Workshop/Summer School on Evolutionary Computing
Lecture Series by Pioneers, August 18-22, 2008, Londonderry

Genetic Programming Applications in Financial Modeling: A Brief Survey

Zheng Yin
University College Dublin
zheng.yin@ucdconnect.ie

Anthony Brabazon
University College Dublin
anthony.brabazon@ucd.ie

Conall O'Sullivan
University College Dublin
conall.osullivan@ucd.ie

Abstract

Genetic Programming (GP) is an automated computational programming methodology, inspired by the workings of natural evolution techniques. This paper reviews its applications in financial modeling cross different financial markets and analyzes GP's potential utility in these areas. The future research directions of GP in financial markets have been highlighted.

1 Introduction

Genetic programming (GP) is a population-based search algorithm which starts from a high-level statement of what is required, and automatically creates a computer program (or rule set) to solve the problem. It offers particular utility in data-driven financial modeling as it allows the user to combine existing domain knowledge with a powerful model induction methodology.

In this paper we provide a brief overview of a range of GP's applications in finance. While a substantial literature on asset pricing and trading applications of GP has emerged over the past decade, multiple other financial applications of GP have also been documented. This paper concludes with some suggestions for future financial applications of GP.

2 GP in Financial Modeling

GP is drawn from the family of Evolutionary Computing. One particularly interesting aspect of genetic programming and grammatical evolution is that both the solution form and associated parameters are co-evolved. This offers particular utility in financial modeling, as the environment is data rich but theory poor. Typically, while many plausible explanatory variables exist but the interrelationship among the relevant variables is poorly understood, although some domain knowledge may exist. This suggests that model induction methodologies will have particular utility. Unlike black-box methodologies such as neural networks, GP offers the potential to generate human-readable rules. This is of particular importance in (for example) financial trading systems where human decision makers will want to have insight into the trading rationale. Another advantage of GP is that it permits the incorporation of domain knowledge, and the generation of 'solutions' of a particular form. This allows the financial user to (for example) seed the evolutionary process with their current trading strategies, in order to see what improvements GP can uncover. More generally, all evolutionary algorithms allow the incorporation of complex fitness functions, which is of particular importance in finance as fitness is generally a complex amalgam of return and risk. Recent years have also seen an explosion in the quantity and quality of electronic financial information available, hence, the practicality of applying GP-type methodologies in finance has increased.

In summary, GP is a powerful model-induction methodology which offers particular advantages in modeling in complex dynamic environments such as financial markets.

3 GP in Financial Forecasting and Asset Pricing

Financial markets are affected by multiple economic, political and social events. The relationship between these factors and financial asset prices is not well understood and, moreover, is changing over time. Most theoretical financial pricing models are based on strong assumptions which are often not met in reality. This offers opportunities for the application of GP in order to 'recover' the underlying data generating models.

GP has been applied to stock market forecasting in the studies of [4], [8], [18], and to option pricing in [3], [13] and [19]. The GP applications above, their choices of fitness function, performance measure, terminal set and function set, have been summarized in table 1 and 2.

The degree of predictability of market prices is subject to much debate. Using GP, [7] introduced a new measure of time series' predictability.

Terminal Set In most applications (table 2), GP was fed with preprocessed data, such as lagged time series and normalized data. In [8] two preprocess steps, data transformation and embedding were implemented before feeding the data to GP. A wavelet analysis has been used to extract indicators from the raw index data in [18].

Fitness Function The choice of fitness function has a critical impact on the behavior of GP. Commonly applied fitness metrics for financial forecasting, such as Mean Squared Error(MSE), try to minimize the distance between predicted points and realized points. GP's applications in [13] and [19] have picked up the fitness function of the combination of absolute error and percentage error as the options' special property. The fitness function in [8] takes account the size of the GP individual solution, in an attempt to relieve the overfitting problem. The fitness function in [18] uses the logic that in a given tolerance how many points have been predicted correctly, and how many points have been misspecified.

Dynamic GP Two applications [20] and [19] have been

Table 1: GP Financial Time Series Forecasting Summary-1

App.	Times Series	Fitness Function	Performance Measures	Comparison Method
[4]	Nikkei225	MSE	MSE, HIT, Profit gain	ANNs
[7]	8 Dow Jones Stocks	SSE	SSE,STD,η,θ,R^2	-
[8]	Nikkei225	RAT	MSE,RSE,GCV,HIT,Profit Gain	Standard GP
[18]	DJIA	w1*RC-w2*RMC-w3*RF	RP, RMC, RC, TP, FP, TN, FN(p-value)	GP with different terminals
[3]	S&P 500 index options	SSE	MSE	Linear Regression, B-S, ANNs
[13]	Simulated Option Price	SAP	Hedging effects, AE, PE	Linear Regression, B-S, ANNs
[19]	FTSE 100 future index options	Sqrt of averaged AE and PE	AE, PE	Standard GP
[20]	US GDP,CPI Inflation	MSE/MAD/CF	RMSE	Benchmark models

DJIA: Dow Jones Industrial Average index. η:The predictability measure, a percentage hypothetical reversed entropy when a shuffled series is put back to its original order. STD: Standard deviation. θ:A complexity statistic measure. PC: price-change. MSE: Mean square error of predicted value and target data. ANNs: Artificial neural networks. SSE: Sum of squared errors. RAT: Rational average error, a combination of an accuracy component reflecting the degree of fitting wildly varying series and a complexity measurement. RSE: Ratio of squared errors. HIT: the Hit percentage. GCV: the generalized cross validation.MAPE: Mean absolute percentage error. RC: Rate of Correctness. RMC: The rate of missing chance. RF:The rate of failure. w1,w2 and w3 are weights for RC, RMC and RF. RP: The rate of precision. TP:True positive. FP:False positive. TN:Ture negative. FN:False negative. B-S: Black-Scholes option pricing model. SAP: Sum of absolute dollar error and percentage errors.AE: Absolute error. PE: Percentage error. Sqrt: Square root. GDP: Gross domestic product. CPI Inflation:Comsumer price index inflation rate. MAD: The mean absolute deviation. CF: A combination of MSE and MAD by a user specified threshold value. RMSE: Root mean squared error

Table 2: GP Financial Time Series Forecasting Summary-2

App.	Terminal Set	Function Set
[4]	Different frequency intraday data	+,-,$\times$,$\div$,Sin,Cos,Exponential
[7]	Lagged return and explanatory variables	+,-,$\times$,$\div$,Sin,Cos,Exponential,Sqrt,Logarithmic
[8]	Preprocessed time series including volume	polynomials
[18]	Wavelets based indicators, Prediction terminals(0 and 1)	I-T-E, And, Or, Not,$<$,$>$,$\geq$,$\leq$,=
[3]	S/E, τ, R_f, $Randomnumber$	+,-,$\times$,$\div$,Sin,Cos, Real Logarithmic, Exponential
[13]	S,E,S/E,τ, max(S-E,0)	+,-,$\times$,$\div$,Natural Logarithmic, Exponential,Sqrt, NCDF
[19]	S/E, $R_f \times \tau$, $\sigma \times \sqrt{\tau}$	+,-,$\times$,$\div$,Square, Exponential, NCDF
[20]	Economic indicators	+,-,$\times$,$\div$,Sqrt, Exponential, Sin,Cos, Logarithmic

I-F-E: if-then-else. C_M:Option daily closing price. S:Underlying stock index daily closing price on the trading day. E: The strike price of the index option.τ, The option's time to maturity in year. R_f, The 3-month treasure-bill annual rate. NCDF: Normal cumulative distribution function. σ: Implied volatility. Economic indicators: 29 economic indicators including employment financial, survey, production and sales and others for US GDP; Unemployment rate and past values of the monthly inflation rate for US CPI Inflation

specially designed for dynamic financial markets where the environments change over time but not completely. In [19] the probabilities of crossover and mutation are adapted during the GP searching process. In [20] the size of data window fed to GP is sliding "on-the-fly"and past knowledge is retained by a "Dormant solution".

4 GP in Financial Trading

GP's applications in financial trading look to combine technical indicators, price and volume time series data, in order to produce a 'trading signal'. GP allows the automation of this process, with the concurrent vast expansion of the search space which can be feasibly searched. A range of GP applications in stock market trading and foreign exchange market trading are summarized in tables 3, 4 and tables 5 and 6.

4.1 Stock Market

In the studies of [6],[9],[15],[16] and [17] GP has been applied for stock market trading.
Terminal Set Typically the raw financial time series is preprocessed using technical indicators such as moving average rules (see [9]). Sometime the technical indicators are embedded in the function set in [16].
Fitness Function Decided by trading purpose, cumulated return, excess return over buy-and-hold strategy and portfolio value at end of trading period have been the most popular forms of fitness function in table 3. Different risk adjusted return measures and cumulated excess return have been compared to be the fitness function in the study of [6].

4.2 Foreign Exchange(FX) Market

The FX Market is the largest and most liquid financial market in the world. GP has been applied in FX market trading in a series of studies including [1], [2], [5], [10], [11], [12], and [14].
Terminal Set The same as in stock market trading, technical indicators are commonly used as the terminals in applying GP for uncovering FX market trading systems. Unlike stock markets, FX markets can see central bank intervention. The US authorities' intervention data has been included in the terminal set in the study of [11] to explore the relation of currency rate change and authority intervention. Other studies have seen the inclusion of macro-economic variables such as interest rates as an explanatory variable for the construction of a FX trading system [5][14].
Fitness Function Maximum cash drawdown has been used as risk criteria in the studies of [10] and [12].
Advanced GP Parallel GP was applied by [2] and [10] in FX market. The parallel implementation produced a notable speed up in training time without compromising the quality of out of sample results.

5 Other Financial Applications of GP

In addition to financial forecasting and trading system design, GP has been applied in the following areas.
Volatility Prediction Volatility is the degree to which financial prices tend to fluctuate. Asset price volatility is a key issue in measuring the financial risk. It helps to optimally allocate assets in a portfolio and derivative security trading. Volatility forecasting is a function-fitting problem. GP's model induction ability has been applied in this area in the study of [23].
Discovering Arbitrage Opportunity Arbitrage trading can be defined in a variety of ways, broadly speaking, these trades seek to make profits by exploiting price differences of identical or similar financial instruments, on different markets or in different forms. GP has been used to discovering arbitrage opportunities in call, put options and futures by [22].

Table 3: GP in Financial Stock Market Trading Summary-1

App.	Trading Markets	Fitness Function	Performance Measures	Comparison Method
[6]	S&P 500 Index(1929-1995)	Cum.ER,S-R,X^*, X_{eff}, $Jensen's\alpha$	Fitness	B-A-H
[9]	DJIA(1969-1981)	RC,$f_{(1)}$, $f_{(2)}$	RC, RMC, RF,AARR,RPR	3 types ANNs, linear classifier
[15]	S&P 500 (1954-2002)	PV,PV with C-P,Consist-f	Avg PV	B-A-H
[16]	14 stocks in TSE 300 index	ER over B-A-H	Fitness	B-A-H
[17]	S&P 500, S&P auto,S&P banks(1990-1999)	Max-diff	Gross%Returns, Risk adjusted % returns	B-A-H

ER:Excess Return. Cum.:Cumulative S-R: Sharpe ratio. $S-R$, $X^*statistic$, $X_{eff}measure/Jensen's\alpha$: Different measures of risk adjusted returns. DJIA: Dow Jones Industrial Average index. PA: Prediction accuracy, the percentage of correct predictions. ARR: Annualized rate of return. AARR:Average Annualized Rate of Return. RPR: Ratio of positive returns.$f_{(1)}$:w1*RC-w2*RMC-w3*RF, w1,w2,w3 are weights for RC,RMC and RF.$f_{(2)}$: constrained $f_{(1)}$ with $R = [P_{min}, P_{max}]$, R defines the minimum and maximum percentage of recommendations that GP has to make. Weights approached by trial and error. 6-Indi.: 6 trading indicators include 12 days Moving Average, 50 days moving average, 5 days filter rule, 63 days filter rules, 5 days trading range breakout rule(the difference of today's price and previous 5 days maximum price), 50 days trading range breakout rule. Avg:Average. Ntr: Number of Trades. PV: Portfolio value produced by the rule at the end of the in-sample period.C-P: complexity-penalizing factor. Consist-f: Consistency of performance fitness function, calculated as the number of periods, modified by the factor with 12 month. Well performing if it beat or equal to buy-and-hold return and the risk-free interest rate. B-H: Buy-and-hold. Max-diff:maximize the difference between buying and selling price. TAIEX:the Capitalization Weighted Stock Index(Taiwan)

Table 4: GP in Financial Stock Market Trading Summary-2

App.	Terminal Set	Function Set
[6]	MA, L_{max}, L_{min},Lag of Stock Index,Stock Index	+,-,×,÷,Norm,C(0,2), If-then,And,Or,$<$, $>$, Not, True, False
[9]	Positive, Negative, Constant, 6-Indi.	I-T-E,And,Or,Not,$<$, $>$
[15]	Tech-Indi.	And,Or,Not,$<$, $>$
[16]	Price,Volume,True,False,Real in[0,250]	+,-,×,÷,and,or,not,$<$, $>$,I-T-E,Norm, Avg, Max,Min,Lag,Volatility,RSI,ROC
[17]	–	–

MA: Moving Average. L_{max}:Local maximum. L_{min}: Local minimum 6-Indi.: 12-day Moving average, 50-day Moving average,5-day filter,63-day filter, 5-day Trading Range Breakout rule, 50-day Trading Range Breakout rule. Tech-Indi.: Prices(opening, closing, high, low of the current month), Moving Average(2,3,6 and 10 month),Rate of change(3 and 12 month),Price resistance markers: two previous 3-month moving average minima and two previous 3-month moving average maxima. Trend line indicators: a lower resistance line based on the slope of the 2 previous minima and a upper resistance line based on the slope of the 2 previous maxima.RSI:Relative strength index. ROC: Rate of change

Portfolio Management One of the classic multiple objective optimization problems in finance is portfolio selection, where the object is to invest a fixed amount of money in a diverse set of assets so as to maximize return while minimizing a risk measure. The study of [21] has tried to use GP as a robust stock selection method in Malaysian stock market.

6 Conclusions

Initially GP was primarily applied for financial forecasting and trading system induction and the sophistication of GP applications in these areas has increased markedly in recent years. GP has also seen increasing applications to other problems in finance, such as derivative pricing and volatility prediction. The widespread application of GP is not surprising given the nature of financial modeling, being typically undertaken in a data rich, theory poor environment. It is also notable that the availability of data and the increasing power of computers alter the relative costs of theory vs inductive modeling methodologies, helping to explain the growth of GP type applications in finance. The financial returns to even small edges in asset pricing / trading systems also explain industry's enthusiasm for the application of GP methods.
In considering future directions for the application of GP in financial markets, one avenue is to combine more domain knowledge into GP terminal/function sets. Another critical issue is the appropriate design of fitness functions using domain knowledge. For example, in creating a trading system for the futures market, we would wish to incorporate account margin requirements in order to ensure that the resulting trading system is practically useful. The other primary direction of research will entail the application of ever-more sophisticated GP methodologies. For example, as parallel and adaptive GP methodologies are improved, these will in turn allow us to attack more complex finance problems.

References

[1] Edmonds A.N., Burkhardt D. and Adjei O., Genetic Programming of Fuzzy Logic Production Rules in *IEEE International Conference on Evolutionary Computation, 1995.*, Vol.2,pp.765-770

[2] Oussaidène M., Chopard B., Pictet O.V. and Tomassini M., Parallel Genetic Programming and its application to trading model induction in *Parallel Computing*, Vol.23, 1997, pp.1183-1198

[3] Chen S.H.,Yeh C.H. and Lee W.C. (1998). Option Pricing with Genetic Programming, in *proceedings of the Third Annual Genetic Programming Conference*, 1998, pp.32-37.

[4] Iba H. and Sasaki T.,Using Genetic Programming to Predict Financial Data in *Proceedings of the Congress on Evolutionary Computation*, Vol.1, 1999, pp.244-251

[5] Neely C.J. and Weller P.A., Technical Trading Rules in the European Moneytary System in *Journal of International Money and Finance* Vol.18, 1999, pp.429-458

[6] Neely C.J., Risk-Adjusted, Ex Ante, Optimal, Technical Trading Rules in Equity Markets in *Federal Reserve Bank of St.Louis Working Papers*, No.1999-015D, 1999(Revised Aug. 2001)

[7] Kaboudan M.A., A Measure of Time Series'Predictability Using Genetic Programming Applied to Stock Returns in *Journal of Forecasting*, vol. 18, 1999, pp. 345-357

[8] Iba H. and Nikolaev N. ,Genetic programming polynomial models of financial data series in *Proceedings of the 2000 Congress on Evolutionary Computation*, Vol.2, 2000, pp.1459-1466

Table 5: GP in Financial FX Market Trading Summary-1

App.	Trading Markets	Fitness Function	Performance Measures	Comparison Method
[1]	US T.Bond, Nikkei, FTSE, S&P, DM	MPLE	Annualized profit	–
[2]	(GBP,DM,ITL,JPY,CHF,FRF,NLG) / USD(1987-1994)	X'_{eff}	Avg return, X_{eff},Avg complexity	–
[5]	DEM/(USD,NLG,FRF,ITL,GBP)(1979-1996)	ER	Mean return,Monthly Std, S-R, Ntr, PtgL, Market timing test	MA, Filter rules
[10]	GBP/USD tick data(1993-1997)	MSR	Cum. profit,Stirling Ratio	–
[11]	(DM,JPY,GBP,CHF)/USD(1975-1998)	ER	Mean annual ER, Std, S-R, Ntr, PtgL	–
[12]	AUS/USD(1993-2001)	MSR	Trading activity, Trading profit	AMA,CCI,RSI
[14]	(DEM,JPY,GBP,CHF)/USD(1996 intra-day data)	ER	Annual return, Ntr, Break-even transaction cost,PtgL, Long return	Linear forecasting models

MPLE:A mixture of final profit and linearity of the equity curve X_{eff}:risk adjusted return by taking account of variance of total return over time. PtgL:percentage of long position. X'_{eff}:is a transformed form of X_{eff}.Std: standard deviation. Break-even transaction cost is the one-way transaction cost which reduces the annual excess return during the test period to zero. Tick data: A new data point or tick, is recorded with every change in price. MSR:Modified Stirling ratio, the ratio of return against modified max drawdown.

Table 6: GP in Financial FX Market Trading Summary-2

App.	Terminal Set	Function Set
[1]	Open/closing price,Highest/lowest price during the day	Lag, Fuzzy operators(And,Or,Not)
[2]	Pre-optimized momentum indicators,RAND in [-2,2]	$+,-,\times,\div$,ABS, Max,Min,And, Or, Not,$<,>$,if
[5]	NER, Interest rate data , Constants	$+,-,\times,\div$,Norm, Average, Max,Min, Lag, And, Or, Not,$<,>$,If-then, I-T-E
[10]	Technic indicators(SMA,AMA,PCB,The stochastic, RSI, CCI)	And, Or, Xor
[11]	NER, US authorities intervention data, Constants between(0,6)	$+,-,\times,\div$,Norm,MA,L_{Max},L_{Min}, Lag of data, And, Or, Not,$<,>$,If-then,True, False
[12]	AMA,CCI,RSI	–
[14]	NER, Interest differential, The hour of day	–

NER: normalized exchange rate time series. SMA:Simple moving average crossover. AMA: adaptive moving averages. PCB:Price channel breakout. RSI: the relative strength index, CCI:The commodity channel index.

[9] Li J. and Tsang E.P.K., Reducing Failures in Investment Recommendations Using Genetic Programming in *Proceedings of Sixth International Conference on Computing in Economics and Finance*, 2000

[10] Dempster M.A.H. and Jones C.M., A Real-Time Adaptive Trading System Using Genetic Programming in *Quantitative Finance* Vol.1, 2001, pp.397-413

[11] Neely C.J. and Weller P.A. ,Technical Analysis and Central Bank Intervention in *Journal of International Money and Finance* Vol.20, 2001, pp.949-970

[12] Lee C.S. and Loh K.Y.,GP-Based Optimisation of Technical Trading Indicators and Profitability in FX Market in *Proceedings of the 9th International Conference on Neural Information Processing, ICONIP'02.*, Vol.3, 2002, pp.1159-1163

[13] Chidambaran N.K., Genetic Programming with Monte Carlo Simulation for Option Pricing in *Proceedings of the 35th Conference on Winter Simulation:Driving Innovation*,2003,pp.285-292.

[14] Neely C.J. and Weller P.A. ,Intraday Techinical Trading in Foreign Exchange Market in *Journal of International Money and Finance* Vol.22, 2003, pp.223-237

[15] Becker L.A. and Seshadri M.,GP-evolved technical trading rules can outperform buy and hold in *Proc. of the Sixth International Conference on Computational Intelligence and Natural Computing*, 2003.

[16] Potvin J.Y., Soriano P. and Vallée M., Generating Trading Rules on the Stock Markets with Genetic Programming in *Computer & Operations Research*, Vol.31, 2004, pp.1033-1047

[17] Fyfe C., Marney J.P. and Tarbert H., Risk Adjusted Returns from Techinical Trading: A Genetic Programming Approach in *Applied Financial Economics*, No.15, 2005, pp.1073-1077

[18] Li J., Shi Z. and Li X.L. Genetic Programming with Wavelet-based Indicators for Financial Forecasting, in *Transactions of the Institute of Measurement and Control*, Vol.28, No.3, 2006, pp.285-297

[19] Yin Z., Brabazon A. and O'Sullivan C., Adaptive Genetic Programming for Option Pricing in *GECCO '07*, pp.2588-2594

[20] Wagner N., Michalewicz Z., Khouja M., McGregor R.R, Time Series Forecasting for Dynamic Environments: The DyFor Genetic Program Model in *IEEE Transactions on Evolutionary Computation*, Vol.11, 2007,pp.433-452

[21] Wei Y. and Clack C.D., Evolving robust GP solutions for hedge fund stock selection in emerging markets, in *Proceedings of the 9th annual conference on Genetic and evolutionary computation* , 2007, pp.2234-2241

[22] Tsang E., Markose S. Er H. and Garcia A., EDDIE for Discovering Arbitrage Opportunities in *Post-Conference Proceedings Keynote Speech, Numerical Methods for Finance Conference*,Dublin, Ireland, June, 2006

[23] Chen S.H. and Yeh C.H.,Using Genetic Programming to Model Volatility in Financial Time Series: The Case of Nikkei 225 and S&P 500 in *Proceedings of the 4th JAFEE International Conference on investments and Derivatives(JIC'97)*, pp.288-306