Genetic Programming and the Efficient Market Hypothesis

Shu-Heng Chen
AI-ECON Research Group
Department of Economics
National Chengchi University
Taipei, Taiwan 11623
E-mail: chchen@cc.nccu.edu.tw

Chia-Hsuan Yeh
AI-ECON Research Group
Department of Economics
National Chengchi University
Taipei, Taiwan 11623
E-mail: g3258501@grad.cc.nccu.edu.tw

Abstract

While search plays an important role in the efficient market hypothesis (EMH), the traditional formalization of the EMH, based on probabilistic independence, fails to capture it. Due to this failure, recent findings of nonlinear tests misled us into concluding that the EMH is rejected. Even though most economists are reluctant to make this conclusion, the traditional formalization leaves us no other choice. This paper reformalizes the EMH with a biologically-based search program, i.e., genetic programming (GP). The GP-based search enables us to model search in the EMH explicitly. Through this, search cost as well as search intensity can be measured objectively, and the notion of predictability and profitability can then be formalized. The GP-based notion of the EMH will be exemplified by testing the EMH with a small, medium and large sample of the S&P 500 stock index.

1 Introduction

> Thus, investors cannot device an investment strategy to yield abnormal profits on the basis of an analysis of past price patterns. (Malkiel, 1987, p.127)

In his survey article, Malkiel gave the intuitive notion of the *efficient market hypothesis* cited above. Based on this intuitive notion, to formalize the EMH, two technical notions are needed, namely, a technical notion of *predictability* and a technical notion of *profitability*. However, despite its long history, the formalization of the EMH has been built solely on a technical notion of *predictability* and it proceeded as if an appropriate technical notion of unpredictability would automatically defy profitability in any sense. In the history of the EMH, there is only one technical notion of predictability been formalized and it was constructed in terms of *probabilistic independence* in probability theory. Mathematically speaking, considering the rate of returns R_t a random variable (function) defined in the L^2 probabilistic Hilbert space and Ω_{t-1} the σ-algebra generated by the history of rate of returns $\{R_j\}_{j=-\infty}^{t-1}$, then the EMH simply says that R_t is independent of any random variables in the Ω_{t-1}. Based on the different information sets Ω_{t-1}, one could have different versions of the EMH (Malkiel, 1987). The one introduced above is called *weak-form efficiency*. Furthermore, considering the expected rate of returns the conditional expectation $E(R_t \mid \Omega_{t-1})$, then the EMH also implies that

$$E(R_t \mid \Omega_{t-1}) = 0. \tag{1}$$

Since Equation (1) is also a consequence of the *random walk* defined in a discrete-time stochastic process, the EMH is often used interchangeably with the random walk hypothesis in finance, even though the former is far from equivalent to the latter.

While this formalization is precise, Chen and Yeh (1995) argued that *it is not computable* in the sense that we can not test it with an effective algorithm. Nevertheless, this problem had not been fully recognized until recently when a series of *nonlinear tests* failed this hypothesis, e.g., Brock, Dechert and Scheinkman (1987), Savit (1988, 1989), Hinich and Patterson (1989), Hsieh (1989), Frank, Gencay and Stengos (1988), Scheinkman and LeBaron (1989), Peters (1991), and Willey (1992). These tests either tell us that there exists nonlinear dependence between R_t and Ω_{t-1} or that $\{R_i\}_{i=-\infty}^{t}$ is a chaotic time series which looks random but is in fact deterministic.

What do these tests imply? If we take Equation (1) as the formalization of the EMH, then the rejection of the EMH naturally follows from the statistical rejection of Equation (1). Unfortunately, this is a certainly a very negative result for the mainstream economists and an unacceptable one too. However, there may be hope of solving this conundrum if we can interpret these results in a different way. For example, Fogler (1995) had the following remarks:

> ...whether a chaotic process or a complex nonlinear process generates a pattern is irrelevant unless one knows the exact functional form generating the process. Even if tests indicate a high embedding dimension, forecasting is impossible unless a specific function form is assumed. (p.16)

Accordingly, the rejection of Equation (1) based on any nonlinearity test does not authorize us to say that the market is not efficient *unless we are able to show the underlying data generating process.* Fogler's remarks powerfully separated the nonlinearity tests from the efficient market hypothesis. But, that also means Equation (1) has to be given up and a new technical notion of unpredictability is needed. To be consistent with Fogler's remarks, the new notion must be *search oriented.* More precisely, to formalize Fogler's notion of unpredictability, a *search algorithm* must be included so that the decision of predictability can be made based on search results. Moreover, since we are searching for the potential regularity in L^2 space and the cardinality of this space is *infinite*, the design of this search algorithm cannot be limited to only a finite number of functions or any special subclass of functions. However, the theoretical technology to design the search algorithm of this sort was not in the classical toolbox of economists. That also partially explains why the search-based notion of unpredictability was not formalized over the last 90 years, while it was already captured by economists' intuition. e.g., Malkiel (1987).

Given the introduction above, the contribution which genetic programming can make to the EMH is clear. Using the novel idea of incorporating the evolutionary operation into the automatic search over the space of programs (L^2 space), genetic programming provides financial economists with a theoretical technology to design a search algorithm and to formalize the notion of unpredictability. This formalization also enables us to define *search intensity* and to give us an objective measurement of *search costs* and *the chance of success in the search.* Based on that, *the expected profits of search* can be derived and the other aspect of the EMH, i.e., the profitability aspect, can also be formalized.

This contribution also distinguishes the application of genetic programming from other applications of biologically inspired computational models to finance. For example, artificial neural nets are mainly used in finance as a problem-solving technique for the nonlinear estimation and forecasting of the asset price or for the computation of option pricing. Regardless, ANNs per se provide little to change or advance our understanding of the fundamental concepts in finance. In contrast, genetic programming not only enriches the finance toolbox, but also enhances our understanding of finance theory from its very root.

Chen and Yeh (1995) conducted a pioneer study to demonstrate the GP-based notion of unpredictability. This formalization of unpredictability was illustrated with examples of predicting chaotic dynamic systems. They applied GP to predict the time series generated by three chaotic dynamic systems which are represented by the *LISP S-Expression* and depicted as a rooted, pointed-labeled tree (GP-tree). They have shown that if the rates of return are generated by a *simple* deterministic chaotic dynamic system, then GP may actually discover it. Here, "*simple*" refers to the *depth* of the GP-tree and has nothing to do with the embedding dimension. Their study illustrated how genetic programming can provide us with an explicit search program upon which an objective measure for predictability can be constructed. Also, it indicated that the use of *Lyapunov exponents* to estimate the predictability of a chaotic time series might not be appropriate.

In the same study, the EMH based on the GP-based notion of unpredictability was exemplified by an application to a mini-size sample of S&P 500 index (sample size = 50). It was found that, while linear models cannot predict better than the random walk, a GP-based search can beat random walk by 50%. However, Chen and Yeh (1995) did not explicitly show how the GP-based notion of unpredictability can be related to the other aspect of the EMH, i.e., *the notion of profitability.* While they could show to Fogler how to find the *underlying data generating process* given the existence of *nonlinear dependence*, they certainly have not answered the essential question *whether predictability implies profitability.* Therefore, their work, at best, only formalized the first half of Malkiel's intuitive notion of the EMH.

This paper will extend Chen and Yeh (1995) to the second half. To do this, we exemplify the notion of profitability by comparing long-range forecasting with large samples and short-range forecasting with small samples. Since different forecasting strategies involve different search cost, i.e., direct *computational cost* and indirect *risk cost*, the GP-based notion of profitability enables us to check whether the better prediction by a certain search strategy was made at high search cost. In other words, we can test the *no-free-lunch property* in Malkiel's intuition of the EMH. This work will be done in Sections 2 and 3. Section 2 will describe the choice of data. Section 3 will present the results of the GP-based search over different sizes of sample along with the analysis.

Section 4, which is motivated by Koza (1992) (pp.245-255), provides an *knowledge-discovery perspective* to study the GP-based formalization of the EMH. In this section, the EMH represented by Equation (1) is regarded as a *benchmark*. We then evaluate the competitiveness of this benchmark by observing how well it can fight against its potential competitors. If, during the evolution, it is beaten by a competitor, then this benchmark is considered not competitive; otherwise, we say it is competitive. It is this sense that we will show that *the EMH is not competitive in the small sample but is quite competitive in the large sample.* Concluding remarks are given in Section 5.

2 The Design of the Data Environment

2.1 The Sample Size

The data environment in which the GP-based search is implemented concerns the daily rate of return of the S&P 500 Index, from 1/2/62 to 9/6/95. There are totally 8,478 observations in this dataset. This dataset is rich enough for us to test the EMH *in the long run* as well as *in the short run.* Distinguishing between short-run tests and long-run tests is motivated by the recent studies of the time series of stock prices which seem to indicate that *even though nonlinear regularities might exist, they are not stable over time.* For example, Peters (1991) made an excellent note on this.

> For too long, we seem to have been divided between "technicians" who believe that the market follows a regular cycle and "quants" who believe there is no cycle at all. The truth lies somewhere in between. There are cycles, but they are not regular, hence may be invisible to standard statistical techniques. (p.62)

LeBaron (1992) also came up with a similar observation.

> Some out-of-sample forecast improvements are demonstrated for the weekly S&P 500 series. The evidence shows that these improvements are difficult to detect since they only occur during certain time periods. ... Although forecast improvements appear significant, they are extremely small and occur only for a fraction of the weeks tested. Stock returns remain, as they should, a relatively difficult series to forecast. (p.381)

Therefore, the application of GP to different sizes of sample does require lots of thought. Suppose that the stock market encounters a sequence of short-term time-variant nonlinear relations, a large sample size may average out all these relations. In this case, a smaller sample size is desirable, whereas for the time-invariant long-term nonlinear relation, a large sample size is needed. In an earlier study, Chen and Yeh (1995) conducted large-scale experiments for a sample with sample-size 50 (about ten weeks). In this paper, we extend the study by testing the EMH with a short-term (sample size=200), medium-term (1000) and long-term (2000) sample of S&P 500. The short-term sample corresponds to data about 10 months long, the medium-term about 4 years and the long-term about 8 years. It will be interesting to see how the validity of the EMH will be affected by the sample size.

2.2 Selecting Data with the MDL

In a data set with 8478 observations, there are many subsets associated with different sample sizes. To test all of them seems to be time-consuming. The alternative is to test a few of them which are representative. If these representative datasets can reject the GP-based test of the EMH, so can the rest. For this purpose, the subsets considered are the ones with the highest complexity. Since there is no unified definition or criterion for *complexity* (Horgan, 1995), our adherence to any particular style is, more or less, arbitrary. Of course, this weakness can be consolidated by simultaneously using different styles of complexity measure. However, in this paper, we only consider the most popular style used in statistics and econometrics, i.e., *Rissanen's minimum description length* (**MDL**).

Rissanen's MDL (minimum description length) is an approximation for *Kolmogorov complexity* which measures the complexity of a set of data by the length of the shortest universal Turing machine program that will generate the data. The measure is well-defined, but not practically computable. The MDL developed by Rissanen (1982) is a way to approximate this uncomputable measure by replacing the universal Turing machine with a class of probabilistic models.

A detailed description of the MDL algorithm used in this paper can be found in Chen and Tan (1996).

Table 1: MDL-Based Data Selection: S&P 500

Sample size	Whole Sample Period (Post-Sample Period)	MDL
200	1/3/92 - 10/16/92 (9/21/92 -)	142.472
1000	3/1/82 - 2/11/86 (9/19/85 -)	697.794
2000	2/5/81 - 1/4/89 (3/22/88 -)	1387.579

Briefly speaking, we first transform the original sequence of $\{R_t\}$ from 1/2/62 to 9/6/95 into a 0-and-1 sequence based on the sign of R_t. Then the MDL is computed for each of the 200, 1000 and 2000 consecutive observations in the 0-and-1 sequence by choosing the *Bernoulli class* and *Markov class* as our model classes. The MDL(x_i^{2000}) of S&P 500 is given in Figure 1. From Figure 1, we can see that, in terms of the 2000-day MDL, the second half of the time series is more complex than the first half. It is consistent with what we have been told in finance that complexity has a tendency to increase with time. Based on the maximum MDL criterion, the periods for different sizes of chosen samples are given in Table 1. The S&P 500 time series of the chosen period for the sample with sample size 1000 and 2000 are drawn in Figures 2 and 3.

Once the in-sample data series are given, the next issue is how to determine the length of the post-sample series. In this paper, we arbitrarily set the in-sample data to be ten times the size of the post-sample data. The periods of post-sample series are also given in Table 1.

3 The Empirical Results

To implement genetic programming, the program GP-Pascal is written in Pascal 4.0 by following the instructions given in Koza (1992). A detailed description of this program can be found in Chen, Lin and Yeh (1995). The chosen parameters to run GP-Pascal are given in Table 2. % and *RLOG* appearing in the function set are the protected division function and the protected natural logarithm function (Koza, 1992, pp.82-83). In this paper, all simulations conducted are based on the terminal set, which includes the ephemeral random floating-point constant R ranging over the interval [-9.99, 9.99] and the rate of return lagging up 10 periods, i.e., R_{t-1},...R_{t-10}.

Table 2: Tableau for Predicting Rates of Return

Population size	500
Number of trees created by complete growth	50
Number of trees created by partial growth	50
Function set	$\{+, -, \times, \%, Sin, Cos, EXP, RLOG\}$
Terminal set	$\{R_{t-1}, R_{t-2}, \cdots, R_{t-10}, R\}$
Number of trees generated by reproduction	50
Number of new lives (immigrants)	50
Number of trees generated by mutation	100
Probability of mutation	0.2
Maximum length of the tree	17
Probability of leaf selection under crossover	0.5
Number of generations	200
Maximum number in the domain of Exp	1700
Criterion of fitness	MAPE

MAPE (*Mean Absolute Percentage Error*) is defined as follows:

$$MAPE = \sum_{i=1}^{m} \frac{\mid \hat{R}_i - R_i \mid}{\mid R_i \mid}, \qquad (2)$$

where m is the sample size and $\hat{R}_i$ is the prediction value of R_i. The choice of the *mean absolute percentage error* as the fitness function is attributed to Makridakis (1993), who suggested a modified form of MAPEs as the most appropriate measure satisfying both theoretical and practical concerns while allowing meaningful relative comparisons. Since the prediction made by the random walk hypothesis is always 0, the corresponding MAPE is always 1. So, *GP is said to beat the RW if its MAPE is less than 1.* In addition, we define $\pi_1(n)$ as the probability that GP can beat the RW in Generation n in the in-sample data and $\pi_2(n)$ the probability that GP can beat the RW in the holdout sample.

Based on those control parameters, multiple runs of simulations were executed for each sample. For each of the simulations, the MAPE is calculated for the in-sample period and the post-sample period. Statistics show that the sample $\pi_1(n)$ are 100% in Generations 50, 100, 150 and 200 for the small and medium sample

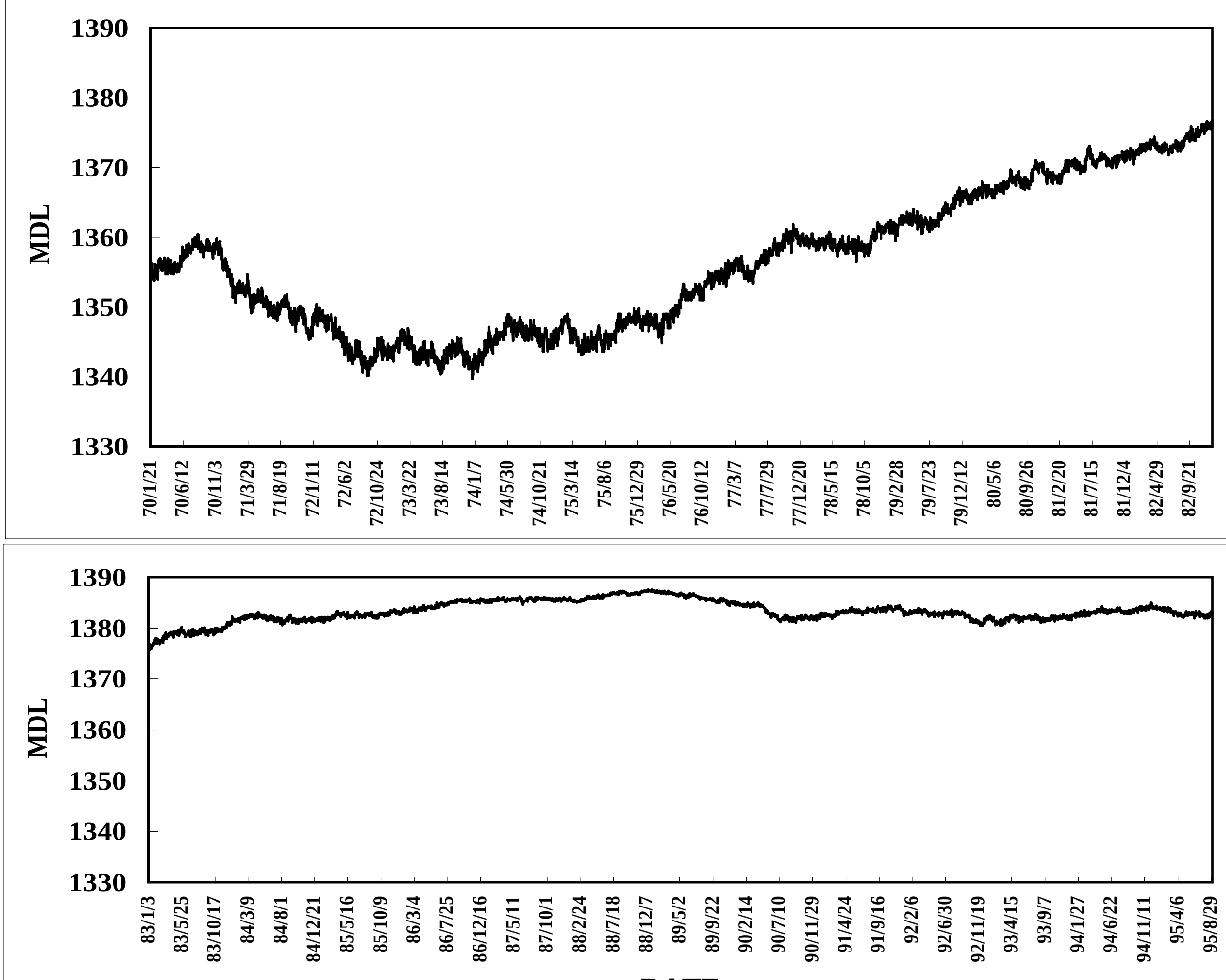

Figure 1 : The MDL of S&P 500 (1962.1.1-1995.9.6) (2000 Daily Returns)

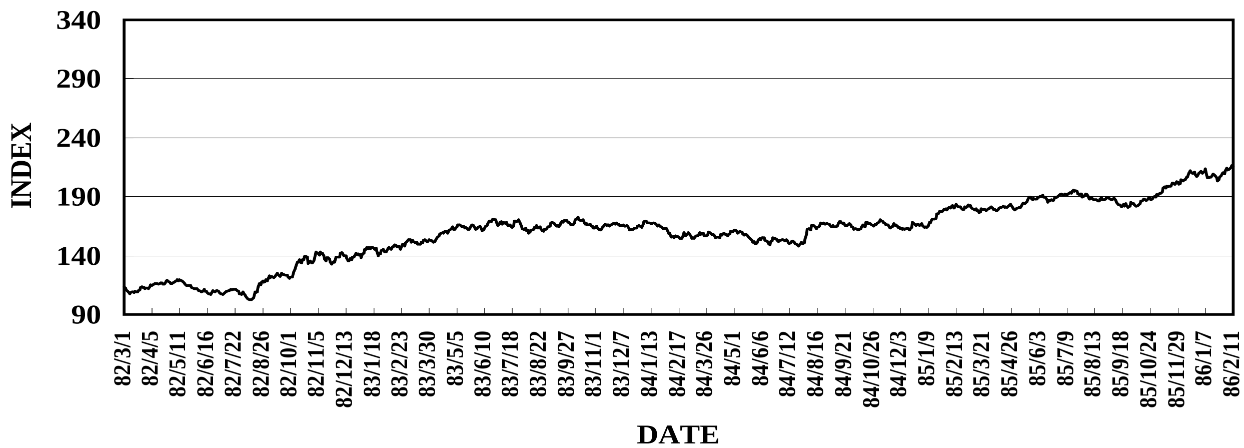

Figure 2 : The Stock Price Indices Corresponding to the Maximum MDL S&P 500 (1000 Daily Returns)

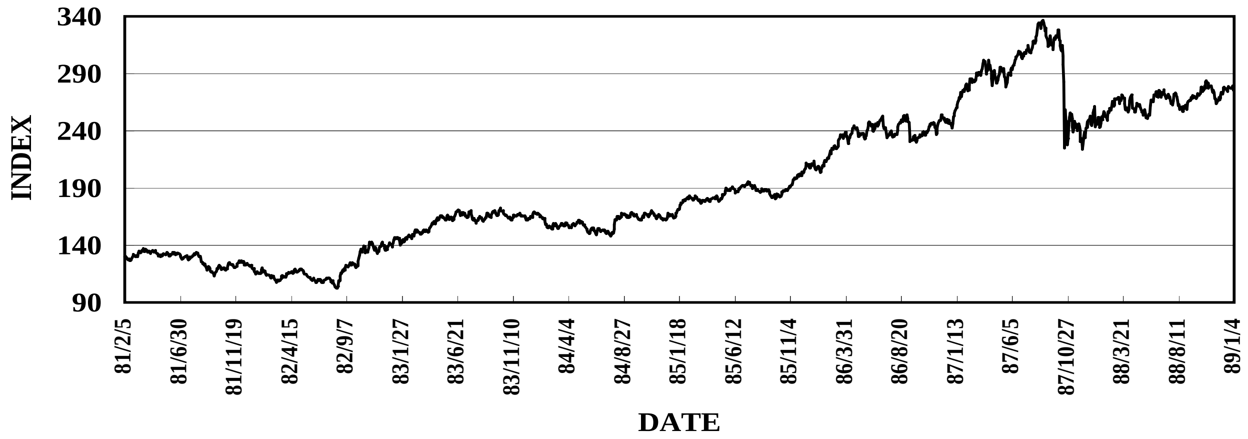

Figure 3 : The Stock Price Indices Corresponding to the Maximum MDL S&P 500 (2000 Daily Returns)

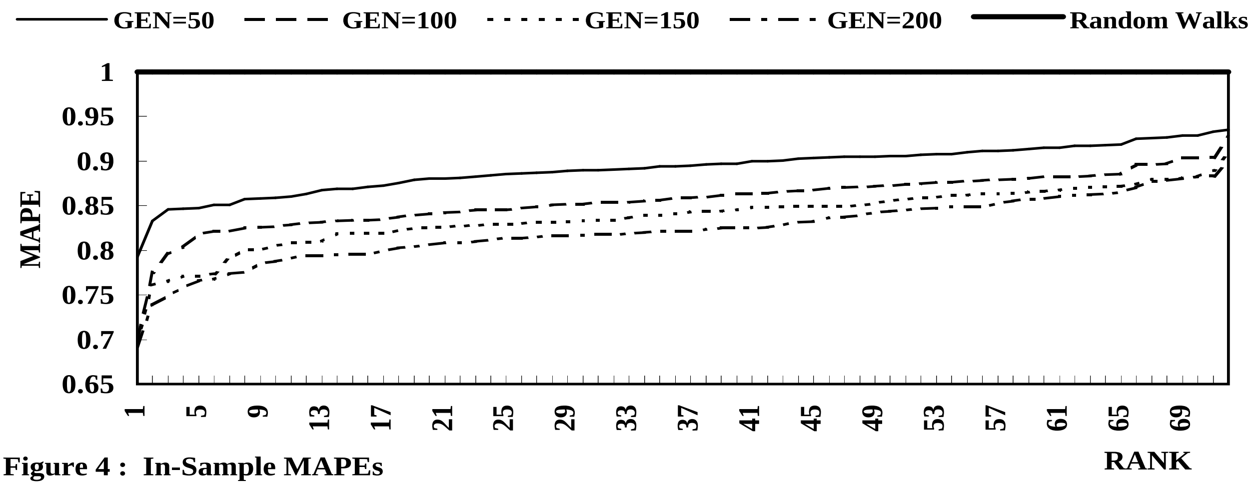

Figure 4 : In-Sample MAPEs (S&P 500)

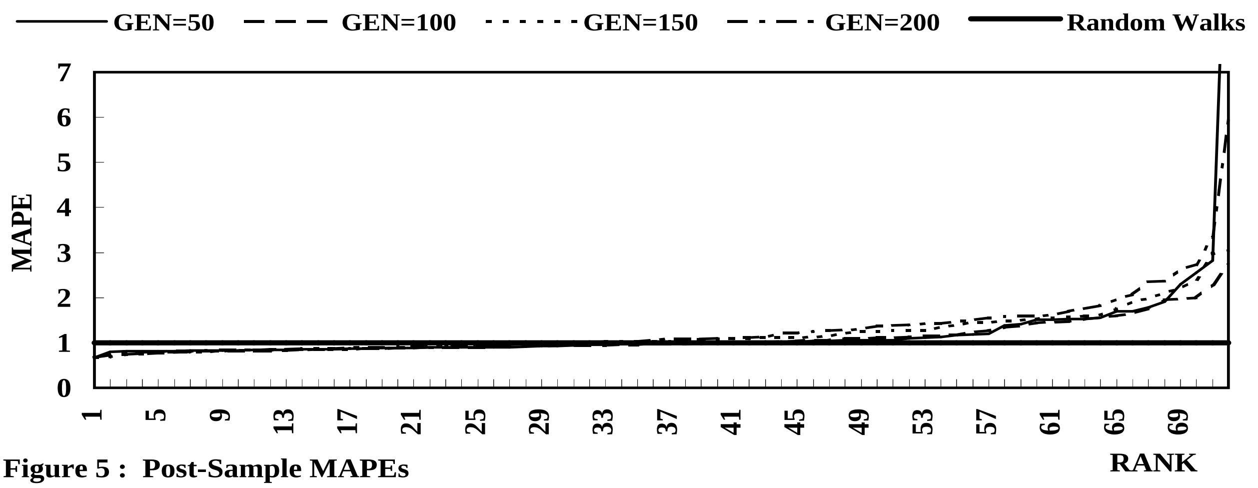

Figure 5 : Post-Sample MAPEs (S&P 500)

Table 3: The In-Sample MAPE of the Best Case and the Worst Case

Sample Size/Gen	50	100	150	200
Mini-Best	0.7925	0.6979	0.6955	0.6919
Small-Best	0.9698	0.9613	0.9567	0.9452
Medium-Best	0.9970	0.9964	0.9964	0.9964
Large-Best	0.9983	0.9980	0.9980	0.9974
Mini-Worst	0.9348	0.9268	0.9077	0.8985
Small-Worst	0.9904	0.9904	0.9870	0.9842
Medium-Worst	0.9999	0.9977	0.9977	0.9977
Large-Worst	1.0000	1.0000	1.0000	1.0000

Table 4: The Post-Sample MAPE of the Best Case and the Worst Case

Sample Size/Gen	50	100	150	200
Mini-Best	0.6712	0.6680	0.6680	0.6680
Small-Best	0.9987	0.9987	1.0000	0.9919
Medium-Best	0.9962	0.9952	0.9951	0.9951
Large-Best	1.0000	1.0000	1.0000	1.0000
Mini-Worst	2.8223	2.7109	3.0470	5.9289
Small-Worst	1.2021	1.0391	1.0383	1.0389
Medium-Worst	1.0239	1.0383	1.0337	1.0333
Large-Worst	1.0019	1.0019	1.0019	1.0075

(sample size=200 and 1000). Moreover, when evolution takes longer, limited improvement can always be made. For example, in the small sample, the MAPE in the best and worst case of these simulations in Generation 50 is 0.9698 and 0.9904 respectively, and in Generation 200, it is improved to 0.9452 and 0.9842. These results are also reported in Table 3.

From Table 3, there are a couple of points worth noticing. First of all, as the sample size gets larger, the difference in the in-sample MAPE between GP-based search and the random walk hypothesis becomes more and more negligible. The difference is particularly striking when we compare the in-sample MAPE performance of the mini sample (sample size = 50) with that of larger samples. To get a more general flavor of the MAPE performance under the GP-based search, the best in-sample and post-sample MAPEs of the mini sample of each simulation in Gen (generation) 50, 100, 150, and 200 are ranked from the lowest to the highest and are shown in Figures 4 and 5 (See Chen and Yeh (1995) for details).

Secondly, the range of the MAPE, i.e., the difference in the MAPE bewteen the best and the worst search tends to narrow as the sample size increases. It may be reasonable to summarize these results by the following *asymptotic-like hypothesis*,

$$MAPE_{gp}(n,m) \rightarrow 1, \quad as \quad m \rightarrow \infty, \tag{3}$$

where n is the number of evolution and m is the size of the training sample.

This hypothesis clarifies some confusing points around the EMH, in particular, the *time horizon* of the EMH. Starting with the pessimistic side, it says that *large sample training is not going to be very helpful.* In fact, our simulations show that it is easier to search for patterns from a smaller sample than from a larger one. This is consistent with the old truism in economics that "if there is a dollar bill on the ground, it is soon picked up.". Some econometricians such as White (1988) suggest that large sample training with recurrent neural nets might give the EMH a real challenge. Our hypothesis says exactly the opposite: it is the large sample that makes the EMH easily acceptable. Turning to the optimistic side, this hypothesis does imply that there is room for profits for *short-range forecasting with the small sample*; but then it also implies that the rule used to forecast is preferred to be updated on a weekly basis rather than on a monthly or yearly basis. As a result, the computational cost associated with this highly frequent revision must be higher than that for *long-range forecasting with the large sample*, and it may be so high that, in terms of net profits, there is no significant difference between long-range and short-range forecasting.

The computational cost is the physical cost. Another important cost concerning forecasting is the *risk cost* which is not directly derived from the physical side but from the utility (psychological) side. In economics and finance, investors are assumed to be *risk aversers*. This being the case, risk can be counted as another cost for them. So, let us turn to the risk side of forecasting strategies, and use the difference in the post-sample MAPE between the best GP-based search and the worst one as a measure of risk. Table 4 summarizes the results on the risk of different styles of the GP-based search.

From Table 4, we can see that the risk under long-range forecasting with large-sample training is much less than that under short-range forecasting with short-sample training. In particular, if we examine the mini sample studied by Chen and Yeh (1995), the risk can be five times higher than that with the large sample. Thus, even though GP-based data mining can be useful for the short-term data, after taking into account the computational cost and the risk cost, its advantages may not be that great.

4 On The Competitiveness of the Random Walk Hypothesis

To get a better grasp of the GP-based notion of the EMH, we propose in this section another perspective to study the EMH by raising the following issues:

- If the efficient market hypothesis is true, can genetic programming automatically discover this truth? In fact, we are asking whether genetic programming can be used to prove the EMH.
- Can the random walk hypothesis survive well in the competitive environment generated by genetic programming?

As to these questions, the major result of our simulations is that *when the sample size is large enough, e.g., sample size = 2000, genetic programming can actually be used to prove the EMH by discovering the random walk hypothesis as the most competitive model.* We shall illustrate this based on Simulation 2000-1, which is the first simulation for the 2000-observation sample. The best model chosen from Gen. 0 is:

$$\mathtt{F}^{0}_{\mathtt{best}} = \mathtt{Log}(\mathtt{R}_{\mathtt{t}-1} - \mathtt{R}_{\mathtt{t}-1}) \tag{4}$$

Since $Log(0)$ is defined as 0 in the program, the best model found by GP in the initial generation is exactly the random walk hypothesis. In other words, the RW hypothesis was discovered at the very beginning of the evolution. Still, if the random walk hypothesis is not competitive, then during the evolution, one could expect that other better models will be discovered and replace the RW. However, this was not the case observed in Simulation 2000-1. The RW hypothesis kept on dominating the evolution till the end while it appeared in different styles. For example, the best model from Gen. 16 is:

$$\mathtt{F}^{16}_{\mathtt{best}} = (((\mathtt{R}_{\mathtt{t}-10} - \mathtt{R}_{\mathtt{t}-10})\%\mathtt{ExpR}_{\mathtt{t}-3})\% - 8.163083) \tag{5}$$

Since R_{t-10} cancels itself in Equation (2), F^{16}_{best} is in fact the zero function. The evolution might generate seemingly complicated functions. After cancellation and reduction, however, they can always be simplified to a zero function. For example, the best model from Gen. 195 is:

$$\begin{aligned}\mathtt{F}^{195}_{\mathtt{best}} = {} & \mathtt{Log}(((\mathtt{ExpR}_{\mathtt{t}-1} - (\mathtt{R}_{\mathtt{t}-8}\%\mathtt{R}_{\mathtt{t}-10})) - (\mathtt{ExpR}_{\mathtt{t}-3} \\ & -(\mathtt{R}_{\mathtt{t}-8}\%\mathtt{R}_{\mathtt{t}-10})))\%\mathtt{Log}((\mathtt{R}_{\mathtt{t}-10} - \mathtt{Cos}(\mathtt{R}_{\mathtt{t}-1} - \mathtt{R}_{\mathtt{t}-1})) \\ & \%(((\mathtt{R}_{\mathtt{t}-1} - \mathtt{R}_{\mathtt{t}-1})\%\mathtt{R}_{\mathtt{t}-1})\%(\mathtt{R}_{\mathtt{t}-1} - \mathtt{R}_{\mathtt{t}-10}))))\end{aligned} \tag{6}$$

Considering the zero function an equivalent class, then we can see from the evolution in the large sample that *the members of the class of zero function dominated the whole evolution.* In other words, the random walk hypothesis is very competitive in the large sample. GP could not easily find anything better than the RW hypothesis. However, this result does not hold for the small sample. In particular, when the sample size is 200, the random walk hypothesis can be the best only in the initial generation (Gen. 0). It would be quickly replaced by other models once the evolution started. For example, the best model from Gen. 0 in Simulations 200-2 and 200-3 are given in Equations (7) and (8) respectively.

$$\mathtt{G}^{0}_{\mathtt{best}} = (\mathtt{R}_{\mathtt{t}-7} * (\mathtt{R}_{\mathtt{t}-6} - \mathtt{R}_{\mathtt{t}-6})) \tag{7}$$

$$\mathtt{H}^{0}_{\mathtt{best}} = \mathtt{Sin}((\mathtt{R}_{\mathtt{t}-3} + \mathtt{R}_{\mathtt{t}-2}) * (\mathtt{R}_{\mathtt{t}-4} - \mathtt{R}_{\mathtt{t}-4})) \tag{8}$$

Clearly, both G^0_{best} and H^0_{best} are the members of the equivalent class of zero function, hence they represent the random walk hypothesis. However, in Gen. 1, they were defeated by the G^1_{best} and H^1_{best} (given below) separately.

$$\mathtt{G}^{1}_{\mathtt{best}} = (\mathtt{R}_{\mathtt{t}-5} * (\mathtt{R}_{\mathtt{t}-7} * \mathtt{R}_{\mathtt{t}-1})) \tag{9}$$

$$\begin{aligned}\mathtt{H}^{1}_{\mathtt{best}} = {} & \mathtt{Sin}(\mathtt{Sin}(\mathtt{CosLogR}_{\mathtt{t}-8} * (\mathtt{R}_{\mathtt{t}-1} - \mathtt{R}_{\mathtt{t}-6}))\%((\mathtt{Log} \\ & \mathtt{R}_{\mathtt{t}-4} - (4.522217\%\mathtt{R}_{\mathtt{t}-3})) + \mathtt{CosExpR}_{\mathtt{t}-2}))\end{aligned} \tag{10}$$

Therefore, the random walk hypothesis is not competitive in the small sample and combined with the finding of Chen and Yeh (1995), we can conclude that *the learnability of the financial data (S&P 500) is only restricted to the small sample.* When the sample gets large, nothing can be learned from it except that it is random.

This conclusion can also be justified by *the complexity of the best discovered model.* That is, instead of asking what rules are discovered by GP, we are inquiring how complex those rules are which are discovered by GP. The length of the *LISP program* is used to measure complexity. The complexity of the best model chosen by GP in Gen 50, 100, 150 and 200 in a few selected simulations is given in Table 5. It is interesting to note that when the sample size is small, in-sample fitness can always be improved by searching for a bigger LISP tree. For example, when the sample size is 200, the length of the best LISP program has a positive relation to the number of generation of the evolution. However, when the sample size is large, the relation disappears.

5 Concluding Remarks

The attempts to formalize the EMH as a consistent, analytical economic theory have met

Table 5: The Length of the Best Model (LISP Program)

Simulation/Gen	50	100	150	200
200-2	75	243	303	315
200-3	24	24	53	126
200-4	48	54	210	465
200-7	210	300	333	339
1000-1	309	387	330	360
1000-3	198	108	111	117
1000-5	114	186	162	198
2000-1	111	69	191	168
2000-3	24	42	153	84

> with less success than the empirical tests of the hypothesis. (Ross, 1987, p.7)

Needless to say, at this moment, any conclusion about the validity of the EMH is premature and this paper should not leave readers any impression that we are pro or against the EMH. After all, the validity of the EMH is not the major concern of this paper. However, the paper does have a purpose. It attempts to convince readers that over the last 90 years the EMH has not been formalized in an appropriate way, and that biologically inspired approaches, in particular, genetic programming, hopefully can make it right. Right not only in terms of formalization but also about questions and analysis.

For example, the right question is not whether there is evidence of nonlinear independence but whether such nonlinear patterns can be discovered at reasonable cost. The right analysis is not to test the large sample with asymptotic econometrics but to do a cost-effective analysis of forecasting different ranges of market with different sizes of sample. Similarly, the right question is not whether there will be a sequence of advanced techniques which can beat the random walk but whether the retraining or updating price of these techniques is competitive; the right analysis is not to test whether the random walk can be beaten by one recurrent neural net once but to check whether the random walk can be beaten by *evolutionary neural nets* continuously with reasonable cost. All the right questions with their corresponding analyses have an *economic sense.* Malkiel's intuition of the EMH is also an economic one rather than a statistical one. It is the absence of this economic sense in the traditional analysis that has divided us between *technicians* and *quants.* To capture the economic sense, we need the theoretical technology which, to some extent, can standardize the search activity so that search intensity and search cost can be communicated among different searchers. Genetic programming is just the candidate we are looking for.

This concludes the summary of what this paper intends to do. The purpose is very limited and the attempt is far from mature. There is little doubt that there are lots of imperfections to be fixed and details to be figured out in the follow-up research. But, if the relevance of genetic programming to the EMH is clear, the next non-trivial issue, which is motivated by Maarten Keijzer, is *what is the relationship between different styles of GP with different computing environments and the validity of the EMH.* For example, will the EMH be rejected if the computer power is doubled? Will the conclusion about the EMH differ if we are using supercomputer system to find the pattern instead of a single processor machine? Will the EMH be rejected if it is tested by parallel genetic programming with tick-by-tick data? List of questions of this sort can go on and on. But what are the answers? At the end of his survey article, Malkiel seemed to offer some insight into such an intriguing issue.

> Undoubtedly with the passage of time and with the increasing sophistication of our data bases and empirical techniques, we will document further departures from efficiency and understand their causes more fully. But I suspect that the end result will not be an abandonment of the profession's belief that stock market is remarkably efficient in its utilization of information. (p.133)

Acknowledgments

The authors are indebted for the research support from NSC grant No.85-2415-H-004-001. This paper was revised from its original version by taking the advice and comments of Christian Johnson, Marten Keijzer and three other anonymous referees in the **GP'96** program committee. The authors are very grateful for their painstaking review of this paper. Of course, all remaining errors are the authors' sole responsibility.

References

[1] Brock, W., W. Dechert, and J. Scheinkman 1987. A Test for Indepedence Based on the Correlation Dimension. Working Paper, University of Wisconsin at Madsion, University of Houston, and University of Chicago.

[2] Chen, S., C. Lin and C. Yeh 1995. On the Model Selection and its Stability of the Natural Vacancy

Rate of Housing: An Application of Genetic Programming. *Journal of Housing Studies.* 3. Pages 73-98.

[3] Chen, S. and C. Yeh 1995. Toward a Computable Approach to the Efficient Market Hypothesis: An Application of Genetic Programming. Working Paper Series, No. 9502. Department of Economics, National Chengchi University.

[4] Chen, S. and C. Tan 1996. Measuring Stock Market Efficiency by Rissanen's MDL. Paper presented at the *Pacific Rim Allied Economic Organizations 2nd Biennial Conference* coordinated by *Western Economic Association International.* Jan. 10-15, 1996. Hong Kong.

[5] Diebold, F. X. and J. A. Nason 1990. Nonparametric Exchange Rate Prediction?. *Journal of International Economics.* 28. Pages 315-332.

[6] Fogler, H. R. 1995, Investment Analysis and New Quantitative Tools. Paper presented at the *1995 Eastern Finance Association Meeting.* Hilton Head Island, South Carolina.

[7] Frank, M. Z., R. Gencay, and T. Stengos 1988. International Chaos?. *European Economic Review.* 32. Pages 1569-1584.

[8] Hinich, M. J. and D. M. Patterson 1985. Evidence of Nonlinearity in Daily Stock Returns. *Journal of Business and Economic Statistics.* Vol. 3, No. 1. Pages 69-77.

[9] Horgan, J. 1995, From Complexity to Perplexity. *Scientific American.* June, Pages 104-109.

[10] Hsieh, D. A. 1989. Testing for Nonlinear Dependence in Daily Foreign Exchange Rates. *Journal of Business.* 62. Pages 339-368.

[11] Koza, J. 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* Cambridge, MA: The MIT Press.

[12] LeBaron, B. 1992. Nonlinear Forecasts for the S&P Stock Index. In M. Casdagli and S. Eubank (eds.). *Nonlinear Modeling and Forecasting.* Pages 381-393. Addison-Wesley Press.

[13] Makridakis, S. 1993, Accuracy Measure: Theoretical and Practical Concerns. *International Journal of Forecasting.* 9. Pages 527-529.

[14] Malkiel, B. G. 1987. Efficient Market Hypothesis, in J. Eatwell, M. Milgate and P. Newman (eds.). *The New Palgrave: Finance.* Pages 127-134. Norton.

[15] Peters, E. E. 1991. A Chaotic Attratcors for the S&P 500. *Financial Analysis Journal.* March-April. Pages 55-62.

[16] Rissanen, J. 1982. A Universal Prior for Integers and Estimation by Minimum Description Length. *Annals of Statistics.* 11. Pages 416-431.

[17] Ross, S. 1987. Finance. In J. Eatwell, M. Milgate and P. Newman (eds.). *The New Palgrave: Finance.* Pages 1-34. Norton.

[18] Savit, R. 1988. When Random Is Not Random: An Introduction to Chaos in Market Prices. *Journal of Futures Markets.* 8. Pages 271-290.

[19] Savit, R. 1989. Nonlinearities and Chaotic Effects in Option Prices. *Journal of Futures Markets.* 9. Pages 507-518.

[20] Scheinkman, J. A. and B. LeBaron 1989. Nonlinear Dynamics and Stock Returns. *Journal of Business.* 3. Pages 311-337.

[21] White, H. 1988. Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns. *Proceedings of the IEEE International Conference on Neural Networks.* II. Pages 451-458.

[22] Willey, T. 1992. Testing for Nonlinear Dependence in Daily Stock Indices. *Journal of Economics and Business.* Vol. 44, No. 1. Pages 63-74.