods (derived through emigrated GDTs) to one of the fixed centroids. This reclustering is conducted in the following way: the behavior vector of each emigrated GDT is compared to each centroid of the nine clusters and it will then be assigned to the one with minimum Euclidean distance. We do this period by period from 1991a to 2007a. 33 SOMs are constructed in this way16, and now these SOMs can be directly compared to each other given that they all sharing the same centroids. Figure 4 presents 4 out of these 34 SOMs. We will then use them to test Hypothesis 1 and 2.

5.2 Testing Methodology for Test 3

Test 3 tests the correlation between the size of the clusters with their earning performance within the same period. Since the test takes place only in each period individually, there is no need for inter-period data processing. After creating and evolving the trading strategies with the GP, we directly apply SOM to the strategies. Then, after having created the clusters with SOM, we can run Test 3.

15 The base period was chosen arbitrarily; however, we found that the results are insensitive to the choice of the base period.
16 2007b does not need reclustering, since we use it as the base period.
6 Results

6.1 Results of a single run of a single dataset

6.1.1 Test 1: The Short-Run Test

The first test regards the short-run behavior of market fractions. In the short run, the fraction of different clusters of strategies is expected to keep swinging over time, which implies a short dominance duration for any cluster. To be operational, a type of strategies is said to be dominant if its fraction is greater than the threshold,

\[ DS = \frac{1 + p}{N + p}, \]

where \( DS \) denotes a threshold, \( N \) the number of clusters and \( p \) is a free parameter to manipulate the degree of dominance. For example, as \( N = 9 \), the threshold of being a dominate type changes with \( p \) as follows. It is 11.11% when \( p=0 \), 20% when \( p=1 \), and 27.27% when \( p=3 \). Clearly, the higher the \( p \), the higher the threshold. If all clusters were having the same number of members, then each cluster would be occupying 11% (1/9) of the population. Hence, the case that \( p = 0 \) corresponds to a threshold that just breaks the tie. However, to be dominant, we may expect a value of \( p \) to be higher than just breaking the tie. Hence, in this paper, \( p \) is set to be 2.

Furthermore, we need to be precise on what we mean by short duration for a dominant type. Here, any specific number may be arbitrary; after all, short is only a matter of degree. We, therefore, first present the statistics of duration observed for each type. Figure 5 summarizes the dominance results over the 34 periods. It presents the minimum, average and maximum of the duration times of each type. What we can observe from Figure 5 is that the longest duration observed is nine periods (four and half year) for type 6. For other types, the longest duration is hardly over two periods. Hence, if we look at the average duration, with the exception of type 6, no type remains dominant for more than 2 consecutive periods, i.e., a year.
Figure 5: Min, average and max number of consecutive times that a strategy remains dominant over the 34 periods for p=2 (Daily Closing Price for TAIEX:1991-2007)

6.1.2 Test 2: The Long-Run Test

The second hypothesis regards the long-run behavior of market fractions. It says that, in the long run, different clusters are equally attractive and thus their market fractions are equal. As we said earlier, we expect to see that the fraction of strategies keeps changing. In Figure 2, for instance, we can see that one strategy is occupying a quite big fraction of the population (around 70%-360 members out of a total of 500). The rest of the strategies have lower percentages. According to the MFH, these percentages should keep changing from period to period so that, in the long run, these percentages should be close to each other. In order to test this, we sum up the cardinality of each cluster over the 34 periods. Figure 6 present the fractions of the nine clusters for the period 1991-2007.

Of course, it is obvious that this distribution is very different from the uniform one. In order to give a measure on how far it is away from the uniform, we use the familiar entropy as a metric. Let us denote the empirical distribution presented in Figure 6 as $f_X$, and the uniform distribution $f_Y$. By definition, $f_Y = \frac{1}{N}$, where $N$ is the number of clusters, which in this case is 9. In order to measure how close $f_X$ is to the uniform distribution $f_Y$, we calculate the entropy of both distributions. For the discrete random variable, entropy is defined as

$$H = - \sum_{i=1}^{N} p_i \ln p_i,$$

where $p_i$ is the fraction of each cluster. It is well known that for the uniform distribution $H(Y) = \ln N$. When $N=9$, it is $\ln 9 \approx 2.2$. The closer $H(X)$ is to 2.2, the closer $X$ is to the uniform distribution. After calculating $X$’s entropy, we find it equal to 1.3, which is only
59% of the entropy of the uniform distribution.

6.1.3 Test 3-Temporal Correlation between Size and Returns

The size (fraction) of each type of trading strategies is expected to be positively correlated to its earning performance.

So far, the above tests were comparing SOMs between different years. Nevertheless, the MFH also talks about the trading strategies within a period. This third test is to test the correlation between profit and the number of members of each trading strategy. To test that, we rank the 9 clusters of one period in terms of their size/fraction. Denote it by $v_f$. Moreover, we find out the maximum profit per cluster, and again, we rank them and denote it by $v_{\pi}$. Therefore, we end up with 2 $1 \times 9$ vectors, vectors $v_f$ and $v_{\pi}$. In order to test the third hypothesis, we run the Spearman's rank correlation test at 5% significance level. Out of the 34 periods, in only 13 of them was the profit significantly positive correlated with the size.

We also run this test one more time, but instead of using the maximum profit per cluster, we use the mode. Again, after running the Spearman correlation test at 5% significance level we get 3 out of 34 periods that the profit is positive correlated with the size. It is obvious that in both of these cases the number of periods with significant positive correlation is very low.

Now that we have seen the test results of a single run for one dataset, it is interesting to see if these results generalize for more runs and more datasets. The next part of this section presents and discusses these summary results.
### Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>Test 1 Average</th>
<th>Max</th>
<th>Test 2 Entropy Ratio</th>
<th>Max</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>1.81</td>
<td>5.5</td>
<td>0.68</td>
<td>14.1</td>
<td>1.3</td>
</tr>
<tr>
<td>DJIA</td>
<td>1.93</td>
<td>5.78</td>
<td>0.66</td>
<td>15.22</td>
<td>1.22</td>
</tr>
<tr>
<td>FTSE 100</td>
<td>1.77</td>
<td>6</td>
<td>0.64</td>
<td>16.38</td>
<td>1.62</td>
</tr>
<tr>
<td>HSI</td>
<td>1.71</td>
<td>4.6</td>
<td>0.7</td>
<td>15.4</td>
<td>1.3</td>
</tr>
<tr>
<td>NASDAQ</td>
<td>1.59</td>
<td>4.11</td>
<td>0.69</td>
<td>14.67</td>
<td>1.78</td>
</tr>
<tr>
<td>NIKEI 225</td>
<td>1.51</td>
<td>3.4</td>
<td>0.79</td>
<td>16.9</td>
<td>1.8</td>
</tr>
<tr>
<td>NYSE</td>
<td>1.93</td>
<td>6.56</td>
<td>0.6</td>
<td>14.4</td>
<td>1.89</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>2.16</td>
<td>6.89</td>
<td>0.64</td>
<td>16.22</td>
<td>1.89</td>
</tr>
<tr>
<td>STI</td>
<td>1.67</td>
<td>3.7</td>
<td>0.75</td>
<td>16.4</td>
<td>1.3</td>
</tr>
<tr>
<td>TAIEX</td>
<td>2.02</td>
<td>8.25</td>
<td>0.55</td>
<td>14.75</td>
<td>1.75</td>
</tr>
</tbody>
</table>

The first two numeric columns are related to Test 1 and present the averages over the 10 runs for the average and maximum dominance duration of the 9 clusters, respectively. The next column presents the ratio of the average realized entropy (over the 10 runs) over the base entropy under the null of the uniform distribution. This ratio gets maximized when it is one. Finally, the last two columns show the averages of the number of periods having significant positive Spearman’s rank correlation between size and profits, where profits are measured by the statistic, maximum or mode.

### 6.2 Summary results for all datasets under 9 clusters (3×3 SOM)

As we saw in the previous section, the experimental results of the three tests seem that they all deviate from what the MFH predict to some extent. Test 1 has one cluster that dominates the market for 9 consecutive periods, kind of too long. Test 2 shows even a larger deviation since the long-term market fraction is very much different from the uniform distribution. Finally, Test 3 shows that the connection between the size of the clusters and the respective earning performance is weak, regardless that the latter is measured by the maximum or the mode. Altogether, the evidence for the MFH is weak. However, so far we have only presented a single run for a single dataset. Table 4 presents the results over 10 runs for all datasets tested. The first two numeric columns are related to Test 1. They present the averages over the 10 runs for the average and maximum dominance duration of the 9 clusters. Furthermore, the next column is related to Test 2 and shows the ratio of the average realized entropy (over the 10 runs) over the base entropy (equal to 2.2). Lastly, the last two columns are related to Test 3; they present the averages of the number of periods having significant positive Spearman’s rank correlation between size and profit, over the 10 runs.

The first observation we can draw from Table 4 is that homogeneity exists across the majority of the results. Let us first start with Test 1. We can see that on average there is no cluster that remains dominant for 2 consecutive periods. This is in line with Test 1. However, the second column tells us that even though on average no cluster dominates for more than 1 period, there is always an outlier that can remain dominant for longer, e.g. 8 consecutive periods for TAIEX. Regarding Test 2, the entropy ratios for all datasets
are somewhat distant from their maximum value. All entropy ratios are in the range 0.55-0.75, which basically is a 25-45% difference from the entropy of the uniform. This essentially means that the distributions are on average different from the uniform and therefore the clusters, in the long run, are not equally attractive, as Test 2 requires. Test 3 also seems to be consistent among the different markets. Periods with positive correlation between size and performance are low for the both proxies of earning performance. As a total, the MFH seems to be relatively weak for the case of 9 clusters.

6.3 Changing the number of clusters

So far, all of our tests have been done for $3 \times 3$ SOMs. It is, however, interesting to investigate how sensitive the results are if we tune the number of clusters. Therefore, we repeat the whole procedure mentioned above for different SOM dimensions: $2 \times 1$, $3 \times 1$, $2 \times 2$, $5 \times 1$, $3 \times 2$, $7 \times 1$ and $4 \times 2$, i.e., from 2 clusters to 8 clusters.

6.3.1 Test 1

Figures 7 present the averages, over 10 runs, of the average (the upper panel) and maximum dominance duration (the lower panel) under different SOMs having numbers of clusters from 2 to 9. What we observe in these figures is that the dominance duration decreases as the number of clusters increases. To see how significant or how interestingly this pattern is, we run a Monte Carlo simulation as follows. Starting with two clusters, we randomly assign a winner (dominant cluster) to either cluster 1 and cluster 2. We then conduct this binomial experiment 34 times. Considering this as one run, we do it for ten runs. Hence, we have 10 artificial series of dominant clusters, each series lasting for 34 runs. We then conduct the same analysis as above by figuring out the average duration and maximum duration of each series, and the average of the whole. We then apply this Monte Carlo experiment with additional number of clusters, from three to nine incrementally. A comparable result is then drawn in Figure 8.

By comparing Figure 7 with Figure 8, we can see that the behavior of the real markets is very different from that of the multinomial experiment. For the latter, the average of maximum duration decays, from above 6 to below 3, but for the former this decaying tendency is shown in none of the ten markets. Instead, they all fluctuate slightly around a horizontal line, and, depending on the market, the line is situated at a interval from four to eight. For the average duration, while both figures features a decaying tendency, the one with financial data decays much slower than the one of the artificial data. Therefore, our result cannot be treated as an incident from a random draw of the multinomial experiments, and in this sense this pattern is not spurious.

The difference in the two panels in Figure 7 can be better explained if we also take into account Figure 5. What seems to happen for all datasets is that there are always a few clusters that have strong (long) dominance over the 34 periods, whereas the rest have very low. The low average dominance duration we see in Figure 7 for the high number of clusters can be therefore explained by the extremely low dominance duration of the majority of clusters.
Figure 7: Summary Results for Test 1-Averages and Maximum, Financial Data
6.3.2 Test 2

As we said earlier, we are interested in obtaining the distance of the entropy of the empirical distribution $f_X$ (fractions of clusters) from the uniform distribution (benchmark). We also said that the closer the entropy of distribution $f_X$ is to the entropy of the uniform, the closer distribution $f_X$ is to the uniform one. After obtaining the entropies over 10 runs for each dataset, we first calculated the average of these runs. We then divided each one of these averages with the benchmark entropy and thus obtained 10 different ratios (one per dataset). Of course, this ratio is maximized when the two entropies are equal, and therefore their ratio is equal to 1. Hence, the higher the ratio, the closer to the uniform the empirical distribution will be. Figure 9 presents these ratios for all datasets.

What we observe from this figure is that the ratios tend to decrease as the number of clusters increases (thus, the difference of the two distributions increases). Such divergence of the two distributions indicates again that the strong dominance for a few clusters continues existing, even in the long run. Therefore, after combining Tests 1 and 2, we can have a quite clear picture. Clusters tend to dominate for long periods and this dominance is usually interchanged among few clusters.

To make this argument even clearer, we also present in Figure 10 the cumulative fractions for TAIEX for different number of clusters. A graph named “Number of Clusters: 2” means that the strategies are divided between 2 clusters. Likewise “Number of Clusters: 3” means that the strategies are divided among 3 clusters, etc. The clusters have been sorted by their size (fraction) in descending order. Therefore, Cluster 1 in the X-axis indicates the cluster with the highest fraction, Cluster 2 the cluster with the second higher fraction, etc. The Y-axis presents the cumulative fraction of the clusters.

From this figure, we can make two observations. First of all, the maximum fraction
Figure 9: Test 2: Difference of the empirical distribution \( x \) (fractions of clusters) from the uniform distribution.

The size of the largest cluster is always above 60%. The second largest cluster has a fraction in the range of 20-30%. Generally, we can see that by the time we calculate the size of the top 4 largest clusters, we have reached around 90%. This therefore justifies our previous argument that few clusters tend to dominate the market in the long run. The rest have a very low size. This finding can be related to Aoki (2002), who shows that 95% of the market participants of his experiments would belong to one of the two largest clusters of agents.

Table 5 gives the result of the minimum number of clusters required to cover a targeted fraction of market participants. The three targeted values given in the table are 90%, 95%, and 99%. Since the purpose is to see whether only a small number of clusters is required, we started with a larger number clusters, namely, nine, and see how much reduction we can make. If the target is to cover 90% of the market participants, then most markets need four to five types, and if target is even higher to 95%, then most markets need five to six types. Remember 95% is the parameter used in Aoki (2002). Hence, our finding about the minimum number required is also larger than his suggested two or three. Two or three types can be sufficient if we have a target somewhat lower than 90%.

The second observation is that the contribution of each ranked cluster decreases when the number of clusters increases, which causes the entire cumulative curve to shift down. For instance, when the number of clusters is 2, the largest cluster has a size of about 70%. However, while we move to a higher number of clusters, we can see that this size gradually decreases and finally reaches below 60%, when the number of clusters is 9. The same happens for the contribution of the rest of the clusters. Hence, each graph moves a
Table 5: Minimum number of clusters whose cumulative fraction is above the required threshold of 90%, 95% and 99%, respectively

<table>
<thead>
<tr>
<th></th>
<th>Threshold 90%</th>
<th>Threshold 95%</th>
<th>Threshold 99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>4</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>DJIA</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>FTSE 100</td>
<td>4</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>HSI</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>NASDAQ</td>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>NIKEI 225</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>NYSE</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>STI</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>TAIEX</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

bit below, when the number of clusters increases.

Finally, these two observations apply to other markets. Their cumulative curves are presented in the appendix (Figures 13). However, NYSE and S&P 500 are the only exceptions: these two markets do not have gigantic clusters. However, their distribution is still not very close to the uniform, as we saw from Figure 9.

6.3.3 Test 3

After observing figures 11 and 12 we can reach to two very interesting conclusions. First of all, evidence for the MFH gets weaker as the number of clusters increases. This is what we also observed at Test 2. This is very interesting. Test 1 is closer to the hypothesis for higher number of clusters, whereas Tests 2 and 3 are closer to the MFH for lower ones. The second observation from these two figures is the steep decrease of the average max and mode when we move from 2 to 3 clusters. Why this happens is at the moment unclear and certainly deserves further investigation.

7 Conclusion

To summarise, this paper presents the MFH and suggests a testing methodology for it. This hypothesis says that the financial time series observed can be described in terms of switching between different types of agents, each with either different beliefs or different trading rules. This switching has been observed in both theoretical and empirical models (see Chen, Chang, and Du, 2010). However, all these models assume that the financial agents can only choose between a finite and prespecified number of trading strategies. We addressed this issue by creating and evolving trading strategies with GP. We then clustered them with SOM and time-invariant SOM, which is a novel tool that allows comparison of SOMs among different periods. We finally tested the MFH with empirical
Figure 10: Cumulative Fraction for TAIEX

Figure 11: Summary Results for Test 3-Maximum
data from 10 international markets. Results show that the hypothesis cannot be always supported and can be affected by the number of clusters. In fact, the MFH seems to be weak for the majority of our datasets. Furthermore, our empirical results suggest that when we have many different types of strategies in the market (e.g. 9 clusters), only a few of them occupy the majority of the agents. These two are very interesting results that yet need more examination. The next step of our research will focus on this.

In addition, the results seem to be sensitive to the number of clusters. Depending on the test, the MFH is stronger or weaker when we have lower or higher number of clusters. This is another issue that requires further investigation.

Future research could also include some changes in our model. Can the results be affected by the periods’ window? A this work, the 17 years dataset was divided in semesters and thus each window was 6 months. It is interesting to see whether a shorter or longer window can affect our results. In addition, at Section 3 we mentioned that complex dynamics have not been taken into account for the design of Test 3. The dominance of popular strategies could reverse very soon after they become popular. We believe that this is another aspect of our work that needs further research.

Finally, we should mention that these results might be sensitive to the nature of the GP algorithm. How much the different algorithms can affect these results is at the moment unclear. Future research could therefore also focus on testing the influence of different GP algorithms on our results.
Appendix A  Appendix

Appendix A.1  Technical Indicators

The following section presents the technical indicators that the GP is using, along with their formulas. We performed a sort of standardization in order to avoid to have a very big range of numbers generated by GP, because this would increase the size of the search space even more.

Moving Average Indicator

\[ MA(L,t) = \frac{P(t) - \frac{1}{L} \sum_{i=1}^{L} P(t - i)}{\frac{1}{L} \sum_{i=1}^{L} P(t - i)} \]

Trade Break Out Indicator

\[ TBR(L,t) = \frac{P(t) - \max\{P(t-1), \ldots, P(t-L)\}}{\max\{P(t-1), \ldots, P(t-L)\}} \]

Filter Indicator

\[ FLR(L,t) = \frac{P(t) - \min\{P(t-1), \ldots, P(t-L)\}}{\min\{P(t-1), \ldots, P(t-L)\}} \]

Volatility Indicator

\[ Vol(L,t) = \frac{\sigma(P(t), \ldots, P(t-L+1))}{\frac{1}{L} \sum_{i=1}^{L} P(t - i)} \]

Momentum Indicator

\[ Mom(L,t) = P(t) - P(t - L) \]

Momentum Moving Average Indicator

\[ MomMA(L,t) = \frac{1}{L} \sum_{i=1}^{L} Mom(L, t - i) \]

Appendix A.2  International Markets

The following table presents the international markets used for our tests, along with their abbreviations.
Table 6: International Markets

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>France</td>
</tr>
<tr>
<td>DJIA</td>
<td>USA</td>
</tr>
<tr>
<td>FTSE 100</td>
<td>UK</td>
</tr>
<tr>
<td>HSI</td>
<td>Hong Kong</td>
</tr>
<tr>
<td>NASDAQ</td>
<td>USA</td>
</tr>
<tr>
<td>NIKEI 225</td>
<td>JAPAN</td>
</tr>
<tr>
<td>NYSE</td>
<td>USA</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>USA</td>
</tr>
<tr>
<td>STI</td>
<td>Singapore</td>
</tr>
<tr>
<td>TAIEX</td>
<td>Taiwan</td>
</tr>
</tbody>
</table>

Appendix A.3 Figures of Cumulative Fractions

In this section we present the figures of cumulative fractions for all ten datasets.

References


Figure 13: Cumulative Fraction for CAC 40