Modeling Financial time Series using Grammatical Swarm

Kamal Adamu, and Steve Phelps

***Abstract*—In this paper we employ a methodology based on *Grammatical Swarm* (GS) in producing models of financial returns. The models produced are used in trading single stocks in high frequency. The performance of the models produced using GS is compared to the performance of models produced using *Grammatical evolution (GE)* and models produced using GS outperform models produced using GE. Further analysis shows that the models produced using GS are better than a random strategy and an AR model picked using Aikake Information Criteria (AIC)**

I. 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] we introduce a methodology based on GE for producing models of financial return and the profitability of the models developed was tested using historical high frequency data. This paper follows the same framework as [1] with the exception that Grammatical Swarm GS is used in producing models of financial return. The performance of the best solutions produced using GS is compared to the performance of the solutions produced using GE. The results show that GS is a capable of producing models of financial return and GS is more suitable for the modeling problem than GE.

The rest of the paper is organised as follows. Section II gives a description of the framework employed in this paper. Results are presented in section IV with a discussion on the results. The paper ends with concluding remarks in section V.

II. 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 predicting the future returns r^e_t from historical high-frequency data.

Kamal Adamu is with the Center for Computational Finance and Economic Agents, University of Essex, Colchester, United Kingdom, email: ksadam@essex.ac.uk

Steve Phelps is with the Center for Computational Finance and Economic Agents, University of Essex, Colchester, United Kingdom, email: sphelps@essex.ac.uk

A. Fitness Evaluation

The fitness of the models developed was measured using the Sharpe ratio which 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 the underlying asset being traded, and I_t is the indicator of the position taken at time t.

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

A position is opened and closed at every time interval t. 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 [2]. 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;
- Only a single share (unit) is traded at each time interval t
- There is no market friction i.e transaction cost, slippage, and market impact (since we trade only a single unit per period market impact can be assumed to be negligible).

In a nutshell, we are concerned with finding models that maximise θ under the assumption of frictionless markets (see Equation 3)

$$\max \theta \tag{3}$$

B. Grammatical Swarm

In this section we give a brief description of our implementation of the Grammatical Swarm of [3]. The GS uses the particle swarm as its search engine. Solutions are represented by vectors (particles) that fly through the search space, in a swarm, looking for promising areas of the search space. Each solution is attracted to previous local best areas they have visited (Personal best), and the the best solution within the entire population (Global best)[4].

Initially, a random population of integer strings x_j, $j \in \{1, 2, 3, ...P\}$, and P is the population size is initialised. A constraint is placed on x_j such that $x_j \in I_0^{+255}$. I_0^{+255} is a set of integers between 0 and 255.

Solutions are updated using Equation 4, and Equation 5. $x_j(t)$, $t \in \{1, 2, 3, ...\}$, is the current position of the vector $x_j(t)$, and $v_j(t)$, $j \in \{1, 2, 3, ...P\}$, is the current velocity of the vector $x_j(t)$. C_1, C_2, and δ are constants. $z_1, z_2 \in N(0, 1)$[1]. xP_j is the current best solution attained by the vector x_j (personal best), and xG is the global best solution within the entire population. The particles, x_j fly through the search space attracted to their personal best xP_j and the global best solution xG.

$$x_j(t) = x_j(t) + v_j(t) \tag{4}$$

$$v_j(t) = \delta v_j(t-1) + C_1 z_1(x_j(t-1) - xP_j) + C_2 z_2(x_j(t-1) - xG) \tag{5}$$

Solutions are mapped from integer strings to a human readable (executable) solutions using a set of production rules (Grammar). The grammar used in developing, r_t^e, is given in Table I.

The intuition behind including sinusoidal functions within the grammar is that, by principle of fourier analysis, a wave can be decomposed into a mixture of sine, and cosine waves[5] and it will be interesting to see if GS is able to come up with the fourier series representation of r(t) (r(t) can be assumed to be a wave). Moreover, returns are cyclical and including sinusoidal functions may capture the cyclical aspect of returns.

III. 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 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.

IV. Results

30 independent experiments were run using the framework described in section II using GS. Each experiment was run using the parameter settings in Tableau IV.

[1]N(0,1) denotes a normal distribution with mean zero and unit variance

TABLE I
PRODUCTION RULES FOR GS, AND GE

P		
A	$<$ expr $>$	$<$ expr $><$ op $><$ expr $>$ $\mid$ ($<$ expr $><$ op $><$ expr $>$)$\mid$ $<$ mo $><$ op $><$ expr $>$ $<$ mo $><$ op $>$ ($<$ expr $>$) $<$ preop $>$ ($<$ expr $>$)$\mid$ $<$ coeff $>$ ($<$ expr $>$)$\mid$ $<$ var $>$
B	$<$ preop $>$	$\sin \mid \cos \mid \tan$
C	$<$ var $>$	$r(t-$ $<$ Win $>$)
D	$<$ mo $>$	MovAvg($<$ Win $>$) Std($<$ Win $>$)
E	$<$ Win $>$	$<$ integer $><$ integer $>$
F	$<$ op $>$	$+ \mid - \mid \times \mid \div$
G	$<$ coeff $>$	$<$ integer $>$ $\mid$ $<$ float $>$
H	$<$ integer $>$	$1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$
I	$<$ float $>$	$<$ integer $>$. $<$ integer $>$

N	Population size	$N = 200$
δ		$\delta = 0.4$
C_1		$C_1 = 0.2$
C_2		$C_2 = 0.2$
UB	Maximum codon value	UB=256
FE	Functional Evaluations	FE=40000

TABLE II
PARAMETER SETTINGS FOR GRAMMATICAL SWARM (GS)

The average of the mean fitness of the population of solutions produced using GS per generation is shown in Figure 1. Unlike in GE where one expects to see a non diminishing improvement in mean fitness over time the mean fitness in GS is choppy and this is so because GS in mainly an exploratory method and there is no strong tendency for solutions to converge to one area of the solution space. This is behavior is also seen in the square root of the average variance of fitness per generation shown in Figure 2

Figure 4 depicts the average fitness of the elitists per generation from the 30 experiments carried out using GS compared to the average fitness of the elitist produced using GE per generation. As expected the average fitness of the elitists improves over generations as the swarm explores the search space. In addition, The average fitness of the elitists produced using GS per generation dominates the average fitness of the elitists produced using GE per generation. This means on average the quality of the best solutions using GS is better than the quality of the best solutions using GE

There is the possibility that the performance of the solutions produced by GS is as a result of sheer luck. In other to see if the solutions produced by GS outperform random strategies (chance), 30 random strategies that make trade decisions based on the flip of a coin were produced. One set of strategies FC (Fair Coin) have an equal likelihood of going long or short. The second set of strategies LBC have a higher likelihood of going long than going short, and the third set of random strategies SBC are short biased. Figure 3 shows a boxplot comparison between the best solutions produced by GS and the random strategies. Table IV shows the difference of means

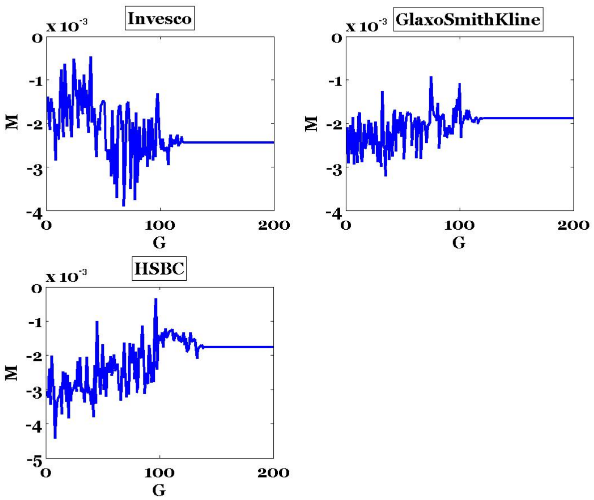

Fig. 1. Average mean fitness M per generation G for grammatical swarm (GS). The period of inactivity is the period when the termination criteria has been met.

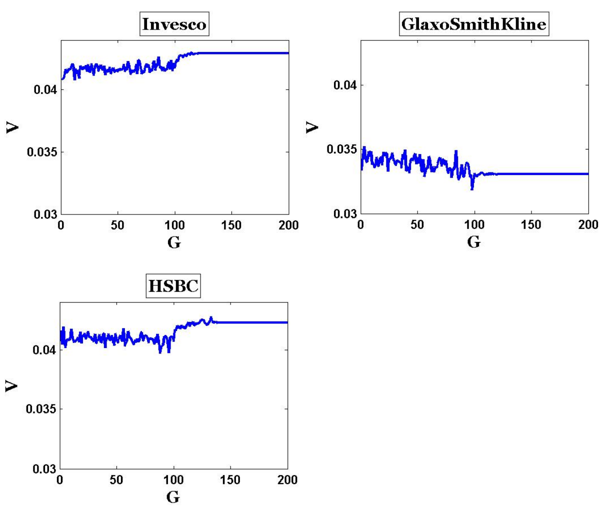

Fig. 2. Square root of average variance of fitness V per generation G for grammatical swarm (GS). The period of inactivity is the period when the termination criteria has been met.

t-statistics for the null hypothesis that the mean of the solutions are equal. The results in Figure 3 and Table IV suggests that the solutions produced by GS outperform random strategies.

In order to see if GS is a better fit for the modeling application than GE, the best solutions produced using GS are compared to the best solutions produced using GE. Figure 5 shows a boxplot comparison between the best solutions produced using GS and the best solutions produced using GE and Table IV shows difference of means t-statistics for the null hypothesis that, the mean fitness of the best solutions produced using GS is the same as the mean of the best solutions produced using GE. The results in Table IV and Figure 5 suggest that GS is a better fit the modeling application than GE.

Figure 6 shows a comparison between the the cumulative return of a GS elitist to the cumulative return of an AR model picked using Aikake Information Criteria (AIC), and Buy and

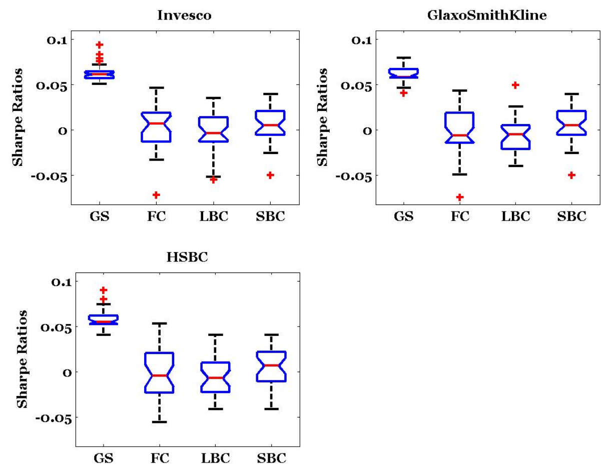

Fig. 3. A comparison between the out of sample Sharpe ratios of solution produced using GS to out of sample Sharpe ratios achieved by a fair coin (FC) a long biased coin (LBC) and a short biased coin (SBC)

Stock	FC	LBC	SBX
Invesco	-13.35	-14.35	-13.92
	(0.00)	(0.00)	(0.00)
GlaxoSmithKline	-12.56	-16.28	-13.51
	(0.00)	(0.00)	(0.00)
HSBC	-11.20	-14.54	-11.94
	(0.00)	(0.00)	(0.00)

TABLE III
DIFFERENCE OF MEANS T STATISTICS FOR THE NULL HYPOTHESIS THAT THE MEAN OF THE OF THE SHARPE RATIOS OF SOLUTIONS PRODUCED USING GS ARE SIGNIFICANTLY BETTER THAN THE MEAN SHARPE RATIO OF A FAIR COIN (FC), A LONG BIASED COIN (LBC), AND A SHORT BIASED COIN (SBC)

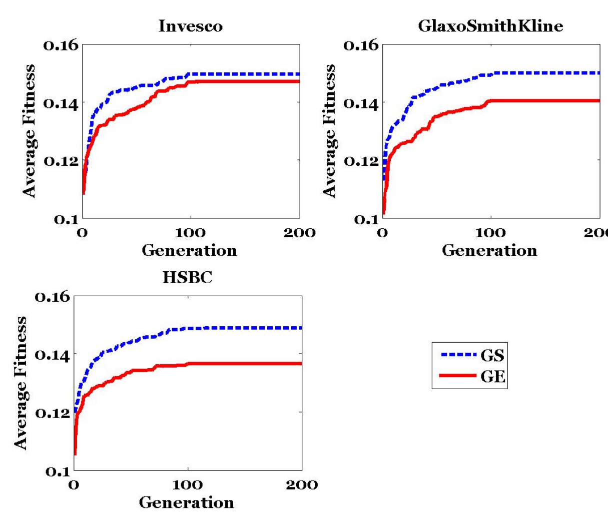

Fig. 4. A Comparison between the average insample fitness per generation of 30 elitists produced using GE to the average fitness per generation of 30 elitists produced using GS

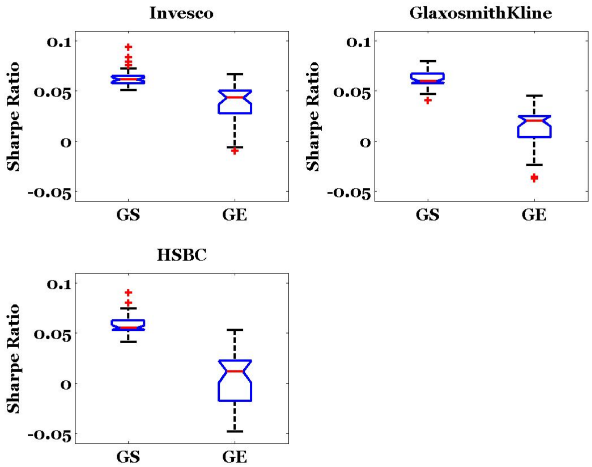

Fig. 5. A comparison between the best solutions produced using GS to the best solutions produced using GE

Stock	T-Statistics	P-value	Result
Invesco	-6.32	0.00	Reject
GlaxoSmithKline	-11.57	0.00	Reject
HSBC	-10.90	0.00	Reject

TABLE IV

DIFFERENCE OF MEAN T-STATISTICS FOR THE NULL HYPOTHESIS THAT THE MEAN SHARPE RATIO OF THE BEST SOLUTIONS PRODUCED BY GE FROM 30 INDEPENDENT RUNS ARE SIGNIFICANTLY BETTER THAT OF GS

Hold Strategy. The GS elitist outperforms an AR model picked using Aikake Information Criteria, and a Buy and Hold. It is clear that the best solution produced using GS was able to predict some large returns and this in turn lead to its high fitness.

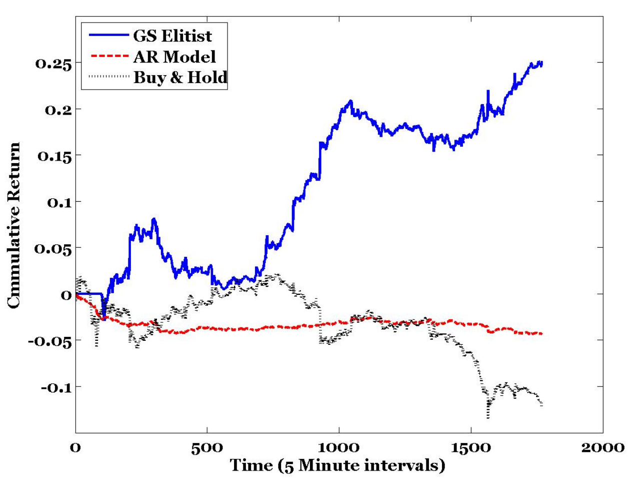

Fig. 6. A comparison of the cumulative return of a GS elitist to the cumulative return of an AR model picked using Aikake Information Criteria (AIC), and Buy and Hold Strategy

The solution that produced the result in Figure 6 is given in Equation 6.

$$r_t^e = \sin(\sin(\frac{1}{T}\sum_{i=1}^{T=36} r_{t-i}) - \sigma_{t-1}^{t-14}) \tag{6}$$

σ_{t-1}^{t-14} in Equation 6 is the standard deviation of $r_{t-1} tor_{t-14}$. The reason behind the superior performance of the model in Equation 6 is unclear and subject of future work

V. CONCLUSION

In this paper, we have demonstrated that GS is capable of producing profitable models of financial return for high frequency trading. Results from controlled experiments carried out suggest that the performance of the best solutions produced using GE is not as a result of sheer chance. Moreover, the results show that GS is a better fit for the modeling application than GE. In addition, the results suggests the best solution produced using GS outperforms an AR model picked using Aikake Information Criteria (AIC). The best solution found is a solution that is able some predictions of large magnitude right and not a solution that gets it right all the time. The reason behind the superior performance of the best solution of GS remains a subject for future work.

REFERENCES

[1] K. Adamu and S. Phelps, “Modeling financial time series using grammatical evolution,” July 2009.

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

[3] M. ONeill and A. Brabazon, “Grammatical swarm: The generation of programs by social programming,” *Natrual Computing, springer*, vol. 5, November 2006.

[4] D. Corne, M. Dorigo, and F. Glover, *New Ideas in Optimization*. McGraw-Hill Publishing Company, 1999.

[5] T. Korner, *Fourier Analysis*. Cambridge University Press, 1988.