Bridging the Gap between Nonlinearity Tests and the Efficient Market Hypothesis by Genetic Programming*

Shu-Heng Chen
AI-ECON Research Group
Department of Economics
National Chengchi University
TEL: 886-2-9387139
FAX: 886-2-9390344
E-mail: chchen@cc.nccu.edu.tw

Chia-Hsuan Yeh
AI-ECON Research Group
Department of Economics
National Chengchi University
TEL: 886-2-9387308
FAX: 886-2-9390344
E-mail: g3258501@grad.cc.nccu.edu.tw

Paper to be presented at the *Conference on Computational Intelligence for Financial Engineering* **CIFEr Conference**, March 24-26, 1996, New York City, Crowne Plaza Manhattan.

Abstract

This paper applies the GP-based notion of unpredictability to the test of the efficient market hypothesis. It extends the study of Chen and Yeh (1995) by testing the EMH with a small, medium and large sample of the S&P 500 stock index. It is found that, in terms of the prediction performace, the probability $\pi_2(n)$ that GP can beat the random walk tends to have a negative relation to the size of the in-sample dataset. For example, $\pi_2(n)$ is 0.5, 0.2, and nil when sample size is 50, 200, and 2000. It therefore suggests that while nonlinear regularities could exist, they might exist in a very short span. As a consequence, the search costs of discovering them might be too high to make the exploitation of these regularities profitable, hence the efficient market hypothesis is sustained.

1 Introduction

Despite its long history[1], the efficient market hypothesis (EMH) has remained at the heart of much of the contemporary debate in financial economics. Throughout its entire history, the EMH was mainly formalized and modified based on the concept of *probabilistic independence*. Technically speaking, it shows that the σ-algebra generated by the history of the rates of return will tell us nothing about the present or future rates of return. In other words, the rate of return at time period t R_t should be independent of any Borel functions of R_s (s < t)[2]. Up to the present, this seems to have been the only way to mathematize the intuitive meaning behind the EMH, i.e., *unpredictability*, but that is not to say that it is scientifically sound. The major problem of this formalization is that there is no effective algorithm such that we can construct the evidence of independence by trying *all* Borel functions of R_s (s < t). Therefore, *the EMH based on this notion of unpredictability is practically uncomputable*, and in practice, the test gradually proceeds from *linear independence* to *non-linear independence* such as the BDS test[3]. While lots of recent studies based on nonlinear tests[4] evidence that there might exist nonlinear dependence in the financial data, these tests alone tell us nothing about the predictability of the the rates of return[5]; nor can they indicate whether we

*The authors are grateful for the research support from NSC grant No.85-2415-H-004-001.

[1] It goes back to Bachelier (1900).

[2] In the literature, this is called *weak-form efficiency*.

[3] See Brock, Dechert and Scheinkman (1987).

[4] See Savit (1988, 1989), Hinich and Patterson (1989), Hsieh (1989), Frank, Gencay and Stengos (1988), Scheinkman and LeBaron (1989), Peters (1991), and Willey (1992).

[5] For example, the tests alone cannot tell us whether low-dimensional chaos is easier to predict than high-dimensional chaos or whether deterministic chaos is easier to predict than stochastic nonlinearity.

can profit from a better prediction. For example, the rejection of nonlinear tests might suggest that there exist nonlinear regularities to be exploited; however, the search costs to find such nonlinear regularities may be so high that the net profits of using this nonlinear regularity might be negative. In this case, rejecting the non-existence of nonlinear dependence does not mean the rejection of the EMH. Hence what we can truly learn from these nonliearity tests is not clear. In other words, *there is a gap between the result of nonlinearity tests and its implication for the efficient market hypothesis.*

This paper contributes to bridging the gap between nonliearity tests and the EMH by providing an approach where both the issues of *predictability* and *profitability* are taken into account. Thus, if the EMH is rejected in our approach, it means what it should mean. We shall call this approach a *computable* approach. Under this approach, the intuitive meaning of *hard to predict* or *very hard to predict* is considered equivalent to *(very) hard to find a rule from past experience under intensive search which can predict the future better than a random walk (RW).* If this is the case, to capture the technical meaning of unpredictability and the EMH, we only need *an explicit search program* in which *the intensity of the search* denoted by a vector x, $x \in R^m$, and *the chance of success in the search* denoted by π can be formalized. Then π as a function of x, i.e., $\pi(x)$ can be considered an objective measure of unpredictability, i.e., an indicator of showing how hard it is to predict. Given the same level of x, the lower the π, the harder it is to predict; or given π, the higher the level required, the harder it is to predict.

The explicit search program considered in this paper is the *genetic programming paradigm* developed by Koza (1992). The search intensity of GP can be revealed by the set of chosen parameters in running genetic programming. For example, by increasing the "*population size*" from 500 to 1000 and/or the "*number of generations*" from 1000 to 2000, we are increasing the search intensity. As to the chance of success, it crucially depends on what we mean by success. For example, success could means MAPE (mean absolute percentage error) = 0, or MAPE < 1, and so on. Compared with the traditional notion, a GP-based search provides an explicit and efficient search program upon which an objective measure for predictability can be formalized in terms of search intensity and chance of success in the search.

In Chen and Yeh (1995), this formalization of unpredictability was illustrated by examples of predicting chaotic dynamic systems. They applied GP to predict the time series generated by different chaotic dynamic systems. They have shown that if the rates of return are generated by a *simple* deterministic chaotic dynamic system, then GP may actually discover it[6]. 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.

2 Selecting Data with the MDLP

We will now give an example to demonstrate how the GP-based notion of unpredictability can be used to test the EMH. These examples are based on the data concerning the daily rate of return of the S&P 500 Index. From 1/2/62 to 9/6/95, there are 8,478 observations in the S&P 500 dataset. Are we going to use all of them? While the efficient market hypothesis puts no restriction on the sample size, the application of GP to different sizes of sample does require lots of thought. This is because GP aims at finding the existence of potential nonlinear regularities. The requirement for the sample size varies with different periodicities. For the time-invariant long-term nonlinear relation, a large sample size is needed. However, the recent studies of the time series of stock prices, such as LeBaron (1992), seem to indicate that, even though nonlinear regularities might exist, they are not stable over time. Therefore, 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. In this paper, a short-term (sample size=200), medium-term (1000) and long-term (2000) sample of S&P 500 with the highest complexity defined by Rissanen's MDL (minimun description length) were chosen and tested.

Rissanen's MDL (minimun 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

[6] If we represent each chaotic dynamic system by the *LISP S-expression* depicted as a rooted, point-labeled tree (GP-tree), the term "*simple*" refers to the *depth* of the GP-tree and has nothing to do with the embedding dimension.

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

Table 2: Tableau for Predicting Rates of Return

Population size	500
The number of trees created by complete growth	50
The number of trees created by partial growth	50
Functional set	$\{+, -, \times, \%, sin, cos, EXP, RLOG\}$
Terminal set	$\{R_{t-1}, R_{t-2}, \cdots, R_{t-10}\}$
The number of trees generated by reproduction	50
The number of new lives	50
The number of trees generated by mutation	100
The probability of mutation	0.2
The maximum length of the tree	17
The probability of leaf selection under crossover	0.5
The number of generations	200
The maximum number in the domain of Exp	1700
Criterion of fitness	MAPE

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 this procedure and its meaning can be found in Chen and Tan (1996). Briefly speaking, we first transform the orginal 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[7]. By this criterion, the period for different sizes of sample are given in Table 1. The time series of the chosen period for different sizes of smaple are drawn in Fig. 2.1- 2.3.

Once the in-sample data series are given, the next issue is how to determine the length of 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 instruction 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.

Based on these parameters, 10 simulations were executed for each sample. For each of the simulation, the MAPE is calculated for the in-sample period and the post-sample period. The in-sample MAPEs of the best model chosen by each simulation in Gen (generation) 50, 100, 150, and 200 are ranked from the lowest to the highest and are shown in Figures 3.1-3.2[8]. *Since the MAPE of the RW is 1, GP is said to beat the RW if its MAPE is less than 1.* Let π_1(n) be the probability that GP can beat the RW in Generation n in the in-sample data. Statistics show that the sample $\pi_1(n)$ are 100% in Generations 50, 100, 150 and 200. However, the difference seems to be negligible. When evolution takes longer, limited improvement can always be made. For example, the MAPE in the best and worst case of these simulations in Generation 50 is 0.9698 and 0.9908 respectively, and in Generation 200, it is improved to 0.9452 and 0.9842.

[7] The MDL(x_i^{200}) of S&P 500 is given in Figures 1.1-1.3.

[8] Due to the size limit, only the case of short-term sample is discussed in this paper.

Figure 1.1 : The MDL of S&P 500
(1962.1.1-1974.12.31) (200 daily returns)

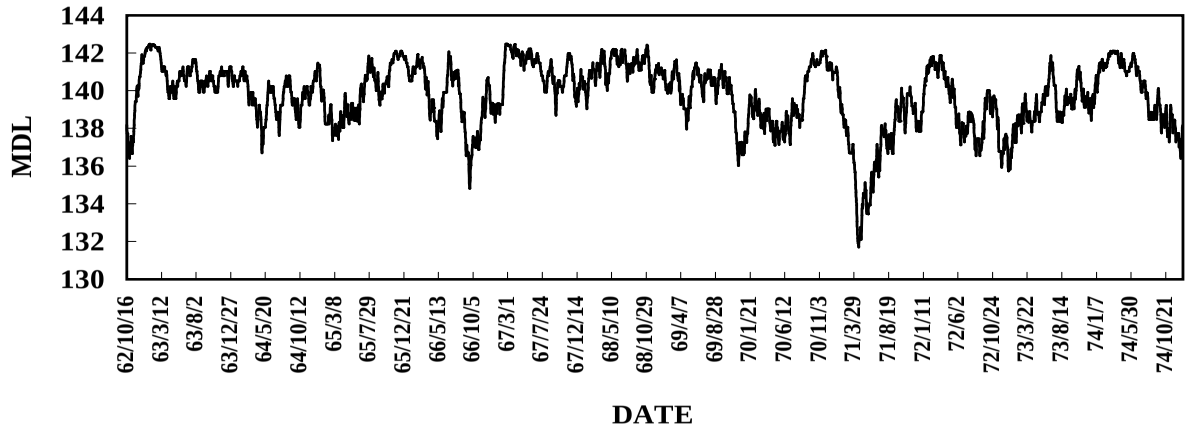


Figure 1.2 : The MDL of S&P 500
(1975.1.1-1986.12.31) (200 daily returns)

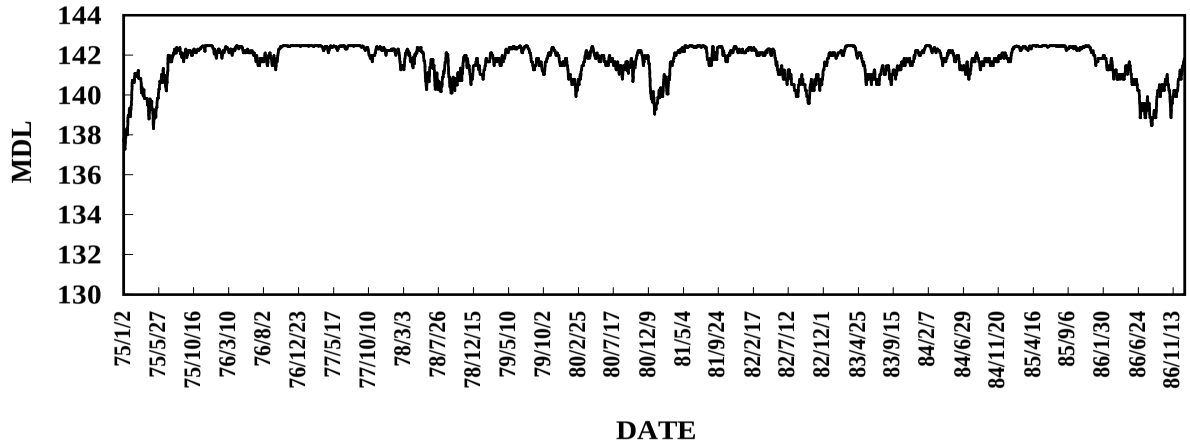


Figure 1.3 : The MDL of S&P 500
(1987.1.1-1995.9.6) (200 daily returns)

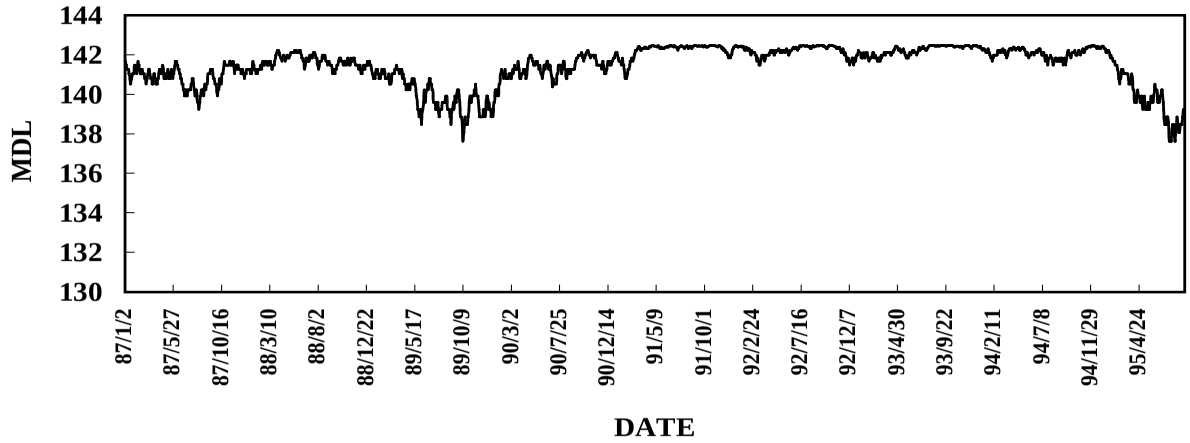


Figure 2.1 : The Stock Price Indices Corresponding to the Maximum MDLS&P 500 (200 daily returns)

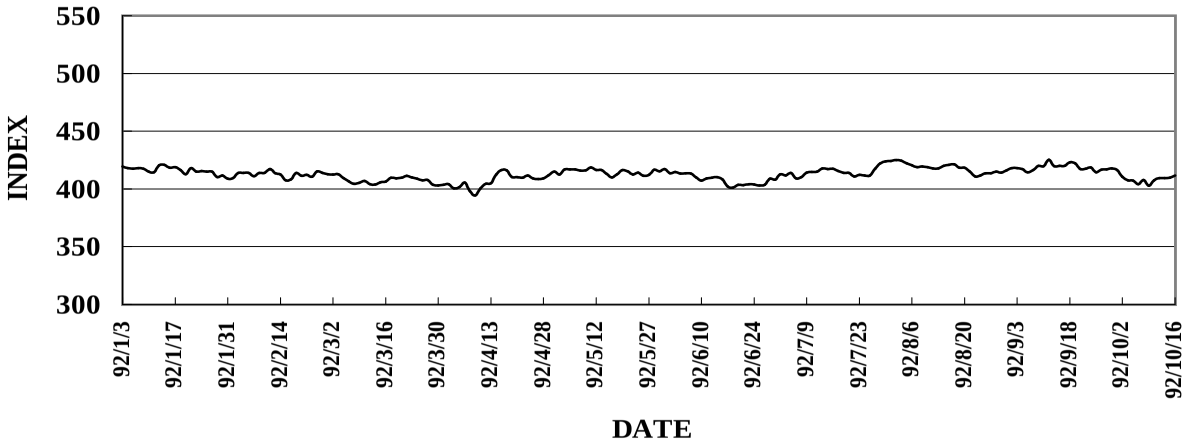


Figure 2.2 : The Stock Price Indices Corresponding to the Maximum MDL S&P 500 (1000 daily returns)

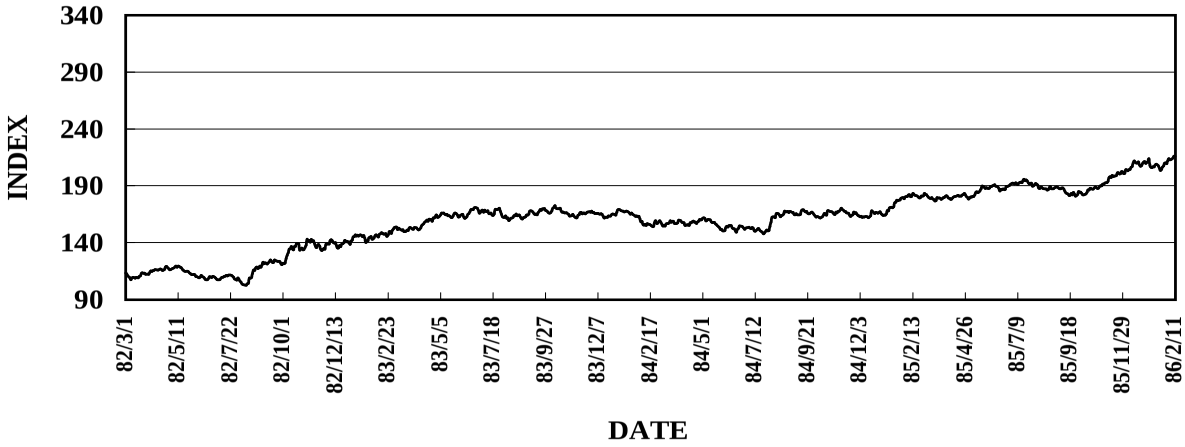


Figure 2.3 : The Stock Price Indices Corresponding to the Maximum MDL S&P 500 (2000 daily returns)

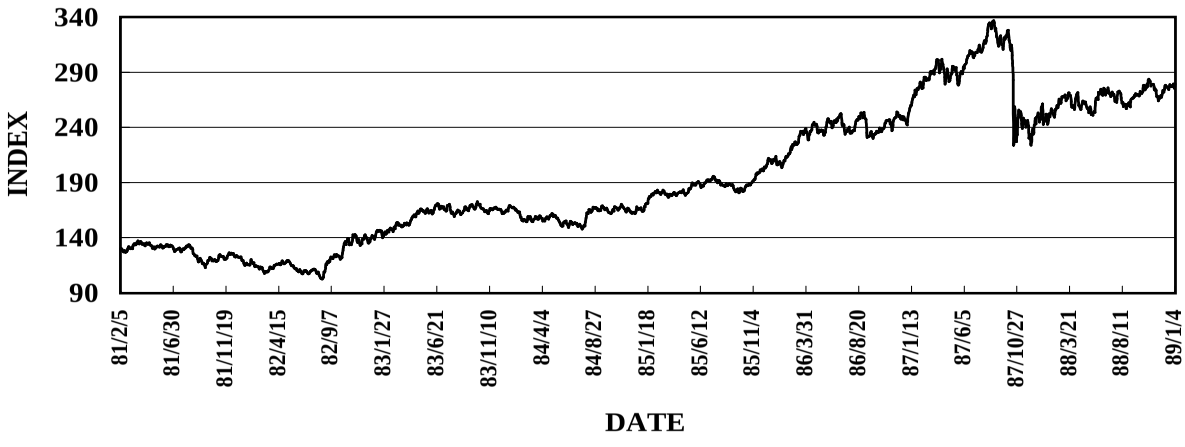


Figure 3.1 : In-Sample MAPEs

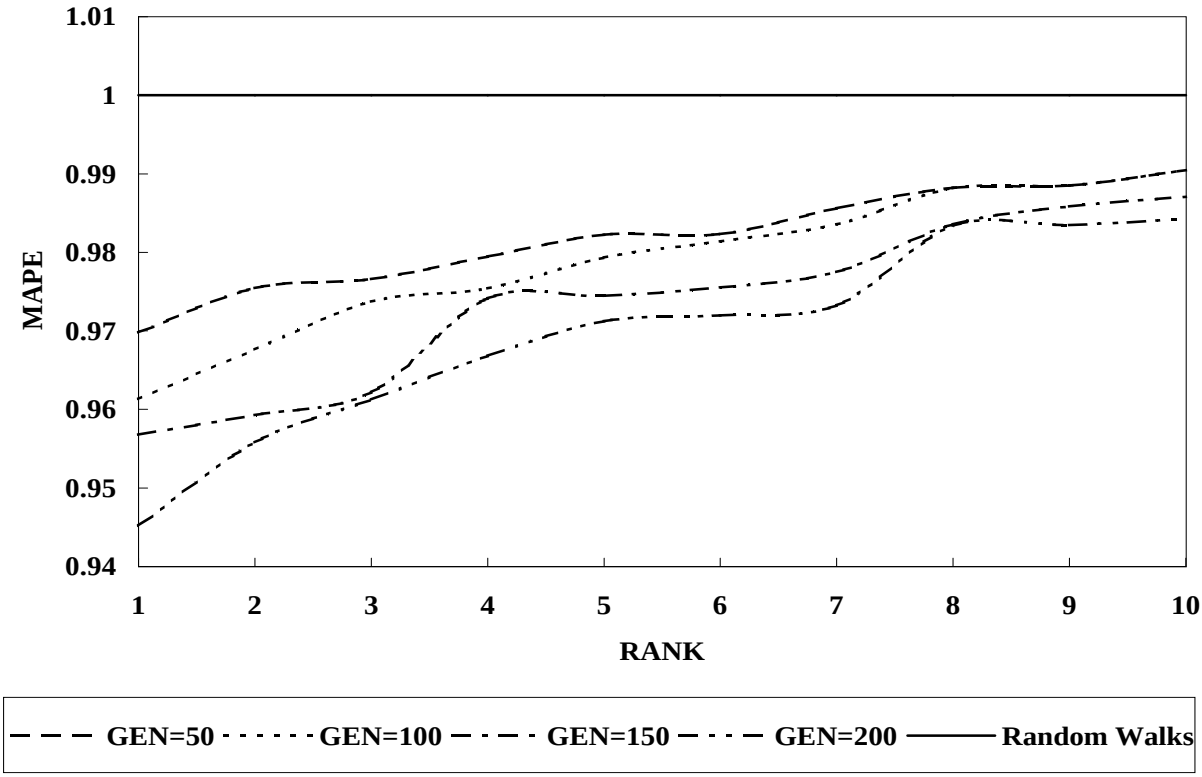
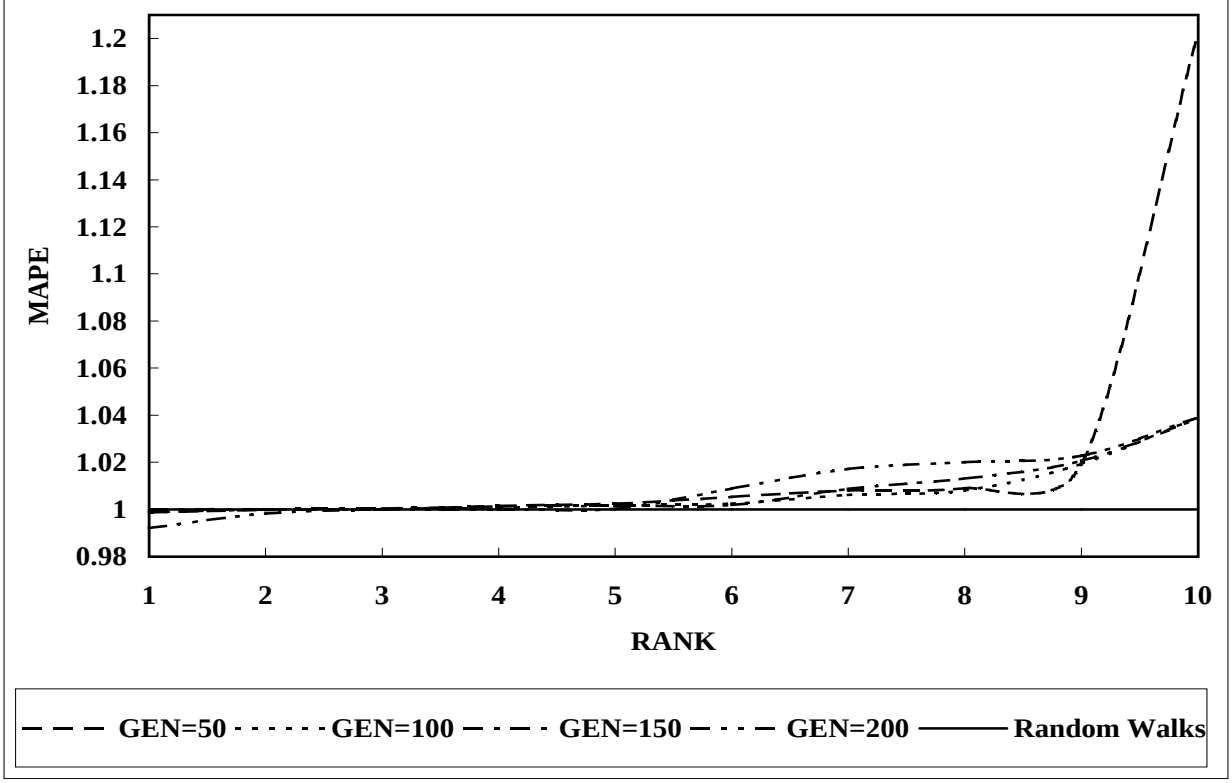


Figure 3.2 : Post-Sample MAPEs



Chen and Yeh (1995) conducted a similar study but within an even mini sample (sample size = 50). It is found that in that sample, the MAPE of the best chosen model of GP at Generation 200 can go down to 0.6. We suspect that GP-based data mining can be useful for the short-term data rather than for the long-term.

Finally, in terms of post-sample performance, we consider $\pi_2(n)$, i.e., the probability that GP can beat RWs in the holdout sample. It is found that, while linear models cannot predict better than the random walk, the GP-based search based on the short-term dataset can beat random walk by 20% in the short-term sample. Moreover, given the ratio of the size of the in-sample dataset to that of post-sample, $\pi_2(n)$ tends to have a negative relation to the size of the in-sample dataset. For example, the statistics of $\pi_2(n)$ for the the mini sample in Chen and Yeh (1995) can be higher up to 0.50. This is another evidence against the use of GP for the long-term dataset. In fact, this just confirms what Peters (1991) and LeBaron (1992) have found: *while nonlinear regularity might exist, the useful ones usually exist in a very short span.*

4 On The Competitiveness of the Random Walk Hypothesis

To gain a better grasp of the GP-based notion of the EMH, we propose in this section a different 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 enviroment generated by genetic programming?

As to these questions, the major result from 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:

$$F^0_{best} = Log(R_{t-1} - R_{t-1}). \tag{1}$$

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 begining 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:

$$F^{16}_{best} = (((R_{t-10} - R_{t-10})\%ExpR_{t-3})\% - 8.163083). \tag{2}$$

Since R_{t-10} cancels iteself in Eq. 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:

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

Considering the zero function as 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 anyone 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 (4) and (5) respectively.

$$G^0_{best} = (R_{t-7}(R_{t-6} - R_{t-6})) \tag{4}$$

Table 3: 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
2000-1	111	69	191	168

$$H^0_{best} = Sin((R_{t-3} + R_{t-2})(R_{t-4} - R_{t-4})) \tag{5}$$

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.

$$G^1_{best} = (R_{t-5}(R_{t-7}R_{t-1})) \tag{6}$$

$$H^1_{best} = Sin(Sin(CosLogR_{t-8}(R_{t-1} - R_{t-6}))\%((LogR_{t-4} - (4.522217\%R_{t-3})) + CosExpR_{t-2})) \tag{7}$$

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 restrcited 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 ask how complex the rules are which are discovered hy GP. The length of the *LISP program* is used to measure the 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 3. 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.

References

[1] Bachelier, L. (1990). Thé orier de la speculation. *Annales Scientifiques de l'Ecole Normale Supé rieure.* **17**, 21-88.

[2] 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.

[3] Chen, S., C. Lin and C. Yeh (1995). "On the Model Selection and its Stability of the Natural Vancancy Rate of Housing: An Application of Genetic Programming," *Journal of Housing Studies.* **3**, 73-98.

[4] Chen, S. and C. Yeh (1995), "Toward a Computable Approach to the Efficient Market Hypothesis: An Application of Genetic Programming," paper presented at the *First International Conference of the Society for Computational Economics: Computing in Economics and Finance*, May 21-24, 1995, Austin, Texas. Forthcoming in the *Journal of Economic Dynamics and Control.*

[5] 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.

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

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

[8] Hsieh, D. A. (1989), "Testing for Nonlinear Dependence in Daily Foreign Exchange Rates," *Journal of Business*, 62, pp.339-368.

[9] 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, pp.69-77.

[10] Koza, J. (1992), *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. The MIT Press.

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

[12] Peters, E. E. (1991), "A Chaotic Attratcors for the S&P 500," *Financial Analysis Journal*, March-April, pp.55-62.

[13] Rissanen, J. (1982), "A Universal Prior for Integers and Estimation by Minimum Description Length," *Annals of Statistics*. **11**, pp.416-431.

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

[15] Savit, R. (1989), "Nonlinearities and Chaotic Effects in Option Prices," *Journal of Futures Markets*, 9, pp.507-518.

[16] Scheinkman, J. A. and B. LeBaron (1989), "Nonlinear Dynamics and Stock Returns," *Journal of Business*, 3, pp.311-337.

[17] Willey, T. (1992), "Testing for Nonlinear Dependence in Daily Stock Indices," *Journal of Economics and Business*, pp.63-74.