Chapter 1

EVOLUTIONARY DECISION TREES FOR STOCK INDEX OPTIONS AND FUTURES ARBITRAGE

How Not To Miss Opportunities

Sheri Markose
Economics Department and Institute of Studies in Finance
University of Essex
scher@essex.ac.uk

Edward Tsang
Computer Science Department
University of Essex
edward@essex.ac.uk

Hakan Er
Economics Department
University of Essex
her@essex.ac.uk

Abstract **EDDIE-ARB** (**EDDIE** stands for Evolutionary Dynamic Data Investment Evaluator) is a genetic program (**GP**) that implements a cross market arbitrage strategy in a manner that is suitable for online trading. Our benchmark for **EDDIE-ARB** is the Tucker (1991) put-call-futures (**P-C-F**) parity condition for detecting arbitrage profits in the index options and futures markets. The latter presents two main problems. (*i*) The windows for profitable arbitrage opportunities exist for short periods of one to ten minutes. (*ii*) From a large domain of search, annually, fewer than 3% of these were found to be in the lucrative range of £500-£800 profits per arbitrage. Standard *ex ante* analysis of arbitrage suffers from the drawback that the trader awaits a contemporaneous signal for a profitable price misalignment to implement an arbitrage in the same direction. Execution delays imply that this naive strategy may fail. A methodology of random sampling is used to train **EDDIE-ARB** to pick up the fundamental arbitrage patterns. The further novel aspect of **EDDIE-ARB** is a constraint satisfaction feature supplementing the fitness function that enables the user to train the **GP** *how not to miss opportunities* by learning to satisfy a minimum and maximum set on the number of arbitrage opportunities being sought. Good **GP** rules generated by **EDDIE-ARB** are found to make a 3-fold improvement in profitability over the naive ex ante rule.

Keywords: Genetic Programming, Machine Learning, Genetic Decision Trees, Arbitrage, Options, Futures, Constraint Satisfaction.

Introduction

This paper presents an extension of a Genetic Programming (**GP**) tool called **EDDIE** (Evolutionary Dynamic Data Investment Evaluator) developed by Tsang et. al.(1998) at the University of Essex for purposes of financial forecasting and engineering (see, Tsang et. al. 2000, for a survey). The application we study here is of **EDDIE** within the context of cross market arbitrage, specifically in the FTSE-100 index options and futures, in a manner that is suitable for online trading. Hence, the **GP** merits the acronym of **EDDIE-ARB**. There are two main problems in conducting the FTSE-100 index market arbitrage in real time. (i) The windows for profitable arbitrage opportunities exist for short periods of one to ten minutes, and (ii) from a large domain of search annually fewer than 3% of these were found to be in the lucrative range of £400 - £800 profits per arbitrage. The static representation of an arbitrage position, as is well known, is based on contemporaneous price signals and entails a riskless position where no explicit role for forecasting appears to exist. However, this standard ex ante analysis of arbitrage suffers from the drawback that the trader awaits a contemporaneous signal for a profitable price misalignment to implement an arbitrage in the same direction. Execution delays imply that this so called naive strategy may fail. Hence, a forecasting tool is necessary to evaluate market conditions and make trading recommendations in anticipation of price misalignments in a way that is robust to execution delays. The recognition or detection of arbitrage positions can be aided by known formulaic conditions derived from theory. But this is not sufficient if temporary market price anomalies exist. **GP**s that perform better than the naive strategy can be trained to detect arbitrage positions from a range of economic variables determining prices in the two markets. However, the problem we encountered is that arbitrage opportunities with sizeable profits though they exist were infrequent. They are as few as only two on average in any trading day. Thus, for the arbitrageur to net a sizeable annual income, the crucial aspect for **EDDIE-ARB** to be successful is that not only must it give correct predictions whilst making recommendations to trade; **EDDIE** must not miss a requisite number of arbitrage opportunities!

EDDIE is built on Genetic Programming (Koza, 1992, 1994; Koza et.al., 1996) which is a powerful variant of genetic algorithms (Holland, 1975, Goldberg 1989, Mitchell, 1996). The strengths of **GP** especially for financial applications arise from its use of decision tree representations instead of strings of chromosomes. The decision tree representations referred to as Genetic Decision Trees (**GDT**s) will be able to handle rule sets of variable size [1] and these rules are easy to understand and evaluate by human users. This makes this approach more attractive than neural networks , most of which are black boxes (Goonatilake and Treleaven, 1995).

The use of Genetic algorithms (**GA**s) and **GP**s is now widespread in financial markets for purposes of forecasting and financial engineering. The efficacy of these adaptive computational methods in financial forecasting and financial engineering is seen to lie in their flexibility to account for potentially complex non-linear relationships which are not captured well by traditional linear and parametric methods. As excellent surveys already exist on application of **GA**s and **GP**s in financial applications, we will only briefly list some of our precursors here. Bauer (1994) reported his **GA** intelligent systems which aimed at finding tactical market timing strategies; Allen & Karjalainen (1995) applied the **GP** technique to find profitable technical trading rules for trading on the S&P 500 index; Chen & Yeh (1997) attempted to formalize the notion of unpredictability in the efficient market hypothesis in terms of search intensity and chance of success in the search conducted by genetic programming; Mahfoud & Mani (1996) presented a new genetic-algorithm-based system and applied it to the task of predicting the future performances of individual stocks; Neely et al. (1997) and Oussaidene et al. (1997) applied genetic programming to foreign exchange forecasting and reported some success.

In earlier work (Tsang et al. 1998, Li & Tsang, 1999a and Li & Tsang, 1999b) **EDDIE** was trained primarily in the art of financial forecasting with different objectives to be satisfied. Initially, **EDDIE**'s focus was on predicting whether a price series will increase by r% or more within the

[1] Typically, in contrast genetic algorithms that operate with strings use strings of fixed lengths.

next n periods. **EDDIE**'s performance was found to compare favourably with random rules, commonly used individual technical rules and C4.5 rule sets with respect to prediction accuracy and average annualised rate of return.

The objective in this paper is to develop and implement **EDDIE-ARB** on intra daily tick data for cross market stock index arbitrage in a manner that is suitable for online trading when windows of profitable arbitrage opportunities exist for short periods from one minute to ten minutes. Recent work by Markose and Er (2000) on the FTSE-100 stock index options and futures clearly indicates that these contracts with even fewer than 40 days to maturity can present arbitrage opportunities. A methodology of random sampling is used to train **EDDIE-ARB** to pick up the fundamental arbitrage patterns. As the arbitrageur maximizes return by exploiting as many profitable arbitrage opportunities and avoiding as many positions that are loss making, we will show how the fitness function of **EDDIE -ARB** will enable the arbitrageur to tune it to obtain the most favourable trade off. In other words, in arbitrage activity avoiding positions that are loss making or minimizing rate of failure **(RF)** alone which was the focus of earlier work by Tsang et. al.(1998) is not sufficient. Thus, a novel aspect of **EDDIE-ARB** is a constraint satisfaction feature supplementing the fitness function that enables the user to train the **GP** how not to miss opportunities by learning to satisfy a minimum and maximum set on the number of arbitrage opportunities being sought. Historical sample data on arbitrage opportunities enables us to set these minimum and maximum bounds. Good **GP/GDT** rules generated by **EDDIE-ARB** are found to make a 3-fold improvement in profitability over the naive ex ante rule.

The rest of the paper is organized as follows. In Section 2, the methodology of Put -Call- Futures index options and futures arbitrage is given This determines the preprocessing requirements of intra daily tick data for purposes of training and testing the **EDDIE -ARB** program to learn from contemporaneous conditions for arbitrage opportunities and to predict these from up to ten minutes in advance. In Section 3, we give some details of the **EDDIE -ARB** program, in particular, how the fitness function has to be supplemented and fine tuned to obtain the favourable trade off between the search intensity for arbitrage opportunities and loss making recommendations. Section 4 reports the empirical results of the application of **EDDIE -ARB** to stock index options and futures arbitrage. Some robustness tests are done to see how well the previously trained **GDT** rules performed in a period when with LifeConnect and electronic trading greatly increased the turnover in FTSE-100 index futures and options markets.

1. Methodology of Put-Call-Futures Stock Index Arbitrage

1.1. Cross Market Arbitrage on the Stock Index

As the FTSE 100 spot index has a stock index futures and a European style index option traded on it, we will briefly outline why the most cost effective arbitrage is one that bypasses the cash/spot leg and involves only the two stock index derivatives with the same maturity date. This is illustrated in Figure 1 along the arrow C between the two derivatives on the underlying spot index. Figure 1 also indicates the three important equilibrium pricing and arbitrage criteria that relate the index options and futures on the same underlying stock index.

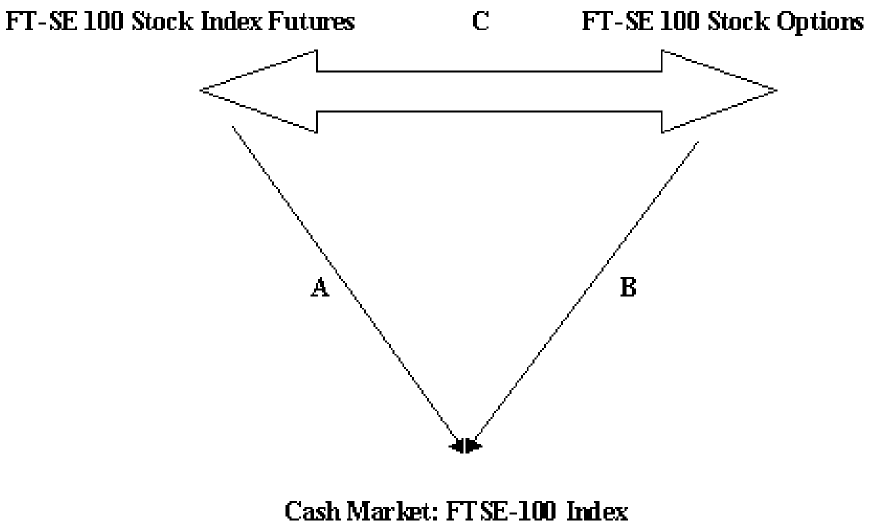

Figure 1.1. Cross Market Arbitrage

A: Index Arbitrage based on the cost of carry fair futures formula. B: Option-Cash Index Arbitrage based on Stoll (1969)Put-Call Parity condition. C: Index Options-Futures Arbitrage based on Tucker(1991) Put-Call- Futures Parity condition.

Along arrow A in Figure 1, which relates the FTSE-index futures to the spot index, we have the so called index arbitrage based on a theoretical cost of carry formula which is used to detect arbitrage profit opportunities (eg. Modest and Sunderasan 1983, Cornell and French 1983, Yadav and Pope1990). However, there are high costs of trading in the stock index portfolio and in particular problems of liquidity can arise when shorting stock in the conduct of a short index arbitrage strategy when the futures price trades at a discount to the spot index. The equilibrium pricing relationship between the index option and the stock index, defined by arrow B in Figure 1, is based on the Stoll (1969) Put-Call (P-C) parity relationship. Here, typically the option price is derived from the underlying stock index. However, it is well known from studies on the efficiency of index options (see, Evnine and Rudd [2] ,1985, Gemmill, 1993 ,Gwilm and Buckle,1999), that the high costs of hedging index option positions with a portfolio based on the constituent shares of the spot index may lead to a number of problems for the tests of market efficiency of index options. In addition, Fung and Chan (1994) have shown that the distribution of stock dividends which affect arbitrage strategies along the arrows A and B do not affect the **P-C-F** arbitrage along arrow C of Figure 1. For the above reason of the high cost of using the spot index in an arbitrage, it is widespread that arbitrage in stock index options follows arrow C in Figure 1 and bypasses the cash/spot index leg.

The Tucker(1991) result that combinations of put and call options with the same exercise price can replicate payoffs of futures positions with the same expiration and on the same cash asset underpins the put-call-futures (**P-C-F**, for short) parity relationship. Thus, arbitrage strategies based on the **P-C-F** parity (arrow C in Figure 1), involves time synchronized triplets of prices on a put, a call and a futures contract as specified above. A short futures position can be replicated by a combination of a short position in a call index option and a long one in a put index option. Conversely, a long futures position is equivalent to a combination of a long call and a short put position. The existence of these synthetic futures contracts involving combinations of options generates conditions for risk free arbitrage profits for the futures and options markets on the same underlying cash asset. For these reasons the put-call-futures parity relationship has also become the basis of recent tests for efficiency in the index options prices (Lee and Nayar, 1993, Fung et. al 1997 and Bae et. al.1998). We will now turn to the construction of a risk free **P-C-F** arbitrage position.

1.2. P-C-F Short Hedge Arbitrage

A risk free arbitrage portfolio can be constructed by combining a short futures contract and a long synthetic futures position by buying

[2]Evnine and Rudd(1985) note that unlike the option on a stock, an option on an index involves more complex procedures for arbitrage, 'it is conceivable that arbitrage forces are not as powerful with index options and hence index options may display greater pricing errors than stock options'.

a call, shorting a put and by borrowing the present discounted value of the futures price and lending the same for the exercise price. Panel A in Table 1 shows how a zero net return is accomplished by this short hedge arbitrage strategy should S_T>E or S_T <E. The upper bound of the FTSE-100 index futures bid price denoted by F_{bt} is given by

$$F_{bt}e^{-r_a\tau} \leq C_{at} - P_{bt} + Xe^{-r_b\tau} + TC \tag{1}$$

Here, $\tau = (T - t)$ is the remaining time to maturity; C_{at} is the call premium at the ask, P_{bt} is the put premium at the bid and TC denotes the transactions costs (that will be specified). Note, the interest rate on the futures price is the offer/ask rate as this amount has to be borrowed by the arbitrageur and that on the exercise price is the bid interest rate as it has to lent. If the condition in (1) is violated then the arbitrageur by definition will make a risk free profit equal to

$$[F_{bt}e^{-r_a\tau} - (C_{at} - P_{bt} + Xe^{-r_b\tau}) + TC] > 0 \tag{2}$$

by shorting the futures and by creating the synthetic long futures given on the R.H.S of (1).

1.3. P-C-F Long Hedge Arbitrage

Likewise we have a risk free arbitrage portfolio by combining a long futures contract and a short synthetic futures position obtained by shorting a call at the bid, being long in the put at the ask and by lending the present discounted value of the futures price and borrowing the same for the exercise price. Panel B shows how a zero net return is obtained by this long hedge arbitrage should S_T>E or S_T<E. The lower bound of the ask futures price is given by

$$F_{at}e^{-r_b\tau} \geq C_{bt} - P_{at} + Xe^{-r_a\tau} - TC. \tag{3}$$

If the present value of the ask futures price is less than the R.H.S of (3), then the arbitrageur buys the futures and sells the portfolio on the R.H.S of (3). This nets a risk free profit equal to

$$[(C_{bt} - P_{at} + Xe^{r_a\tau} - TC) - F_{at}e^{-r_b\tau}]. \tag{4}$$

Note, here that the bid interest rate applies to the amount for the futures price that is lent in the arbitrage and the ask/offer interest rate applies to the value of exercise price as that has to be borrowed. As trade prices are not specified in terms of whether a buy or sell is involved the condition for a profitable short arbitrage is given as

$$F_te^{-r\tau} - [C_t - P_t + Xe^{-r\tau}] - TC > 0. \tag{5}$$

Table 1.1.

A:		**Cash Flows at Expiration (at T)**	
P-C-F Short Arbitrage TRANSACTIONS UNDERTAKEN	**Cash Flows Today(at t)**	$S_T > X$	$S_T < X$
Short a futures at F_t	0	$F_t - S_T$	$F_t - S_T$
Borrow $F_t e^{r(t-T)}$	$+F_t e^{r(t-T)}$	$-F_t$	$-F_t$
Long call	$-C_t$	$S_T - X_t$	0
Short put	$+P_t$	0	$-(X - S_T)$
Lend $Xe^{r(t-T)}$	$-Xe^{r(t-T)}$	$+X$	$+X$
	$F_t e^{r(t-T)}$ -($C_t - P_t + Xe^{r(t-T)}$)	0	0
B:		**Cash Flows at Expiration (at T)**	
P-C-F Long Arbitrage TRANSACTIONS UNDERTAKEN	**Cash Flows Today(at t)**	$S_T > X$	$S_T < X$
Long a futures at F_t	0	$S_T - F_t$	$S_T - F_t$
Lend $F_t e^{r(t-T)}$	$-F_t e^{r(t-T)}$	$+F_t$	$+F_t$
Short call	C_t	$-(S_T - X)$	0
Long put	$-P_t$	0	$X - S_T$
Borrow $Xe^{r(t-T)}$	$+Xe^{r(t-T)}$	$-X$	$-X$
	($C_t - P_t + Xe^{r(t-T)}$) -$F_t e^{r(t-T)}$)	0	0

Here, r is the mid interest rate. In trade data with long **P-C-F** arbitrage, the condition for its profitability is given as

$$[C_t - P_t + Xe^{-r\tau}] - Fe^{-r\tau} - TC > 0. \qquad (6)$$

In the following, analysis is confined to the short **P-C-F** arbitrage.

2. Data, Transactions Costs and Ex Post Evidence of Efficiency Violations

2.1. LIFFE Intradaily Tick Data for FTSE-100 Index Options and Futures

We first use the intraday historical tick data on the FTSE-100 European style index option traded on the LIFFE from March 1, 1991 to

June 30, 1998. Intraday tick data for a further period from June 30,1998 to March 30, 1999 was earmarked to do additional robustness tests to see if the introduction of the LIFFEConnect electronic trading had changed the fundamental conditions for the **GP**.

The tick data records contain both two way bid/ask quote data and the transactions/trade price data. Each record is time stamped with the following information: (i) option identity ,ie. whether a call or put ;(ii) the maturity date; (iii) the exercise price;(iv) the price of contract (bid and ask for quoted prices and transaction/trade price if indicated as a trade record);(v) the time synchronized underlying FTSE-100 stock index price.

The risk free interest rate used for borrowing funds in the hedge portfolio is the London Interbank Offer Rate(LIBOR) maturing on the day closest to the expiration date of the option. The tick data on the FTSE 100 is obtained from Data Stream. The bid or lending interest rate is conventionally taken to be 1/8 of the LIBOR rate. The mid interest rate is the mean of the above two rates.

To process the tick data for arbitrage relating to the put-call-futures parity, we follow a three way matching criteria for puts, calls and futures with the same nearby maturity. For trade data, all calls and puts with the same exercise price and traded within the same minute are matched. This pair is matched further with a futures contract traded within a minute of the time stamp of the call-put pair. In the LIFFE tick trade data from January 1991 to June 1998, 8073 short arbitrage trade price triplets satisfying condition (5) with zero transactions costs were found.

2.2. Transactions Costs

It is conventional for transactions costs for arbitrage to be estimated for at least three categories of traders- the private investor, institutional investors and for broker/market makers. The first two incur commission/broker fees in addition to the exchange fees. Further, there are the bid-ask spreads, these have been estimated to be approximately £57 [3]. This study focuses on the arbitrage profits that can accrue to the broker/market maker who is estimated to have transactions costs of less than .1 % value of the value FTSE100 index futures contract. The latter works out to about £60 per **P-C-F** arbitrage. Note that analysis of marking to market and the management of margin requirements which is important in the practical implementation of the **P-C-F** arbitrage is

[3]See, Markose and Er (2000) for a full discussion on the derivation of the bid ask spreads in the index option and futures markets.

omitted here. The assumption is that the arbitrageur can post treasury bills for the margin requirements and hence does not incur interest rate costs on this.

2.3. Ex Post Efficiency Violations

Table 2 gives the ex post analysis of efficiency violations for short arbitrage positions [4] using the 8073 **P-C-F** price triplets for the sample period (1991-1998) at costs appropriate to the market maker/broker. Some 42% of the sample yields profitable short **P-C-F** arbitrage opportunities. The profits reported are those that accrue if the arbitrageur could have obtained as quotes the trade prices [5] recorded at these points in time.

Table 1.2. Ex Post Analysis of Efficiency Violations For Short Arbitrage At Transactions Costs of .1% of FTSE-100 Index

	Number of Mispricings Short Arbitrage	Ex PostProfits Mean (Eq. 5) Short Arb. (£s)
All	3357	368.89
	42%*	(29.79)
50-80 Days Before	1145	496.87
	14% *	(67.79)
20-50 Days Before	1206	316.55
	15%*	(45.65)
Spot Month	1006	285.96
	12%*	(29.02)

N.B: ( ) gives the 5% confidence interval; * gives % of total sample of 8073

In Table 2, we see that on average for all periods to maturity the profits from the arbitrage is substantial and statistically significant. The spot month nets on average £285 and the 20-50 day period to maturity and further yields over £316. We will now turn to the question of whether arbitrageurs, in a real time setting, can correctly identify and exploit profitable short arbitrage opportunities shown in Table 2 .

[4]Markose and Er(2000) find that the P-C-F short arbitrage shows more efficiency violations than the long arbitrage.

[5]Markose and Er (2000) discuss issues relating to whether these reported profits in trade prices are robust to the well known direction of trade problem.

3. Ex Ante Analysis of Arbitrage Profits and Application of EDDIE-ARB

3.1. The Objective of EDDIE-ARB

The standard ex ante analysis of arbitrage (see, Bae et. al., 1998) is based on the following scenario. The naive premise is that the arbitrageur waits for a contemporaneous profit signal in the category of either short or long arbitrage (given by equations 1-6) above and then continues with arbitrage trades in the same direction in a given time interval. It will be assumed that after a time execution delay of about a minute, the arbitrageur proceeds to execute his arbitrage position within a time interval of ten minutes. He, thus, faces execution price risk, should the trio of the **P-C-F** prices diverge. The question is can a **GP** be trained to improve on the naive strategy as not every contemporaneous **P-C-F** profit signal will continue to be profitable after the execution delay and more importantly can a profit signal be predicted in advance of the required period.

The case for the mechanization of the complex **P-C-F** arbitrage can easily be made. The domain of search involves four asset prices (put price, call price, futures price, and the interest rate), three maturity dates for the derivatives and over thirty exercise prices for the index options. With data arriving continuously, the detection of contemporaneous profitable arbitrage opportunities with the three way **P-C-F** price matching is a considerable task. With the assumed time delay in execution of an arbitrage from an observed contemporaneous profit signal an effective forecasting tool is needed to assess the success rate of such a strategy. Thus, the objective we set **EDDIE-ARB** in the generation of decision rules is as follows: **at any point in time corresponding with the occurrence of a matched P-C-F price triplet** [6] **, predict which of these have adjacent 10 minute intervals in which profitable arbitrage is possible in a given direction after a one minute execution delay**. Note, the recommended arbitrage positions are judged profitable or not by the criteria given in equations (1-6). However, as indicated in the introduction, **EDDIE- ARB** imple-

[6] Note that here as the prediction of profitable arbitrage in the next ten minutes is not constrained to be one that follows from a profitable contemporary price signal as in the naive rule, the **GP** is in principle trained to anticipate arbitrage opportunities within a given time interval from any point in time. With electronic trading , any observed **P-C-F** arbitrage opportunity can be executed immediately. As quotes do not 'live' longer than 1-3 minutes even with electronic trading, the methodology given here for generating robust **GDT**s that anticipate profitable arbitrage in a given direction with a ten minute lead is advantageous.

ments a trade off between missed profit opportunities and loss making recommendations.

3.2. Background of EDDIE

Like other standard **GPs** , **EDDIE** maintains a population (set) of candidate solutions, each of which is a decision tree for financial forecasting. Candidate solutions are selected randomly, biased by their fitness, for involvement in generating members of the next generation. General mechanisms (referred to as genetic operators, e.g. reproduction, crossover, mutation) are used to combine or change the selected candidate solutions to generate offspring, which will form the population in the next generation. For details of **GAs** and **GPs**, readers are referred to Holland (1975), Goldberg (1989) and Koza (1992).

In the **GP EDDIE**, a candidate solution is represented by a genetic decision tree (**GDT**). The basic elements of **GDTs** are rules and forecast values. A single rule consists of one useful indicator for prediction, one relational operator such as "greater than", or "less than", etc, and a threshold (real value). Such a single rule interacts with other rules in one **GDT** through logic operators such as "Or", "And", "Not", and "If-Then-Else". Forecast values in this example are either a positive position (i.e. positive return within specified time interval can be achievable) or negative position (i.e. negative return within a specified time interval will be achievable).

3.3. The FGP Fitness Criterion

Since **GDTs** are used to predict whether a profitable arbitrage can exist from any point in time and within the next prespecified minutes, the prediction actually can be categorised as a two-class classification problem. Each time point can be classified into either a positive position or a negative position. For each **GDT**, we define **RC** (Rate of Correctness), **RMC** (Rate of Missed Opportunities), and **RF** (Rate of Failure) as its prediction performance criteria. Formula for each criterion is given through a contingency table (Table 3) as follows:

3.4. Linear Fitness Function

In earlier work by Tsang et. al.(1998), the fitness function mainly used in **EDDIE** was the rate of correctness **RC**. By using the following fitness function, the user may satisfy individual objectives by adjusting the weights w_rc, w_rmc and w_rf:

$$f_1 = w\_rc * \mathbf{RC} - w\_rmc * \mathbf{RMC} - w\_rf * \mathbf{RF} \tag{7}$$

Table 1.3. A contingency table for two-class classification prediction problem

Predicted Negative Positions (N_-)	**Predicted Positive Positions (N_+)**	
# of True Negative (**TN**)	# of False Positive (**FP**)	**Actual Positive Positions (O_-)**
# of False Negative (**FN**)	# of True Positive (**TP**)	**Actual Positive Positions (O_+)**

RC=$\frac{TP+TN}{O_+ + O_-}=\frac{TP+TN}{N_+ + N_-}$; **RMC**=$\frac{FN}{O_+}$; **RF**=$\frac{FP}{N_+}$;
Where O_+=FN+TP; O_-=TN+FP;N_-=TN+FN;N_+=FP+TP

It involves three performance values, i.e. **RC**, **RMC** and **RF**, each of which is assigned a different weight. Obviously, the performance of a **GDT** is no longer assessed by **RC** only, but by a synthetic value, which is the weighted sum of its three performance rates. By proper adjustment of the three weights, the experimenter is able to put his emphasis on one aspect of the performance of the **GDT** than on the others. In order to achieve a low **RF**, one may assign a higher value to w_rc and w_rf and a smaller or zero value to w_rmc. To a certain extent, the fitness function $f_{(1)}$ does allow us to reduce **RF**. However, the fitness function in (7) appears to have two drawbacks: 1) **EDDIE**'s performance is very sensitive to the three weights; 2) results are unstable. For example, in one of our series of preliminary experiments in which we used the following three weights:

$w\_rc = 1$; $w\_rmc = 0$ and $w\_rf = \alpha$ where $0 < \alpha \leq 1$.

We found that merely altering α in the linear fitness function did not generate reliable performance of the generated **GDT**s. In some cases even a small positive α resulted in **GDT**s that did achieve a lower **RF** (even zero over training period) by making no positive recommendations over the test period. This was probably due to over-fitting. In contrast, a reduction of α from an α as high as 0.62 resulted in generating **GDT**s that it did not show any improvement on **RF**. We refer to this as the no-effect problem. For example, among 10 runs, only two runs generated a **GDT** that predicted a few correct positive positions in the test period. The remaining 8 runs showed either the over-fitting or no-effect problem.

3.5. Putting constraints into GP

We can further improve the linear fitness function $f_{(1)}$ by introducing constraints into it. We introduce a new parameter to **EDDIE**, $\Re = [P_{min}, P_{max}]$, which defines the minimum and maximum percentage of recommendations that we instruct **EDDIE** to make in the training data (with the assumption that the test data exhibits similar characteristics, an assumption that is made in most machine learning methods). We denote the new fitness function by $f_{(2)}$.
Choosing appropriate values for $\Re$ and the weights for $f_{(2)}$ remains a non-trivial task, which we approached by trial and error. When appropriate parameters were chosen, **EDDIE** managed to reduce **RF** and avoid the over-fitting and no-effect problems. In particular, the intensity of search by **EDDIE-ARB** so as not to miss arbitrage opportunities was set at the minimum and maximum of the number of profitable opportunities found on average in the sample data set. With this in place, the α in (7) was typically left at a low level so as not to increase the **GDT**s' rate of failure, **RF**, as the range of $\Re = [P_{min}, P_{max}]$ was increased. The results pertaining to this are discussed in the next section.

4. Experimentation Results

4.1. Preprocessing of the P-C-F arbitrage data

The objective of these tests is to compare the effectiveness of **GDT**s generated by **EDDIE** against a naive arbitrage rule which executes an arbitrage trade whenever there is a contemporaneous profit signal. For the purpose of training **GDT**s, the following data pertaining to the short **P-C-F** arbitrage was fed into **EDDIE**: the strike price, the call premium, the put premium, the underlying index value, futures price, time to maturity of the contract and profit after transactions costs, **TC**. The total number of observations is 8073. **EDDIE** did not give any recommendations on this set. The data was then altered to be more informative in terms of economic theory. The strike price, was converted into the 'moneyness' variable , viz. as the ratio of the underlying to the strike price. This variable is introduced as in, at and out of the money for call and put options as this is known to have an impact on the arbitrage profits. The second variable added was basis, the formula for basis is as follows:
basis=(futures-spot)/futures.

This variable controls for the mispricing in the futures-spot leg of the arbitrage. The difference between the call and the put price, (C-P) is also added to the model. Finally, the following variables, ie. the underlying, (C-P) and profit after transaction costs, were input as a percentage of the futures price.

The results of a preliminary run on this data set are summarized below:

Table 1.4. Trial Run Results

Training	4178	rows	From	3	to	4180	
29 Jan 91	To	20 Dec 96	Contingency Table for Selected Data				
RC	**RMC**	**RF**	0	1			
94.18%	64.66%	80.38%	3894	168	0	4062	
			75	41	1	116	
			3969	209	5.00%	4178	2.78%
Testing	3895	rows	From	4181	to	8075	
2 Jan 97	To	18 Jun 98	Contingency Table for Selected Data				
RC	**RMC**	**RF**	0	1			
27.55 %	27.72%	93.01%	867	2743	0	3610	
			79	206	1	285	
			946	2949	75.71%	3895	7.32%

RC: Rate of Correctness; **RMC**: Rate of Missed Chances; **RF**: Rate of Failure

Table 1.4 (continued)

	EDDIE	**Naive**
Total Profit	122727	147257.2
Number of Recommendations	2949	1848
Average Profit	41.62	79.68

As the rate of failure, **RF**, was so high at 93% and the prediction accuracy of the recommendations made by **EDDIE** was so low, we decided to further process the data. Indeed, we recommend the following methodology for training good **EDDIE-ARB** rules due to some typical problems associated with arbitrage opportunities and the time execution delay.

Table 1.5. PROFIT DISTRIBUTION FOR SHORT P-C-F ARBITRAGE (January, 1991 - June, 1998)

Date	**Total Profits/ Losses**	**No**	**Mean** (£s)	**Total Profit** (£s)	**No**	**Mean** (£s)	**Total Losses** (£s)	**No**	**Mean** (£)
3 - 91	-887.71	30	-29.59	60.17361	4	15.04	-947.89	26	-36.46
6 - 91	525.01	10	52.50	594.809	8	74.35	-69.80	2	-34.90
9 - 91	-2357.48	71	-33.20	163.5158	5	32.70	-2520.99	66	-38.20
12 - 91	-3267.67	108	-30.26	487.4149	13	37.49	-3755.08	95	-39.53
3 - 92	-1689.68	76	-22.23	576.8741	14	41.21	-2266.56	62	-36.56
6 - 92	-1938.43	68	-28.51	0	0	0.00	-1938.43	68	-28.51
9 - 92	331.89	144	2.30	3414.901	45	75.89	-3083.01	99	-31.14
12 - 92	-346.46	40	-8.66	1118.904	9	124.32	-1465.36	31	-47.27
3 - 93	-2212.62	77	-28.74	393.1021	14	28.08	-2605.73	63	-41.36
6 - 93	-3346.30	106	-31.57	205.1266	5	41.03	-3551.42	101	-35.16
9 - 93	-6097.95	138	-44.19	0	0	0.00	-6097.95	138	-44.19
12 - 93	-3558.71	79	-45.05	0	0	0.00	-3558.71	79	-45.05
3 - 94	-2849.51	71	-40.13	19.84893	1	19.85	-2869.36	70	-40.99
6 - 94	-1826.17	149	-12.26	1913.429	46	41.60	-3739.60	103	-36.31
9 - 94	-10955.10	305	-35.92	240.5704	20	12.03	-11195.70	285	-39.28
12 - 94	-6611.33	236	-28.01	876.9749	17	51.59	-7488.30	219	-34.19
3 - 95	-8672.75	268	-32.36	941.3809	38	24.77	-9614.12	230	-41.80
6 - 95	9639.84	209	46.12	16376.27	35	467.89	-6736.43	174	-38.72
9 - 95	-4449.78	111	-40.09	0	0	0.00	-4449.78	111	-40.09
12 - 95	2573.72	135	19.06	6722.296	27	248.97	-4148.58	108	-38.41
3 - 96	1629.49	228	7.15	10461.15	8	1307.64	-8831.67	220	-40.14
6 - 96	-1012.07	159	-6.37	4328.274	6	721.38	-5340.34	153	-34.90
9 - 96	-3383.22	527	-6.42	15530.46	74	209.87	-18913.70	453	-41.75
12 - 96	11038.08	833	13.25	40913.9	108	378.83	-29875.80	725	-41.21
3 - 97	-21460.20	772	-27.80	4006.288	109	36.75	-25466.50	663	-38.41
6 - 97	74697.27	855	87.37	95803.59	241	397.53	-21106.30	614	-34.38
9 - 97	7688.20	435	17.67	19901.87	101	197.05	-12213.70	334	-36.57
12 - 97	36055.49	494	72.99	42875.33	282	152.04	-6819.86	212	-32.17
3 - 98	534927.60	483	1107.51	536097.6	409	1310.75	-1170.13	73	-16.03
6 - 98	394354.80	856	460.69	398458.8	706	564.39	-4103.79	148	-27.73

5. Methodology for Training EDDIE-ARB With Historical Tick Arbitrage Data

I. The distribution of contemporaneous **P-C-F** arbitrage signals is given in Table 5 in terms of contracts and profitability evaluated using equation (5) with about .1% FTSE-100 as transactions costs. In the LIFFE FTSE-100 **P-C-F** arbitrage triplets, it is clear that arbitrage opportunities were sparse in the early years from 1991-1994 with

December contracts being the most voluminous. As trading volume in the index options increased, we have a three fold increase of **P-C-F** arbitrage opportunities by 1994 and by the end of 1998, there is over a tenfold increase since the inception of electronic index options trading in LIFFE. On average the numbers of profitable **P-C-F** arbitrage opportunities are far outnumbered by loss making **P-C-F** opportunities in all years. However, in the years after 1995, the average total profitability of **P-C-F** positions become positive with the loss making arbitrages generating smaller losses than the gainful ones. In other words, the returns to **P-C-F** arbitrage as already seen from Table 2 are significant if the arbitrageur can successfully 'pick' the cherry.

II. The naive rule recommends that any profitable **P-C-F** signal is followed by an arbitrage in the same direction in a 9 minute window after a one minute execution delay. In the early years, the sparseness of **P-C-F** arbitrage price triplets meant that the few contemporaneous profitable signals that existed do not have follow ups in the given 10 minute window. Many of those that did have loss making follow ups. In fact, only 20% of the total sample of 8073 **P-C-F** triplets have any followups at all. It is only after 1996 that there are profitable follow ups from any **P-C-F** time stamp.

III. The above problem indicates that **EDDIE** has to be trained in the 20% of the total sample of short arbitrage **P-C-F** price triplets that have follow ups in the given window of opportunity. Thus, the sample size is reduced to 1641. Unlike the naive arbitrage strategy, **EDDIE** has to predict profitable follow up trades in the prescribed time window from any given time stamp of a **P-C-F** price triplet.

IV. The follow up based sample of 1641 lines is divided into test and training areas using a randomized procedure rather than one that is time based on a time series. Each **P-C-F** triplet is assigned one random number between 0 to 1. If the random number for the triplet is less than 0.63 then that triplet with all its adjacent followups in the next ten minutes is included in the training part of the sample. All of the remaining triplets, i.e. the triplets with a random number greater than 0.63 along with their adjacent ten minute followups, were included in the test part. This sampling procedure ensured that we have enough observations in the training part for **EDDIE** to function properly (the recommended minimum for training this class of **GPs** is 1000 observations). This sampling has also resulted in an equal distribution of arbitrage opportunities in training and test areas to counter the problems raised in point I.

For the 1641 sample points in which any **P-C-F** price triplet had a followup in the next 10 minutes, the naive strategy produced the results given in Table 6.

Table 1.6. Naive Strategy Trade Recommendations: Performance (January,1991- June 30 1998)

	Total Trade Signals	Overall Profit/ Loss	Profitable Signals	Total Profits	Average Profits
Training	290	274.76	159	82858.97	521.13
Testing	196	338.10	108	68092.68	630.49

Table 1.6 (continued)

	Loss Making Signals	Total Losses	Average Losses
Training	131	-3177.41	-24.26
Testing	88	-1824.52	-20.73

*All profit/losses calculated on the basis of 0.1% of the FTSE-100 transactions costs.

From Table 6 , we see that in the test sample, the naive rule made 196 recommendations to trade of which 88 were wrong. Hence the rate of failure is roughly 45%. The average profit on an arbitrage is £338 and the total profits net losses were £66268.16. The problem we were confronted with was to get the **GDT**s generated by **EDDIE-ARB** not only to yield higher average net profit per arbitrage than the naive rule, but to find enough profitable arbitrage positions that netted total profits of over £66268.16. As we will see, **EDDIE-ARB** had to be fitted with an additional constraint, $\Re = [P_{min}, P_{max}]$, where these parameters reflected the intensity of search that was desired without much deterioration of its performance in terms of rate of failure, **RF**.

6. EDDIE-ARB Evaluated

The **EDDIE-ARB** runs trained on this randomized data set provided good results. The results are summarized in Table 7.

Table 1.7. EDDIE-ARB **GDT**s : Training and Testing Results with Different $\Re = [Pmin, Pmax]$

$\Re = [5\% - 10\%]$

	RF	**RMC**	**RC**	**No of Signals**	**MEAN PROFIT**	**TOTAL PROFIT**
RULES	**TRAIN TEST**	**TRAIN TEST**	**TRAIN TEST**	**TRAIN TEST**	**TEST**	**TEST**
GDT1	0	0.624	0.8499	91	1046.49	62789.57
	0	0.6226	0.8441	60		
GDT2	0	0.5868	0.8588	100	929.54	64138.22
	0	0.566	0.8583	69		
GDT3	0	0.5992	0.8559	97	968.46	63918.51
	0	0.5849	0.8535	66		
GDT4	0	0.5992	0.8559	97	968.46	63918.51
	0	0.5849	0.8535	66		
GDT5	0	0.6157	0.8519	93	1009.24	60554.29
	0	0.6226	0.8441	60		
GDT6	0	0.5868	0.8588	100	968.46	63918.51
	0	0.5849	0.8535	66		
GDT7	0	0.5992	0.8559	97	904.25	65105.89
	0	0.5472	0.863	72		
GDT8	0	0.5868	0.8588	100	904.39	64211.48
	0	0.5535	0.8614	71		
GDT9	0	0.5909	0.8579	99	903.54	65055.18
	0	0.5472	0.863	72		
GDT10	0	0.5868	0.8588	100	968.46	63918.51
	0	0.5849	0.8535	66		
MEAN	0	0.5975	0.8563	97.4	957.1	63752.9
	0	0.5799	0.8548	66.8		
SD	0	0.0131	0.0032	3.1693	47.64	1297.61
	0	0.0275	0.0069	4.3665		

As we move from a conservative setting to a more ambitious one in the search intensity parameters $\Re = [P_{min}, P_{max}]$, **RF** increases and **RMC** decreases. As expected in the low setting of $\Re = [5\%, 10\%]$, the **GDT**s produced zero **RF**, but having missed over 84% of the profitable arbitrage opportunities, these rules failed to produce total profits greater than the naive rule. In every setting after that, most decision rules generated by **EDDIE-ARB** beat the naive strategy which provided 196 arbitrage signals, £66,268.16 total profit, and £338.10 average profit in the test area. Average profit generated by **GDT**s are highest when the most conservative setting is applied for the tests. However, the

Table 1.7 (continued)

$\Re = [10\% - 15\%]$

	RF	RMC	RC	No of Signals	MEAN PROFIT	TOTAL PROFIT
RULES	**TRAIN TEST**	**TRAIN TEST**	**TRAIN TEST**	**TRAIN TEST**	**TEST**	**TEST**
GDT1	0 0	0.4752 0.4591	0.8857 0.885	127 86	771.68	66364.11
GDT2	0.0338 0.01	0.4091 0.3774	0.8966 0.9039	148 100	668.46	66846.35
GDT3	0 0	0.438 0.4465	0.8946 0.8882	136 88	753.78	66332.2
GDT4	0.0726 0.05	0.5248 0.522	0.8648 0.863	124 80	842.08	67366.07
GDT5	0 0	0.4959 0.478	0.8807 0.8803	122 83	796.08	66074.3
GDT6	0 0	0.5165 0.4843	0.8757 0.8787	117 82	804.37	65958.05
GDT7	0.0152 0.0112	0.4628 0.4465	0.8867 0.8866	132 89	744.31	66243.73
GDT8	0.0877 0.0286	0.5702 0.5723	0.8529 0.8535	114 70	892.51	62475.56
GDT9	0 0	0.5207 0.4969	0.8748 0.8756	116 80	826.75	66140.26
GDT10	0 0	0.4669 0.4591	0.8877 0.885	129 86	770.43	66256.81
MEAN	0.0209 0.01	0.488 0.4742	0.88 0.88	126.5 84.4	787	66005.7
SD	0.0333 0.0168	0.0472 0.0515	0.0135 0.0139	10.395 7.7201	61.12	1308.05

total profit increases as the ambitious level increases. The parameters $\Re = [20\%, 25\%]$, roughly reflect the percentage of profitable arbitrage opportunities in the test area of the data with follow ups. The best **GDT**(4) in this setting produces 154 arbitrage signals with a 25% rate of failure (ie. 27 wrong predictions) with very large net total profits of £112683.86 and average profits per arbitrage of £731.71. For the full breakdown of the performance of **GDT**(4),see **Table 8**.

Our final analysis is to put the best performing **GDT**(4) to a robustness test to see how it performs in a period in which electronic trading was introduced into the LIFFE index markets. The robustness test pe-

Table 1.7 (continued)

$\Re = [15\% - 20\%]$

	RF	**RMC**	**RC**	**No of Signals**	**MEAN PROFIT**	**TOTAL PROFIT**
RULES	**TRAIN TEST**	**TRAIN TEST**	**TRAIN TEST**	**TRAIN TEST**	**TEST**	**TEST**
GDT1	0.3415 0.3077	0.4421 0.434	0.8241 0.8283	205 130	420.91	54718.71
GDT2	0.2199 0.2031	0.3843 0.3585	0.8658 0.8693	191 128	521.28	66723.61
GDT3	0.2135 0.2214	0.376 0.3585	0.8688 0.8646	192 131	508.36	66595.46
GDT4	0.2414 0.2241	0.4545 0.434	0.8489 0.8504	174 116	388.43	45057.48
GDT5	0.2513 0.3023	0.3967 0.434	0.8559 0.8299	195 129	515.24	66465.95
GDT6	0.1623 0.15	0.3388 0.3585	0.8877 0.8819	191 120	560.2	67223.75
GDT7	0.2246 0.2593	0.4008 0.3711	0.8618 0.852	187 135	491.91	66408.02
GDT8	0.2328 0.2188	0.4008 0.3711	0.8598 0.863	189 128	519.97	66555.53
GDT9	0.2083 0.2406	0.3719 0.3648	0.8708 0.8583	192 133	499.77	66469.69
GDT10	0.1961 0.2391	0.3223 0.3396	0.8827 0.863	204 138	488.41	67401.27
MEAN	0.2292 0.2366	0.3888 0.3824	0.8626 0.8561	192 128.8	491.4	63361.9
SD	0.0467 0.0462	0.0407 0.0366	0.0179 0.0167	8.705 6.5794	50.5	7464.85

riod was from March 30,1998- March 30, 1999, when three additional contracts, viz. the September 1998, December 1998 and the March 1999 contracts were studied. There were 2140 **P-C-F** time matched price triplets in the trade tick data, reflecting the increased turnover in the index markets from 1998. Of these almost as many had followups in the next ten minutes, again a reflection of the increased turnover. But, only 328 were profitable (at .1% of FTSE-100 as transactions costs) in next ten minutes from any **P-C-F** time matched price triplets. The naive strategy makes a total of 273 recommendations in the robustness period of which 27 were wrong, ie. a 10% error rate. The interesting point is

Table 1.7 (continued)

$\Re = [20\% - 25\%]$

	RF	**RMC**	**RC**	**No of Signals**	**MEAN PROFIT**	**TOTAL PROFIT**
RULES	**TRAIN TEST**	**TRAIN TEST**	**TRAIN TEST**	**TRAIN TEST**	**TEST**	**TEST**
GDT1	0.4492 0.4346	0.3058 0.3208	0.7903 0.789	305 191	415.2	79303.87
GDT2	0.4023 0.3734	0.3554 0.3774	0.8101 0.8126	261 158	495.89	78351.04
GDT3	0.458 0.4302	0.3595 0.3585	0.7833 0.789	286 179	442.11	79137.55
GDT4	0.252 0.1753	0.2397 0.2013	0.8807 0.9071	246 154	731.71	112683.86
GDT5	0.373 0.3554	0.3471 0.327	0.8231 0.8252	252 166	397.27	65947.63
GDT6	0.3373 0.3092	0.3182 0.3396	0.84 0.8409	249 152	439.22	66760.98
GDT7	0.4633 0.4462	0.3058 0.3208	0.7823 0.7827	313 195	406.01	79171.06
GDT8	0.4498 0.4353	0.3884 0.3962	0.7863 0.7843	269 170	459.08	78043.79
GDT9	0.3485 0.377	0.2893 0.283	0.839 0.8205	264 183	430.58	78796.89
GDT10	0.4327 0.4167	0.3554 0.3396	0.7962 0.7969	275 180	436.09	78497.09
MEAN	0.3966 0.3753	0.3264 0.3264	0.8131 0.8148	272 172.8	465.3	79669.4
SD	0.0687 0.0827	0.0433 0.0542	0.0324 0.038	23.0314 15.1936	97.68	12701.92

that average profit of £952 per arbitrage position from the naive strategy, see Table 8, is substantially higher in the period after mid 1998. Hence, clearly, though the markets offer fewer arbitrage opportunities, those that exist are highly profitable. The remarkable result from Table 8 is that the best **GDT**(4) rule trained in the period when market turnover was low produced a very robust result for the period after the introduction of electronic trading. The **GDT**(4) rule made 249 recommendations to trade of which only 13 were wrong, ie. an error rate of about 5% which is less than that for the naive rule. Thus, though the total net profit remained on par with that of the naive rule, **EDDIE-**

Table 1.8. Robustness Test :**GDT**(4) Performance vs. Naive Strategy

	Total Trade Signals	Overall Profit/ Loss	Profitable Signals	Total Profits	Average Profits
Training	246	617.82	184	154768.16	841.13
Testing	154	731.71	127	113847.68	896.44
Robustness Period*					
GDT(4)	249	1043.35	236	260507.03	1103.84
Robustness Period*					
Naive Rule	273	952.88	246	261480.83	1062.93

Table 1.8 (continued)

	Loss Making Signals	Total Losses	Average Losses
Training	62	-2784.61	-44.91
Testing	27	-1163.82	-43.10
Robustness Period*			
GDT(4)	13	-712.509	-54.81
Robustness Period*			
Naive Rule	27	-1345.46	-49.83

*The robostness test period was March 30,1998-March 30,1999.

ARB can clearly 'pick' the cherry quite competently with the average return on an arbitrage exceeding that for the naive rule.

7. Conclusion

The main conclusion is that the conduct of a complex arbitrage in a fast moving market requires a mechanized strategy that cannot rely only on arbitrage recommendations that arise from contemporaneous profit signals. **EDDIE-ARB** that uses genetic programming was developed to predict arbitrage opportunities in the FTSE-100 index futures and options market for a period of up to ten minutes. This gives an arbitrageur a time advantage that cannot be obtained on the basis of contemporaneous profit signals. Technically, the success of **EDDIE-ARB** lies in the unique fitness function that was developed that enabled the genetic program to search more intensively without much deterioration in the rate of failure of its recommendations.

Acknowledgments

We acknowledge the receipt of a RPF grant from the University of Essex for this research.

References

Allen, F. Karjalainen, R. (1995). Using Genetic Algorithms to find Technical Trading Rules. Working paper at Rodney L. White Center for Financial Research.

Alexander, S.S., (1964). Price movement in speculative markets: trend or random walks, No. 2, in Cootner, P. (ed.), The random character of stock market prices, MIT Press, Cambridge, MA, 338-372.

Backus, J.W., (1959). "The syntax and semantics of the proposed international algebraic language of Zurich", ACM-GAMM conference, ICIP, Paris, June.

Bae, K.H., Chan, K., Cheung, Y.L.(1998). The profitability of index futures arbitrage: Evidence from bid-ask quotes, Journal of Futures Markets, 18, 743-763.

Bauer, R. J. Jr., (1994).Genetic Algorithms and Investment Strategies. New York, John Wiley Sons, Inc.

Brock, W., Lakonishok, J. LeBaron, B., (1992). Simple technical trading rules and the stochastic properties of stock returns, Journal of Finance, 47, 1731-1764.

Chen, S-H Yeh, C-H. , (1997). Toward a computable approach to the efficient market hypothesis: An application of genetic programming, Journal of Economic Dynamics and Control, 21, 1043-1063.

Cornell, B., and French, K.(1988). Taxes and the Pricing of Stock Index Futures. Journal of Finance, 38, 675-694.

Evnine, J. Rudd A.(1985). Index Options - Early Evidence. Journal of Finance.

Fama, E.F. Blume, M.E., (1966). Filter rules and stock-market trading, Journal of Business 39(1), 226-241.

Fung, Joseph, K.,W. Chan Kam,C.(1994). On the Arbitrage Free Relationship Between Index Futures and Index Options: A Note. Journal of Futures Markets, Vol. 14, 957-962

Fung, Joseph, K.,W., Cheng,L., T.W., Chan Kam,C.(1997).The Intraday pricing Efficiency of Hong Kong Hang Seng Index Option and Futures Markets, Journal of Futures Markets, Vol. 17, 797-815.

Gemmill, G.(1993). Options Pricing, Maidenhead, UK, Mc Graw-Hill.

Gwilm, O.P, Buckle M.(1999). Volatility Forecasting in the Framework of the Option Expiry Cycle. The European Journal of Finance, 5,73-94.

Goldberg, D.E., (1989).Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley.

Holland, J.H., (1975), *Adaptation in natural and artificial system*, University of Michigan Press.

Koza, J.R., (1992). Genetic Programming: on the programming of computers by means of natural selection. MIT Press.

Koza, J.R., (1994).Genetic Programming II: Automatic Discovery of Reusable Programs. MIT Press.

Koza, J., Goldberg, D., Fogel, D. Riolo, R. (ed.). Proceedings of the First Annual Conference on Genetic programming, MIT Press, 1996.

Lee, J.H., Nayar, N.(1993). A transactions data analysis of arbitrage between index options and index Futures. The Journal of Futures Markets, Vol. 13, 889-902

Li, J. Tsang, E.P.K, (1999a). Improving technical analysis predictions: an application of genetic programming, Proceedings of The 12th International Florida AI Research Society Conference, Orlando, Florida, May 1-5, 1999, 108-112.

Li, J. Tsang, E.P.K, (1999b). Investment decision making using FGP: a case study,Proceedings of Congress on Evolutionary Computation (CEC'99), Washington DC, USA, July 6-9 1999.

Mahfoud, S. Mani, G., (1997). Financial Forecasting Using Genetic Algorithms, Journal of Applied Artificial Intelligence Vol.10, 6, 543-565.

Markose, S. and H. Er, 2000. The Black Effect(1976) Effect and Cross Market Arbitrage in FTSE-100 Index Futures and Options, Working Paper, No. 522, Economics Department, University of Essex.

Mitchell, M., (1996). An Introduction to Genetic Algorithms. MIT Press.

Modest, D. Sunderesan, M.(1983). The relationship between spot and futures prices in stock index futures markets:some preliminary evidence, Journal of Futures Markets, 3, 15-41.

Stoll, H.R. (1969). The relationship between put and call option prices. Journal of Finance. 25: 801-824

Neely, C., Weller, P. Ditmar, R., (1997). Is technical analysis in the foreign exchange market profitable? A genetic programming approach, Journal of Financial and Quantitative Analysis, 32, 405-26.

Oussaidene, M., Chopard, B., Pictet, O. Tomassini, M., (1997). Practical aspects and experiences - Parallel genetic programming and its application to trading model induction, Journal of Parallel Computing Vol. 23, No. 8, 1183-1198.

Saad, E., Prokhorov, D., and Wunsch, D., (1998). Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks, IEEE Transactions on Neural Networks, vol. 9. 1456-1470. Sweeney, R.J., (1988), "Some new filter rule tests: Methods and results," Journal of Financial and Quantitative Analysis, 23, 285-300.

Tucker, A.L. (1991).Financial Futures, Options and Swaps, West Publishing Company, St. Paul, MN.

Tsang, E.P.K , Li, J., Markose, S., Hakan, E., Salhi, A., and Iori,G., 2000. EDDIE in Financial Decision Making, Journal of Management Economics, http: //www.econ.uba.ar/www/servicos/publicaciones/journal3/index.htm

Tsang, E.P.K., Li, J. Butler, J.M., (1998). EDDIE beats the bookies, International Journal of Software, Practice Experience, Wiley, Vol.28 (10), 1033-1043.

Yadav, P.K and P. Pope, 1990. Stock Index Futures arbitrage: International Evidence, Journal of Futures Markets, 10, 573-603.