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 across 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 inter-relationship 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...
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(FOREX) 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 indicators 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 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

<table>
<thead>
<tr>
<th>App.</th>
<th>Trading Markets</th>
<th>Fitness Function</th>
<th>Performance Measures</th>
<th>Comparison Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>[16]</td>
<td>Dow Jones Industrial Average Index</td>
<td>Max-diff</td>
<td>GR(%),Return, Risk adjusted % returns</td>
<td>B-A-H</td>
</tr>
</tbody>
</table>

ER: Excess Return. Cum.: Cumulative. S-R: Sharpe ratio. AARR: Average Annualized Rate of Return. ARR: Annualized rate of return. TAIEX: Capitalization Weighted Stock Index (Taiwan).

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

<table>
<thead>
<tr>
<th>App.</th>
<th>Terminal Set</th>
<th>Function Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>[6]</td>
<td>MA, Fmax, Lmax, Log of Stock Index, Stock Index</td>
<td>&gt;, &lt;-, =, Norm, Divide, And, Or, C(0,2), If-then, And, Or, Not, True, False</td>
</tr>
<tr>
<td>[15]</td>
<td>Tech-Ind.</td>
<td>And, Or, Not, True, False</td>
</tr>
<tr>
<td>[16]</td>
<td>Price, Volume, True, False, Real, MA, L(250)</td>
<td>&gt;, &lt;-, =, and/or/not, &gt;, &lt;, P-E, Norm, Avg, Max, Min, Log, Volatility, RSI, ROC</td>
</tr>
<tr>
<td>[17]</td>
<td>Price, Volume, True, False, Real</td>
<td>MA, Fmax, Lmax, Log of Stock Index, Stock Index</td>
</tr>
</tbody>
</table>

MA: Moving Average. Lmax: Local maximum. L(250): Local minimum 6-ind.: 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-Ind.: 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 objective 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 allow 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 strategy 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

Copyright IEEE 2008 32
modified max drawdown. 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

MPL: A mixture of final profit and linearity of the equity curve

Long position. Strength index, CCI: The commodity channel index.


Xeff′ is a transformed form of eff. Std: Standard deviation. Break-even transaction cost is the one-way transaction cost which reduces the annual excess return during the best 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 5: GP in Financial FX Market Trading Summary-1

<table>
<thead>
<tr>
<th>App.</th>
<th>Terminal Set</th>
<th>Function Set</th>
<th>Performance Measures</th>
<th>Comparison Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>Open/closing price, Highest/lowest price during the day</td>
<td>Lag, Fuzzy operators, And, Or, Not</td>
<td>Xeff, Eff</td>
<td>–</td>
</tr>
<tr>
<td>[2]</td>
<td>Pre-optimized momentum indicators EMA, MA, And, Or, Not</td>
<td>&lt;, &gt;, Min, Max, Ave, St-Dev, Ave, St-Dev, Ave, St-Dev</td>
<td>Xeff, Eff</td>
<td>–</td>
</tr>
<tr>
<td>[3]</td>
<td>Technical indicators SRA, SMA, PCL, AUS, MSB, CAT, QML</td>
<td>And, Or, Not</td>
<td>Xeff, Eff</td>
<td>–</td>
</tr>
<tr>
<td>[4]</td>
<td>US authorities intervention data, Constants between(0,6)</td>
<td>And, Or, Not</td>
<td>Xeff, Eff</td>
<td>–</td>
</tr>
</tbody>
</table>


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

<table>
<thead>
<tr>
<th>App.</th>
<th>Terminal Set</th>
<th>Function Set</th>
<th>Performance Measures</th>
<th>Comparison Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>Open/closing price, Highest/lowest price during the day</td>
<td>Lag, Fuzzy operators, And, Or, Not</td>
<td>Xeff, Eff</td>
<td>–</td>
</tr>
<tr>
<td>[2]</td>
<td>Pre-optimized momentum indicators EMA, MA, And, Or, Not</td>
<td>&lt;, &gt;, Min, Max, Ave, St-Dev, Ave, St-Dev, Ave, St-Dev</td>
<td>Xeff, Eff</td>
<td>–</td>
</tr>
<tr>
<td>[3]</td>
<td>Technical indicators SRA, SMA, PCL, AUS, MSB, CAT, QML</td>
<td>And, Or, Not</td>
<td>Xeff, Eff</td>
<td>–</td>
</tr>
<tr>
<td>[4]</td>
<td>US authorities intervention data, Constants between(0,6)</td>
<td>And, Or, Not</td>
<td>Xeff, Eff</td>
<td>–</td>
</tr>
</tbody>
</table>


