Modelling Financial Time Series using Grammatical Evolution

Kamal Adamu*
Center for Computational Finance and Economic Agents
University of Essex
Colchester, CO4 3SQ
ksadam@essex.ac.uk

Steve Phelps
Center for Computational Finance and Economic Agents
University of Essex
Colchester, CO4 3SQ
sphelps@essex.ac.uk

Abstract

The traditional models of price, and its statistical signatures are often based on limiting assumptions, such as linearity. Moreover, the model developer is faced with the model selection problem, and model uncertainty. In this paper we introduce a method based on *Grammatical Evolution (GE)* to evolve models for predicting financial returns, and we examine the profitability of these models. Our empirical analysis demonstrates that for some securities our method is able to produce models of return that are lead to more profitable trading compared with an Autoregressive model picked using Aikake Information Criterion (AIC), under the assumption of frictionless markets.

1 Introduction

The traditional models of price, and its statistical signatures are often based on limiting assumptions, such as linearity. Moreover, the model developer is faced with the model selection problem, and model uncertainty. In a previous paper [1], Grammatical Evolution (GE) was shown to have consistently produced algo traders that trade based on technical trading rules, that are significantly better than a strategy with zero intelligence.

Grammatical Evolution (GE) is an Evolutionary Automatic Programming (EAP) technique for automatically generating symbolic solutions in an arbitrary language [2, 3]. Programs produced using GE are human readable, allowing for a better understanding of the problem domain. In addition, the separation between search space and solution space brings with it many advantages such as neutral evolution, and easy deployment of the optimization tool [4].

This paper employs GE in evolving models for predicting financial returns, and examines the profitability of these models under the assumption of frictionless markets. Results from experiments carried out show that, under the assumption of frictionless markets, GE is able to produce profitable models of return for trading the stock of Invesco in high-frequency that are significantly better an AR model chosen using Aikake Information Criteria (AIC). In addition we perform a control experiment in which we demonstrate that our evolved models yield superior trading performance compared with a strategy that makes blind trade decisions based on the flip of a coin.

The rest of the article is organized as follows, section (2) gives an overview of the data processing employed, and section (3) explains the framework used in developing the models and the fitness function used. The main findings of the paper are reported in section (4), and the article ends with a conclusion and recommendation for future work in section (5).

2 Data

This section gives an insight into the data preprocessing employed in this paper, and explores the statistical properties of the data. High frequency tick data for GlaxoSmithKline, Invesco, and HSBC was filtered, sampled at five minutely intervals, and interpolated at one minutely intervals using cubic spline interpolation. The period of study is the period between 1 March and 30th March 2007. Moreover, results from a Ljung Box test suggest we should reject the null hypothesis, that there is autocorrelation within the return series of the stocks considered.

3 Experimental Framework

In our framework, agents make intra-day single-period trading decisions based on the same decision rule that is prespecified as follows:

if $(r^e_{(t+1)}) > k$ **then**
  Go Long
else
  Go Short
end if

where $r^e_{(t+1)}$ denotes the predicted return at time $t+1$, and k is a free parameter. This is the same as saying if the predicted return is above a required amount, k (set to zero in this paper), then go long, otherwise, go short. The key problem that we address in this paper is deriving a model for prediction the future returns r^e_t from historical high-frequency data.

The objective function used in this paper is the Sharpe ratio and it was calculated using equation (1). T in equation (1) is the length of the trading period, and σ is the standard deviation of the return obtained by a solution in a trading period. The risk free rate, r_f, has been omitted from the Sharpe ratio because it is assumed that at high frequency r_f is negligible. r_{raw}, is the raw return, or return on underlying asset being traded, and I_t is the indicator of the position taken at time t.

$$f = \frac{\frac{1}{T}\sum_{t=1}^{T} r_{raw}(t) \times I_{t-1}}{\sigma} \tag{1}$$

The assumption made is, a position is opened and closed at every time interval. In other words, if at time t=1 a long position is taken, the position is closed at t=2, and another position is opened at t=2. I_t is positive for a long position, and negative for a short position as illustrated in equation (2).

$$I_t = \begin{cases} +1 & \text{if Long} \\ -1 & \text{if Short} \end{cases} \tag{2}$$

Solutions that associate periods of positive return with going long, and periods of negative return with going short will have a relatively high Sharpe ratio, and vice versa [5]. Furthermore, the following assumptions are implicit in the fitness evaluation:

- the solution being evaluated has an infinite amount of fund available, and there is no restriction on short selling;
- solutions trade at one minutely intervals, opening a position at interval t, closing it at time $t+1$, and opening another position at that interval $(t+1)$; and
- there are no elements of market friction such as, transaction cost, slippage, and market impact is negligible (since we trade only a single unit per period).

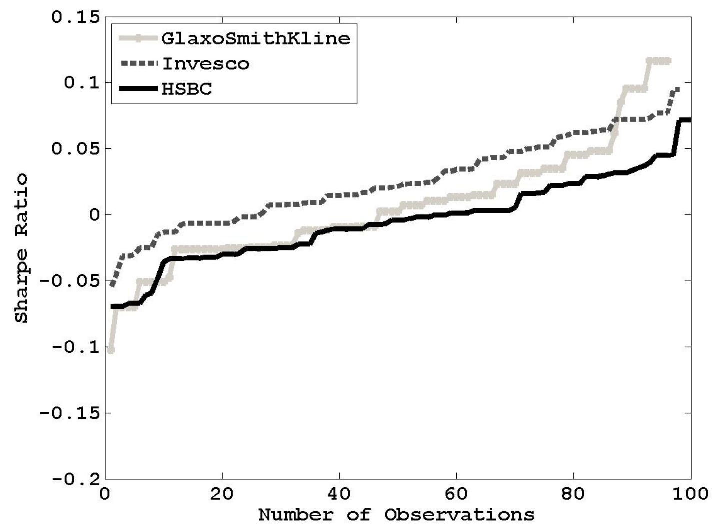

Figure 1: The out-of-sample distribution of elitists for GlaxoSmithKline, Invesco, and HSBC

4 Results and Discussion

Two benchmark strategies are used in this paper, MC (Monte Carlo) [5], and an AR (Autoregressive) model picked using Aikake Information Criteria (AIC) [6]. MC is a zero intelligence strategy that has an equal likelihood of going long or short at any time interval t. The intuition behind benchmarking GE against MC is to enable us conjecture if GE is able to produce solutions that are significantly better than a strategy with zero intelligence. The AR model is used to conjecture if the models produced by GE obtain better profit than a traditional econometric model (AR model in this case).

The random initialization procedure in GE implies that independent initializations would produce a different set of initial solutions, however, it is expected that evolutionary operators would drive the search towards the optimum. Effectively, a good heuristic should consistently produce an optimum solution with relatively little dispersion between different initializations. 100 independent elitist models were produced using GE, for the stocks of GlaxoSmithkline, Invesco, and HSBC. 100 independent experiments were equally performed for MC, placing random trades at each interval, t, with the trades having equal likelihood of being long or short. The mean Sharpe ratio of the elitists is compared to that of MC using a difference of means t-statistics. For GE, each experiment involved training the system offline using a population size of 100, and 200 generations.

4.1 Results

Figure 1 depicts the distribution of out-of-sample Sharpe ratio for the models produced using GE, for trading the stocks of GlaxoSmithKline, Invesco, and HSBC. The distributions were developed by performing 100 experiments of 200 generations each, and for each experiment i, $i \in \{1, 2, 3, ...100\}$, recording the out-of-sample Sharpe ratio of the elitist at the end of 200 generations. The x axis in Figure 1 represents the number of experiments while the y axis represents the Sharpe ratio. The t-statistics for GE, and MC is tabulated in Table 1. The t-statistics in Table 1 is for the null hypothesis, that the mean of MC is the same as the mean of the GE.

4.2 Discussion

Analyzing Figure 1, it can be seen, that the top 30 solutions produced for GlaxoSmithKline, Invesco, and HSBC have Sharpe ratio greater than zero implying the distribution of the best solutions produced is good. The results from a difference of mean t-test in Table 1, however, shows at 95% confidence level, on average, only the solutions produced for Invesco are statistically better than

Table 1: Difference of mean t-statistics for Sharpe ratio of the different methods across different stocks

Stock	t-statistic	Result
GlaxoSmithKline	+1.2405	Accept
Invesco	+4.6558	Reject
HSBC	-1.1810	Accept

flipping a coin. Moreover, Figure 1 shows, that close to 50% of the solutions produced for GlaxoSmithKline, and HSBC have a Sharpe ratio less that zero. This could be interpreted to mean, giving the experimental setting, there is 50% chance of getting a poor solution for GlaxoSmithKline, and HSBC. As expected, only 22% of the solutions produced for Invesco. Furthermore, a comparison between the mean performance of the solutions produced for the stocks using GE and an AR model picked using AIC shows, that the solutions produced for Invesco, are better than an AR model.

5 Conclusion

This paper has shown, that under the assumption of frictionless markets and given other system settings, GE produced profitable models, that are on average statistically better than a strategy with zero intelligence and an AR model, for trading the stock of Invesco in high frequency, but not for GlaxoSmithKline, or HSBC. Moreover, the distribution of the solutions produced for all stocks is positively skewed, implying the distribution of the best solutions produced is good. Implicit in this paper however, is the assumption that the investor is operating in frictionless markets under no constraints.

Acknowledgements

The authors are grateful for the contribution of Tikesh Ramtohoul[1] in the data processing carried out, and the guidance of Dietmar Maringer[2] in the initial stages of this work.

References

[1] K. Adamu, Y.-S. Chu, J. Jiang, A. Kablan, and T. Ramtohoul, “Using evolutionary algorithms for intraday trading using ccfea sets simulator (single stock case).” 2008.

[2] M. ONeill and C. Ryan, *Grammatical Evolution, Evolutionary Automatic Programming in an Arbitrary Language*. Kluwer Academic Publishers, 2003.

[3] M. ONeill and A. Brabazon, *Biologically Inspired Algorithms for Financial Modeling*. McGraw-Hill Publishing Company, 2006.

[4] M. Dacorogna, R. Gencay, and U. Muller, *An Introductory to High Frequency Finance*. Academic Press, 2001.

[5] T. Masters, “Monte carlo evaluation of trading systems.” 2006.

[6] R. Tsay, *Analysis of Financial Time Series,Second Edition*. John Wiley, 2004.

[1] Tikesh Ramtohul is a teaching assistant at the University of Basel, Switzerland
[2] Dietmar Maringer is a professor at the University of Basel, Switzerland