Genetic Algorithms: Genesis of Stock Evaluation

Rama Prasad Kanungo[*]

[*]Rama Kanungo is the Research Co-ordinator of the Asian Accounting, Finance and Business Research Unit, Cardiff Business School, Colum Drive, Cardiff, CF10 3EU, UK.
Email:kanungomba@hotmail.com, Tel.: + 44 (0) 29 20206037

Abstract

The uncertainty of predicting stock prices emanates pre-eminent concerns around the functionality of the stock market. The possibility of utilising Genetic Algorithms to forecast the momentum of stock price has been previously explored by many optimisation models that have subsequently addressed much of the scepticism. In this paper the author proposes a methodology based on Genetic Algorithms and individual data maximum likelihood estimation using *logit* model arguing that forecasting discrepancy can be rationalised by combined approximation of both the approaches. Thus this paper offers a methodological overture to further investigate the anomalies surrounding stock market. In the main, this paper attempts to provide a temporal dimension of the methods transposed on recurrent series of data over a fixed window conjecture.

Introduction

Functional optimisation is the key underlying rationale of Genetic Logarithms. Irrespective of controlled variation Genetic algorithms eliminate uncertainty and imprecise momentum of any unfit system and derive representative degree of correctness. Genetic Algorithms were espoused by Holland (1975) during 70s envisaging the conceptual framework of Darwinian survival of fittest strategy. Genetic Algorithms from herein, referred as GAs throughout the text. The application of GAs in differentiating optimal value of multi-dimensional functions has been received high credence in evolutionary algorithms (Baricelli, 1962; Baker, 1985; Bramlette, 1991 and Altenberg, 1994).

GAs, Probability Density Function and Individual Data Maximum Likelihood Estimation

Essentially, complex multi-parameter functions exhibit threshold maxima and minima, which GAs represents in terms bit strings in real numbers. for example the value attributed at that point X(0 Π) for a simple probability linear function f(y)= y+ X(0 Π) can be evaluated either at the threshold minima or maxima. The fitness of a string is the function value at that point X (0 Π) (Riolo, 1992). The process is very identical to distribution of a function of a random variable.

If y is derived from x and the function represents linear probability distribution, the expression can be represented as the probability of that Y=y(x) equals the probability that X=x; i.e. when several values of y, then probability of Y is the sum of corresponding probabilities for x.

Whereas, the random variable is a discreet transform of the variable y, all the mean value assumes respective interval, such as;

Prob ($Y=\mu_1$) = P ($-\infty < X \leq a$),

Prob ($Y=\mu_2$) = P ($a \leq X \leq b$),

Prob ($Y=\mu_3$) = P ($b \leq X \leq c$)

and the probability distribution continues up until n th term.

If x is a continuous random variable with probability density function $f_x(x)$ and y =g(x) is a continuous monotonic function of x, the density of y is obtained by using the change of variable to find combine density function of y.

Here

$$\textbf{Prob } (y \leq b) = \int_{-\infty}^{b} f\,(g^{-1}(y))\ \left| g^{-1'}(y) \right| dy$$

rearranging it we can write,

$$\textbf{Prob } (y \leq b) = \int_{-\infty}^{b} f_y(y)\, dy$$

The term $g^{-1'}(y)$ is the Jacobian of the transformation of x to y. Customarily the Jacobian is non-zero to assume non-zero value for y. The probability density of f(y) within the interval of discreet random variable reflects that GAs can be used in a same manner to identify any sequence following selection, crossover and mutation process, starting with a randomly generated population of n l-bit chromosomes, calculating the fitness of f(y) of each chromosome y in population and repeating that until n offspring have been created. Here the probability of selection becoming the increasing functions of dimensional fitness used in probability density function of f(y). Now with crossover probability or crossover rate, i.e., P_c we can continue crossover to generate two forms of offsprings, whereas as crossovers do not produce identical patterns of their respective parents. At this point mutating the offsprings at each locus with subset probability of P_m and reiterating the process with the new chromosomes in the new

population an optimal fitness value can be obtained. This process at the end results a highly fit chromosome giving the best expected value of the y.

GAs are highly effective to identify signals and eliminate noisy data set, particularly over a long period and lagged time series where unstructured nature and hidden relationship in variables are not correctly identified. Furthermore, least square approximation and probability density function do not always provide a robust calculation to establish the maxima threshold of parameters. GAs have unique attributes to address such anomalies. Packard (1990) utilising GAs established a predictive methodology to examine dynamic models. He envisaged that when a series of observation are generated from a dynamic system or process they usually form a set of pairs. The binary observation of such series can be represented as,

$\mathbf{\{(x^{\rightarrow 1}, y^1),\ldots\ldots..(x^{\rightarrow 1}, y^N)\},}$

where $x^{\rightarrow 1} = (x^i_1, x^i_2,\ldots.x^i_N)$ are nth number of independent variables and y^i is a dependent variable having probability of $(1\leq i \leq N)$.

In the uncertain and dynamic stock market share prices fluctuate due to multiple associative parameters. In such instance the independent variable might be the value of particular stock at a given time, i.e., $x^{\rightarrow} = (x(t_1), x(t_2),\ldots..x(t_n))$ whereas the dependent variable $y= x(t_{n+k})$ representing value of stock at some t+k time. This illustrates a single vector representative of each dependent variable to independent variable, but in a dynamic system each dependent variable has their associated independent variables. Observations obtained in a specific space assign sets of conditions for every independent variable. Herein the condition C would be

C= {(£ 25≤ Stock Price of Firm A on day 1)

Λ (£30 ≤ Stock Price of Firm A on day 2≤ £ 32)

Λ (£ 27≤ Stock Price of Firm A on day 3≤ £30)},

where Λ is the logical operator equivalent of text 'AND'. At this point condition C represents a subset when three observed conditions are met with a probability density function $f(x, C) \approx (x^{\rightarrow 1}, 25 \leq Con \leq 32)$. These three conditions can be arranged in a matrix form to observe the determinant value of each probability, suppose the stock price on day 1 is denoted by s_1 and s_2 for day 2 and so on, then the matrix form of each stock price variance and covariance would be;

$$\begin{vmatrix} \mathbf{Var(s_1)} & \mathbf{Cov(s_1\, s_2)} & \mathbf{Cov(s_1, s_3)} \\ \mathbf{Cov(s_1\, s_2))} & \mathbf{Var(s2)} & \mathbf{Cov(s_2\, s_3)} \\ \mathbf{Cov(s_1\, s_3)} & \mathbf{Cov(s_2\, s_3)} & \mathbf{Var(s_3)} \end{vmatrix}$$

Applying Gaussian elimination individual variance of stock price for a specific day can be calculated and each value can be used an approximation of stock price of that day to arrive at an optimal value specific to that date. In the above case searching the space condition that can return the subsets of data points whose dependent variable values would be close to uniform density distribution. Here GAs identify a condition set, where the set were followed by days on which the Firm A's stock rises to approximate high of £ 30. This allows rationalising that if the conditions sustain, the prices will go up. The fitness of each individual condition C is calculated by running all the data points $(x^{\rightarrow}, y)$ in the training set through C and for each $x^{\rightarrow}$ that satisfies C, collecting corresponding y. After that if the y values are close approximation of a certain value V, then condition C is a robust predictor of y. At this point $x^{\rightarrow}$ also satisfies C. Mayer and Packard (1992) proposed an alternate approach to identify regions of predictability in time series generated by Mackey-Glass equation (1977), i.e.,

$$\mathbf{dx/dy=\{ax(t-\tau)/(1+[x(t-\tau)]^{c}\}-bx(t)}$$

Whereas, x (t) is the independent variable at time t and a**,** b, c, τ are constants. If we are assuming different stock prices for different days we can have subsets of each 5 days or subsets of each 10 days for each corresponding y^i value say for example we investigating 24 days of price change, then i= 24.

Furthermore, they fixed the function of the condition as,

$f(C) = -\log_2(\sigma/\sigma_0) - \alpha/N_C$

Where σ is the standard deviation of the set y^i for data point satisfying condition C, σ_0 is the standard deviation of the distribution of y^i over entire data set, N_C is the number of data points satisfying the condition C and α is the constant. Previously we have discussed that a matrix form of variance values can be employed to identify the best predictor approximation by using Gaussian elimination. Furthermore the first term of the above function measures the amount of information in the distribution of y^i for all the data points satisfying conditions C, the second term represents the error variance in distribution. More the number of points satisfying the conditions C, more the reliability of predictor and C is supposed to have higher fitness values. Mayer and Packard followed a sequence to reach at the best predictor approximation, such as; initialised the sample with random set of conditions C, calculating fitness of each subset satisfying conditions C, ranked the measures in terms of higher value, and discarded the lower fitness individuals and replaced them with new conditions C* obtained by applying crossover and mutation to remaining conditions C. They continued the sequence to find the ideal offsprings. In the stock market example this sequence will help to manifest a higher fitness value of the observed price at a given future time t.

Mayer and Packard's best predictor approximation exhibits close similarities with individual data maximum likelihood estimation. In individual data maximum likelihood estimation[*], probability distribution function has been represented as;

Prob[y*> 0] = Prob[β'x +ε> 0]

= Prob[ε>- β'x]

where $y^* = \beta'x + \varepsilon$ for the conditions y= 1 if $y^* > 0$

y= 0 if $y^* \leq 0$

β'x is known as index function , here the assumption of unit variance is normalised and assumption of zero for threshold is likewise if model contains a constant term which we have in this case. Now if the distribution is systematic and normal as well as logistic, then

Prob[y*> 0] = Prob[ε< β'x]

= F (β'x)

The model with probability F ($\beta'x_i$) and each observation is sampled as individual draw from a Bernoulli distribution, i.e., binomial with one draw leads to joint probability or a likelihood function such as;

$$\textbf{Prob}[Y_1=y_1, Y_2=y_2,\ldots\ldots\ldots\ldots Y_n=y_n] = \prod_{y=0}\{1- F (\beta'x_i)\} \prod_{y=1}\{ F (\beta'x_i)\}\ldots\ldots\ldots..(1)$$

Representing the probability function of RHS with L,

we can rewrite,

$$\textbf{L}= \prod_i[F (\beta'x_i)]^{yi} [1- F (\beta'x_i)]^{1- yi} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .(2)$$

This is the likelihood for sample of n observations. In GAs such joint probability function can be compared with conditions C subsets with different offsprings after crossover and repeated mutation. GAs identify sample of n observations that consists of a finite pool of individual data. Thus GAs and estimation with individual data treat

[*] Analytical discussion *on individual data maximum likelihood estimation* in this section has been cited from Green (1990).

each observation as a single parameter with binomial with one draw. In this instance Eq. (2) is denoted as the likelihood for a sample of n observations. Further extending it by obtaining logs; we get

$$\ln L = \sum_i [y_i \ln F(\beta' x_i) + (1-y_i) \ln (1 - F(\beta' x_i))] \quad \text{.....(3)}$$

By converting it into first order condition for maximisation the model became

$$\partial \ln L / \partial \beta = \sum_i \left[y_i f_i / F_i + (1-y_i) \, -f_i/(1-F_i) \right] x_i = 0 \quad \text{.....(4)}$$

The model with probabilities $F(\beta' x_i)$ where subscript i denotes the density of distribution.

As far as a logistic model is concerned we know,

$$\text{Prob}[Y=1] = e^{\beta' x} / 1 + \beta' x = \Lambda(\beta' x) \quad \text{.....(5)}$$

which represents logistic distribution, where Λ represents logistic cumulative distribution function. The density function of a cumulative distribution is represented by

$$d\Lambda[\beta' x] / d(\beta' x) = e^{\beta' x} / (1+e^{\beta' x})^2$$

The above model equals to $\Lambda(\beta' x)\,(1 - \Lambda(\beta' x))$ .....(6)

In the instance of linear probability model the Eq.(4) would become highly nonlinear and requires further linearization as we are concerned about the individual estimation. A simpler approach to address this issue for a logit model is to insert both Eq.(5) and Eq.(6) into Eq.(4). After collapsing all three equations it gives,

$$\partial \ln L / \partial \beta = \sum_i (y_i - \Lambda_i) x_i \quad \text{.....(7)}$$

whereas, x_i contains a constant term. Also in the terms of least square normal equations the term $y_i - \Lambda_i$ can be seen as a residual. However for normal distribution, the log likelihood is denoted by

$$\ln L = \sum_{y=0} \ln(1-\Phi_i) + \sum_{y=1} \ln \Phi_i \quad \text{.....(8)}$$

here Φ_i stands for standard normal density of i th term.

Hence the first order conditions for maximisation of L are,

$$\partial \ln L / \partial \beta = \sum_{y=0} (-\phi_i / 1- \Phi_i) x_i + \sum_{y=1} (\phi_i / \Phi_i) x_i \quad \ldots\ldots (9)$$

Therefore converting individual variables into first order log likelihood we can obtain effect of changes in these variables on the predicted probability.

The author proposes that each individual variable would be converted by utilising Eq.(9) and would be used in GAs as chromosome syntax for any n variables to obtain an optimal solution. Each variable would have bitstrings length N, whereas a **1** at a position **a** means that variable is used in the network denoted by the bitstrings taken as chromosome syntax. The fitness value of for each bitstring B is weighted by training a neural network defining B for a number of times, i.e., mutation and crossover.

During each training time, generated minimal error would be logged on test set. After N times of training the cumulative average of those minimal errors would be used to determine another fitness value. This process obtains higher fitness value for the lower error predictors. Once fitness values have been determined, those fitness values would be assigned and this would create a new sample having best survived offsprings replacing weaker offsprings of the previous sample.

If at least two crossover operators would be used, any finite sample N would yield higher fitness value for each bitstrings. In this case we can select two bitstrings B_1 and B_2 and any two crossover sites at random. The first offspring B_1*essentially inherits the part between the cross sites from B_1 and the other parts from B_2. Similarly the second offspring B_2* inherits the part between cross sites from B_2 and B_1. Similarly the second crossover operator would also select two parents B_1 and B_2 randomly.

Further, a random number x [0, (x/2)] is generated form the crossover site. Now x times a string position p would be selected on a probabilistic assumption where every time the values of B_1 and B_2 at position p would be swapped. In this context only one mutation operator is suffice to generate optimal solution to N sample population. A parent is selected randomly assigning a sting position p so that value at position p is inverted for subsequent mutation operator if any is selected for further extension.

This process can be repeated to achieve accuracy up to 99.9% interval confidence over n finite sample population. To examine the proposed method the author has selected 24 days stock price of a firm A*. Each variable were input into GAs crossover site as bitstrings, following a network training representing each one as formal neurons. Mainly a formal neuron is the basic element in the training network, represented by n-dimensional vector $[x_1,\ldots.x_{24}]^T$ with a constant component $x_0=1$. The weighted sum of neurons is,

$$\mathbf{w^T x = w_0 + S1 \leq i \leq n\ w_i x_i},$$

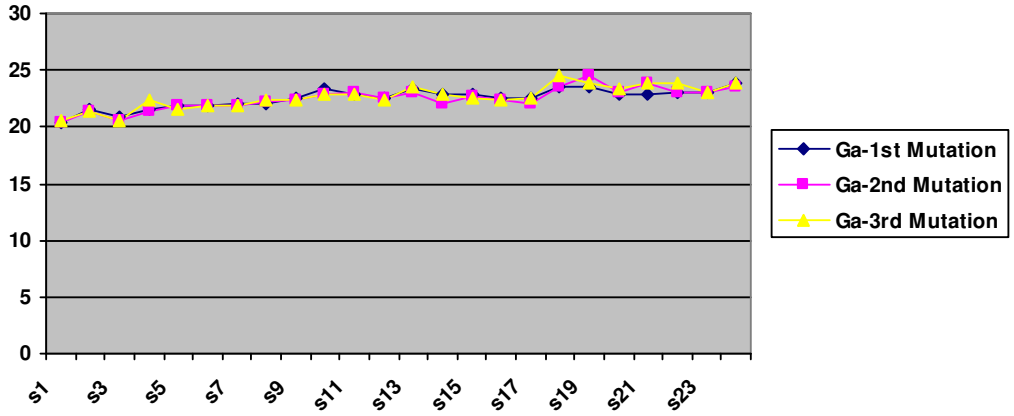
where $x=[1,x_1,\ldots x_{24}]$ and $w=[w_0,\ldots.w_{24}]^T$.

Here w is the weighted vector which is stored in each neurons. Such neurons are calcified as n-dimensional neurons assuming two different vector values, i.e., y=1 for class 1 vector and y=-1 for class two vector. Interestingly GAs produced only 4.67% of type I error and 0.09% of type II error. However the significance level was decided at 5% level and the model indicated high statistical significance.

* Stock prices were obtained from FT fact sheet.

Following 1st, 2nd and 3rd mutation it was observed that the fluctuation of price is not too distributed rather parsimonious. The following graph represents three nodes of mutations.

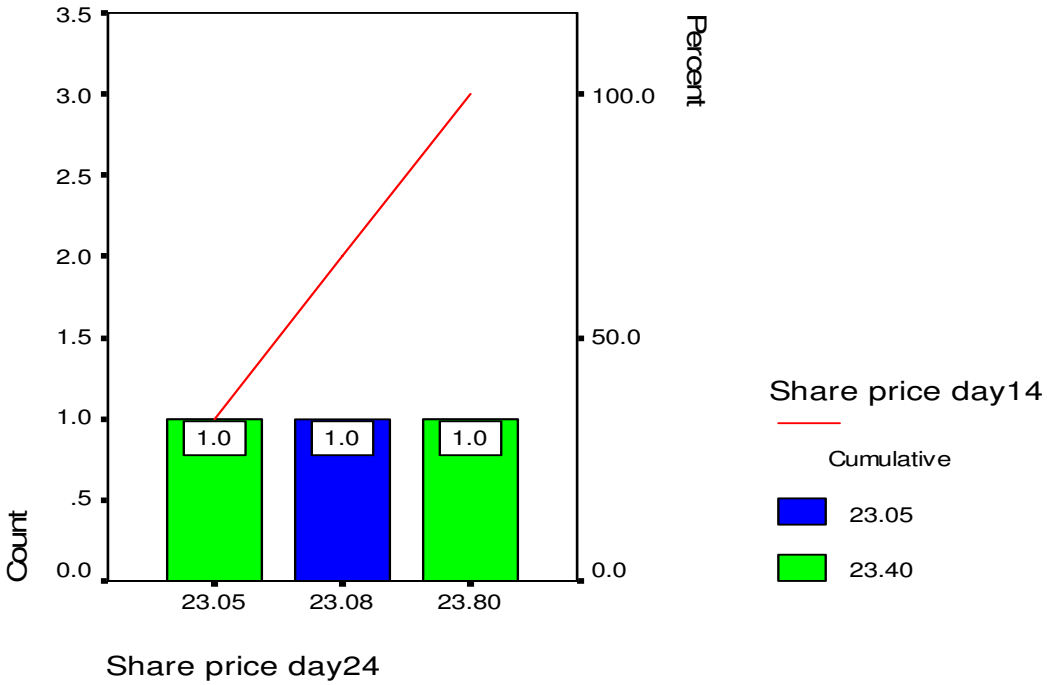
Graph I: Line Graph of Share Values over 24 days following 1st mutation, 2nd mutation and 3rd mutation



Furthermore, the chart indicates that following three subsequent mutations and taking each surviving best price over 24 days window the variance in 1st, 2nd and 3rd mutation does not differ significantly. This leaves enough reason to argue that the similarities might have stemmed from the effect of each survival price which must be best in their respective categories. This somehow underpins that in each sub-window the mutation prices serve best during that temporal period. A follow-up mutation would reveal the similar trend. Moreover the plausibility behind the causality is another concern of this approach. The volatility of stock market could be the reason to infer the causality. However many other variables, i.e., analyst coverage, market information and index adjustment equally affect the market in deciding the causal trend.

To investigate the causality of variance consistency a Pareto graph was generated which is presented below.

Graph II: Pareto graph of Share Prices over 24 days following 1st mutation, 2nd mutation and 3rd mutation



Following a three tier mutation process and stacking each day share price over subsequent day taking the final day share price s24 as the maximum share on closing date the author noticed a very flat and similar cumulative variance over the 24 days window. Further, counting on day 12th share price assuming it as the hypothetical price of mean day of the share sequence the chat indicates that 23.05 % of reasoning behind the share price could be the cause of 76.95 % anomalies, though the count percentage maintains a consistency.

However to understand the effect of the higher anomalies a time series cross correlation was computed which evidently indicates that prices on each nodes, i.e., s1, s12 and s24 do exist in a nonlinear fashion. Interestingly the mid node value is mostly negative identifying a periodic time lag over 24days.

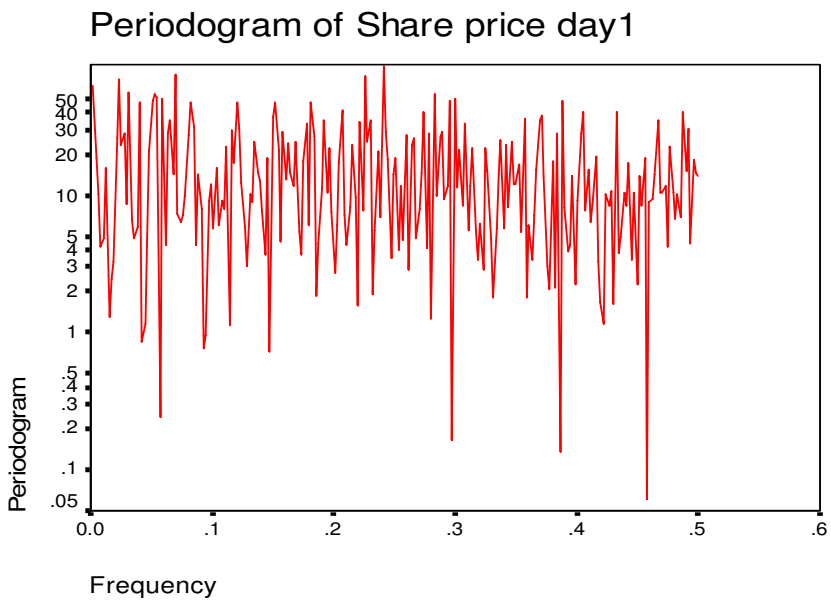
Time Series Cross Correlation between s1, s12 and s24

```
Cross Correlations:   S1
                      S12

Transformations:  natural log

     Cross   Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                  ├────┼────┼────┼────┼────┼────┼────┼────┤
 -1   .192   .707                     |****
  0  -.996   .577  ********************|
  1   .088   .707                     |**

Plot Symbols:      Autocorrelations *     Two Standard Error Limits.

Total cases:  3    Computable 0-order correlations:  3
—


Cross Correlations:   S1
                      S24

Transformations:  natural log

     Cross   Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                  ├────┼────┼────┼────┼────┼────┼────┼────┤
 -1   .680   .707                     |**************
  0  -.693   .577        **************|
  1  -.152   .707                  ***|

Plot Symbols:      Autocorrelations *     Two Standard Error Limits.

Total cases:  3    Computable 0-order correlations:  3
—


Cross Correlations:   S12
                      S24

Transformations:  natural log

     Cross   Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                  ├────┼────┼────┼────┼────┼────┼────┼────┤
 -1  -.693   .707        **************|
  0   .756   .577                     |***************
  1   .063   .707                     |*

Plot Symbols:      Autocorrelations *     Two Standard Error Limits.

Total cases:  3    Computable 0-order correlations:  3
```

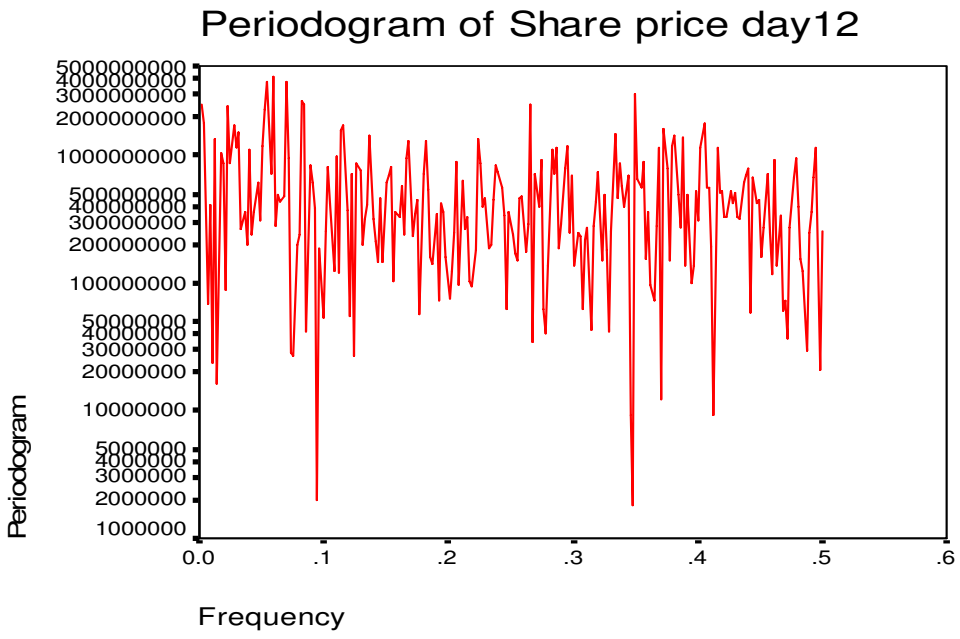
Furthermore a spectral frequency chart was generated to provide a straight forward view of the day 1 and day12 share values. This indeed explains a higher

lower bound value than higher values. It is noteworthy that lower bounds are extended over longer periods.

Spectral Frequency of Share price on Day one



Spectral Frequency of Share price on Day 12



Conclusion

The certainty of prediction adopting GAs within economic and financial system has been resourcefully acknowledged, particularly in parallelisation, relaxed function evaluation and fuzzy sets. This article advances that it can be competently used along with individual data estimation to predict optimal solution of any finite set of population. However further empirical investigation is imperative to examine the effectiveness of this proposed method.

Bibliography

Altenberg, L. (1994) 'The Evolution of Evolvability in Genetic Programming' in K.E. Kinnear Jr. ed. *Advances in Genetic Programming*, MIT Press.

Baker, J.E. (1985) 'Adaptive Selection Methods for Genetic Algorithms' in J.J. Grefenstette ed. *Proceedings of the First International Conference on Genetic Algorithms and Their Applications*, Erlbaum.

Baricelli, N. A. (1962) 'Numerical Testing of Evolution Theories', ACTA *Biotheoretica*, Vol. 16, pp 69-126.

Bramlette, M. F.(1991) 'Initialisation, mutation and selection methods in Genetic algorithms for function optimisation' in R.K. Belew and L.B. Booker ed. *Proceedings of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufmann.

Green, H. G. (1990) '*Econometric Analysis*', Macmillan Publishing: NY.

Holland, J.H. (1975) '*Adaptation in Neutral and Artificial Systems*', University of Michigan Press, 2nd Edition: MIT Press.

Mackey, M.C. and Glass, L. (1977) *Science* 197:297.

Mayer, T.P. and Packard, N.H.(1992) 'Local forecasting of high-dimensional chaotic dynamics' in M. Casdagli and E. Eubank ed. *Nonlinear Modelling and Forecasting*, Addison-Wesley.

Packard, N.H. (1990) 'A Genetic Learning Algorithms for the Analysis of Complex Data', *Complex Systems*, Vol. 4, No. 5, pp 543-572.

Riolo, R.L. (1992) 'Survival of the fittest bits', *Scientific American*, Vol. 267, No.1, pp 114-116.