Can We Believe That Genetic Algorithms Would Help without Actually Seeing them Work in Financial Data Mining?: Part 1, Theoretical Foundations*

Shu-Heng Chen[1]

National Chengchi University, Taipei, Taiwan 11623

Abstract. Genetic algorithms (GAs) have, time and again, shown some promising features when applied to optimization problems. The theoretical foundations of these successful applications however are rather limited, in particular, when the problem embodies a *dynamic* rather than a *static* landscape. In this paper, dynamic landscapes are treated as *random variables*, and we sort out a few *stochastic properties* which may impinge upon the performance of GAs in financial data mining. Several tests of these properties are then proposed and *a priori* evaluation of the potential of GAs can be made based on these proposed tests.

1 Motivation and Introduction

"*If you had everything (computationally), where would you put it (financially)?*" This is the opening question raised in Leiweber and Arnott (1995) and is the question faced by all investment managers in the era of high-performance computing. Efforts made to answer this question have introduced a new field entitled *computational intelligence in finance*. While many techniques have been claimed to *be helpful* in this area, knowledge about *why they are helpful* is rather limited so that these techniques are usually crowned with the unpleasant term *black box*. As a consequence, the justification of using computational intelligence in finance has yet to be established. In this paper, efforts are devoted to narrowing the gap between theory and practice. We would like to know *why we should believe in computational intelligence without actually seeing it work*.

Our contemplation of this issue starts with a specific type of computational intelligence, namely, *genetic algorithms* (GAs), and a specific type of financial application, namely, *trading strategies*. For a concrete example about the application of genetic algorithms to trading strategies, we refer to Bauer (1994). These restrictions facilitate our work to concretize the properties that can make GAs helpful. We believe that the properties developed here should be easily extended to other techniques in various applications.

* Research support from MasterLink Securities Corporation is gratefully acknowledged.

2 Some Fundamental Concepts and Theorems

A *trading strategy* d is formally defined as a mapping:

$$d : \Omega \to \{0,1\}. \tag{1}$$

In this paper, Ω is assumed to be a collection of *finite-length binary strings.* This simplification can be justified by a *data-preprocessing procedure* which *transforms* the raw data into *binary strings.* Since a different data-preprocessing procedure may generate different Ωs, all theorems presented in this paper are expected to be read with respect to all Ωs, or we can *assume* that there exists a *universal-like* Ω and all theorems are working with that Ω. By assuming the existence of a *Universal* Ω, we avoid the trouble of constructing an effective coding scheme which functions as a universal Turing machine.

The *range* of the mapping d is simplified as a 0-1 action space. In terms of simple *market-timing* strategy, "0" means to "*sell*" and "1" means to "*buy*". Of course, there is no reason to restrict ourselves to such limited choices. "*Just hold and wait to see*", for example, could be another alternative. In any case, the theorem presented in this paper has little to do with the specific action space chosen.

Given the Ω (or the Universal Ω), consider *a collection of trading strategies* $\mathcal{D}$:

$$\mathcal{D} = \{d \mid d : \Omega \to \{0,1\}\} \tag{2}$$

Denote the return earned by following a trading strategy d ($d \in \mathcal{D}$) over the time interval $[s,t]$ by r_s^t. r_s^t can be considered as a mapping:

$$r_s^t : \{0,1\}^{t-s+1} \to \mathcal{R} \tag{3}$$

On the other hand, r_s^t can be considered as a *random variable* mapping the sample space $\mathcal{D}$ to $\mathcal{R}$, and let f_s^t be the histogram of r_s^t or the density function of the random variable r_s^t, then the important property which can help us to answer why genetic algorithms would help in financial data mining is the behavior of r_s^t and f_s^t, and without losing generality, simply r^t ($\equiv r_0^t$) and f^t ($\equiv f_0^t$).

We say that r^t holds a *no-free-lunch property* (briefly, **NFL**) if

$$\lim_{t\to\infty} r^t \xrightarrow{a.s} c_t, \tag{4}$$

where c_t is a time-dependent deterministic variable and "a.s" refers to "*almost surely*". If the **NFL** property holds, then clearly,

$$\lim_{t\to\infty} f^t \xrightarrow{deg} c_t, \tag{5}$$

where "deg" means "*degenerates to*".

The term *no free lunch* is frequently used in economics, and recently it has been adopted by Wolpert and Macready to name their well-celebrated theorem in machine learning (Wolpert and Macready, 1997). Our **NFL** property is named in the same spirit. It simply means that in the long run, all trading strategies perform equally well or equally bad. Alternatively speaking, if a trading strategy d performs exceptionally well during a specific period of time, then it will be compensated by an extremely poor performance during some other time, and vice versa. Furthermore, if we treat c_t in Equation (4) as returns from of a *riskless asset*, then what the **NFL** property says is that, in the long run, all trading strategies can earn no more, or no less for that matter, than what the riskless asset does. This is analogous to the *efficient market hypothesis* or the *no-arbitrage condition* in finance. With these understandings, we can start to introduce our first theorem about GAs in financial data mining.

Theorem 1. *If the NFL property holds, then the GA used in a* non-adaptive manner *is not expected to work in the long run.*

Note that *Theorem 1* only states that the non-adaptive GA can fail to work in the long run. This includes all versions of GAs as long as they are used in a non-adaptive manner. Therefore, the choice of styles and control parameters running GAs *does not matter* as predicted by Theorem 1. However, *Theorem 1* itself is not sufficient to claim whether or not *adaptive GAs*, i.e., the trading strategies trained *recursively* by GAs, will work. Also, it is mute on the short-run behavior of any styles of GAs.

If *the NFL property fails to hold*, then the limit behavior of r^t, if existent, is a random variable. To see whether genetic algorithms would be helpful when $r^\infty (\equiv \lim_{t\to\infty} r^t)$ is a random variable, let us define a term called the *well-ordered property.*

Definition 2. $\mathcal{D}$ is a *well-ordered set* if for all pairs of (d_1, d_2) in the product space $\mathcal{D} \times \mathcal{D}$ and for all t large enough, either

$$r_0^t(d_1) \geq r_0^t(d_2) \tag{6}$$

or

$$r_0^t(d_1) < r_0^t(d_2), \tag{7}$$

but not both.

If $\mathcal{D}$ is a well-ordered set, then it is equivalent to saying that the *landscape* (*fitness function*) determined by r_0^t is *essentially static*. In particular, all local and global optima are time-invariant when t is large enough. In this case, by the convergence theorem of GAs given by Rudolph (1994, 1996), we have the following theorem to support the use of GAs in financial data mining.

Theorem 3. *If $\mathcal{D}$ is a* **well-ordered set**, *then the non-adaptive GAs would work provided that*

1. *The size of the training set is appropriately chosen (i.e., t is large enough).*
2. *The version of GAs employed can guarantee the convergence to the global optimum.*

Theorem 3 provides the conditions under which we can be very positive about the use of genetic algorithms in financial data mining. Of the three conditions in Theorem 3, the easiest one to satisfy is the condition on sample size. As a matter of fact, one of the advantages of using financial data is its chunk size, e.g., the tick by tick data. It is also not difficult to satisfy the condition of the global convergence. For example, one can add the *elitism* operator to their GAs, though this operator does not guarantee a reasonable speed of convergence. The most difficult condition to meet is the well-ordered property. To see how strong this condition is, let us introduce the notation $Corr(r_0^t, r_0^{t+\Delta})$ to be the correlation coefficient of the random variables r_0^t and $r_0^{t+\Delta}$. An alternative definition for the well-ordered property is given below.

Definition 4. $\mathcal{D}$ is a *well-ordered set* if there exists a T such that for all $t > T$,

$$Corr(r_0^t, r_0^{t+\Delta}) = 1, \quad \forall \Delta \in \aleph^+. \tag{8}$$

In other words, a well-ordered set implies a *perfect* correlation. Under a perfect correlation, we can surely say that the global optimum d^* found from r_0^t is also the global optimum of the fitness function $r_0^{t+\Delta}$ (for $t > T$). The perfect correlation may be too strong. For example, if

$$Corr(r_0^t, r_0^{t+\Delta}) > 0.9, \quad \forall \Delta \;\; and \;\; t > T, \tag{9}$$

we can form an *equally weighted portfolio* over the top k strategies founded from r_0^t. While not all of these k strategies are expected to perform well on an arbitrarily extended landscape $r_0^{t+\Delta}$, due to the high correlation coefficient, we can still expect that the portfolio can perform reasonably well. We, therefore, provide the following modified version of Theorem 3.

Theorem 5. *The non-adaptive GAs would work provided that for t large enough and for all Δ in $\aleph^+$,*

$$Corr(r_0^t, r_0^{t+\Delta}) \;\; is \;\; high, \tag{10}$$

and the higher the better.

Theorems 1, 3 and 5 provide us with the theoretical foundations for the *non-adaptive genetic algorithms* in financial data mining. Theorem 1 gives the reason why non-adaptive GAs may fail to work, while Theorems 3 and 5 explain explicitly the conditions under which non-adaptive GAs would help. Alternatively, the predictable failure of the non-adaptive GAs can be summarized by Theorem 6.

Theorem 6. *In the long run, all non-adaptive GAs are doomed to fail provided that*

- $r^\infty \longrightarrow c_t$ *or*
- *Statistically speaking, $\mathcal{D}$ is not well-ordered enough.*

Theorem 6 delineates the domain where non-adaptive GAs can fail. Within this domain, the issue left is then to decide whether or not there exists a subdomain on which *adaptive* GAs can work. The main additional function of the adaptive GA lies in its *retraining design*. The *retraining schedule* can be determined in an *active* (non-supervising) manner or *passive* (supervising) manner. We do not intend to further differentiate the many variants of each manner. The interested reader can find some examples in Chen and Lin (1997). Here, we would like to restrict our attention to the *general condition* under which these types of GAs tend to work. For convenience, we shall term this condition *temporal correlation* defined as follows.

Definition 7. Two fitness function determined by $r_t^{t+\Delta_1}$ and $r_s^{s+\Delta_2}$ is said to be *connected* if either

$$s = t + \Delta_1, \tag{11}$$

or

$$t = s + \Delta_2, \tag{12}$$

where $\Delta_1, \Delta_2 \in \aleph^+$.

The *temporal correlation* is then the statistical correlation of any fitness functions which are connected. More precisely, it is a function of three parameters:

$$Corr(t, \Delta_1, \Delta_2) \equiv Corr(r_t^{t+\Delta_1}, r_{t+\Delta_1}^{t+\Delta_1+\Delta_2}) \tag{13}$$

To facilitate our discussion, we may need a few more notations, first, the set Γ.

$$\Gamma \equiv \{(t, \Delta_1, \Delta_2) \mid t \in [0, \infty), \Delta_1, \Delta_2 \in \aleph^+\}. \tag{14}$$

Second, let

$$\Gamma_\rho \equiv \{(t, \Delta_1, \Delta_2) \mid (t, \Delta_1, \Delta_2) \in \Gamma, \mid Corr(t, \Delta_1, \Delta_2) \mid \leq \rho\} \tag{15}$$

Now, we are ready to give the first theorem on adaptive GAs. First, let us consider an extreme case.

Theorem 8 (The Random-Walk Theorem). *In general, adaptive GAs would not work if, for all triples* $(t, \Delta_1, \Delta_2) \in \Gamma$*,*

$$Corr(t, \Delta_1, \Delta_2) = 0 \tag{16}$$

or simply,

$$\Gamma = \Gamma_0 \tag{17}$$

Theorem 8 basically says that, no matter how we divide our data into subsamples, what has been learned from the previous training set can no longer be useful to the testing set. In other words, the training sample is always uncorrected to the testing sample. In finance, Theorem 5 corresponds the *random walk hypothesis* on which the *efficient market hypothesis* is built. In fact, Theorem 5 can be considered as the strongest version that can support those finance people who believe in the **EMH** and claim that machine learning would play no role in getting excess returns. Given the Random Walk Theorem as the extreme and the most adverse circumstances for adaptive GAs, any deviations from it can only be better. For example, consider the following deviation.

Theorem 9. *The adaptive GAs would work if there exists a pair* (Δ_1, Δ_2) *such that, for all* $t \in [0, \infty)$, $Corr(t, \Delta_1, \Delta_2)$ *is high enough, and the higher the better.*

If Theorem 9 holds, then the main job left for machine learning is to search for this pair. The search for Δ_1 is equivalent to the determination of the *size of the training sample*, and the search for Δ_2 is equivalent to *the design of the retraining plan*. In the literature, it is quite often that these two numbers are arbitrarily chosen or chosen based on some rules of thumb. If the condition embodied in Theorem 9 holds, then there should be some systematic ways designed for the determination of the pair (Δ_1, Δ_2).

While Theorem 9 is obviously a large deviation of the efficient market hypothesis, the money left on the ground may not be so easy to pick up. As opposed to Theorem 9, Theorem 10 provides a closer description of the more sophisticated situation.

Theorem 10. *The adaptive GAs would work if, for all* $t \in [0, \infty)$, *there exists a pair* (Δ_1, Δ_2) *depending on* t *such that* $Corr(t, \Delta_1(t), \Delta_2(t))$ *is high enough, and the higher the better.*

In Theorem 10, while the pairs (Δ_1, Δ_2) exists for all t, they are not fixed. Different t may imply different Δ_1 and Δ_2. Therefore, any adaptive GAs with a fixed size of training sample and a fixed updating frequency may fail.

3 Concluding Remarks

By asking *why GAs may or may not work in principle*, this paper distinguish itself from the existing empirical literature of financial data mining in that we do not evaluate the performance of GAs just by their luck. Instead, we try to provide some properties which can help us explain why some kinds of GAs can performs well, while others cannot, in particular, the properties about the success or failure of non-adaptive GAs and adaptive GAs. In the next paper, we will test some of these properties and, based on that, to gauge the performance of different styles of GAs.

References

1. Bauer, R. (1994), *Genetic Algorithms and Investment Strategies*, Wiley.
2. Chen, S.-H. and W.-Y. Lin (1997), "Financial Data Mining with Adaptive Genetic Algorithms," in T. Philip (ed.), *Proceedings of the ISCA 10th International Conference*, pp. 154-159.
3. Leinweber, D. J. and R. D. Arnott (1995), "Quantitative and Computational Innovation in Investment Management," *Journal of Portfolio Management*, Vol. 21, Issue 2, pp. 8-15.
4. Rudolph, G. (1994), "Convergence Properties of Canonical Genetic Algorithms," *IEEE Transaction on Neural Networks,* Vol. 5, No. 1, pp. 96-101.

5. Rudolph, G. (1996), "Convergence of Evolutionary Algorithms in General Search Space," *Proceedings of 1996 IEEE International Conference on Evolutionary Computation*, pp.50-54.
6. Wolpert, D. H. and W. G. Macready (1997), "No Free Lunch Theorems for Optimization," *IEEE Transactions on Evolutionary Computation*, Vol. 1, No. 1, pp. 67-82.