RESEARCH PAPERS IN MANAGEMENT STUDIES

UNIVERSITY OF CAMBRIDGE
Judge Institute of Management
cambridge's business school

INTRADAY FX TRADING: AN EVOLUTIONARY REINFORCEMENT LEARNING APPROACH

M A H Dempster and Y S Romahi

WP 3/2002

The Judge Institute of Management Studies
Trumpington Street
Cambridge CB2 1AG

Intraday FX Trading: An Evolutionary Reinforcement Learning Approach

M A H Dempster and Y S Romahi

WP 3/2002

Professor M A H Dempster & Y S Romahi
Centre for Financial Research
The Judge Institute of Management
University of Cambridge
Trumpington Street
Cambridge
CB2 1AG
UK
Tel.: +44 1223 339652
Email: {mahd2,yr206}@cam.ac.uk

INTRADAY FX TRADING: AN EVOLUTIONARY REINFORCEMENT LEARNING APPROACH

M.A.H. DEMPSTER AND Y.S. ROMAHI
CENTRE FOR FINANCIAL RESEARCH
JUDGE INSTITUTE OF MANAGEMENT
UNIVERSITY OF CAMBRIDGE
MAIL: {MAHD2,YR206}@CAM.AC.UK

We have previously described trading systems based on unsupervised learning approaches such as reinforcement learning and genetic algorithms which take as input a collection of commonly used technical indicators and generate profitable trading decisions from them. These contrast with traditional supervised learning approaches which use labeled trading data directly.

In general, there are two distinct approaches to solving the reinforcement learning problem which search in either value function space or policy space. Temporal difference reinforcement learning methods and evolutionary algorithms are well known examples of these approaches and both were explored in our previous work. This article demonstrates the advantages of applying evolutionary algorithms to the reinforcement learning problem using a hybrid credit assignment approach. In earlier work, the temporal difference reinforcement learning suffered from problems with overfitting the in-sample data. This motivated the present research which attempts to introduce generalisation by creating a hybrid evolutionary based RL system.

Technical analysis has been shown previously to have predictive value regarding future movements of foreign exchange prices and this article presents methods for automated high-frequency FX trading based on evolutionary reinforcement learning about signals from a variety of technical indicators. These methods are applied to GBPUSD, USDCHF and USDJPY exchange rates at various frequencies. We find that the evolutionary reinforcement learning approach is indeed able to consistently outperform the standard RL approach. Statistically significant profits are made consistently at transaction costs of up to 2bp for the hybrid system while the standard RL is only able to trade profitably up to about 1bp slippage per trade. It is also shown that for non-zero slippage a system that also allows a neutral out-of-market position consistently outperforms one which is always in the market.

1 Introduction

Kaelbling *et al.* [11] illustrate the challenges facing reinforcement learning in scaling up to realistic tasks. Of relevance to building a financial trading system is the issue of rarely occurring states. Previous work by the authors [8] examined the issue of searching value function space (through an RL based approach) vs searching the policy space (using an evolutionary algorithm). The two approaches were shown to provide complementary strengths though the RL showed classic signs of overfitting as a result of the rarely occurring states. This led to the currently proposed system that combines a *genetic algorithm* (GA) with a *reinforcement learning* (RL) framework to bring about a *hybrid credit assignment* approach. This paper examines the hybrid approach and contrasts it with standard RL.

In Section 2 we give a brief literature review of relevant earlier work. The *stochastic optimization* problem to be solved by all the compared methods is defined in Section 3, while the following sections, 4 and 5, briefly describe how each approach can be applied to solve this optimization problem approximately. In Section 6, computational experiments are outlined and their results given. Section 7 concludes with a discussion of these results and some potential further avenues of research.

Reinforcement learning has to date received limited attention in the financial literature and this paper demonstrates that RL methods show significant promise. The results for the hybrid approach developed indicate that generalization and incorporation of constraints limiting the ability of computational learning algorithms to overfit improves out-of-sample performance.

2 Literature Review

Despite a century long history amongst investment professionals the technical analysis methods introduced by Dow at the turn of the last century initially met a high degree of academic scepticism culminating in a belief in the *efficient markets* or *random walk* hypothesis. As evidence has increased that markets are less efficient than was originally believed academics have only recently begun to make serious attempts to study the assumptions behind technical analysis [13, 14]. Lo and McKinley [13] state that financial markets are predictable but rather than this being a symptom of inefficiency as is commonly believed they see predictability as the "oil that lubricates the gears of capitalism".

A number of researchers have examined net returns due to various trading rules in the foreign exchange markets [12, 22, 17, 19, 5, 4, 7, 18]. The general conclusion is that trading rules are able to earn significant returns net of transaction costs and that this cannot be easily explained as compensation for bearing risk.

The application of *computational learning* techniques to technical trading and finance has experienced significant growth in recent years. *Neural networks* have received the most attention in the past and have shown varying degrees of success. However recently there has been a shift in favour of user-transparent, non-black box evolutionary methods like genetic algorithms and genetic programming. An increasing amount of attention in the last several years has been spent on these genetic approaches which have found financial applications in option pricing [2, 3] and as an optimization tool in technical trading applications [19, 10, 6].

Pictet *et al.* [23] employ a GA to optimize a class of exponentially weighted moving average rules, but run into serious overfitting and poor out-of-sample performance. They report 3.6% to 9.6% annual returns net of transaction costs. Neely and Weller [17] report that for their GA approach, although strong evidence of predictability in the data is measured out-of-sample when transaction costs are set to zero, no evidence of profitable trading opportunities arise when transaction costs are applied and trading is restricted to times of high market activity.

Reinforcement learning has to date received only limited attention in financial applications. The reinforcement learning technique is strongly influenced by the theory of *Markov decision processes* (MDPs) which evolved from attempts to understand the problem of making sequences of decisions under uncertainty when each decision can depend on the previous decisions and their outcomes.

As fundamental research in reinforcement learning advances, applications to finance have started to emerge. Neuneier [21, 20] has demonstrated Q-Learning in an asset allocation framework applying it to the German DAX stock index. Moody *et al.* [16] examine a *recurrent* reinforcement learning algorithm that seeks to optimize an online estimate of the Sharpe ratio. They also compare the *recurrent RL* approach to that of Q-learning. Dempster *et al.* [7] similarly explore GAs and RLs, as well as the exact solution of an appropriate Markov decision problem and a simple heuristic, in an asset allocation framework.

The main shortcoming of this previous work however is that most technical analysts active in the foreign exchange market are *traders* and also operate at the high frequency level. In fact even technical traders who look for patterns in daily data alone often use tick data for confirmatory entry signals. In subsequent work [8] the authors contrast evolutionary methods with reinforcement learning within such a *trading* framework [6] and this framework will also be used in the sequel.

3 The Problem Defined

3.1 Modelling trading

This paper considers agents that trade fixed position sizes in a single exchange rate. This setting can be generalized to more sophisticated agents that trade varying quantities of a currency, several currencies or indeed manage multiple portfolios.

Traditionally, trading strategies have been evaluated as *asset allocation* strategies in the academic literature (eg. in [19]). The agent has a current lump sum of money and must choose at each timestep whether to allocate this money to be held in the home currency or the foreign currency (possibly earning the overnight interest rate in the chosen currency). Any profit or loss made is added to or subtracted from the lump sum to be allocated in the next timestep, *i.e. reinvested.*

High frequency traders, however, typically are able to draw on a fixed credit line from which they may borrow in either the home or the foreign currency. The money borrowed is then converted to the other currency at the current market rate to hold cash in one currency and a debt in the other. When the trader wishes to close his position he converts his cash at the new (hopefully advantageous) exchange rate and pays any profit into or shortfall from his account. Thus he places a series of fix-sized bets.

More formally, a trade with proportional transaction cost c, exchange rates (expressed per unit of home currency) of F_t at trade entry and $F_{t'}$ at trade exit, drawing on a credit line of C units of home currency and taking a long position in the foreign currency (and a corresponding short position in the home currency) will yield a profit of

$$C\left[\frac{F_t}{F_{t'}}(1-c)^2-1\right]. \tag{1}$$

If a short position is taken in the foreign currency (and correspondingly long in the home) then C/F_t units of foreign currency are drawn from the credit line and the profit is

$$C\left[(1-c)-\frac{F_{t'}}{F_t}\frac{1}{(1-c)}\right]. \tag{2}$$

The asymmetry of these equations is apparent and results from the profit or loss on a short position in the foreign currency being credited in the home currency. Both formulae involve transaction costs being paid per unit on two currency conversions (see [6] for a discussion of the *slippage* c).

In this paper, we examine two approaches. The first system is continuously forced to be in the market while the second system is able to maintain a neutral out-of-market position. These are termed the *2 state* and *3 state* systems respectively.

3.2 Technical indicators

We consider a set of technical indicators, to be used as input for our trading strategies and employ eight commonly used indicators with parameters suggested by [1] as in [6, 7]. These are Price Channel Breakout, Adaptive Moving Average, Relative Strength Index, Stochastics, Moving Average Convergence/Divergence, Moving Average Crossover, Momentum Oscillator and Commodity Channel Index. Each indicator produces two signals: buy (long) or not buy, and sell (short) or not sell. These sixteen binary signals together define the *market state* $\mathbf{s}_t \in \mathcal{S} = \{0,1\}^{16}$.

3.3 Trading strategies

We can consider the market state $\mathbf{s}$ represented by the indicator signals to be a *stochastic process* driven by the *exchange rate* process $\mathbf{F}$ and make the required trading decisions by solving the *stochastic optimization problem* defined by the maximization of expected return over the *trading* *horizon* T net of transactions costs, *viz.*

$$\mathbb{E}\sum_{i=1}^{\mathbf{N}_T} r_i(\mathbf{F}_{t_i}, \mathbf{F}'_{t_i}), \tag{3}$$

where $\mathbf{N}_T$ denotes the random number of trades to the horizon each with return $r(\mathbf{F}_t, \mathbf{F}_{t'})$ in the home currency.

The systems we consider attempt to find approximate solutions to this problem. They attempt to discover a *trading strategy* $\phi : \mathcal{S} \times \{l, s\} \to \{l, s\}$ that maps the current market state $\mathbf{s}_t$ and current position (long, short or neutral) to a new position (long, short or neutral). It should be noted that although our trading strategies ϕ are formally *Markovian* (feedback rules), the technical indicators require a number of periods of previous values of $\mathbf{F}$ to compute the corresponding $0-1$ entries in $\mathbf{s}_t$.

The objective of the trading strategies developed in this paper is thus to maximize the expected home currency (dollar) return (after transaction costs) using the model of § 3.1.

3.4 Evaluation

Since we do not have an explicit probabilistic model of how exchange rates evolve, we adopt the familiar approach of dividing our data series into an *in-sample* region, over which we optimize the performance of a candidate trading strategy, and an *out-of-sample* region, where the strategy is ultimately tested.

4 Applying RL to the Technical Trading Problem

The ultimate goal of reinforcement learning based trading systems is to optimize some relevant measure of trading system performance such as profit, economic utility or risk-adjusted return. This paper follows the approach of [7] which is summarised here.

Reinforcement learning systems consist of an *agent* interacting with an *environment.* At each time step t the agent *perceives* the state of the environment $s_t \in \mathcal{S}$ and chooses an *action* $a_t \in \mathcal{A}$ from the set of available actions in state s_t. As a consequence of this action the agent observes the new state of the environment s_{t+1} and receives a *reward* r_t. This can be defined as a *dynamic programming* problem where the objective is to find the *policy* π (state to action mapping) that maximises the *optimal value function* V^* given by

$$V^*(s) = \max_a \mathbb{E}\left\{\boldsymbol{r}_{t+1} + \gamma V^*(\boldsymbol{s}_{t+1}) | \boldsymbol{s}_t = s\right\}, \tag{4}$$

where γ is the *discount factor* representing the preference given to immediate over future rewards.

The value of state s can be considered in terms of the values of each action a that can be taken from that state assuming that policy π is followed subsequently. This value Q^* is referred to as the Q*-value* and is given by

$$Q^*(s, a) = \mathbb{E}\{\boldsymbol{r}_{t+1} + \gamma \max_{a'} Q^*(\boldsymbol{s}_{t+1}, a') | \boldsymbol{s}_t = s, \boldsymbol{a}_t = a\}. \tag{5}$$

The optimal value function expresses the obvious fact that the value of a state under an optimal policy must equal the expected return for the best action from that state, *i.e.*

$$V^*(s) = \max_a Q^*(s, a).$$

The functions Q^* and V^* provide the basis for learning algorithms expressed as solutions of Markov decision problems.

We use Watkins's Q*-learning algorithm* [25] that estimates the Q*-value function* using data from the previous learning episode. The Q-learning *update* is the backward recursion

$$Q(s_t, a_t) \leftarrow Q(s_{t_c}, a_{t_c}) + \alpha[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_{t_c}, a_{t_c})], \tag{6}$$

where the current *state-action pair* $(s_t, a_t) := (s_{t_c}, a_{t_c})$, that from the previous learning episode. At each *iteration* (episode) of the learning algorithm, the action-value pairs associated with all the states are updated and over a large number of iterations their values converge to those optimal for (5) [24].

For our trader the state s_t is the market state as defined by the technical indicators and the set of actions $\mathcal{A}$ in the *2-state system* is whether to take a long or short position in the foreign currency (and is not state dependent). In the *3-state system* a neutral position is also a possible action in all states. Following Maes and Brookes' [15] suggestion that immediate rewards are most effective the reward function r_{t+1} is the differential change in value of the agent's portfolio from time t to $t + 1$.

5 Evolutionary Reinforcement Learning Approach

In [8], it was demonstrated that a Q-Learning based system suffers from overfitting the in-sample dataset. Its in-sample performance was significantly superior to that of the genetic algorithm while its performance out-of-sample tended to be inferior. It was therefore clear that the inputs to the RL system needed to be constrained and the notion of a hybrid evolutionary reinforcement learning system was thus introduced. The role of the GA here is to choose some optimal subset of the underlying indicators that the RL system will then use.

The form of GA utilised is the binary string form due to Holland [9]. Each bit in the bitstring represents whether or not the corresponding indicator is being fed into the RL. An initial population of 50 was used and the GA was allowed to evolve for 100 generations. Selection is based on the *roulette wheel* approach. However across every generation we also introduce *elitism*. Thus a number of the top individuals in each population are allowed to survive into the next generation.

With regards to fitness evaluation, the in-sample period was broken down into 8 months of true in-sample data with a further 4 months of data in the *evaluation* period which is used to evaluate individuals within the GA's population of potential solutions. The return over this second period is used as the fitness function of the GA. Once the optimal bitstring is found, the subset of indicators that the bitstring represents is fed into the RL system described earlier (see Figure 1).

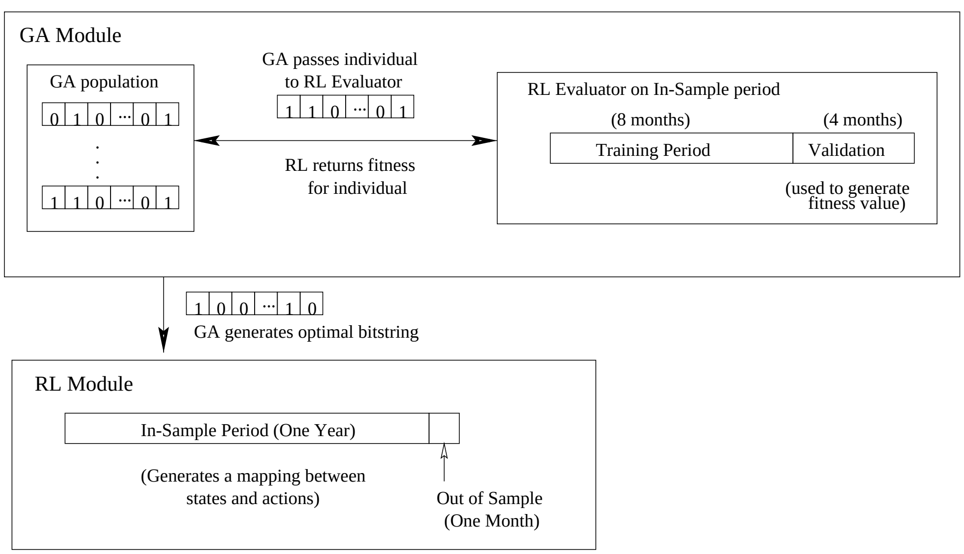

Figure 1: Hybrid System

6 Numerical experiments

The results reported here were obtained by applying the approaches described above to GBPUSD, USDCHF and USDJPY midpoint exchange rate data of 1 minute frequency from January 1994 to January 1998 using a moving window of 1 year for training (fitting) followed by 1 month out-of-sample testing (cf. [7]).

There are several issues that we wish to highlight in these results. Primarily, we try to answer the question as to whether or not using evolutionary learning as part of the credit assignment approach in the reinforcement learning framework improves performance of the system. The resolution of this question is important both in shedding light on the development of successful trading systems and in highlighting basic research issues of interest to the RL community.

Previous work by the authors [8] showed that although the RL consistently outperformed the GA at no slippage; once slippage was introduced the RL system showed classic signs of overfitting. We therefore felt that we could improve upon the original results by constraining the inputs fed into the RL system and this led to the introduction of the hybrid approach discussed above.

At no slippage all four systems tested are able to trade profitably consistently at a 15 minute trading frequency (as shown in Figure 2). However it can be seen that the hybrid approach consistently did as well or better than the basic RL. It is important to note that typical transactions costs in the foreign exchange markets faced by traders for market makers are of the order of 2bp. Considering Figure 2 the performance of the two methods start to diverge at 2bp. It is immediately clear that once we introduce slippage the ability of the system to take neutral positions becomes important. Figure 3 also demonstrates that at lower frequencies this is also a desirable property.

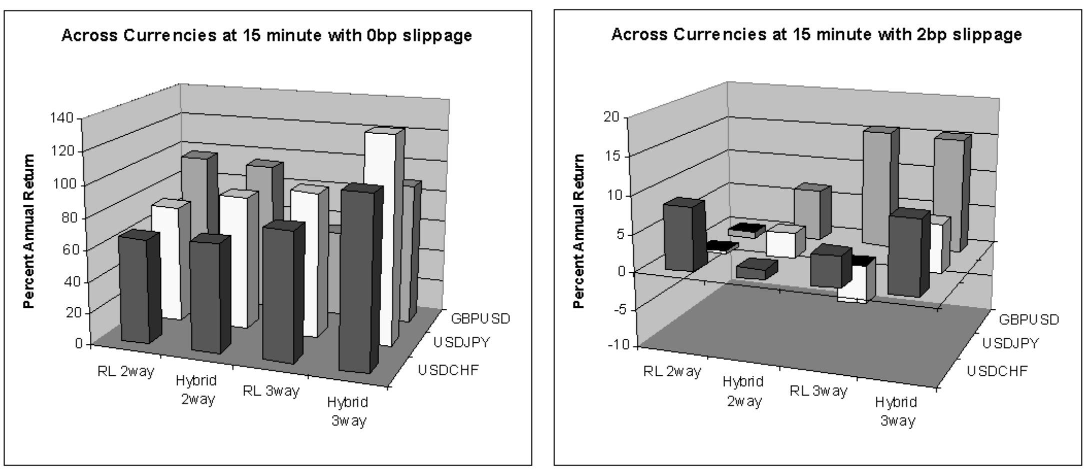

Figure 2: 15 minute trading frequency at 0bp & 2bp across methods and currencies

More importantly however we consider the performance of the hybrid systems compared to their respective standard RL systems. At 2bp slippage at both the 15 minute frequency (Figure 2) and the 1 hour frequency (Figure 3) it is clear that the hybrid systems consistently outperform the standard RL. Furthermore at lower frequency trading (consider Figure 4) performance of these systems is attenuated. Our previous results [8] demonstrated that frequency selection is best left to the learning algorithm which adapts the trading frequency to the slippage rather than forcing a user-constrained choice. The algorithms studied in [8] in fact adapt remarkably well as slippage increases. In general, the results hold across all three currency pairs examined.

The Sharpe ratio is a measure commonly used to evaluate portfolio models given by

$$\frac{\hat{\mu}_{R_{\text{month}}}}{\hat{\sigma}_{R_{\text{month}}}}, \tag{7}$$

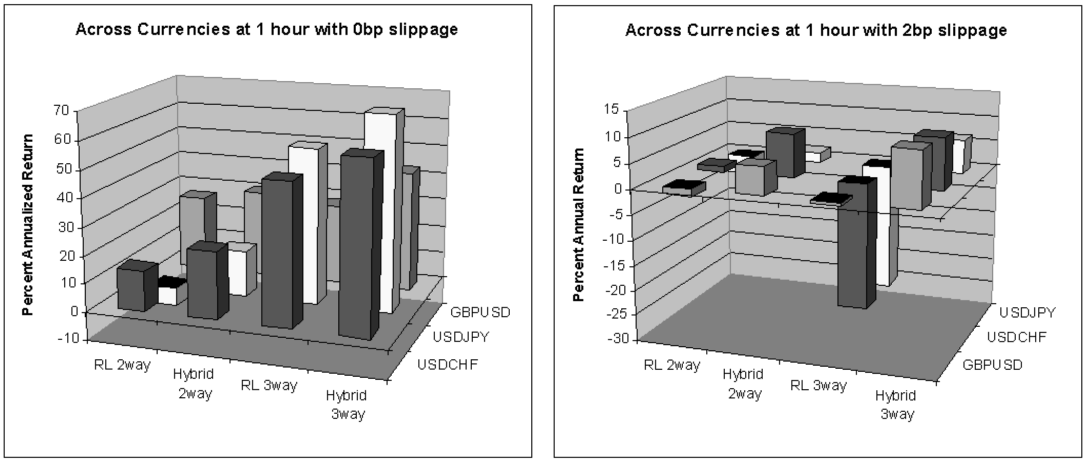

Figure 3: 1 hour trading frequency at 0bp & 2bp across methods and currencies

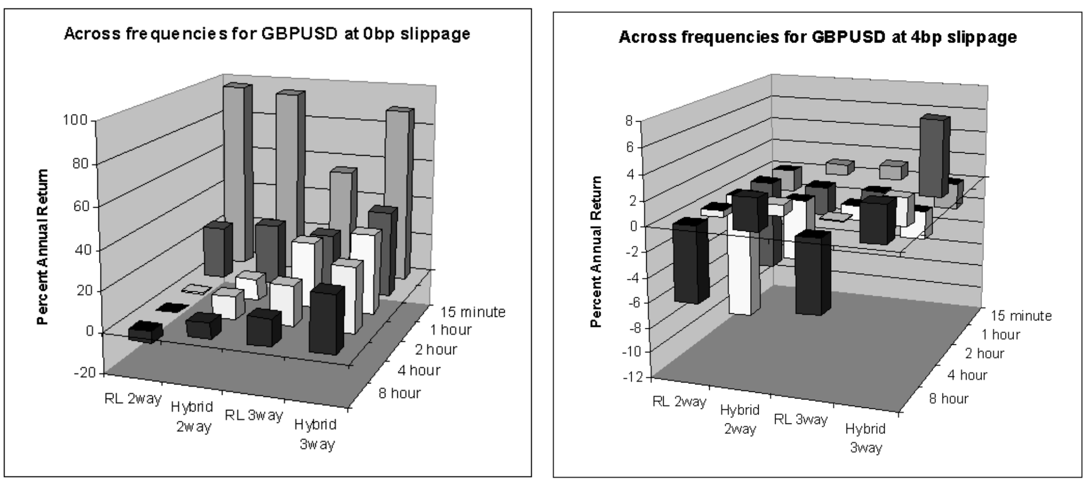

Figure 4: Across frequencies: GBPUSD at 0bp & 4bp

where $\hat{\mu}_{R_{month}}$ and $\hat{\sigma}_{R_{month}}$ denote the mean and standard deviation of monthly out-of-sample returns over the test period of 36 months. The Sharpe ratios shown in Table 1 demonstrate that on the dataset used we are able to gain substantial risk-adjusted returns up to and including a slippage value of 2bp. At 4bp the results were mixed and by 10bp a trend was no longer visible - although there remained pockets of profitability.

Having motivated the hybrid (evolutionary RL) approach and demonstrated that it is indeed able to outperform the RL, we now examine the statistical significance of these results. To this end we utilize the following simple non-parametric binomial test [6]. We take as the null hypothesis that out-of-sample cumulative trading profits and losses are periodically sampled from a continuous time stationary ergodic process with state distribution having median zero. Under this null hypothesis, profits and losses are equally likely with probability 1/2. It follows that over n monthly out-of-sample periods the number of profitable months n_+ is binomially distributed with parameters n and $\frac{1}{2}$. We therefore test the two-tailed alternative hypothesis that median profit and loss is non-zero with the statistic n_+.

The significance of our results is given in Table 2. (Note that significance levels were not included for the 10bp case as these results were uniformly not significant.) To illustrate how these

Table 1: Out-of-sample annualized Sharpe ratios - 15 minute trading

	GBPUSD	USDCHF	USDJPY
RL 2 state - 0bp	2.82	1.28	0.82
RL 3 state - 0bp	1.62	2.08	1.36
Hybrid 2 state - 0bp	2.26	1.32	1.24
Hybrid 3 state - 0bp	**2.21**	**2.22**	**1.82**
RL 2 state - 2bp	-0.04	0.24	0
RL 3 state - 2bp	0.76	0.07	-0.07
Hybrid 2 state - 2bp	0.45	0.15	0.12
Hybrid 3 state - 2bp	**0.7**	**0.25**	**0.13**
RL 2 state - 4bp	-0.21	-0.22	-0.01
RL 3 state - 4bp	0.05	0.18	0.09
Hybrid 2 state - 4bp	0.2	0.11	0
Hybrid 3 state - 4bp	**-0.04**	**0.16**	**0.12**
RL 2 state - 10bp	0	-0.17	-0.17
RL 3 state - 10bp	0.06	0.07	0.08
Hybrid 2 state - 10bp	-0.02	-0.39	-0.07
Hybrid 3 state - 10bp	**-0.01**	**0.23**	**0**

numbers were calculated, consider the monthly out-of-sample returns shown in Figure 5. We find that with n=36, n_+=22, the p-value is 90.88%, giving us significance at the 10% level. Considering the significance levels given for the 15 minute trading frequency in Table 2, we can see that the hybrid approach shows significant promise. When we consider lower frequencies we find returns are more volatile and the results are no longer as consistently significant. These values have not been included for lack of space. The significant profitability of the hybrid 3 state system at 4 basis points in spite of a negative (cumulative) return shows the importance of risk management to stop a small number of large drawdowns. (It also suggests the use of more powerful nonparametric tests for profitability which we are currently developing that take account of signed return *magnitudes*.)

Table 2: Significance level of the 15 minute trading results (N/S=Not Significant)

	GBPUSD	USDCHF	USDJPY
RL 2 state - 0bp	0.01%	0.01%	0.01%
Hybrid 2 state - 0bp	0.01%	0.01%	0.01%
RL 3 state - 0bp	0.01%	0.01%	0.01%
Hybrid 3 state - 0bp	**0.01%**	**0.01%**	**0.01%**
RL 2 state - 2bp	N/S	N/S	N/S
Hybrid 2 state - 2bp	5%	N/S	25%
RL 3 state - 2bp	0.1%	10%	25%
Hybrid 3 state - 2bp	**0.01%**	**1%**	**5%**
RL 2 state - 4bp	N/S	N/S	N/S
Hybrid 2 state - 4bp	N/S	N/S	25%
RL 3 state - 4bp	N/S	20%	15%
Hybrid 3 state - 4bp	**10%**	**10%**	**5%**

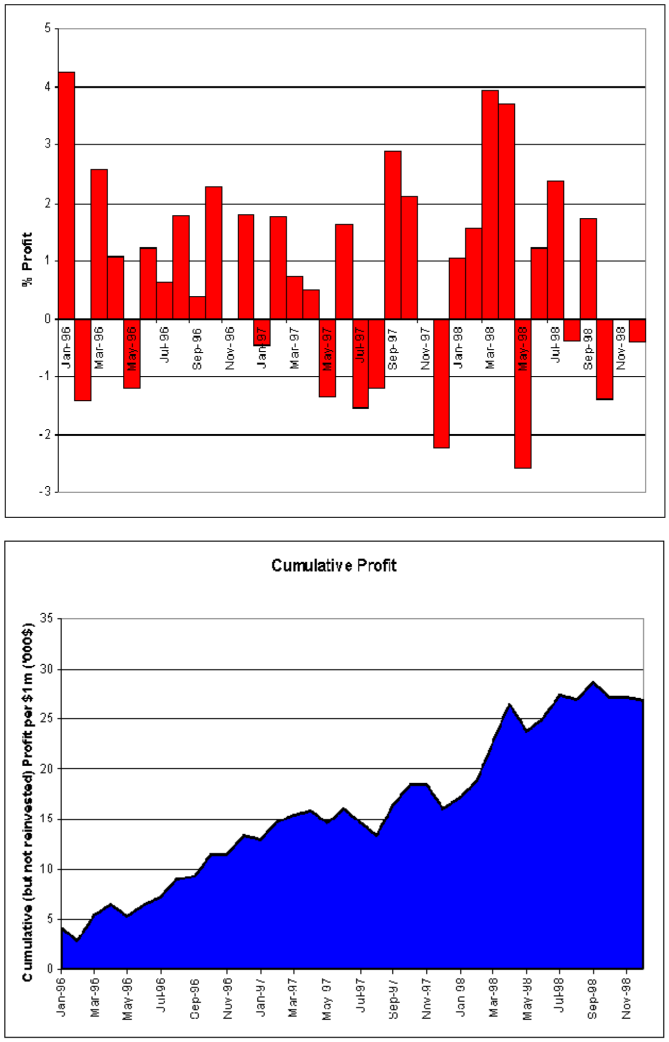

Figure 5: Evolutionary RL at 15 minute trading with 2bp slippage

7 Discussion and Future Work

We have shown that the use of computational learning techniques in high frequency foreign exchange trading shows significant promise. The fact that the techniques investigated here return positive results both in-sample and out-of-sample implies that there is useful information in the technical indicators that can be exploited. This is consistent with the tenets of technical analysis and contradictory to the Efficient Market Hypothesis. Furthermore, the evolutionary RL's relatively good out-of-sample performance demonstrates that using a combination of technical indicators leads to better performance than using the individual indicators themselves. In fact, Dempster and Jones [6, 10] demonstrate that these indicators are largely unprofitable on a different data set when considered in isolation. At low slippage values, annual returns of 10-20% are not uncommon. However, these slippage values are only typically available to market makers. Investment managers for example, who more typically face slippage of up to 10bp, would be unable to utilize the methods outlined here in the manner described. In general, we have shown that by constraining the inputs to the RL system using a GA, we have improved the performance of the underlying system. Results were furthermore statistically significant and similar across currencies.

The next step is to consider different optimization functions, in particular, exploring risk adjusted rather than raw return and overlaying the system with cash management such as stop losses. Another current avenue of research is the use of alternative reinforcement learning approaches such as the *recurrent* reinforcement learning approach described by Moody [16]. We are also exploring the incorporation of trade volume data into the learning algorithms. Generalizing the work to more sophisticated agents that trade several currencies simultaneously is currently being considered as well.

References

[1] S. Achelis, *Technical Analysis from A to Z*, McGraw-Hill, New York, 2001.

[2] S. Chen and W. Lee, *Option pricing with gas: A second report*, IEEE International Conference on Neural Networks, 1 (1997), pp. 21–25.

[3] N. Chidambaran, C. Jevons Lee, and J. Trigueros, *An adaptive evolutionary approach to option pricing via genetic programming*, Proceedings of the 6th International Conference on Computational Finance, (1999).

[4] M. A. H. Dempster and C. M. Jones, *Can technical pattern trading be profitably automated: 2. The head and shoulders.* Working Paper No. 12/99, Judge Institute of Management, University of Cambridge, 1999.

[5] ———, *The profitability of intra-day FX trading using technical indicators.* Working Paper No. 35/00, Judge Institute of Management, University of Cambridge, 2000.

[6] ———, *A real-time adaptive trading system using genetic programming*, Quantitative Finance, 1 (2001), pp. 397–413.

[7] M. A. H. Dempster, T. W. Payne, Y. Romahi, and G. Thompson, *Computational learning techniques for intraday fx trading using popular technical indicators*, IEEE Transactions on Neural Networks, Special Issue on Computational Finance, 12 (2001), pp. 744–754.

[8] M. A. H. Dempster, T. W. Payne, and Y. S. Romahi, *Intraday FX trading: Reinforcement learning vs evolutionary learning.* Working Paper No. 23/01, Judge Institute of Management, University of Cambridge, 2001, December 2001.

[9] J. H. Holland, *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbour, MI, 1975.

[10] C. M. Jones, *Automated technical foreign exchange trading with high frequency data*, PhD thesis, Centre for Financial Research, Judge Institute of Management Studies, University of Cambridge, June 1999.

[11] L. P. Kaelbling, M. L. Littman, and A. W. Moore, *Reinforcement learning: A survey*, Journal of Artificial Intelligence Research, 4 (1996), pp. 237–285.

[12] R. Levich and L. Thomas, *The significance of technical trading rule profits in the foreign exchange market: A bootstrap approach*, Journal of International Money and Finance, 12 (1993), pp. 451–474.

[13] A. W. Lo and A. C. MacKinlay, *Stock market prices do not follow random walks: Evidence from a simple specification test*, Review of Financial Studies, 1 (1988), pp. 41–66.

[14] A. W. Lo and A. C. MacKinlay, *A Non-Random Walk Down Wall Street*, Princeton University Press, Princeton, NJ, 1999.

[15] P. Maes and R. Brooks, *Learning to coordinate behaviors*, Proceedings of the 8th National Conference on Artificial Intelligence, (1990), pp. 796–802.

[16] J. Moody, L. Wu, Y. Liao, and M. Saffell, *Performance functions and reinforcement learning for trading systems and portfolios*, The Journal of Forecasting, 17 (1998), pp. 441–470.

[17] C. Neely and P. Weller, *Intraday technical trading in the foreign exchange market.* Working Paper 99-016A, Federal Reserve Bank of St. Louis, November 1999.

[18] C. J. Neely, *Technical analysis in the foreign exchange market: a layman's guide*, in Federal Reserve Bank of St. Louis Review, September/October 1997, pp. 23–38.

[19] C. J. Neely, P. A. Weller, and R. Dittmar, *Is technical analysis in the foreign exchange market profitable? a genetic programming approach*, Journal of Financial and Quantitative Analysis, 32 (1997), pp. 405–426. Also available as Federal Reserve Bank of St. Louis Working Paper 96-006C.

[20] R. Neuneier, *Optimal asset allocation using adaptive dynamic programming*, in Advances in Neural Information Processing Systems, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, eds., vol. 8, The MIT Press, 1996, pp. 952–958.

[21] ———, *Enhancing Q-learning for optimal asset allocation*, in Advances in Neural Information Processing Systems, M. I. Jordan, M. J. Kearns, and S. A. Solla, eds., vol. 10, The MIT Press, 1998.

[22] C. L. Osler and P. H. K. Chang, *Methodical madness: Technical analysis and the irrationality of exchange-rate forecasts*, Economic Journal, 109 (1999), pp. 636–661.

[23] O. V. Pictet, M. M. Dacorogna, B. Chopard, M. Oudsaidene, R. Schirru, and M. Tomassini, *Using genetic algorithm for robust optimization in financial applications*, Neural Network World, 5 (1995), pp. 573–587.

[24] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, The MIT Press, 1998.

[25] C. Watkins, *Learning from Delayed Reward*, PhD thesis, Kings College, University of Cambridge, 1989.