Predicting Currency Exchange Rates by Genetic Programming with Trigonometric Functions and High-Order Statistics

Roy Schwaerzel
Department of Computer Science
The University of Texas at San Antonio
San Antonio, TX 78249-0667
rschwaer@cs.utsa.edu

Tom Bylander
Department of Computer Science
The University of Texas at San Antonio
San Antonio, TX 78249-0667
bylander@cs.utsa.edu

Abstract: This paper describes an extension of the traditional application of Genetic Programming in the domain of the prediction of daily currency exchange rates. In combination with trigonometric operators, we introduce a new set of high-order statistical functions in a unique representation and analyze each system performance using daily returns of the British Pound and Japanese Yen. We will demonstrate that the introduction of high-order statistical functions in combination with trigonometric functions will outperform other traditional models such as Genetic Programming with the basic function set and ARMA models. Performance will be measured on hit percentage, average percentage change, and profit.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning – *connectionism and neural nets, parameter learning, language acquisition.*

General Terms: Algorithms, Performance, Experimentation.

Keywords: Genetic Programming, Finance, Prediction/Forecasting, Time Series Analysis, Trigonometric Function Set, High-Order Statistical Function Set.

1. Introduction

Although GP has shown to be amenable for a wide variety of optimization problems, the impact of using high-order statistical functions as building blocks for time series analysis has not been effectively explored. Most current GP implementations [1, 2] are restricted to a basic set of operators (e.g., addition, subtraction, multiplication, protected division, etc.) to model the data. For time series analysis, these operators can be enhanced and extended by incorporating a class of statistical measures, defined as the High-Order Statistical Function Set (HSFS). The HSFS includes building blocks that characterize the statistical behavior and nature of financial time series and will include statistical moments (i.e., mean, standard deviation, skewness, and kurtosis). In addition, because financial time series can be periodic, trigonometric functions can characterize this aspect. This research will analyze the performance improvement of the HSFS by itself and in conjunction with trigonometric functions.

2. GP Implementation

We selected the GPC++ - Genetic Programming C++ Class Library [3], which provided us with a basic tool to implement our extended function set. We modified the code to allow the selection of the previous 20 data points to be part of the new terminal set, added the Extended Function Set (sin, cos, tan, log, and exp), and our High-Order Statistical Function Set (mean, standard deviation, skewness, and kurtosis).

Although the implementation of the trigonometric functions were straight forward, only allowing for one operand per function, the implementation of the HSFS required information about the data points this statistics should be calculated on. Therefore, we defined two parameters, the LAG and the LENGTH. The LAG parameter is a value between 1 and 20 specifying how many time steps back from the prediction this statistics should be calculated on. For example, LAG=1 would include the previous data point, LAG=5 would include the data point one week prior the prediction. The LENGTH parameter specifies the number of data points to include for the statistics, starting from the LAG value backwards. Thus, Mean(1,10) would calculate the average of the last 10 trading days.

3. Experimental Setup

The daily rates and returns of two major currencies - the British Pound and the Japanese Yen – against the US Dollar for the period from January 1, 1990 to September 16, 2005 were selected for our experiments and analysis. The data was collected and certified by the Federal Reserve Bank of New York. The above time interval contains 4,100 weekdays with 3,953 actual trading days. Both time series were divided into the training and the testing data sets as specified in Table 3.1. In the case that a weekday was not a trading day, the closing rate of the previous trading day is used as the closing rate for that particular day.

Table 3.1. Training and Testing Data Sets Summary for the British Pound (GBP) and the Japanese Yen (JPY) Exchange Rates

	Training Data Set		Testing Data Set	
	GBP	JPY	GBP	JPY
Time period	1/1/1990 - 3/29/1995		3/30/1995 - 9/16/2005	
Weekdays	1,368 days		2,732 days	
Trading days	1,318 days		2,635 days	
No Change days	15 days	23 days	25 days	14 days

This analysis focuses on daily percentage returns rather than on daily rates, because this transformation converts the nonstationary time series into a stationary time series.

4. Experiments and Results

GP with four different function set selections[1] were used to model both selected time series**.** All other parameters (e.g., population size, number of generations) were kept constant for all experiments.

GECCO'06, July 8–12, 2006, Seattle, Washington, USA.
ACM 1-59593-186-4/06/0007.

Thirty independent runs were performed on each of above experiments using a population size of 1,000 individuals over 100 generations. We measured performance using MSE (mean square error), hits (number of correct predictions), hit percentage, APC (average percentage change), and profit.

Table 4.1. Average Performance for British Pound including the 95% confidence interval ranges (over 30 runs)

Method	MSE	Hits	HIT [%]
GP+EFS+HSFS	0.247198	1,420.10 ± 4.31	54.38 ± 0.17
GP+EFS	0.248566	1,389.40 ± 2.92	53.21 ± 0.11
GP+HSFS	0.241333	1,378.57 ± 5.36	52.80 ± 0.21
Basic GP	0.247727	1,341.10 ± 5.63	51.36 ± 0.22

Table 4.2. Average Performance for Japanese Yen including the 95% confidence interval ranges (over 30 runs)

Method	MSE	Hits	HIT [%]
GP+EFS+HSFS	0.223588	1,364.97 ± 5.29	52.08 ± 0.20
GP+EFS	0.223636	1,359.10 ± 4.94	51.85 ± 0.19
GP+HSFS	0.223826	1,348.83 ± 4.81	51.46 ± 0.18
Basic GP	0.223402	1,345.70 ± 7.28	51.34 ± 0.28

Tables 4.1 and 4.2 summarize the average performance, including the 95% confidence intervals over the 30 runs, of the various GP configurations and currencies. The genetic program with the lowest MSE on the trainings set was selected for the final comparison with our benchmarks.

Table 4.3. Performance Matrix for British Pound

Method	MSE	Hits	HIT	APC	Profit
GP+EFS+HSFS	0.243471	1,442	55.23%	0.042591%	88.78%
GP+EFS	0.244004	1,407	53.89%	0.037322%	75.61%
GP+HSFS	0.245323	1,402	53.70%	0.024652%	47.99%
Basic GP	0.245221	1,367	52.36%	0.020227%	38.71%
AR(15)	-	1,325	50.75%	0.013110%	25.58%
ARMA(5,5)	-	1,299	49.75%	0.012944%	25.37%
AR(10)	-	1,319	50.52%	0.009986%	20.61%
ARMA(20,20)	-	1,331	50.98%	0.006139%	14.83%
Buy & Hold	-	-	-	-	12.15%

Table 4.4. Performance Matrix for Japanese Yen

Method	MSE	Hits	HIT	APC	Profit
GP+EFS+HSFS	0.221574	1,389	53.00%	0.051799%	79.78%
GP+EFS	0.222053	1,386	52.88%	0.048613%	72.03%
GP+HSFS	0.222036	1,385	52.84%	0.033564%	40.05%
Basic GP	0.222195	1,379	52.61%	0.027045%	26.78%
AR(10)	-	1,331	50.78%	0.022277%	21.06%
AR(15)	-	1,314	50.13%	0.016418%	11.64%
ARMA(5,5)	-	1,271	48.49%	-0.004357%	-13.85%
ARMA(20,20)	-	1,295	49.41%	-0.008389%	-19.63%
Buy & Hold	-	-	-	-	-20.71%

Tables 4.3 and 4.4 summarize the performance of the GP models, the ARMA models, and the simple Buy & Hold Strategy. We want to point out that all GP models demonstrated a significantly better performance using any of the performance measures.

5. Discussion and Conclusions

We have shown that the Genetic Programming models outperformed the ARMA model and the simple Buy & Hold strategy using the HIT, the APC, and the Profit performance measures. The traditional ARMA modeling provided better results than the Buy & Hold strategy for both currency exchange rates. A more detailed analysis of the performance over the 10-year testing period shows that the addition of the trigonometric functions was adding better prediction power than the proposed high-order statistical functions. However, the combination of the trigonometric and statistical functions was providing slightly better predictions than any of the individual function sets.

The analysis of the chosen time series becomes further interesting, because the British Pound exchange rate was in an overall upwards trend, whereas the Japanese Yen exchange rate was in a downward trend. Investing in the British Pound using a simple Buy & Hold strategy produced a 12.15% profit over the analyzed 10-year period. The Genetic Programming model utilizing both function sets was able to produce a profit of 88.78% over the same testing period. However, investing in the Japanese Yen using the simple Buy & Hold Strategy resulted in a loss of 20.71%. Our Genetic Programming approach with the combined function sets was still able to realize a calculated profit of 79.78% over the same testing period. We realize that the 79.78% profit is a rather low return for a 10-year period. However, the focus of this research has been the analysis of the impact of various function sets in the predictive power of a model, rather than integrating a larger number of other optimization approaches.

Our results suggest that the implementation of trigonometric functions and our high-order statistical functions may be a valuable add-on for other genetic programming application in this or other domains. The incorporation of additional statistical measures (e.g., Exponential Moving Average) may further improve the performance of our approach. We intend to perform further research in this area.

6. References

[1] Iba, H. and Sasaki, T. Using Genetic Programming to Predict Financial Data. In *Proceedings of the 1999 Congress on Evolutionary Computation, CEC99,* IEEE Press, Piscataway, NY, 1999, 244-251.

[2] Chen, Shu-Heng (ed.) Genetic Algorithms and Genetic Programming in Computational Finance, Boston, MA, Kluwer, 2002.

[3] Weinbrenner, T. *"GPC++ - Genetic Programming C++ Class Library"*, Institute for Mechatronics, Technical University of Darmstadt, Germany, 1997.

[1] **Basic GP:** Traditional GP approach using the Basic Function Set; **GP+EFS:** Traditional GP approach using the Basic and the Extended Function Set; **GP+HSFS:** Traditional GP approach using the Basic and the High-Order Statistical Function Set; **GP+EFS+HSF:** Traditional GP approach using the Basic, the Extended, and the High-Order Statistical Function Set